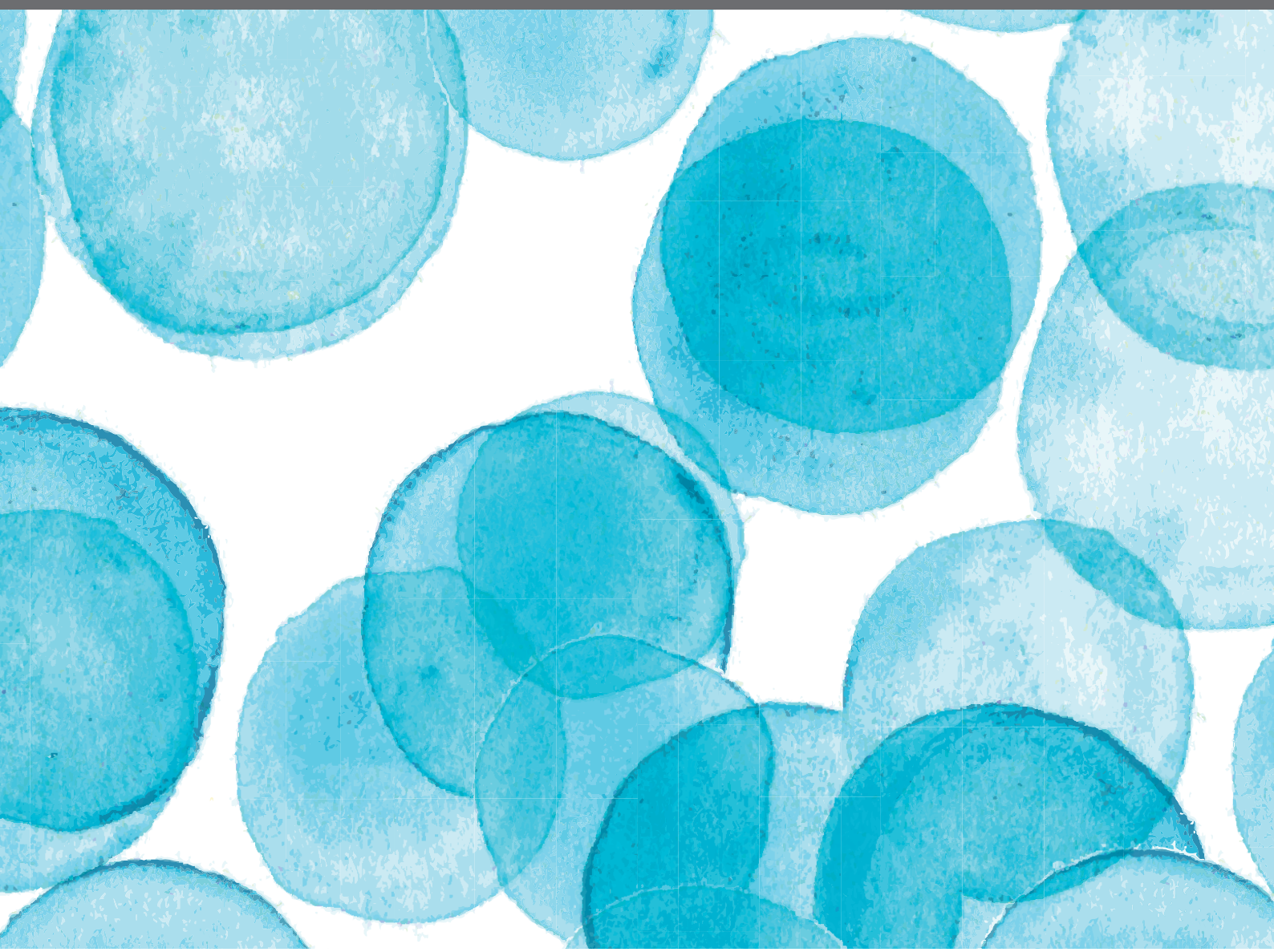
EVOLUTION & GENOMIC ADAPTATION OF EMERGING AND RE-EMERGING RNA VIRUSES

EDITED BY: Kai Huang, Justin Jang Hann Chu and Josanne Hinke Verhagen

PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-985-3
DOI 10.3389/978-2-88971-985-3

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

EVOLUTION & GENOMIC ADAPTATION OF EMERGING AND RE-EMERGING RNA VIRUSES

Topic Editors:
Kai Huang, University of Texas Medical Branch at Galveston, United States
Justin Jang Hann Chu, National University of Singapore, Singapore
Josanne Hinke Verhagen, Erasmus Medical Center, Netherlands

Citation: Huang, K., Chu, J. J. H., Verhagen, J. H., eds. (2021). Evolution & Genomic Adaptation of Emerging and Re-emerging RNA Viruses. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-985-3

Table of Contents

05 ***Editorial: Evolution & Genomic Adaptation of Emerging and Re-emerging RNA Viruses***
Kai Huang and Justin Jang Hann Chu

08 ***Development of a Visible Reverse Transcription-Loop-Mediated Isothermal Amplification Assay for the Detection of Rift Valley Fever Virus***
Qiuxue Han, Shengnan Zhang, Dongping Liu, Feihu Yan, Hualei Wang, Pei Huang, Jinhao Bi, Hongli Jin, Na Feng, Zengguo Cao, Yuwei Gao, Hang Chi, Songtao Yang, Yongkun Zhao and Xianzhu Xia

18 ***The V617I Substitution in Avian Coronavirus IBV Spike Protein Plays a Crucial Role in Adaptation to Primary Chicken Kidney Cells***
Yi Jiang, Mingyan Gao, Xu Cheng, Yan Yu, Xinyue Shen, Jianmei Li and Sheng Zhou

32 ***Addressing Non-linear System Dynamics of Single-Strand RNA Virus–Host Interaction***
Alessandra Romano, Marco Casazza and Francesco Gonella

52 ***Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil***
Paola Cristina Resende, Edson Delatorre, Tiago Gräf, Daiana Mir, Fernando Couto Motta, Luciana Reis Appolinario, Anna Carolina Dias da Paixão, Ana Carolina da Fonseca Mendonça, Maria Ogrzewalska, Braulia Caetano, Gabriel Luz Wallau, Cássia Docena, Mirleide Cordeiro dos Santos, Jessylene de Almeida Ferreira, Edivaldo Costa Sousa Junior, Sandro Patroca da Silva, Sandra Bianchini Fernandes, Lucas Alves Vianna, Larissa da Costa Souza, Jean F. G. Ferro, Vanessa B. Nardy, Cliomar A. Santos, Irina Riediger, Maria do Carmo Debur, Júlio Croda, Wanderson K. Oliveira, André Abreu, Gonzalo Bello and Marilda M. Siqueira

66 ***Molecular Evolution of Human Norovirus GII.2 Clusters***
Xingguang Li, Haizhou Liu, Brittany Rife Magalis, Sergei L. Kosakovsky Pond and Erik M. Volz

76 ***Evolutionary Mechanism of Immunological Cross-Reactivity Between Different GII.17 Variants***
Yueting Zuo, Liang Xue, Junshan Gao, Yingyin Liao, Yanhui Liang, Yueting Jiang, Weicheng Cai, Zhiwei Qin, Jiale Yang, Jumei Zhang, Juan Wang, Moutong Chen, Yu Ding and Qingping Wu

86 ***Isolation and Identification of a Recombinant Porcine Epidemic Diarrhea Virus With a Novel Insertion in S1 Domain***
Dongliang Li, Yongtao Li, Yunchao Liu, Yumei Chen, Wenqiang Jiao, Hua Feng, Qiang Wei, Jucai Wang, Yuhang Zhang and Gaiping Zhang

98 ***Contributions of Genetic Evolution to Zika Virus Emergence***
Su-Jhen Hung and Sheng-Wen Huang

105 ***Genomic Epidemiology of SARS-CoV-2 From Mainland China With Newly Obtained Genomes From Henan Province***
Ning Song, Guang-Lin Cui and Qing-Lei Zeng

117 ***Recurrent Dissemination of SARS-CoV-2 Through the Uruguayan–Brazilian Border***
Daiana Mir, Natalia Rego, Paola Cristina Resende, Fernando Tort, Tamara Fernández-Calero, Verónica Noya, Mariana Brandes, Tania Possi, Mailen Arleo, Natalia Reyes, Matías Victoria, Andres Lizasoain, Matías Castells, Leticia Maya, Matías Salvo, Tatiana Schäffer Gregianini, Marilda Tereza Mar da Rosa, Letícia Garay Martins, Cecilia Alonso, Yasser Vega, Cecilia Salazar, Ignacio Ferrés, Pablo Smircich, Jose Sotelo Silveira, Rafael Sebastián Fort, Cecilia Mathó, Ighor Arantes, Luciana Appolinario, Ana Carolina Mendonça, María José Benítez-Galeano, Camila Simoes, Martín Graña, Fernando Motta, Marilda Mendonça Siqueira, Gonzalo Bello, Rodney Colina and Lucía Spangenberg

128 ***Molecular Epidemiology, Evolution and Reemergence of Chikungunya Virus in South Asia***
Nadim Sharif, Mithun Kumar Sarkar, Rabeya Nahar Ferdous, Shamsun Nahar Ahmed, Md. Baki Billah, Ali Azam Talukder, Ming Zhang and Shuvra Kanti Dey

142 ***Host Adaptive Evolution of Avian-Origin H3N2 Canine Influenza Virus***
Fucheng Guo, Ayan Roy, Ruichen Wang, Jinjin Yang, Zhipeng Zhang, Wen Luo, Xuejuan Shen, Rui-Ai Chen, David M. Irwin and Yongyi Shen

156 ***Coxsackievirus A16 in Southern Vietnam***
Le Nguyen Truc Nhu, Le Nguyen Thanh Nhan, Nguyen To Anh, Nguyen Thi Thu Hong, Hoang Minh Tu Van, Tran Tan Thanh, Vu Thi Ty Hang, Do Duong Kim Han, Nguyen Thi Han Ny, Lam Anh Nguyet, Du Tuan Quy, Phan Tu Qui, Truong Huu Khanh, Nguyen Thanh Hung, Ha Manh Tuan, Nguyen Van Vinh Chau, Guy Thwaites, H. Rogier van Doorn and Le Van Tan

164 ***Meta-Analysis and Structural Dynamics of the Emergence of Genetic Variants of SARS-CoV-2***
Nicolas Castonguay, Wandong Zhang and Marc-André Langlois

183 ***Gene Segment Interactions Can Drive the Emergence of Dominant Yet Suboptimal Gene Constellations During Influenza Virus Reassortment***
Sanja Trifkovic, Brad Gilbertson, Emily Fairmaid, Joanna Cobbin, Steven Rockman and Lorena E. Brown

frontiers in Microbiology

EDITORIAL
published: 12 November 2021
doi: 10.3389/fmicb.2021.777257

Editorial: Evolution & Genomic Adaptation of Emerging and Re-emerging RNA Viruses

Kai Huang[1,2*†] *and Justin Jang Hann Chu*[3,4,5*†]

[1] *Department of Pathology, University of Texas Medical Branch, Galveston, TX, United States,* [2] *Galveston National Laboratory, University of Texas Medical Branch, Galveston, TX, United States,* [3] *Laboratory of Molecular RNA Virology and Antiviral Strategies, Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore,* [4] *Infectious Diseases Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore,* [5] *Collaborative and Translation Unit for HFMD, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, Singapore, Singapore*

Keywords: RNA virus, molecular epidemiology, phylogeny, evolution, emerging/re-emerging infections

Editorial on the Research Topic

Evolution & Genomic Adaptation of Emerging and Re-emerging RNA Viruses

OPEN ACCESS

Edited and reviewed by:
Nejat Duzgunes,
University of the Pacific, United States

***Correspondence:**
Kai Huang
kahuang@utmb.edu
Justin Jang Hann Chu
miccjh@nus.edu.sg

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *15 September 2021*
Accepted: *26 October 2021*
Published: *12 November 2021*

Citation:
Huang K and Chu JJH (2021) Editorial: Evolution & Genomic Adaptation of Emerging and Re-emerging RNA Viruses. Front. Microbiol. 12:777257. doi: 10.3389/fmicb.2021.777257

Emerging and re-emerging infectious diseases are defined as diseases caused by unidentified and reappearing pathogens (NIAID, 2018). RNA viruses are mostly responsible for most of such infectious disease outbreaks (Nichol et al., 2000). One of the key reasons is due to genomic alterations such as spontaneous mutation recombination or reassortment which occurs during adaptation and evolution processes (Nichol et al., 2000). When those genomic changes have been accumulated to a certain level or when the changes are on the antigenic or receptor-binding region, the host immune systems are no longer able to recognize the new variants, resulting in global viral outbreaks (De Wit et al., 2016; Nelemans and Kikkert, 2019; Kikkert, 2020). SARS-CoV-2 and its variants has claimed more than 4.56 million lives from December 2019 until the writing of this manuscript. The high mortality rate emphasizes the importance of continuously monitoring these emerging viruses' evolution and genomic adaptation (WHO, 2021).

The purpose of this Research Topic serves to provide an open access platform for an international global team of multidisciplinary researchers and scientists to share their findings on the viral genomic feathers and changes, to provide a platform enabling public health officials to warn the global community of potential and existing epidemics and pandemics, and to develop an analytical platform for researchers to evaluate the outbreak risks, to prepare and to control future pandemics (Huang et al., 2021). In this Research Topic, the authors will present their most recent genomic investigations on these viruses. A total of 28 manuscripts including original research and review have been received, of which 15 were eventually accepted for journal publications after rigorous peer review processes. Based on the viruses involved, these 15 articles can be briefly classified into three groups: positive sense single-strand RNA (+ssRNA) viruses; negative-sense single-strand RNA (-ssRNA) viruses; diagnosis and dynamic single-strand RNA Virus-Host interactions.

The first group +ssRNA viruses are the Group IV viruses in the Baltimore classification system (Baltimore, 1971; Cann, 2016), including viruses from *Coronaviridae, Picornaviridae, Flaviviridae, Togaviridae*. Within this group viruses, Coronavirus is the most researched virus in this topic. Three

original research articles discussed the SARS-CoV-2 coronaviruses identified in China, Brazil, and Uruguay, respectively. In China, Song et al. isolated SARS-CoV-2 viruses from the Henan Province, which is adjacent to the Hubei Province, the region which has the highest mortality rate due to the SARS-CoV-2 pandemic within China. They analyzed the samples from different locations to estimate the virus's most recent common ancestor (TMRCA) and evolutionary rate. In Brazil, Resende et al. analyzed 190 SARS-CoV-2 viruses isolated from 13 Brazilian states and found the B.1.1.33-like viruses circulating in Brazil might have been transmitted from Europe or domestically erupted a few weeks before regional outbreaks. Their analysis also indicates public health interventions were successful because the median effective reproductive number (Re) dropped by 66%. In Uruguay, Mir et al. investigated the local virus source and the transmission rate(s). Based on the 122 viruses recovered at Brazilian–Uruguayan border area, they found that the SARS-CoV-2 viruses in the Uruguay border were introduced multiple times independently from Brazil (lineage B.1.1.28 and B.1.1.33). The researchers also revealed in their research that the synonymous and non-synonymous single nucleotide polymorphisms (SNP) are the genetic variations responsible to define the lineages. Castonguay et al. completed the fourth SARS-CoV-2 meta-analysis, which systematically tracked the evolutionary trajectory of SARS-CoV-2 over time, identified emerging mutations, and modeled the structural changes and corresponding molecular interactions. The fifth coronavirus is the avian infectious bronchitis virus (IBV). Jiang et al. identified a critical mutation to determine the host tropism alteration, which is a valuable key in understanding why the coronavirus could jump from one host to another. Porcine Epidemic Diarrhea virus (PEDV) is the last coronavirus discussed in this Research Topic. Li et al. isolated a PEDV, which has been detected to contain a unique insertion in the S1 protein binding by the recombination test. They predicted the structure of this new recombinant and proved its biological correlation by showing this strain had higher pathogenicity than the other viruses isolated in the piglet *in vivo* challenge.

Following the *Coronaviridae, Picornaviridae* is the second most popular virus family within this topic. In two articles, researchers investigated the Norovirus and Coxsackievirus, which are enteroviruses within the family *Picornaviridae*. Zuo et al. identified the new Norovirus GII.17 variants, which surpassed the predominant GII.4 genotype causing the Kawasaki variant outbreaks in 2014–2015. Serological analysis showed weak cross-protection to these new variants so attention should be taken to prevent future outbreaks. The corresponding mutated amino acids on antigenic sites were also identified. The second study led by Li's team focused on the Norovirus GII.2 clusters, which caused unprecedented endemic outbreaks in 2016–2017. Eight distinct clusters with increased genetic diversity were characterized with an absence of elevated evolutionary rate. Additionally, the selection pressure was detected, suggesting the outbreak was probably not related to an evolutionary adaptation. The second virus in the *Picornaviridae*, is the Coxsackievirus A16 (CVA16) and was reported by Nhu et al. This molecular epidemiology of CVA16 in Southern Vietnam was documented for the first time. They found the Vietnamese CVA16 strains belong to a single genogroup B1a and these viruses displayed a less pronounced genetic alternation compared to the enterovirus A71 (EV-A71). Zika virus (ZIKV) of the *Flaviviridae* family and Chikungunya virus (CHIKV) of the *Togaviridae* family are the last two members of ssRNA viruses mentioned in this article collection. Su-Jhen Hung and Sheng-Wen Huang reviewed the amino acid substitutions and factors contributing to the pathogenicity and transmission of ZIKV. Sharif et al. systematically reviewed the evolution, epidemiology, phylogeny for CHIKV.

Influenza virus is the major representative of negative-sense ssRNA viruses, the Group V Baltimore viruses outlined in this topic. Influenza had caused at least 5 pandemics since the 20th century (Kilbourne, 2006). Reassortments of segmented genes from different sources were generally believed to be responsible for these prior pandemics. Trifkovic et al. investigated the gene reassortment driving force using the vaccine seed production model. They found a selective preference for genome compatibility between different viruses. They also found a powerful driving force to determine the dominant progeny besides the ones they initially thought. In the second influenza paper, Guo et al. characterized the interspecies transmission process from the original avian host to a new canine host. They identified 54 substituted amino acids fixed during the interspecies transmission. By analyzing the selection pressure and codon usage, they found the canine influenza viruses are better adapted to avian hosts, supporting the theory of their avian origin.

In the third Research Topic group, the virus diagnosis and host-virus interactive network were discussed. Based on the conservative RNA gene segment, Han et al. developed an easy and fast isothermal amplification assay to detect RVFV with high sensitivity and specificity. The test uses a strip that can detect RVFV at as low as ∼200 copies/μL, 100-fold, which is more sensitive than real-time RT-PCR assay without cross-reactivity to viruses causing similar symptoms. The test requires no specific equipment and can be done in 1 h with results visible in 5 min. Such a quick test with high sensitivity and specificity is quite prominent in the detection of emerging pathogens. In the second paper in the group, Romano et al. proposed a multidisciplinary framework using Systems Thinking (ST) to study the dynamics of host-virus interactive networks for the first time. In the Systems Thinking theory, changes of one component in the network will lead to the corresponding changes of other components until it reaches a new balanced stationary status. Using the System Dynamics (SD) modeling, the authors computationally simulated the dynamic trajectory of host-virus interactions and identified the leverage points to minimize virion release.

In summary, this Research Topic provides cutting edge methodologies, research, and practices in the *Evolution & Genomic Adaptation of Emerging and Re-emerging RNA Viruses*. We heartily acknowledge and are grateful for the contributions and outstanding works of all the authors and reviewers of this Research Topic. We believe this collection will raise the awareness of

emerging/re-emerging RNA viruses and enable proficient monitoring processes in order to decrease mortality rates by having an efficient system in place to control the next SARS-CoV-2 outbreak.

AUTHOR CONTRIBUTIONS

All authors contributed to co-editing the special Research Topic, edited, revised, and approved the editorial.

FUNDING

This work was supported by the Centers for Excellence in Influenza Research and Surveillance (CEIRS) Program, Grant Number HHSN272201400008C, from the National Institute of Allergy and Infectious Diseases (NIAID).

ACKNOWLEDGMENTS

We are grateful to Lear Dobbins for the careful proofreading.

REFERENCES

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev*. 35, 235–241. doi: 10.1128/br.35.3.235-241.1971

Cann, A., (2016). *Principles of Molecular Virology*. Available online at: https://ebookcentral.proquest.com/lib/nyulibrary-ebooks/detail.action?docID=5754488 (accessed September 6, 2021).

De Wit, E., Van Doremalen, N., Falzarano, D., and Munster, V. J. (2016). SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol*. 14, 523–534. doi: 10.1038/nrmicro.2016.81

Huang, K., Chu, J. J. H., and Verhagen, J. H. (2021). *Evolution and Genomic Adaptation of Emerging and Re-Emerging RNA Viruses*. Available online at: https://www.frontiersin.org/research-topics/15569/evolution-genomic-adaptation-of-emerging-and-re-emerging-rna-viruses (accessed September 6, 2021).

Kikkert, M. (2020). Innate immune evasion by human respiratory RNA viruses. *J. Innate Immun*. 12, 4–20. doi: 10.1159/000503030

Kilbourne, E. D. (2006). Influenza pandemics of the 20th century. *Emerg. Infect. Dis*. 12, 9–14. doi: 10.3201/eid1201.051254

Nelemans, T., and Kikkert, M. (2019). Viral innate immune evasion and the pathogenesis of emerging RNA virus infections. *Viruses* 11:961. doi: 10.3390/v11100961

NIAID (2018). *NIAID Emerging Infectious Diseases/Pathogens*. Available online at: https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens (accessed September 6, 2021).

Nichol, S. T., Arikawa, J., and Kawaoka, Y. (2000). Emerging viral diseases. *Proc. Natl. Acad. Sci*. 97, 12411–12412. doi: 10.1073/pnas.210382297

WHO (2021). *WHO Coronavirus (COVID-19) Dashboard*. Available online at: https://covid19.who.int/ (accessed September 6, 2021).

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 November 2020
doi: 10.3389/fmicb.2020.590732

Development of a Visible Reverse Transcription-Loop-Mediated Isothermal Amplification Assay for the Detection of Rift Valley Fever Virus

Qiuxue Han[1,2†], Shengnan Zhang[2†], Dongping Liu[3], Feihu Yan[2,4], Hualei Wang[5], Pei Huang[2,6], Jinhao Bi[2,6], Hongli Jin[2,5], Na Feng[2], Zengguo Cao[2,5], Yuwei Gao[2,4], Hang Chi[2,4], Songtao Yang[2,4], Yongkun Zhao[2,4] and Xianzhu Xia[1,2,4]**

[1]Institute of Laboratory Animal Science, Chinese Academy of Medical Sciences (CAMS) and Comparative Medicine Center, Peking Union Medical College (PUMC), Beijing, China, [2]Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Institute of Military Veterinary Medicine, Academy of Military Medical Sciences, Changchun, China, [3]The Nanjing Unicorn Academy of Innovation, Institute Pasteur of Shanghai, Chinese Academy of Sciences, Nanjing, China, [4]Jiangsu Co-innovation Center for Prevention and Control of Important Animal Infectious Disease and Zoonoses, Yangzhou University, Yangzhou, China, [5]College of Veterinary Medicine, Jilin University, Changchun, China, [6]Animal Science and Technology College, Jilin Agricultural University, Changchun, China

OPEN ACCESS

Edited by:
Justin Jang Hann Chu,
National University of Singapore,
Singapore

Reviewed by:
Tetsuro Ikegami,
University of Texas Medical Branch at Galveston, United States
Jin Zhou,
Tsinghua University, China

***Correspondence:**
Yongkun Zhao
zhaoyongkun1976@126.com
Xianzhu Xia
xiaxzh@cae.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 10 August 2020
Accepted: 28 October 2020
Published: 13 November 2020

Citation:
Han Q, Zhang S, Liu D, Yan F, Wang H, Huang P, Bi J, Jin H, Feng N, Cao Z, Gao Y, Chi H, Yang S, Zhao Y and Xia X (2020) Development of a Visible Reverse Transcription-Loop-Mediated Isothermal Amplification Assay for the Detection of Rift Valley Fever Virus. Front. Microbiol. 11:590732. doi: 10.3389/fmicb.2020.590732

Rift Valley fever (RVF) is a severe infectious disease, which can through mosquito bites, direct contact and aerosol transmission infect sheep, goats, people, camels, cattle, buffaloes, and so on. In this paper, a conserved region of the S RNA segment of Rift Valley fever virus (RVFV) ZH501 strain was used as target sequence. The RVFV RT-LAMP-VF assay was successfully established combined reverse transcription-loop-mediated isothermal amplification with a vertical flow visualization strip. The detection limit is up to 1.94×10^0 copies/μl of synthesized RVFV-RNA. RNA extracted from cell culture of an inactivated RVFV-BJ01 strain was also used as templates, and the detection limit is 1.83×10^3 copies/μl. In addition, there was no cross-reactivity with other viruses that can cause similar fever symptoms. The RVFV-LAMP-VF assay exhibited very high levels of diagnostic sensitivity, which had 100-fold more sensitive than RVFV real-time RT-PCR assay. Accordingly, the RVFV RT-LAMP-VF assay developed in this study is suitable for the rapid and sensitive diagnosis of RVFV without specialized equipment and can rapidly complete detection within 60 min, and the results are visible by vertical flow visualization strip within 5 min.

Keywords: Rift Valley fever virus, reverse transcription-loop-mediated isothermal amplification, nucleic acid visualization, visual detection, inactivated RVFV-BJ01 strain

INTRODUCTION

Rift Valley fever virus (RVFV) is an arthropod-borne, zoonotic virus (genus *Phlebovirus*, family Phenuivirudae) that is a significant threat to domestic ruminants and humans (Elliott, 1997). RVFV was first discovered in 1930 in Kenya's Great Rift Valley and spread in Africa (Daubney et al., 1931). The first outbreak outside of Africa was on the Arabian Peninsula in 2000 (Madani et al., 2003; Rolin et al., 2013). As global warming leads to an expanded distribution

of insect vectors and increased international trade, the risk of RVFV introduction into countries, where RVFV is not endemic is increasing (Li et al., 2019). Thus, there is urgent demand for the development of safe, rapid, and accurate diagnostic assays.

Techniques for the diagnosis of RVFV include virus isolation, nucleic acid techniques, detection of viral antigen, and specific antibodies. RVFV can be isolated from whole tissues, blood, or serum during the febrile stage of the disease (Crespo Leon et al., 2005). Various serological assays are used to detect antibodies against RVFV (Paweska et al., 2003). Enzyme-linked immunosorbent assays have been extensively validated for the serodiagnosis of RVFV, and an indirect enzyme-linked immunosorbent assay based on the recombinant nucleocapsid protein of RVFV has been recently developed for the detection of specific antibodies in human and animal sera (Paweska et al., 2007, 2008; van Vuren et al., 2007). The virus neutralization test is regarded as the gold standard, and the classical method takes 4–7 days. A RVFV-4seGFP based VNT can be completed in just 48 h (Wichgers Schreur et al., 2017). Moreover, since it requires live virus and can be done only in biocontainment facilities, the virus production requires biocontainment facilities to limit the risk of exposure of laboratory personnel to infection (Paweska et al., 2003). Highly sensitive PCR assays for the detection and quantification of RVFV have been reported, including reverse transcriptase PCR (Garcia et al., 2001; Sall et al., 2002) and real-time RT-PCR (Drosten et al., 2002; Tercero et al., 2019), which require sophisticated equipped laboratories and might be beyond the resources and capabilities of many developing countries. Research on other diagnostic methods has been reported, including optical fiber immunosensor (OFIS), competitive ELISA (Upreti et al., 2018), lateral flow tests (LFT; Cetre-Sossah et al., 2019), fluorescence microsphere immunoassay (FMIA; Ragan et al., 2018), and reverse transcription recombinase polymerase amplification assay (RT-RPA; Euler et al., 2012). The above methods either have low sensitivity, stringent precision requirements and high requirements on operators or present high cost disadvantages. In resource-poor areas, low cost, easy operation, and intuitive reading of results are more important. At present, there is still a lack of a sensitive, intuitive reading method, which is more suitable for on-site testing.

After being infected with RVFV, the initial incubation period is usually 2–6 days, and then enter the fever period, which usually lasts 3–4 days (Mansfield et al., 2015). During the fever period, high levels of viremia occur in humans and animals, and viral RNA can be detected (Ikegami and Makino, 2011). Relevant studies have shown that neutralizing antibodies also begin to appear about the 4th day after the onset of RVF symptoms. IgG and IgM antibodies can be detected at least 6 days after natural infection in humans (Paweska et al., 2005). In a sheep model, IgG and IgM antibodies to RVFV can be detected as early as 4 days after experimental infection (Sobarzo et al., 2007; van Vuren et al., 2007).

For the detection of DNA (Ihira et al., 2004; Okamoto et al., 2004) and RNA (Hong et al., 2004; Mori et al., 2006; Yoda et al., 2007) viruses, differentiation of viral serotypes and subtypes (Parida et al., 2005; Nakagawa and Ito, 2006), and rapid diagnosis of bacterial infections (Iwamoto et al., 2003), the loop-mediated isothermal amplification (LAMP) method has been shown to be highly accurate and sensitive. This procedure requires primers targeting four or six regions of the RNA for amplification. LAMP amplifies target nucleic acid under isothermal conditions, usually between 60 and 65°C. It does not require specialized equipment, offering an inexpensive diagnostic method (Notomi et al., 2000; Le Roux et al., 2009).

This paper reports the development and diagnostic evaluation of an RVFV RT-LAMP with vertical flow visualization strip (RT-LAMP-VF) targeting the S RNA segment of RVFV. We compared the sensitivity of the established RVFV RT-LAMP-VF assay with RVFV real-time RT-PCR.

MATERIALS AND METHODS

Viruses, Synthetic RNA Samples and Mock Infected Blood Sample

The inactivated RVFV-BJ01 strain was gifted by professor Yuhai Bi (Chinese Academy of Sciences, CAS) from the Key Laboratory of Pathogenic Microbiology and Immunology. The target RNA gene was synthesized by Takara Biotechnology (Dalian, China) Co., Ltd., which was selected from a highly conserved region of 270 bp (from 1,336 nt to 1,605 nt) of the RVFV S segment (accession number DQ380149.1), and 70 bp bases were added before and after the gene to prevent RNA degradation.

The mock infected blood sample is made by mixing inactivated RVFV cell culture and fresh volunteer blood at a 1:1 volume ratio. The total RNA was extracted with the TIANamp Virus RNA Kit (TIANGEN, Beijing, China) from 140 µl mixture and eluted with 60 µl of RNase-free water. Above operation were performed in a biosafety level 2+ laboratory.

Design of Specific Primers for the RVFV RT-LAMP-VF Assay

To establish an RVFV RT-LAMP-VF assay, we designed primers targeting the viral S segment. For primer design, all sequences in the NCBI database (accession numbers DQ380149.1, DQ380156.1, DQ380158.1, KU167025.1, EU312118.1, DQ380176.1, DQ380167.1, DQ380174.1, DQ380172.1, DQ380173.1, DQ380163.1, EU312120.1, DQ380154.1, EU312119.1, DQ380164.1, DQ380144.1, EU312112.1, EU312114.1, EU312116.1, EU312122.1, DQ380177.1, EU312128.1, DQ380170.1, EU312144.1, and KX785330.1) of the full-length RVFV S segment were aligned using MEGA7.0 (**Figure 1**). Finally, we confirmed the S segment of ZH501 (accession number DQ380149.1) as the target gene and designed RVFV RT-LAMP-VF primers by selecting a highly conserved region of 270 bp from the S segment of ZH501 strain. Six primers (**Table 1**), including two inner primers (FIP and BIP), two outer primers (F3 and B3), and two loop primers (LF and LB), were designed using LAMP primer software Primer

TABLE 1 | Primer set for the RT-LAMP-VF assay.

Primer name	Primer position	Sequence (5′-3′)
F3	1351–1369	TCATCCCGGGAAGGATTCC
B3	1568–1586	CAGTGGGTCCGAGAGTTTG
FIP (F1c + F2)	1444–1463+	CGTGGCAACAAGCCCAGGAG
	1389–1408	GTTGATGAGAGCCTCCACAG
BIP (B1c + B2)	1492–1512+	GCATCCTTCTCCCAGTCAGCC
	1542–1560	AGGGTTTGATGCCCGTAGA
LF	1416–1440	FITC-GATGATGAAAATGTCGAAAGAAGGC
LB	1513–1536	Biotin-CCACCATACTGCTTTAAGAGTTCG

Explorer V5[1] (2009) and synthesized by Takara Biotechnology (Dalian, China) Co., Ltd. LF and LB primers were labeled by FITC and biotin, respectively, and these two labels can be combined with the corresponding antibodies on the vertical flow visualization strip.

RNA Extraction

The RVFV-BJ01 strain was cultured in Vero cells, and the cell culture was collected and inactivated by using beta-propiolactone (BPL; Sigma-Aldrich, St. Louis, Minnesota, United States). The total RNA was extracted with the TIANamp Virus RNA Kit (TIANGEN, Beijing, China) from 140 μl cell fluid and eluted with 60 μl of RNase-free water. All extraction work was performed in a biosafety level 2+ laboratory.

Pretreatment of Synthesized RNA Transcripts

The RVFV S segment target sequence (from 1,266 to 1,675 nt) (accession number DQ380149.1) was synthesized by Takara Biotechnology (Dalian, China) Co., Ltd. for *in vitro* transcription. First, linearize the plasmid template, use the Takara MiniBEST Agarose Gel DNA Extraction Kit Ver.4.0 to cut the gel to recover the above digestion products, use the Takara *in vitro* Transcription T7 Kit for *in vitro* transcription reaction, and perform the DNase treatment on the RNA transcripts, using Guide-it IVT RNA Clean-Up Kit for refining, DNase treatment again, refining again. After DNase treatment, the RNA transcripts were purified and quantified. The RNA transcripts were measured with a spectrophotometer, and the copy number was calculated using the following formula: copies/μl = $6.02 \times 10^{23} \times 10^{-9} \times$ concentration/(fragment length × 340). A Nanodrop 2000 ultraviolet visible spectrophotometer was used for quantitative detection of RNA. The concentration of the synthesized RNA transcript was 1,365 ng/μl, which corresponded to 1.94×10^{11} copies/μl. The synthesized RNA transcripts were stored at −80°C after purification.

RVFV RT-LAMP-VF Assay Reaction and Product Detection

Using synthetic RNA of RVFV as a standard template, the isothermal amplification reactions were performed at five different temperatures (59, 61, 63, 65 and 67°C) for 50 min with a 0.4 μM primer (FIP/BIP) concentration. The amplified products were analyzed with the vertical flow visualization strip (Ustar Biotech Co., Ltd., Hangzhou, China) to screen for the optimal temperature. Three replications were performed for each trial. The RNA in the reaction system was replaced with RNase-free water for negative control samples.

Based on the optimum reaction temperature, using the synthetic RNA of RVFV as a standard template, the isothermal amplification reactions were amplified for 20, 30, 40, 50, 60 and 70 min with a 0.4 μM primer (FIP/BIP) concentration, respectively, to obtain the optimal time. The amplified products were analyzed with the vertical flow visualization strip. Three replications were performed for each trial. The RNA in the reaction system was replaced with RNase-free water for negative control samples.

Under the conditions of optimal amplification temperature and time, using synthetic RNA of RVFV as a standard template, the concentration of FIP/BIP primers was set to 0.2, 0.4, and 0.6 μM (LF/LB and F3/B3 concentrations changed proportionally) for constant temperature amplification. The amplified products were analyzed with the vertical flow visualization strip, and the optimal amplification primer concentration was determined. Three replications were performed for each trial. The RNA in the reaction system was replaced with RNase-free water for the negative control samples.

Quantitative Detection of Inactivated RVFV-BJ01 Strain by Real-Time RT-PCR Assay

Synthetic RVFV-RNA was used as a standard template to establish a real-time RT-PCR method. The real-time RT-PCR primer sequences are shown in **Table 2**, and the 25 μl reaction mixtures comprised 5 μl of the synthetic RVFV-RNA, 12.5 μl of 2 × Onestep RT-PCR Buffer, 0.5 μl of TakaRa Ex TaqHS as shown in the instructions (Takara Biotechnology, Dalian, China), 0.5 μl of Primer Mix II, 1.0 μl of RVFV-F/R, 2.0 μl of RVFV-Probe, and 2.5 μl RNase-Free water. The optimized amplification conditions were 40 cycles of reverse transcription at 42°C for 5 min; heat denaturation at 95°C for 10 s; 95°C for 90 s, 95°C for 30 s, and 60°C for 45 s. The RNA content of inactivated RVFV cell culture (RVFV-BJ01 strain) was quantified by real-time RT-PCR. The RNA in the reaction system was replaced with RNase-free water as negative control.

RVFV RT-LAMP-VF Assay Sensitivity and Specificity Evaluation

The synthetic RVFV-RNA was used to assess the sensitivity of the RT-LAMP assay. The synthetic RVFV-RNA with a concentration of 1.94×10^{11} copies/μl was diluted 10-fold to 1.94×10^{-2} copies/μl. The template was amplified under the optimal reaction conditions to assess the detection limit of the RVFV RT-LAMP assay.

The RNA, which was extracted from inactivated RVFV cell culture was used to assess the sensitivity of the RVFV

[1]http://primerexplorer.jp/lampv5e/index.html

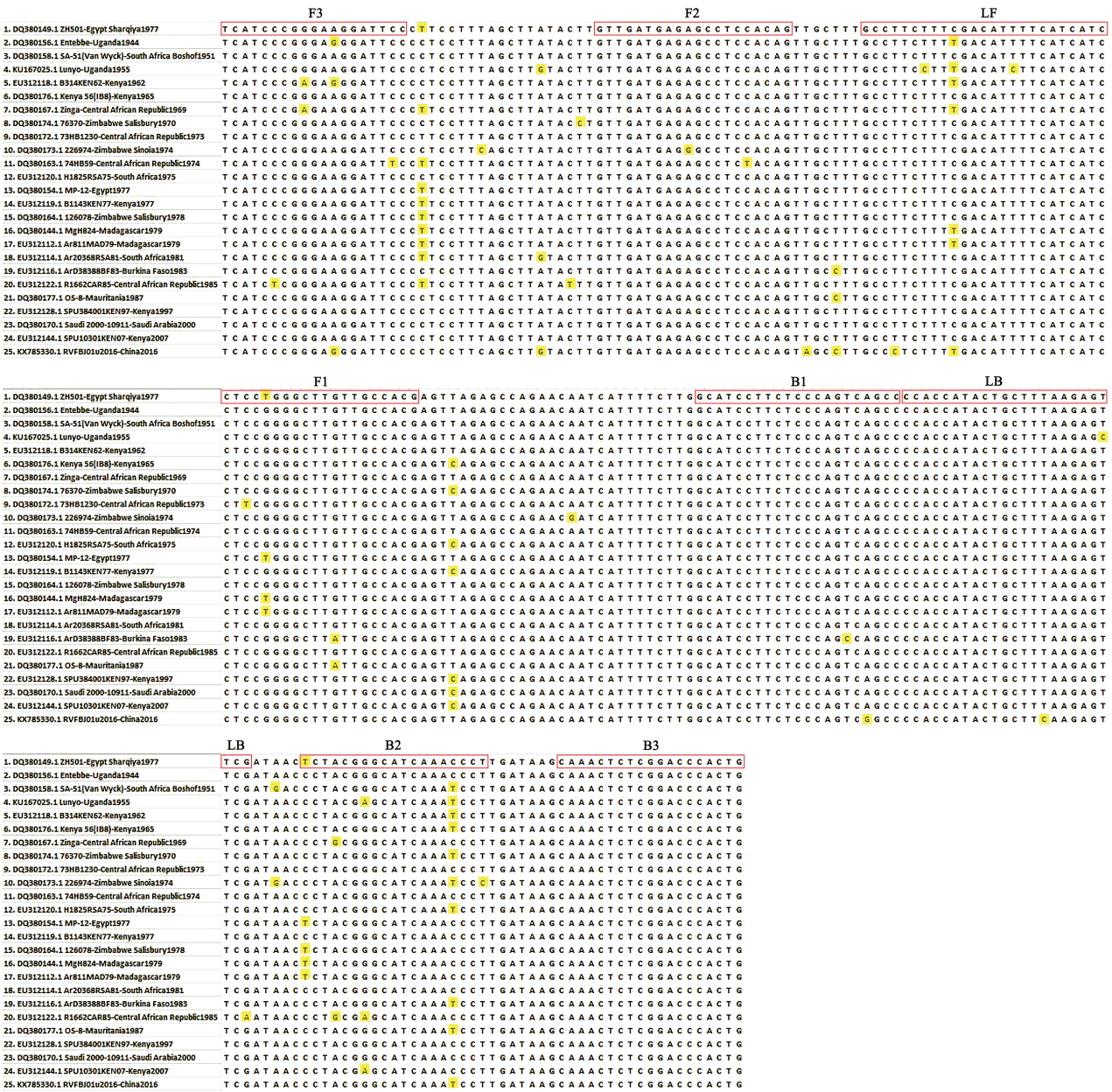


FIGURE 1 | Primer position of the RVFV RT-LAMP-VF assay. The RVFV S segment was retrieved from GenBank and aligned using MEGA 7.0 software.

RT-LAMP-VF assay. The RNA content from inactivated RVFV cell culture with a concentration of 1.83×10^7 copies/μl was diluted 10-fold to 1.83×10^1 copies/μl. The template was amplified under the optimal reaction conditions to assess the detection limit of the RVFV RT-LAMP-VF assay.

The synthesized RVFV-RNA and total RNA extracted from inactivated RVFV cell culture (RVFV-BJ01 strain), inactivated JEV, inactivated H3N2 influenza virus, and recombinant viruses rNDV-EBOV-GP and rSRV9-MGP (expressing MARV G protein) were amplified under the optimal conditions to evaluate the specificity of the RVFV RT-LAMP-VF assay.

Above, the RNase-free water was used to replace the target RNA in the system for negative control samples.

Analytical Sensitivities of RT-LAMP-VF Assay and Real-Time RT-PCR

The diluted synthesized RVFV-RNA was used as the template. According to the reaction conditions shown above, we compared the analytical sensitivities of the RT-LAMP-VF and real-time RT-PCR with respect to the detection of decreasing numbers of RNA copies under the optimal amplification conditions.

TABLE 2 | Primer set for the RVFV real-time RT-PCR assay.

Primer name	Primer position	Sequence (5′-3′)
RVFV-F	1335–1352	TCGTGATAGAGTCAACTC
RVFV-R	1478–1496	GATGCCAAGAAAATGATTG
RVFV-Probe	1454–1474	FAM-TGGCTCTAACTCGTGGCAACA-TAMRA

Three replications were performed for each trial. The RNA in the reaction system was replaced with RNase-free water for negative control samples.

Using RVFV Nucleic Acids to Evaluate the RT-LAMP-VF Assay

The RNA extracted from inactivated RVFV cell culture was present at 1.83×10^7 copies/μl as tested by real-time RT-PCR. Inactivated RVFV cell culture was mixed with fresh volunteer blood at a 1:1 volume ratio as a mock infected blood sample, and then the total RNA in the mock infected blood sample was extracted by the QIAamp viral RNA minikit. These samples were amplified under the RVFV RT-LAMP-VF optimal amplification conditions, and three replications were performed for each trial. The RNA extracted from the blood sample was used as the template for the negative control. RNase-free water was used to replace the synthetic RNA for blank control samples.

RESULTS

Optimization of Reaction Conditions for the RVFV RT-LAMP-VF Assay

Synthetic RVFV-RNA samples were used as the template to optimize the amplification conditions for the RVFV RT-LAMP-VF assay. To screen the optimal amplification temperature, we amplified the standard templates at different temperatures (57, 59, 61, 63, 65, and 67°C) for 50 min. The results revealed that the optimum amplification temperature was achieved at 63°C. Analysis of the results revealed no significant differences in three replications. The lowest standard template copy value, 0.935×10^2 copies/μl, was detected at 63°C, and so 63°C was considered as the optimum amplification temperature for the RVFV RT-LAMP-VF assay.

To determine the optimum amplification time for the RVFV RT-LAMP-VF assay, synthetic RVFV-RNA samples were amplified for different durations (20, 30, 40, 50, 60 and 70 min) at 63°C, respectively. The results displayed were observed by using the vertical flow visualization strip. The lowest standard template copy value, 1.94×10^0 copies/μl, can be detected at both 60 and 70 min amplification. Therefore, 60 min was selected as the optimal amplification time for the RVFV RT-LAMP-VF assay.

To determine the optimal primer concentration in the amplification, the FIP/BIP concentration was set as 0.2, 0.4, and 0.6 μM (LF/LB and F3/B3 concentrations changed proportionally) to amplify at 63°C for 60 min. As the results indicate, the same results were obtained at 0.4 and 0.6 μM primer concentrations under equivalent conditions. Thus, 0.4 μM for FIP/BIP, 0.2 μM for LF/LB and 0.1 μM for F3/B3 were selected as the optimal primer concentrations for the RVFV RT-LAMP-VF assay.

In summary, the RVFV RT-LAMP-VF assay was established by amplifying for 60 min at 63°C with a FIP/BIP concentration of 0.4 μM.

Quantitative Analysis of Inactivated RVFV-BJ01 Strain by Real-Time RT-PCR

The 10-fold diluted synthesized RVFV-RNA was used as the template to establish the standard curve of the RVFV real-time RT-PCR assay. Then, the RNA content of inactivated RVFV-BJ01 was quantified. As the results show in **Figure 2**, the CT value of inactivated RVFV-BJ01 was 16.84. According to the standard curve $Y = -3.420*\log(x) + 41.68$, the RNA content of inactivated RVFV-BJ01 was 1.83×10^7 copies/μl.

Sensitivity and Specificity of the RT-LAMP-VF Assay

Ten-fold dilutions of synthesized RNA (ranging from 1.94×10^4 to 1.94×10^{-2} copies/μl) were used to evaluate the sensitivity of the RVFV RT-LAMP-VF assay. As **Figure 3A** shows, the assay limit of detection for synthesized RNA was 1.94×10^0 copies/μl within 60 min.

Ten-fold dilutions of the RNA extracted from inactivated RVFV cell culture (ranging from 1.83×10^7 to 1.83×10^1 copies/μl) were used to evaluate the sensitivity of the RVFV RT-LAMP-VF assay. As **Figure 3B** shows, the assay limit of detection for inactivated RVFV-BJ01 RNA was 1.83×10^3 copies/μl within 60 min.

Subsequently, the RNAs from viruses with similar clinical symptoms to RVFV and other laboratory stock viruses were used to evaluate the specificity of the RVFV RT-LAMP-VF assay. As seen from the results shown in **Figure 3C**, the result is positive only when the synthetic RVFV-RNA and inactivated RVFV-BJ01 RNA were used as templates. Consequently, the RVFV RT-LAMP-VF assay had good specificity.

Compared Sensitivity of RVFV RT-LAMP-VF Assay and Real-Time RT-PCR

To compare the sensitivity of the RVFV RT-LAMP-VF assay and real-time RT-PCR, the synthesized RNA standard was used as the template, which was diluted 10 times to 1.94×10^{-2} copies/μl. The limit of detection by RVFV RT-LAMP-VF was 1.94×10^0 copies/μl (**Figure 3A**), and the limit of detection by real-time RT-PCR was 1.94×10^2 copies/μl (**Figure 4**). These data demonstrated that the RVFV RT-LAMP-VF assay was 100-fold more sensitive than the real-time RT-PCR assay.

Diagnostic Evaluation of the RVFV RT-LAMP-VF Assay

After mixing the inactivated RVFV cell culture with fresh blood of a volunteer, total RNA in the mixture was extracted by using a viral RNA extraction kit. The total RNA was amplified at 63°C with 0.4 μM primer concentration for 60 min.

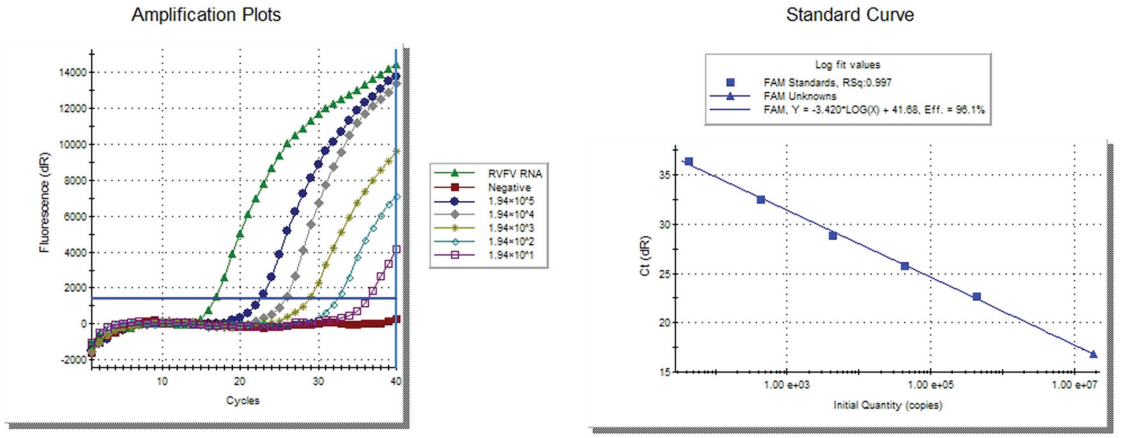


FIGURE 2 | Quantitative detection of inactivated RVFV-BJ01 strain by real-time RT-PCR assay. The left side is the amplification curve and the right side is the standard curve. Dilute the synthesized RNA standard from 1.94 × 10^5 copies/μl to 1.94 × 10^1 copies/μl as templates for quantitative detection of RVFV RNA.

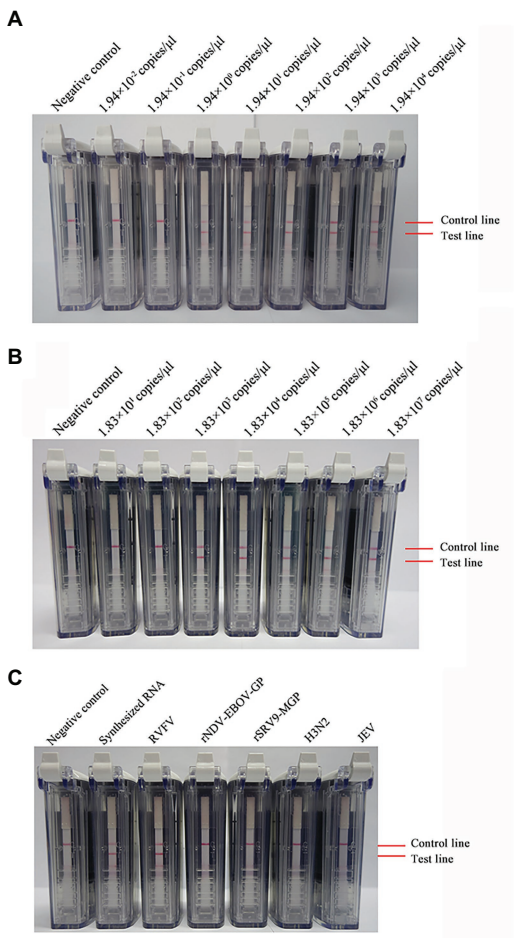


FIGURE 3 | Detection of the sensitivity and specificity of the RVFV RT-LAMP-VF assay. Sensitivity evaluation of the RVFV RT-LAMP-VF assay by using a series of synthesized RVFV-RNA **(A)** and RNA extracted from inactivated RVFV-BJ01 strain **(B)**. Specificity evaluation of the RVFV RT-LAMP-VF assay **(C)**.

As **Figure 5** shows, RNA in the mixed sample can be detected. It is preliminarily indicated that the RVFV RT-LAMP-VF assay can be applied to clinical sample detection.

DISCUSSION

RVFV is an important zoonotic disease and poses a potential bioterrorism threat. Many kinds of mosquito vectors could transmit RVFV, and the effects of global climate change could facilitate the spread of arthropod-borne viruses into nonendemic countries. The potential for further spread of RVFV outside its traditional geographic boundaries has resulted in increased international demand for validated molecular tools for the rapid and accurate diagnosis of RVFV.

LAMP is a novel nucleic acid detection method invented in 2000 (Notomi et al., 2000). This method relies on strand displacement DNA polymerase with high strand displacement activity and a set of two specially designed inner and two outer primers. First, the inner primers initiates LAMP then the target sequence is amplified to form a new single-strand of DNA. Under the action of strand displacement DNA polymerase, the outer primer releases a single-stranded of DNA. The released single strand and the special structure of the inner primer will fold to form cauliflower-like structures. After a series of amplification, a product with multiple ring structures will be formed (Cao et al., 2016; Huang et al., 2018). For the RVFV RT-LAMP-VF assay, in order to be combined with the vertical flow visualization strip, FITC and biotin were labeled at the 5' ends of the loop primers LF and LB, respectively. In addition, given that this experiment uses RNA as templates, so AMV Reverse Transcriptase (Promega) was added to the reaction system. Because LAMP recognizes the target by six distinct sequences, it is expected to amplify the target sequence with high specificity.

The vertical flow visualization test was used to read the results, using the principle of double antibody sandwich.

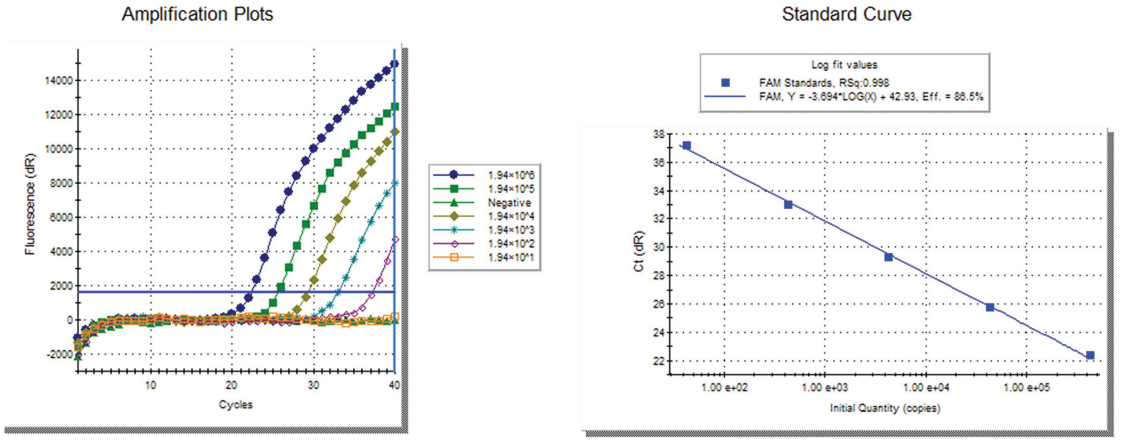


FIGURE 4 | Compared sensitivity of RVFV RT-LAMP-VF assay and RVFV real-time RT-PCR. Ten-fold diluted the synthesized RNA standard as templates for RVFV real-time RT-PCR. The left side is the amplification curve and the right side is the standard curve.

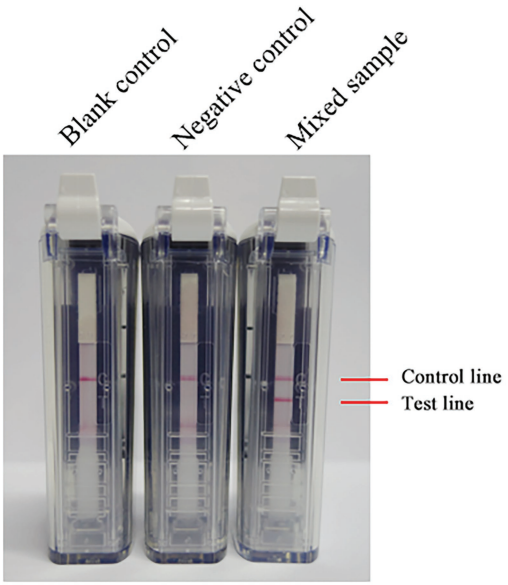


FIGURE 5 | The RNA of inactivated RVFV cell culture (RVFV-BJ01 strain) was used to evaluate the RVFV RT-LAMP-VF assay. Inactivated RVFV cell culture was mixed with fresh volunteer blood at a 1:1 volume ratio as a mock infected blood sample to simulate clinical samples.

It consisted of test strips and diluent. The quality control line and the detection line of the strip were marked with anti-streptavidin antibody and anti-FITC antibody, respectively. At the same time, gold particles covered with streptavidin were incubated on the binding pad of the strip. The primers LF and LB were covered with FITC and biotin, respectively. After the amplified product was added to the device, the gold particles would be combined with the biotin on the product through streptavidin under the action of the diluent and migrated upward to the detection line under the siphon action of the strip. The FITC on the product would bind to the anti-FITC antibody on the detection line, and a large number of gold particles would aggregate to form a visible red detection line.

RVFV RT-LAMP has been established, but it still has some technical shortcomings, such as false positive problems and required special equipment (turbidity meter). When reading the results, the agarose gel electrophoresis method is mostly used, presenting the problem of false positives (Peyrefitte et al., 2008; Le Roux et al., 2009). In addition, there are subjective differences when reading the results of precipitation and staining methods. In this study, the RVFV RT-LAMP-VF technique for rapid and accurate detection of RVFV RNA was investigated. In addition to the high levels of analytical and diagnostic accuracy and speed of detection, another important practical advantage of the RVFV RT-LAMP-VF assay is that it uses simple and relatively inexpensive equipment, which renders it promising for use in resource-poor settings. In addition, only basic molecular and technical skills are required for the assay procedure, and the results could be identified by the naked eye. The primer design for the RVFV RT-LAMP-VF assay is more complex than for the conventional PCR-based assays and, in this study, we designed the RVFV RT-LAMP-VF primers to target the S segment of RVFV, which encodes N protein and is highly conserved among RVFV strains. In **Figure 1**, we can see that there are unmatched nucleotides (highlighted part) between different RVFV strains, which are also present in the primer sequence. The influence of these differences on the sensitivity of the assay may exist, and the specific degree of influence remains to be demonstrated. However, in order to avoid the difference of nucleotides from affecting the detection sensitivity as much as possible, when screening conservative sequences in the early stage, we selected strains from different species, different regions, and different ages for comparison.

The synthesized RVFV-RNA was used as the template to determine the optimal amplifications, including temperature and time and the concentration of primers. Among the results demonstrated in this paper, 63°C was considered as the optimum

amplification temperature and 60 min was selected as the optimal amplification time; 0.4 μM FIP/BIP, 0.2 μM LF/LB, and 0.1 μM F3/B3 were selected as the optimal primer concentrations.

In this study, we compared the sensitivity of the RVFV RT-LAMP-RV assay and RVFV real-time RT-PCR assay. The results of the study showed that the RVFV RT-LAMP-VF assay (1.94×10^0 copies/μl) was 100-fold more sensitive than the RVFV real-time RT-PCR assay (1.94×10^2 copies/μl). RVFV RT-LAMP-VF was subjected to specific evaluation by detecting viruses with similar clinical symptoms (like JEV, H3N2 influenza virus, EBOV, and MARV). The results indicated no cross-reaction with inactivated JEV, inactivated H3N2 influenza virus, and recombinant virus rNDV-EBOV-GP and rSRV9-MGP.

Because the collected clinical RVFV samples were mostly blood, we mixed volunteer blood with inactivated RVFV cell culture at a 1:1 volume ratio to simulate a clinical sample, which was subjected to the RVFV RT-LAMP-VF assay. The results indicated that the sensitivity and specificity of RVFV RT-LAMP-VF were not interfered with by other components in the blood sample.

RT-LAMP is an inexpensive and sensitive detection method for viral RNA and is suitable for field surveys, where specialized equipment is often unavailable. RT-LAMP assays for Ebola virus (Kurosaki et al., 2007), Zika virus (Kurosaki et al., 2017), and Chikungunya virus (Lopez-Jimena et al., 2018), for which diagnostics are necessary in remote areas, have been developed. Under the ravages of SARS-CoV-2, LAMP-based detection methods have also been developed (Yan et al., 2020). Large-scale epidemiological studies are important to predict virus epidemics and transmission routes, thereby facilitating the development of countermeasures against viruses.

One specific conclusion is that the established RVFV RT-LAMP-VF assay has a detection limit of 1.94×10^0 copies/μl RNA transcripts and 1.83×10^3 copies/μl viral RNA. This method offers the advantages of high sensitivity, strong specificity, visual reading, and easy operation. Hence, the RVFV RT-LAMP-VF assay has the potential for rapid detection on point-care-testing. Since the entire amplification process is carried out under constant temperature conditions, only a simple water bath or metal bath is needed. In addition, after the reaction solution is added to the PCR tube, there is no need to open the cap to read the result, which avoids false positives caused by aerosol contamination, and the vertical flow visualization strip has the advantages of simple operation and easy portability. This makes this method highly practical, especially for insufficient equipment areas. Based on the above experimental results, from the perspective of epidemiological investigation, epidemic prevention, and control and technology reserve of Rift Valley fever, the establishment of the RVFV RT-LAMP-VF assay is of great significance. With the development of technology, a digital diagnostic method with rapid execution, high accuracy, sensitivity and specificity, low cost, and suitability for on-site diagnosis will become a trend.

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

The studies involving human participants were reviewed and approved by Use Committee of the Chinese People's Liberation Army. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YZ, YG, and XX designed the experiments. QH, SZ, PH, FY, JB, ZC, and HC performed the experimentation. QH, DL, HW, NF, and HJ analyzed the data. QH and SZ wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Prevention and Control of Transboundary Animal Diseases (Grant No. 2017YFD0501804) and the Key R & D Program of Jiangsu Province (Social Development) Project (BE2019625).

ACKNOWLEDGMENTS

We thank Dr. Yuhai Bi (CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology) for providing inactivated RVFV-BJ01.

REFERENCES

Cao, Z., Wang, H., Wang, L., Li, L., Jin, H., Xu, C., et al. (2016). Visual detection of West Nile virus using reverse transcription loop-mediated isothermal amplification combined with a vertical flow visualization strip. *Front. Microbiol.* 7:554. doi: 10.3389/fmicb.2016.00554

Cetre-Sossah, C., Pedarrieu, A., Juremalm, M., van Vuren, P. J., Brun, A., Ould El Mamy, A. B., et al. (2019). Development and validation of a pen side test for Rift Valley fever. *PLoS Negl. Trop. Dis.* 13:e0007700. doi: 10.1371/journal.pntd.0007700

Crespo Leon, F., Gutierrez Diez, F., Rodriguez Ferri, F., Leon Vizcaino, L., Cuello Gijon, F., Gimeno, E. J., et al. (2005). The translation into Spanish of the OIE manual of diagnostic tests and vaccines for terrestrial animals (mammals, birds and bees): problems, solutions and conclusions. *Rev. Sci. Tech.* 24, 1095–1104. doi: 10.20506/rst.24.3.1639

Daubney, R., Hudson, J. R., and Garnham, P. C. (1931). Enzootic hepatitis or rift valley fever. An undescribed virus. Disease of sheep cattle and man from East Africa. *J. Pathol. Bacteriol.* 34, 545–579. doi: 10.1002/path.1700340418

Drosten, C., Gottig, S., Schilling, S., Asper, M., Panning, M., Schmitz, H., et al. (2002). Rapid detection and quantification of RNA of Ebola and

Marburg viruses, Lassa virus, Crimean-Congo hemorrhagic fever virus, Rift Valley fever virus, dengue virus, and yellow fever virus by real-time reverse transcription-PCR. *J. Clin. Microbiol.* 40, 2323–2330. doi: 10.1128/JCM.40.7.2323-2330.2002

Elliott, R. M. (1997). Emerging viruses: the Bunyaviridae. *Mol. Med.* 3, 572–577. doi: 10.1007/BF03401814

Euler, M., Wang, Y., Nentwich, O., Piepenburg, O., Hufert, F. T., and Weidmann, M. (2012). Recombinase polymerase amplification assay for rapid detection of Rift Valley fever virus. *J. Clin. Virol.* 54, 308–312. doi: 10.1016/j.jcv.2012.05.006

Garcia, S., Crance, J. M., Billecocq, A., Peinnequin, A., Jouan, A., Bouloy, M., et al. (2001). Quantitative real-time PCR detection of Rift Valley fever virus and its application to evaluation of antiviral compounds. *J. Clin. Microbiol.* 39, 4456–4461. doi: 10.1128/JCM.39.12.4456-4461.2001

Hong, T. C., Mai, Q. L., Cuong, D. V., Parida, M., Minekawa, H., Notomi, T., et al. (2004). Development and evaluation of a novel loop-mediated isothermal amplification method for rapid detection of severe acute respiratory syndrome coronavirus. *J. Clin. Microbiol.* 42, 1956–1961. doi: 10.1128/jcm.42.5.1956-1961.2004

Huang, P., Wang, H., Cao, Z., Jin, H., Chi, H., Zhao, J., et al. (2018). A rapid and specific assay for the detection of MERS-CoV. *Front. Microbiol.* 9:1101. doi: 10.3389/fmicb.2018.01101

Ihira, M., Yoshikawa, T., Enomoto, Y., Akimoto, S., Ohashi, M., Suga, S., et al. (2004). Rapid diagnosis of human herpesvirus 6 infection by a novel DNA amplification method, loop-mediated isothermal amplification. *J. Clin. Microbiol.* 42, 140–145. doi: 10.1128/JCM.42.1.140-145.2004

Ikegami, T., and Makino, S. (2011). The pathogenesis of Rift Valley fever. *Viruses* 3, 493–519. doi: 10.3390/v3050493

Iwamoto, T., Sonobe, T., and Hayashi, K. (2003). Loop-mediated isothermal amplification for direct detection of *Mycobacterium tuberculosis* complex, M. avium, and M. intracellulare in sputum samples. *J. Clin. Microbiol.* 41, 2616–2622. doi: 10.1128/JCM.41.6.2616-2622.2003

Kurosaki, Y., Martins, D. B. G., Kimura, M., Catena, A. D. S., Borba, M., Mattos, S. D. S., et al. (2017). Development and evaluation of a rapid molecular diagnostic test for Zika virus infection by reverse transcription loop-mediated isothermal amplification. *Sci. Rep.* 7:13503. doi: 10.1038/s41598-017-13836-9

Kurosaki, Y., Takada, A., Ebihara, H., Grolla, A., Kamo, N., Feldmann, H., et al. (2007). Rapid and simple detection of Ebola virus by reverse transcription-loop-mediated isothermal amplification. *J. Virol. Methods* 141, 78–83. doi: 10.1016/j.jviromet.2006.11.031

Le Roux, C. A., Kubo, T., Grobbelaar, A. A., van Vuren, P. J., Weyer, J., Nel, L. H., et al. (2009). Development and evaluation of a real-time reverse transcription-loop-mediated isothermal amplification assay for rapid detection of Rift Valley fever virus in clinical specimens. *J. Clin. Microbiol.* 47, 645–651. doi: 10.1128/JCM.01412-08

Li, M., Wang, B., Li, L., Wong, G., Liu, Y., Ma, J., et al. (2019). Rift Valley fever virus and yellow fever virus in urine: a potential source of infection. *Virol. Sin.* 34, 342–345. doi: 10.1007/s12250-019-00096-2

Lopez-Jimena, B., Wehner, S., Harold, G., Bakheit, M., Frischmann, S., Bekaert, M., et al. (2018). Development of a single-tube one-step RT-LAMP assay to detect the Chikungunya virus genome. *PLoS Negl. Trop. Dis.* 12:e0006448. doi: 10.1371/journal.pntd.0006448

Madani, T. A., Al-Mazrou, Y. Y., Al-Jeffri, M. H., Mishkhas, A. A., Al-Rabeah, A. M., Turkistani, A. M., et al. (2003). Rift Valley fever epidemic in Saudi Arabia: epidemiological, clinical, and laboratory characteristics. *Clin. Infect. Dis.* 37, 1084–1092. doi: 10.1086/378747

Mansfield, K. L., Banyard, A. C., McElhinney, L., Johnson, N., Horton, D. L., Hernandez-Triana, L. M., et al. (2015). Rift Valley fever virus: a review of diagnosis and vaccination, and implications for emergence in Europe. *Vaccine* 33, 5520–5531. doi: 10.1016/j.vaccine.2015.08.020

Mori, N., Motegi, Y., Shimamura, Y., Ezaki, T., Natsumeda, T., Yonekawa, T., et al. (2006). Development of a new method for diagnosis of Rubella virus infection by reverse transcription-loop-mediated isothermal amplification. *J. Clin. Microbiol.* 44, 3268–3273. doi: 10.1128/JCM.00803-06

Nakagawa, N., and Ito, M. (2006). Rapid subtyping of influenza a virus by loop-mediated isothermal amplification: two cases of influenza patients who returned from Thailand. *Jpn. J. Infect. Dis.* 59, 200–201.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., et al. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 28:E63. doi: 10.1093/nar/28.12.e63

Okamoto, S., Yoshikawa, T., Ihira, M., Suzuki, K., Shimokata, K., Nishiyama, Y., et al. (2004). Rapid detection of varicella-zoster virus infection by a loop-mediated isothermal amplification method. *J. Med. Virol.* 74, 677–682. doi: 10.1002/jmv.20223

Parida, M., Horioke, K., Ishida, H., Dash, P. K., Saxena, P., Jana, A. M., et al. (2005). Rapid detection and differentiation of dengue virus serotypes by a real-time reverse transcription-loop-mediated isothermal amplification assay. *J. Clin. Microbiol.* 43, 2895–2903. doi: 10.1128/JCM.43.6.2895-2903.2005

Paweska, J. T., Burt, F. J., Anthony, F., Smith, S. J., Grobbelaar, A. A., Croft, J. E., et al. (2003). IgG-sandwich and IgM-capture enzyme-linked immunosorbent assay for the detection of antibody to Rift Valley fever virus in domestic ruminants. *J. Virol. Methods* 113, 103–112. doi: 10.1016/S0166-0934(03)00228-3

Paweska, J. T., Burt, F. J., and Swanepoel, R. (2005). Validation of IgG-sandwich and IgM-capture ELISA for the detection of antibody to Rift Valley fever virus in humans. *J. Virol. Methods* 124, 173–181. doi: 10.1016/j.jviromet.2004.11.020

Paweska, J. T., van Vuren, P. J., Kemp, A., Buss, P., Bengis, R. G., Gakuya, F., et al. (2008). Recombinant nucleocapsid-based ELISA for detection of IgG antibody to Rift Valley fever virus in African buffalo. *Vet. Microbiol.* 127, 21–28. doi: 10.1016/j.vetmic.2007.07.031

Paweska, J. T., van Vuren, P. J., and Swanepoel, R. (2007). Validation of an indirect ELISA based on a recombinant nucleocapsid protein of Rift Valley fever virus for the detection of IgG antibody in humans. *J. Virol. Methods* 146, 119–124. doi: 10.1016/j.jviromet.2007.06.006

Peyrefitte, C. N., Boubis, L., Coudrier, D., Bouloy, M., Grandadam, M., Tolou, H. J., et al. (2008). Real-time reverse-transcription loop-mediated isothermal amplification for rapid detection of Rift Valley fever virus. *J. Clin. Microbiol.* 46, 3653–3659. doi: 10.1128/JCM.01188-08

Ragan, I. K., Davis, A. S., McVey, D. S., Richt, J. A., Rowland, R. R., and Wilson, W. C. (2018). Evaluation of fluorescence microsphere immunoassay for detection of antibodies to Rift Valley fever virus nucleocapsid protein and glycoproteins. *J. Clin. Microbiol.* 56, e01626–e01717. doi: 10.1128/JCM.01626-17

Rolin, A. I., Berrang-Ford, L., and Kulkarni, M. A. (2013). The risk of Rift Valley fever virus introduction and establishment in the United States and European Union. *Emerg. Microbes Infect.* 2:e81. doi: 10.1038/emi.2013.81

Sall, A. A., Macondo, E. A., Sene, O. K., Diagne, M., Sylla, R., Mondo, M., et al. (2002). Use of reverse transcriptase PCR in early diagnosis of Rift Valley fever. *Clin. Diagn. Lab. Immunol.* 9, 713–715. doi: 10.1128/cdli.9.3.713-715.2002

Sobarzo, A., Paweska, J. T., Herrmann, S., Amir, T., Marks, R. S., and Lobel, L. (2007). Optical fiber immunosensor for the detection of IgG antibody to Rift Valley fever virus in humans. *J. Virol. Methods* 146, 327–334. doi: 10.1016/j.jviromet.2007.07.017

Tercero, B., Terasaki, K., Nakagawa, K., Narayanan, K., and Makino, S. (2019). A strand-specific real-time quantitative RT-PCR assay for distinguishing the genomic and antigenomic RNAs of Rift Valley fever *phlebovirus*. *J. Virol. Methods* 272:113701. doi: 10.1016/j.jviromet.2019.113701

Upreti, D., Cernicchiaro, N., Richt, J. A., Wilson, W. C., Clavijo, A., and Davis, A. S. (2018). Preliminary evaluation of diagnostic accuracy and precision of a competitive ELISA for detection of antibodies to Rift Valley fever virus in cattle and sheep sera. *J. Virol. Methods* 262, 6–11. doi: 10.1016/j.jviromet.2018.09.002

van Vuren, P. J., Potgieter, A. C., Paweska, J. T., and van Dijk, A. A. (2007). Preparation and evaluation of a recombinant Rift Valley fever virus N protein for the detection of IgG and IgM antibodies in humans and animals by indirect ELISA. *J. Virol. Methods* 140, 106–114. doi: 10.1016/j.jviromet.2006.11.005

Wichgers Schreur, P. J., Paweska, J. T., Kant, J., and Kortekaas, J. (2017). A novel highly sensitive, rapid and safe Rift Valley fever virus neutralization test. *J. Virol. Methods* 248, 26–30. doi: 10.1016/j.jviromet.2017.06.001

Yan, C., Cui, J., Huang, L., Du, B., Chen, L., Xue, G., et al. (2020). Rapid and visual detection of 2019 novel coronavirus (SARS-CoV-2) by a reverse transcription loop-mediated isothermal amplification assay. *Clin. Microbiol. Infect.* 26, 773–779. doi: 10.1016/j.cmi.2020.04.001

Yoda, T., Suzuki, Y., Yamazaki, K., Sakon, N., Kanki, M., Aoyama, I., et al. (2007). Evaluation and application of reverse transcription loop-mediated isothermal amplification for detection of noroviruses. *J. Med. Virol.* 79, 326–334. doi: 10.1002/jmv.20802

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 December 2020
doi: 10.3389/fmicb.2020.604335

The V617I Substitution in Avian Coronavirus IBV Spike Protein Plays a Crucial Role in Adaptation to Primary Chicken Kidney Cells

*Yi Jiang[1,2,3], Mingyan Gao[1,2], Xu Cheng[1,2], Yan Yu[1,2], Xinyue Shen[1,2,3], Jianmei Li[1,2] and Sheng Zhou[1,2]**

[1] Poultry Institute, Chinese Academy of Agricultural Sciences, Yangzhou, China, [2] Jiangsu Institute of Poultry Science (CAAS), Yangzhou, China, [3] Laboratory of Animal Infectious Disease, College of Veterinary Medicine, Yangzhou University, Yangzhou, China

The naturally isolated avian coronavirus infectious bronchitis virus (IBV) generally cannot replicate in chicken kidney (CK) cells. To explore the molecular mechanism of IBV adapting to CK cells, a series of recombinant viruses were constructed by chimerizing the S genes of CK cell-adapted strain H120 and non-adapted strain IBYZ. The results showed that the S2 subunit determines the difference in cell tropism of the two strains. After comparing the amino acid sequences of S protein of CK cell-adapted strain YZ120, with its parental strain IBYZ, three amino acid substitutions, A138V, L581F, and V617I, were identified. Using YZ120 as the backbone, one or more of the above-mentioned substitutions were eliminated to verify the correlation between these sites and CK cell tropism. The results showed that the CK cell tropism of the YZ120 strain depends on the V617I substitution, the change of L581F promoted the adaptation in CK cells, and the change at 138 position was not directly related to the CK cell tropism. Further validation experiments also showed that V617I had a decisive role in the adaptation of IBV to CK cells, but other areas of the virus genome also affected the replication efficiency of the virus in CK cells.

Keywords: infectious bronchitis virus, CK cells, tropism, amino acid mutation, spike

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Vicky Lin van Santen,
Auburn University, United States
Xin Yang,
Sichuan University, China
Abdeljelil Ghram,
Pasteur Institute of Tunis, Tunisia

***Correspondence:**
Sheng Zhou
dragonsheng@163.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *09 September 2020*
Accepted: *27 November 2020*
Published: *18 December 2020*

Citation:
Jiang Y, Gao M, Cheng X, Yu Y, Shen X, Li J and Zhou S (2020) The V617I Substitution in Avian Coronavirus IBV Spike Protein Plays a Crucial Role in Adaptation to Primary Chicken Kidney Cells. Front. Microbiol. 11:604335. doi: 10.3389/fmicb.2020.604335

INTRODUCTION

Avian coronavirus IBV is a positive-sense RNA enveloped virus, which belongs to the order Nidovirales, family Coronaviridae, genus Gammacoronavirus (Cavanagh, 2007). Like other coronaviruses (CoVs), IBV consists of basic structural proteins, including spike (S), membrane (M), nucleocapsid (N), envelope (E) protein, and the genome encodes other accessory and non-structural proteins. The CoV spike protein is a class I fusion protein (Bosch et al., 2003), the ectodomain consists of S1 and S2 domains, and it plays a major role in the process of viral infection (Cavanagh, 1983; Belouzard et al., 2012). The S1 subunit contains the receptor-binding domain (RBD) located on the N-terminal domain (NTD) or C-terminal domain (CTD), which specifically binds the host receptors, such as a variety of proteins and sugars (Belouzard et al., 2012; Li, 2012). The S2 subunit, responsible for virus-cell and cell-cell fusion, contains fusion peptides (FP), two

heptad repeat HR1 and HR2, transmembrane span (TM), and cytoplasmic tail (CT) (Heald-Sargent and Gallagher, 2012).

Although CoVs primarily infect the respiratory and gastrointestinal tracts of a wide range of animal species, such as bats, civet cats, camels, humans, swine, chickens, turkeys, etc. (Hulswit et al., 2016), different CoVs or different serotypes of strains from the same CoV have a restricted tissue and cell tropism, which is largely determined by the spike protein (Graham and Baric, 2010; Belouzard et al., 2012). CoVs enter the cells via the plasma membrane pathway or endosomal pathway (Millet and Whittaker, 2015). Substitution of S protein may lead to changes in the cell or tissue tropism of CoVs (Kuo et al., 2000). During the infection process, the CoV spike protein can be activated and cleaved into S1 and S2 subunits at different cleavage sites by various host proteases (Belouzard et al., 2012; Xia et al., 2018). The S1 subunit is involved in the host receptor recognition, whereas the S2 subunit anchored in the virus membrane is important for membrane fusion. Only a few amino acid alterations in the RBD of the S1 subunit can change the host species tropism of the CoV (Krempl et al., 1997; Li et al., 2005; Qu et al., 2005). Four amino acid substitutions in the S2 subunit of MHV-A59 can extend the host range in the non-permissive mammalian cell types of this virus, which indicates that the S2 subunit also plays an essential role in the viral cell and tissue tropism (Baric et al., 1999; McRoy and Baric, 2008). In addition, changes in the proteolytic cleavage sites in the S protein, located at the S1/S2 boundary or immediately upstream of the FP, are associated with altered tropism. The introduction of two single amino acid substitutions (S746R and N762A) at the S1/S2 boundary of the S protein of the bat coronavirus HKU4 were found to be crucial for the adaptation to human cells (Yang et al., 2015). In addition, PEDV (porcine epidemic diarrhea virus) acquires the ability of replication in non-enteric tissues by a single-amino acid substitution at the S2' cleavage site (Li, 2015).

Infectious bronchitis virus shows strict cell and tissue tropism. Almost all IBV strains cannot infect mammalian cell lines, with the exception of the Beaudette strain, which was generated by several hundred passages in embryonated eggs (Bickerton et al., 2018b). A large number of newly isolated strains from diseased animals do not have the ability to replicate in primary chick kidney (CK) cells and need to be passaged in embryos for adaptation (Fang et al., 2005). The spike protein has been demonstrated to be a determinant of cell tropism by a recombinant strain BeauR-M41(S), which is based on the genome backbone of Beaudette strain and with the replacement of the ectodomain of the S protein with that of M41 strain. The rescued strain acquired the cell tropism of M41, and lost the ability to replicate in Vero, BHK-21, and CEF cells (Casais et al., 2003). In the infection process, the IBV S protein is cleaved into S1 and S2 subunits at the S1/S2 cleavage site. Sialic acid is recognized as the receptor bound by the IBV S1 subunits (Schultze et al., 1992; Winter et al., 2006, 2008), specific amino acids in more than one part of the QX-RBD S1 protein are required to establish binding to kidney tissue (Bouwman et al., 2020). However, when the S1 subunit of the Beaudette strain was replaced with that of the H120 or M41 strain, the recombinant strains retained the ability to replicate in Vero cells, indicating that in this case the cell tropism of IBV is determined by the S2 subunit (Wei et al., 2014; Bickerton et al., 2018a,b). In further studies, the lack of the S2' cleavage motif was shown to result in the loss of the ability of the Beaudette strain to replicate in Vero cells (Bickerton et al., 2018a,b), and adding a furin S2' cleavage site to a QX-type IBV, resulted in a virus with neurotropism (Cheng et al., 2019). In a previous study, we constructed two recombinant strains, rH120-S/YZ and rIBYZ-S/H120, which express the heterologous S gene of the CK cell adapted strain H120 or non-adapted strain IBYZ along with the backbone genome. The results showed that the S gene replacement with the corresponding sequence of CK cell non-adapted strain leads to the loss of replication in CK cells, while replacement with one of the adapted strains would provide the infection ability, demonstrating the S protein to be a determinant of CK cell tropism of IBVs (Jiang et al., 2017). In the present study, we constructed several recombinant strains expressing different regions of the S gene from the IBYZ strain, and with the background genome of H120 strain, which aimed to confirm that S1, S2, or S1/S2 cleavage site is involved in the determination of tropism extended to CK cells. Additional studies used the same method of mutant construction and explored the correlation between the tropism extended to CK cells and the mutation sites on S gene. Consequently, the S amino acid sequences of the CK cell adapted YZ120 strain and its parental strain IBYZ were aligned and analyzed to identify the relevant amino acid residues in the S gene that were modified in different genome backbones of H120, IBYZ, and H120-S/YZ strains; also, their importance for CK cell tropism was determined.

MATERIALS AND METHODS

Virus Strains

The IBV strains used were (i) rH120, a molecular clone of vaccine H120 strain, which is a widely used vaccine strain in the market at present; (ii) rIBYZ, a molecular clone of strain ck/CH/IBYZ/2011 (GenBank KF663561.1), which was isolated from an IBV infected flock by our lab in Jiangsu Province, China in 2011; (iii) rYZ120 (hereafter referred to as rYZ120-S (138+, 581+, 617+), a molecular clone of CK cell adapted strain YZ120, which is derived from rIBYZ strain after multiple passages in chicken embryo (**Figure 1**); (iv) rH120-(S1/S2)/YZ, rIBYZ-(S1/S2)/H120, expressing the reciprocal S1/S2 cleavage site in rH120 or rIBYZ strains; (v) rH120-S1/YZ, rH120-S2/YZ, expressing chimeric S proteins composed of either the S1 subunit derived from IBYZ strain and S2 subunit from H120 or the S1 subunit derived from H120 and the S2 from IBYZ strain, with the backbone genome of H120 strain; (vi) rYZ120-S (138−, 581−, 617−), rYZ120-S (138−, 581+, 617−), rYZ120-S (138−, 581−, 617+), and rYZ120-S (138−, 581+, 617+), recombinants in which one or more amino acid sites at positions 138, 581, and 617 of the S protein were changed; (vii) rH120-S(I614V), in which the amino acid isoleucine at position 614 of the S protein was changed to valine; (viii) rIBYZ-S (V617I) and rH120-S

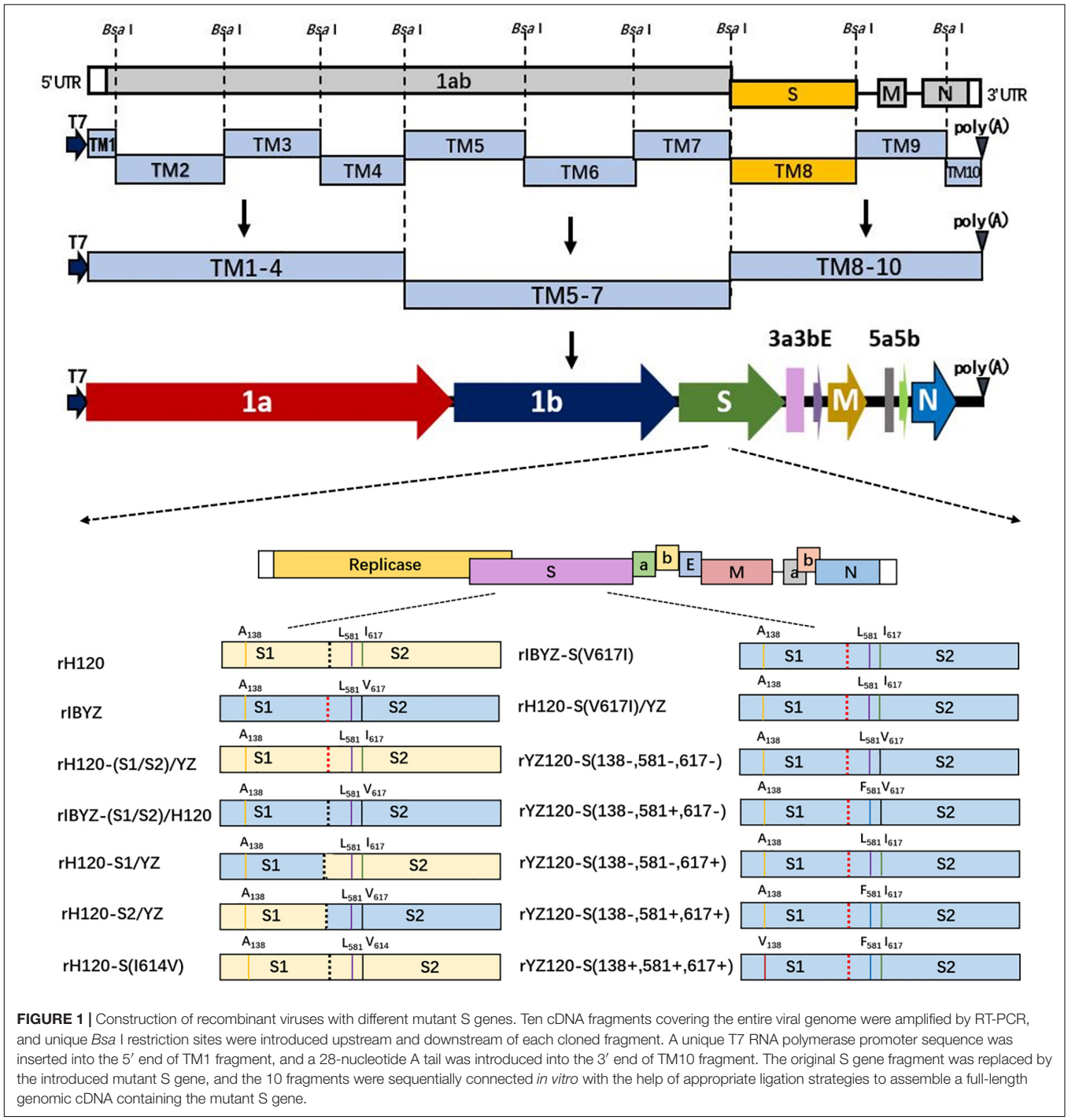


FIGURE 1 | Construction of recombinant viruses with different mutant S genes. Ten cDNA fragments covering the entire viral genome were amplified by RT-PCR, and unique *Bsa* I restriction sites were introduced upstream and downstream of each cloned fragment. A unique T7 RNA polymerase promoter sequence was inserted into the 5′ end of TM1 fragment, and a 28-nucleotide A tail was introduced into the 3′ end of TM10 fragment. The original S gene fragment was replaced by the introduced mutant S gene, and the 10 fragments were sequentially connected *in vitro* with the help of appropriate ligation strategies to assemble a full-length genomic cDNA containing the mutant S gene.

(V617I)/YZ, expressing isoleucine at position 617 of the S protein of IBYZ, with the backbone genome of rIBYZ or rH120 strain.

Construction of IBV Recombinant Strains

The plasmids pH120S, pIBYZS, and pYZ120S harbored the inserted S gene of the H120 vaccine strain, IBYZ strain, and YZ120 strain, respectively, which were constructed during the establishment of the reverse genetic system. By overlapping PCR technology, the furin cleavage site of S1/S2 protein of H120 and IBYZ strains were cross-replaced to construct recombinant plasmids pYZ (S1/S2)/H120 and pH120(S1/S2)/YZ. Using In-Fusion PCR cloning system (Clontech, United States), the S1 or S2 gene of the H120 strain was replaced with the corresponding region of the IBYZ strain to construct the recombinant plasmids pH120S1YZS2 and pYZS1H120S2, which contained the chimeric S genes. By overlapping PCR technology, point mutations were introduced into the specific regions of the S gene of pH120S, pIBYZS, and pYZ120S to construct the S gene mutation plasmids

pH120S (I614V), pIBYZS (V617I), and pYZ120S (138−, 581−, 617−). The strategy to construct the full-length cDNA clones of IBVs are described in the schematic illustration presented in **Figure 1**. The genome RNAs of recombinant viruses were synthesized by T7 RNA polymerase *in vitro* and transfected into BHK-21 cells, and the recombinant viruses were rescued (Zhou et al., 2010, 2011). The recombinant viruses were propagated in allantoic cavities of 11-day-old specific-pathogen-free (SPF) embryonated chicken eggs, and allantoic fluid was collected at 40 h post infection (hpi) and stored at −80°C.

Preparation of Primary Chicken Kidney (CK) Cells

Primary CK cells were prepared from 8-week-old chicks. Kidneys were collected, washed with phosphate buffer saline (PBS), and cut up. The resulting kidney pieces were digested with 0.25% trypsin, and 1% EDTA for 45 min at 37°C. The reaction was stopped with fetal calf serum (FCS). The cells were filtered through a sieve and collected by centrifugation at 1000 × *g* for 5 min. The kidney cells were resuspended in Medium 199 plus 3% FCS and incubated in plastic tissue flasks at 37°C with 5% CO_2. After 48-h incubation, CK cells were ready to be used for viral infection.

Replication Kinetics of rIBVs in Chicken Embryos

A RT-qPCR method was established based on a highly conserved area in the 5′-UTR of the IBV genome. Primers were designed according to the sequences of IBV from GenBank. The upstream primer was 5′-CCGTTGCTTGGGCTACCTAGT-3′, and the downstream primer was 5′-CGCCTACCGCTAGATGAACC-3′. The amplification product was cloned into pMD18-T (Takara) and the concentration of the plasmid was measured. A gradient dilution of 5×10^8–5×10^2 copies/μL was used as template for quantitation test. By plotting the cycle threshold (CT) values against the copies of the plasmid, the standard curve was generated (Jiang et al., 2020).

To examine viral replication in chicken embryos, 11-day-old embryonated SPF chicken eggs (6 eggs/group) were each inoculated with different recombinant strains at a dose of 10^7 viral RNA copies in 100 μL. Allantoic fluids were collected from each of the 6 eggs of each group at 6-h intervals, RNA was extracted using TRIzol, and reverse transcribed into cDNA by using a Primer-Script RT Master Mix Perfect Real-Time Kit (TaKaRa, Otsu, Shiga, Japan), and RT-qPCR was performed using SYBR® Premix Ex TaqTM II (TaKaRa) on an Applied Biosystems 7500 Fast Real-time PCR System.

Replication Kinetics of rIBVs in CK Cells

CK cells were grown on 6-well plates for 48-h. Before the infection, the cells were washed with PBS. Subsequently, CK cells were inoculated with 100 μL of virus suspension containing 10^7 viral RNA copies/well (3 wells/strain). After 60-min adsorption, the excess virus was removed by three PBS washes. Then Medium 199 was added to the infected cells. 100 μL/well viral supernatant of each group was harvested separately at different time points post-infection. Total RNA was extracted and processed as described above.

Indirect Immunofluorescence Assay (IFA)

At 48-h post-infection, CK cells were fixed with 1 mL methanol pre-cooled at −20°C and permeabilized using 0.5% Triton X-100 in PBS. The infected cells were identified by incubation with a 1:100 dilution of serum harvested from chickens at 14 days post-injection with rIBYZ and M41 strains, followed by detection with anti-chicken IgY (IgG) (whole molecule)-FITC antibody produced in rabbit (Sigma–Aldrich, Germany; dilution 1:640). Subsequently, the nucleus was stained with 4′6-diamidino-2-phenylindole (Beyotime Biotechnology, China) for 5 min, and the immunolabeled cells were imaged using an Inverted Microscope for Industry Leica DMi8 (Leica, Germany).

Animals and Ethics Statement

Specific pathogen-free (SPF) chicken embryos were purchased from the Beijing Merial Vital Laboratory Animal Technology Co., Ltd, China. All the animals in this study were cared for in strict accordance with the animal ethics guidelines and established protocols, and the experimental protocols were approved by the Animal Welfare and Ethical Censor Committee of the Poultry Institute, Chinese Academy of Agriculture Sciences.

Statistical Analysis

GraphPad Prism 7 software (GraphPad Software Inc., La Jolla, CA, United States) was used for statistical analysis. In the case of replication kinetics, the data were analyzed using a two-way analysis of variance (ANOVA) to detect any significant difference between various groups.

RESULTS

All Recombinant Viruses Were Successfully Rescued and Replicated Effectively in Chicken Embryos

The BHK-21 cell supernatant contained viruses that were harvested at 48 h after transfection and propagated in 10-day-old SPF chicken embryos. Recombinant viruses were verified by sequencing their entire genomes. Allantoic fluid containing virus (10^7 copies of viral RNA) was inoculated into SPF chicken embryos and harvested at different time points. All the strains could infect and proliferate in the chicken embryo and reach the peak point at 24–36 h post-infection with similar replication curves (**Figure 2**).

S1/S2 Cleavage Site Is Irrelevant to CK Cell Tropism of IBV H120

After analyzing the S gene of H120 and IBYZ strains, different motifs of the S1/S2 cleavage site were identified (H120: $RRFRR_{537}$/S, IBYZ: $HRRRR_{540}$/S). The S1/S2 cleavage site motifs were reciprocally expressed in the S protein of rH120 and rIBYZ strains, and the replication characteristics of the recombinants were analyzed by indirect immunofluorescence and replication

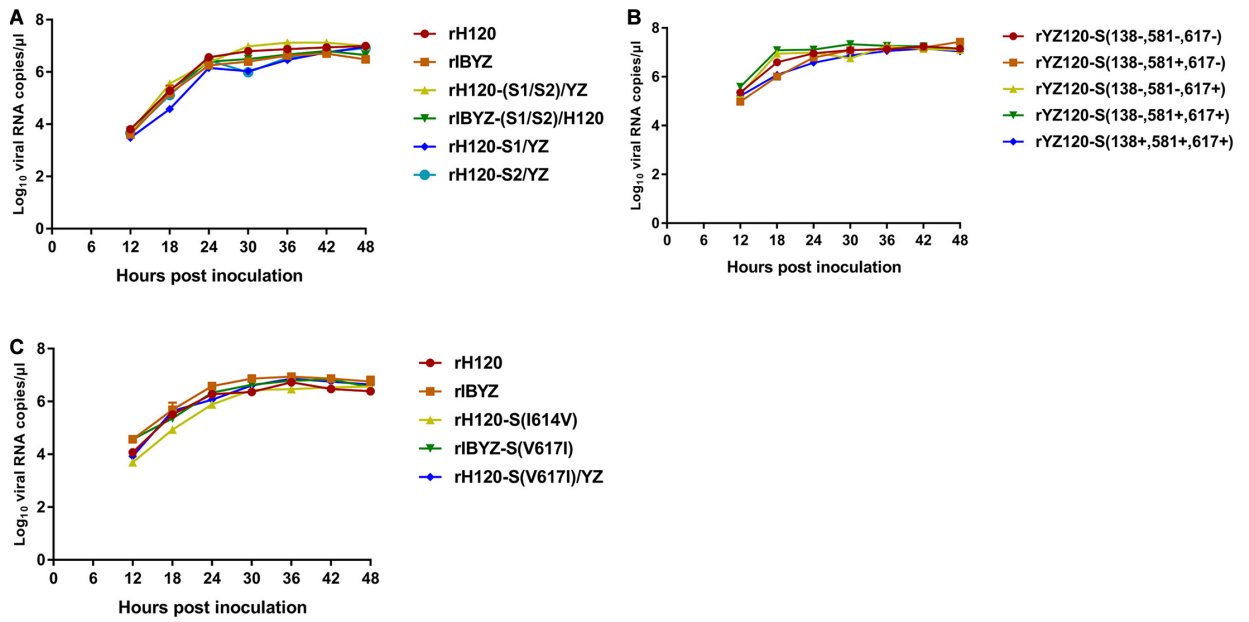


FIGURE 2 | RNA replication curves of recombinants in SPF chicken embryo. Eleven-day-old embryonated SPF chicken eggs were inoculated with 100 μL containing 10^7 viral RNA copies of **(A)** rH120, rIBYZ, rH120-(S1/S2)/YZ, rIBYZ-(S1/S2)/H120, rH120-S1/YZ and rH120-S2/YZ; **(B)** rYZ120-S(138–, 581–, 617–), rYZ120-S(138–, 581+, 617–), rYZ120-S(138–, 581–, 617+), rYZ120-S(138–, 581+, 617+) and rYZ120-S(138+, 581+, 617+); **(C)** rH120, rIBYZ, rH120-S(I614V), rIBYZ-S(V617I) and rH120-S(V617I)/YZ. The allatonic fluid was harvested at 12, 18, 24, 36, 42, and 48 h post-infection. Viral RNA copies were quantified by Real-time RT-PCR. Error bars indicate the standard deviation.

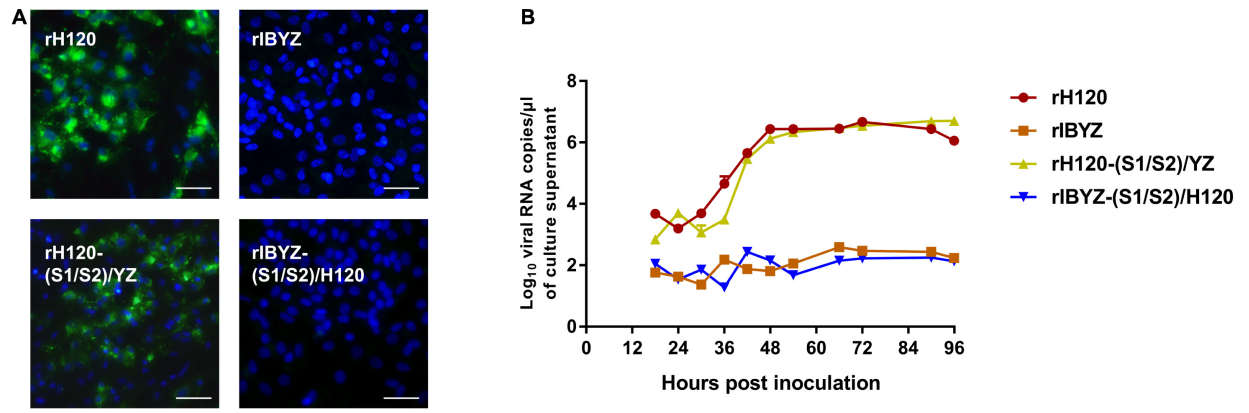


FIGURE 3 | Replication characteristics of rH120-(S1/S2)/YZ and rIBYZ-(S1/S2)/H120 in CK cells. **(A)** CK cells were grown in 6-well plates for 48 h and infected with 100 μL containing 10^7 viral RNA copies of rH120, rIBYZ, rH120-(S1/S2)/YZ, and rIBYZ-(S1/S2)/H120. After 48 h of infection, the infected cells were immunolabeled with anti-IBV serum and secondary antibody anti-chicken IgY (IgG) (whole molecule)-FITC antibody. Nuclei were labeled with DAPI (blue). Bar, 50 μm. **(B)** RNA replication curves for the recombinants in primary CK cells. CK cells in 6-well plates were inoculated with rH120, rIBYZ, rH120-(S1/S2)/YZ and rIBYZ-(S1/S2)/H120; the supernatant was harvested at 18, 24, 30, 36, 42, 48, 54, 66, 72, 90, and 96 h post-infection. Viral RNA copies were quantified by real-time RT-PCR. Y axis indicates log10 viral RNA copies/μL culture supernatant. Error bars indicate the standard deviation.

curves in CK cells. Analysis of the replication kinetics in CK cells indicated that the replacement of the S1/S2 cleavage site motif with the corresponding sequence of CK cell non-adapted strain IBYZ did not alter the replication ability in CK cells. In addition, the S1/S2 cleavage site from the CK cell-adapted strain did not allow the rIBYZ-(S1/S2)/H120 strain to acquire the replication ability in CK cells (**Figure 3B**). The immunofluorescence analysis of CK cells infected with rH120-(S1/S2)/YZ demonstrated that the strain could infect and spread to neighboring cells, as observed for the parental virus. However, no visible green fluorescence was observed in the CK cells infected with rIBYZ-(S1/S2)/H120 strain at 48 hour post infection (hpi), as it was observed for the parental strain rIBYZ (**Figure 3A**). Overall, the replacement of the S1/S2 cleavage motif did not affect the ability of the parental strain to infect the CK cells.

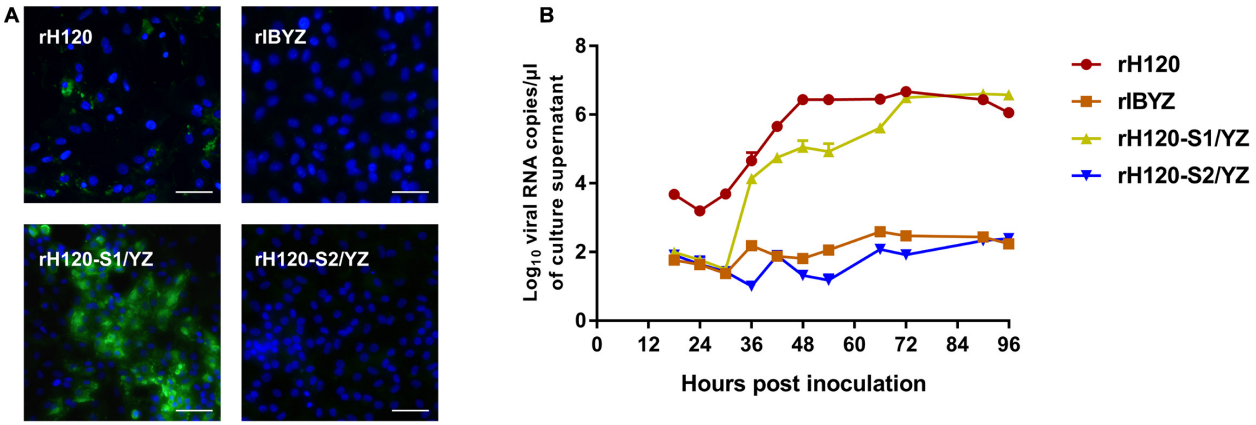


FIGURE 4 | Replication characteristics of rH120-S1/YZ and rH120-S2/YZ in SPF chicken embryo and CK cells. **(A)** CK cells were grown in 6-well plates for 48 h and infected with 100 μL containing 10^7 viral RNA copies of rH120, rIBYZ, rH120-S1/YZ, and rH120-S2/YZ. After 48 h infection, the infected cells were immunolabeled with anti-IBV serum and secondary antibody anti-chicken IgY (IgG) (whole molecule)–FITC antibody. Nuclei were labeled with DAPI (blue). Bar, 50 μm. **(B)** RNA replication curves for the recombinants in primary CK cells. CK cells in 6-well plates were inoculated with rH120, rIBYZ, rH120-S1/YZ, and rH120-S2/YZ; the supernatant was harvested at 18, 24, 30, 36, 42, 48, 54, 66, 72, 90, and 96 h post-infection. Viral RNA copies were quantified by real-time RT-PCR. *Y* axis indicates log10 viral RNA copies/μL culture supernatant. Error bars indicate the standard deviation.

S2 Subunit Plays a Major Role in CK Cell Tropism of IBV H120

Within the genomic background derived from the H120 strain, the recombinant rH120-S1/YZ or rH120-S2/YZ, with the replacement of the S1 or S2 subunit coding sequence with the corresponding sequence of CK cell non-adapted strain IBYZ, were constructed using a reverse genetic system to determine the subunit responsible for the CK cell tropism of rH120 strain. The S1/S2 cleavage site motif (RRFRR/S) from the H120 strain was retained in both recombinant strains. Immunofluorescence analysis of rH120-S1/YZ-infected CK cells at 48 hpi showed that green fluorescence was observed in some infected cells, which was similar to that of the rH120 group (**Figure 4A**). The apparently lower density of infected cells in the H120-inoculated cells is likely due to shedding of infected cells, as indicated by fewer nuclei. The replacement of the rH120 S2 subunit with the S2 subunit from IBYZ caused the infectivity loss in CK cells of rH120-S2/YZ. Analysis of the replication kinetics of recombinant strains in CK cells confirmed the results of IFA. The viral RNA copies of rH120-S1/YZ in CK cells increased rapidly from 30 to 36 hpi and reached a level similar to that of the parental strain rH120 at 72 hpi. Like the rIBYZ strain, rH120-S2/YZ, in which the S2 subunit was replaced with that of IBYZ strain, did not show replication in CK cells (**Figure 4B**). These results demonstrated that virus containing the S2 subunit of IBYZ could not infect CK cells, whereas virus containing the S2 subunit of H120 could.

Chicken Embryo Adapted Virus YZ120 Acquires the Ability to Infect CK Cells

The YZ120 strain was obtained from the parent strain IBYZ through 120 passages in chicken embryos. In contrast to IBYZ strain-inoculated cells, CK cells inoculated with the YZ120 strain exhibited high levels of infected cells 48 hpi, as indicated by immunofluorescence (**Figure 5A**). The results of replication kinetics showed that YZ120 strain replicates effectively in CK cells and reached the peak viral RNA copies of 10^7 copies/μL at 72 hpi (**Figure 5B**). To begin to understand whether the extended tropism of YZ120 is related to its S protein, the nucleotide and predicted amino acid sequences of the S genes of IBYZ and YZ120 strains were aligned and compared using the Clustal W multiple alignment algorithm. Three nucleotide changes, each resulting in an amino acid substitution, were identified. A single nucleotide substitution, C20783T, was detected in the S1 gene of the YZ120 strain, resulting in an amino acid change, A138V. The other two nucleotide substitutions, C22111T and G22219A, occurred in the S2 gene of the YZ120 strain, resulting in amino acid changes L581F and V617I, respectively (**Figure 5C**).

A138V Is Not Essential for the Invasion of YZ120 Into CK Cells

To investigate whether one or more of the amino acids changes were required for CK cell tropism, rYZ120-S (138+, 581+, 617+), a molecular clone of CK cell-adapted strain YZ120 was generated by the reverse genetic system. Because of the results showing that the S2 subunit was the critical factor in the lack of ability of the non-adapted IBYZ strain to replicate in CK cells, we first constructed the recombinant strain rY120-S (138−, 581+, 617+), with a rYZ120 genomic backbone. One codon, encoding the amino acid residue located at position 138 in the S1 subunit was mutated to encode alanine instead of valine. Similar replication kinetics patterns were observed for rY120-S (138−, 581+, 617+) and rYZ120-S (138+, 581+, 617+) in CK cells, although the viral RNA copy numbers of the parental strain rYZ120-S (138+, 581+, 617+) were slightly lower than that of the mutation strain at time points at 24 and 72 hpi (**Figure 6B**). IFA confirmed the findings of the replication kinetics studies and showed similar results for the proportion of infected cells at 48 hpi, indicating that both recombinant strains were able to

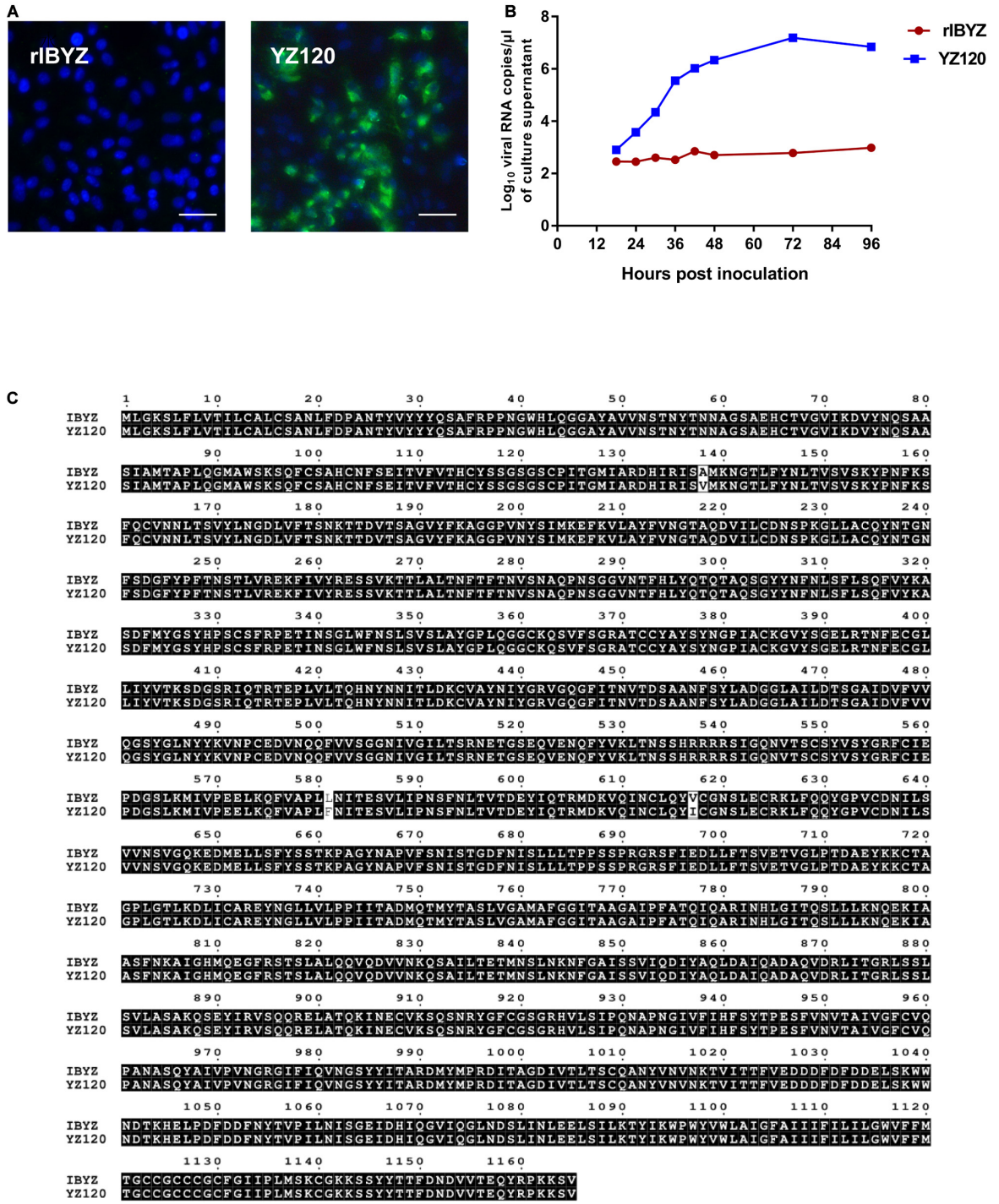


FIGURE 5 | Adaptation of rYZ120 to CK cells. **(A)** CK cells were grown in 6-well plates for 48 h and infected with 100 μL containing 10^7 viral RNA copies of rIBYZ and YZ120 strains. After 48 h infection, the infected cells were immunolabeled with anti-IBV serum and secondary antibody anti-chicken IgY (IgG) (whole molecule)-FITC antibody. Nuclei were labeled with DAPI (blue). Bars, 50 μm. **(B)** RNA replication curves for the recombinants in primary CK cells. CK cells in 6-well plates were inoculated with rIBYZ and YZ120; and the supernatant was harvested at 18, 24, 30, 36, 42, 48, 72, and 96 h post-infection. Viral RNA copies were quantified by real-time RT-PCR. Three replicates were performed, and the average was taken. *Y* axis indicates log10 viral RNA copies/μL culture supernatant. Error bars indicate the standard deviation. **(C)** Comparison of the amino acid sequences of the S protein of YZ120 and IBYZ strains. The sequences were aligned using ClustalX 2.1 and compared using Multiple Sequence Alignment by ClustalW (https://www.genome.jp/tools-bin/clustalw) and ESPript 3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi). Amino acids shaded in black represent identical amino acid residues found in each sequence; non-highlighted residues represent differing amino acids.

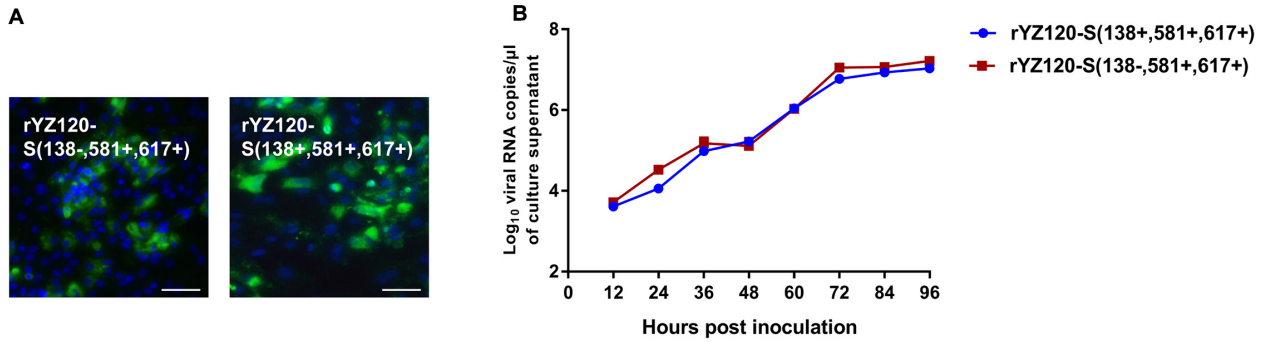


FIGURE 6 | Replication characteristics of rYZ120-S(138–, 581+, 617+) and rYZ120-S(138+, 581+, 617+) in CK cells. **(A)** CK cells were grown in 6-well plates for 48 h and infected with 100 μL containing 10^7 viral RNA copies of rYZ120-S(138–, 581+, 617+) and rYZ120-S(138+, 581+, 617+). After 48 h of infection, the infected cells were immunolabeled with anti-IBV serum and secondary antibody anti-chicken IgY (IgG) (whole molecule)-FITC antibody. Nuclei were labeled with DAPI (blue). Bar, 50 μm. **(B)** RNA replication curves for the recombinants in primary CK cells. CK cells in 6-well plates were inoculated with rYZ120-S(138–, 581+, 617+) and rYZ120-S(138+, 581+, 617+); the supernatant was harvested at 12, 24, 36, 48, 60, 72, 84, and 96 h post-infection. Viral RNA copies were quantified by real-time RT-PCR. *Y* axis indicates log10 viral RNA copies/μL culture supernatant. Error bars indicate the standard deviation.

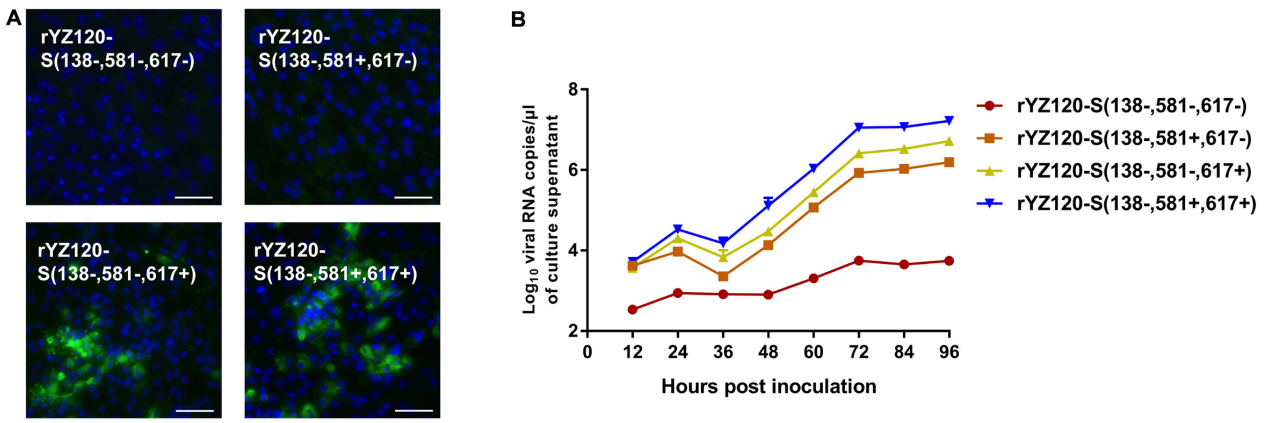


FIGURE 7 | Replication characteristics of rYZ120-S(138–, 581–, 617–), rYZ120-S(138–, 581+, 617–), rYZ120-S(138–, 581–, 617+) and rYZ120-S(138–, 581+, 617+) in CK cells. **(A)** CK cells were grown in 6-well plates for 48 h and infected with 100 μL containing 10^7 viral RNA copies of rYZ120-S(138–, 581–, 617–), rYZ120-S(138–, 581+, 617–), rYZ120-S(138–, 581–, 617+) and rYZ120-S(138–, 581+, 617+). After 48 h of infection, the infected cells were immunolabeled with anti-IBV serum and secondary antibody anti-chicken IgY (IgG) (whole molecule)-FITC antibody. Nuclei were labeled with DAPI (blue). Bar, 50 μm. **(B)** RNA replication curves for the recombinants in primary CK cells. CK cells in 6-well plates were inoculated with rYZ120-S(138–, 581–, 617–), rYZ120-S(138–, 581+, 617–), rYZ120-S(138–, 581–, 617+) and rYZ120-S(138–, 581+, 617+); the supernatant was harvested at 12, 24, 36, 48, 60, 72, 84, and 96 h post-infection. Viral RNA copies were quantified by real-time RT-PCR. *Y* axis indicates log10 viral RNA copies/μL culture supernatant. Error bars indicate the standard deviation.

infect CK cells (**Figure 6A**). These findings demonstrated that the amino acid substitution A138V is not necessary for the invasion of YZ120 into CK cells.

The V617I Substitution in the S Protein of rYZ120 Is Essential for Its Ability to Infect CK Cells, While L581F Also Promotes CK Cell Infection

In order to determine whether the other two amino acid changes are correlated to CK cell tropism, S2 gene mutation strains with one or two codon substitutions (F581L, I617V) were constructed in the rYZ120-S (138+, 581+, 617+) genomic background (**Figure 1**). In contrast to the rYZ120-S (138+, 581+, 617+) infected cells (**Figure 6A**), IFA showed no visible green fluorescence in rYZ120-S (138−, 581−, 617−) and rYZ120-S (138−, 581+, 617−) inoculated cells, indicating that the substitution of I617V led to the loss of viral infection ability in CK cells (**Figure 7A**). In rYZ120-S (138−, 581−, 617+) inoculated cells, only small foci of fluorescence were observed, which indicated that although F581 is not the key factor of CK cell tropism change in the YZ120 strain, substitution of F581L reduced the efficiency of CK cell infection (**Figure 7A**). Analysis of replication kinetics of the recombinant strains in CK cells confirmed that substitution of both F581L and I617V led to the loss of the ability of the YZ120 strain to infect primary CK cells, while the rYZ120-S (138−, 581−, 617+) strain, with a leucine residue at position 581 substituted for phenylalanine, still retains the ability to infect CK cells, but the efficiency is significantly reduced (**Figure 7B**). In contrast to the findings

of the immunofluorescence studies, analysis of viral replication kinetics indicated that the recombinant strain rYZ120-S (138−, 581+, 617−) also had the ability to infect CK cells, although viral RNA replication was significantly lower than that in cells infected with the other two recombinant viral strains (**Figure 7B**). This conflicting result was explained when sequence analysis of the S gene amplified from virus in the culture supernatant of rYZ120-S (138−, 581+, 617−) infected cells showed that the codon encoding amino acid 617 of the S gene of virus recovered from the cells actually encoded isoleucine rather than valine. Thus, the rYZ120-S (138−, 581+, 617−) strain had acquired the ability to infect CK cells by mutation changing it to rYZ120-S (138−, 581+, 617+). Taken together, these results demonstrated that I617 residue in the S2 subunit may play a dominant role in extension of the tropism of the YZ120 strain to CK cells, and F581 also promotes the infection of the YZ120 strain in CK cells.

Ability of CK Cell Infection Is Limited by I614V Substitution in rH120 S Protein

The sequence of S gene from the H120 strain was analyzed and aligned against that of the IBYZ strain using ClustalW multiple alignment algorithm. Comparison of the amino acid sequence of the S gene from H120 and IBYZ strains showed that three amino acid sites in the H120 S protein that corresponded to the substituted sites on the YZ120 S protein were A138, L578, I614, in which only residue I614 was in accordance with that on the YZ120 S protein. Therefore, the recombinant strain rH120-S (I614V), which contains the amino acid substitution I614V in the S protein, was generated using a reverse genetic system. In contrast to the parental strain rH120-infected CK cells (**Figure 8A**), IFA showed no visible green fluorescence in rH120-S (I614V) strain inoculated cells at 48 hpi. The results of replication kinetics also indicated that the substitution I614V in the S protein of rH120 strain induced the loss of the infection ability of the virus in CK cells (**Figure 8B**).

Substitution V617I Provides IBYZ and H120-S/YZ Strains a Limited Ability to Infect CK Cells

To further verify the key role of V617I in IBV CK cell tropism, we constructed the recombinant rIBYZ-S (V617I), in which an isoleucine residue was substituted for the valine residue located at position 617 of the S protein. A small number of fluorescent cells were observed 48 h after inoculation of CK cells with rIBYZ-S (V617I) (**Figure 9A**), and analysis of the viral RNA replication kinetics confirmed that the level of viral RNA replication was extremely low (**Figure 9B**). Then, the V617I substitution in the S protein was introduced into rH120-S/YZ to generate rH120-S(V617I)/YZ, and the rescued recombinant strain could effectively infect CK cells, while the infection and replication efficiency were between that of rIBYZ-S (V617I) and rYZ120 strains. Thus, the results showed that the substitution at site 617 could make IBYZ or H120-S/YZ strain acquire the ability to infect primary CK cells; however, the infection efficiency was significantly lower than that of YZ120 strain, which might be related to the differences in the genome backbone among the three strains (**Figures 9A,C**).

DISCUSSION

Previously, by replacing the S gene of rH120 and rIBYZ, we found that the S protein determines the CK cell tropism of the virus (Jiang et al., 2017), as also described previously by Casais et al. (2003). Following previous results, rH120-(S1/S2)/YZ and rIBYZ-(S1/S2)/H120, i.e., reciprocal substitution of S1/S2 cleavage site motifs in H120 and IBYZ strains were constructed, which indicated that the CK cell tropism was independent of the S1/S2 cleavage site of the H120 or IBYZ strain. Further studies demonstrated the key role of the S2 subunit in CK cell tropism of IBVs, based on the differences observed in CK cell infection ability between rH120-S1/YZ and rH120-S2/YZ. In the subsequent studies, multiple passage strain YZ120 was obtained by 120 passages in embryo, which could proliferate in CK cells. Compared to the parent strain IBYZ, three residue substitutions were found in the S protein of the YZ120 strain. Using the genomic backbone of YZ120 strain, we carried out single or multiple point substitutions on those three amino acid residues. After evaluation by IFA and proliferation curve tests of IBV-infected CK cells, we demonstrated the key role of V617I in the CK cell tropism of the YZ120 strain, which was confirmed using other recombinants, in which the amino acid encoded at S codon 617 had been changed from that in the genomic backbone of rH120 (at position 614), rH120-S/YZ, and rIBYZ strains.

Some CoVs have two cleavage sites (S1/S2 and S2') in the S protein, which can be activated by host cell proteases in the process of infection. An S1/S2 site is located at the border between the S1 and S2 subunits, while an S2' site is located upstream of the putative FP in the S2 subunit. The S protein of some CoVs is hydrolyzed early at the S1/S2 cleavage site by acid pH-activated serine proteases, such as furin in the trans-Golgi network of virus-producing cells (Heald-Sargent and Gallagher, 2012). In MERS-CoV, the S1/S2 cleavage site is processed during viral entry by furin (Kleine-Weber et al., 2018); however, this site is not universally required for MERS S-driven host cell entry. Similarly, "uncleavable" state by site-directed substitution of the S protein in MHV retains the virus infectivity (Gombold et al., 1993; Yamada et al., 1997). In IBVs, the cleavage of the spike protein at the S1/S2 site is not necessary for attachment but promotes the infectivity in cells (Yamada et al., 2009). rH120 strain is a molecular clone of vaccine strain H120, which was generated from the H strain isolated in 1956 by 120 embryo passages. This strain can replicate in CK cells, while the rIBYZ strain, which is a molecular clone of a QX-like IBV isolate, is unable to replicate. The S1/S2 cleavage motifs in H120 and IBYZ are RRFRRS and HRRRRS, respectively, which have the consensus R-X-[KR]-R↓motif of furin at the P4-P1 position (Tian et al., 2011). Thus, the ability of both sites to be cleaved by furin might explain why the replacement of the S1/S2 cleavage site motif from rIBYZ and rH120 strains could not change the CK cell tropism, and the difference in CK cell tropism between rIBYZ and rH120 strains is independent of the S1/S2 cleavage site motifs.

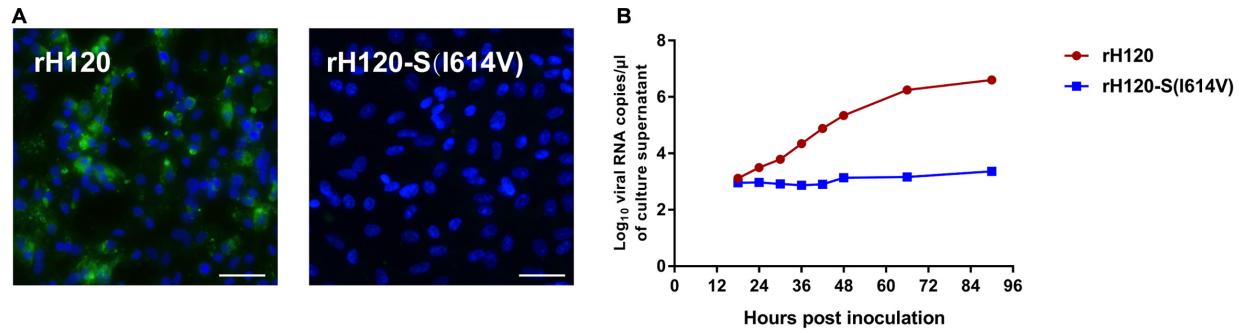


FIGURE 8 | Replication characteristics of rH120 and rH120-S(I614V) in CK cells. **(A)** CK cells were grown in 6-well plates for 48 h and infected with 100 μL containing 10^7 viral RNA copies of rH120 and rH120-S(I614V). After 48 h infection, the infected cells were immunolabeled with anti-IBV serum and secondary antibody anti-chicken IgY (IgG) (whole molecule) – FITC antibody. Nuclei were labeled with DAPI (blue). Bar, 50 μm. **(B)** RNA replication curves for the recombinants in primary CK cells. CK cells in 6-well plates were inoculated with rH120 and rH120-S(I614V); the supernatant was harvested at 18, 24, 30, 36, 42, 48, 66, 90 h post-infection. *Y* axis indicates log10 viral RNA copies/μL culture supernatant. Viral RNA copies were quantified by Real-time RT-PCR. Error bars indicate the standard deviation.

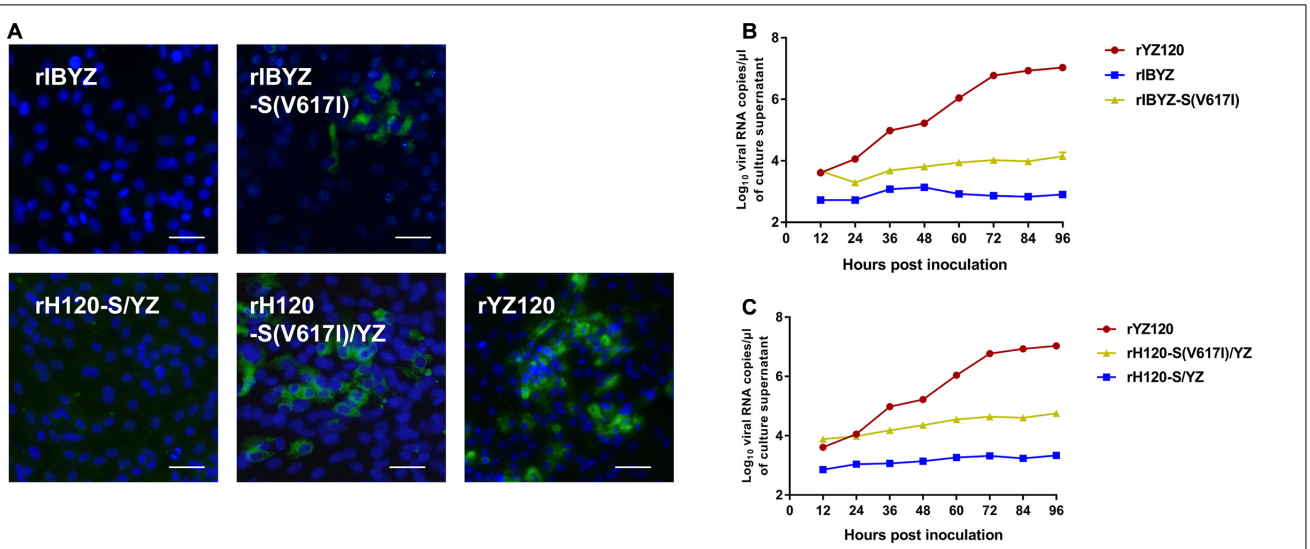


FIGURE 9 | Replication characteristics of rIBYZ, rIBYZ-S (V617I), rH120-S/YZ, rH120-S (V617I)/YZ, and rYZ120 in CK cells. **(A)** CK cells were grown in 6-well plates for 48 h and infected with 100 μL containing 10^7 viral RNA copies of rIBYZ, rIBYZ-S (V617I), rH120-S/YZ, rH120-S (V617I)/YZ, and rYZ120. After 48-h of infection, the infected cells were immunolabeled with anti-IBV serum and secondary antibody anti-chicken IgY (IgG) (whole molecule)-FITC antibody. Nuclei were labeled with DAPI (blue). Bars, 50 μm. RNA replication curves for the recombinants **(B)** rIBYZ, rIBYZ-S (V617I), and rYZ120, **(C)** rH120-S/YZ, rH120-S (V617I)/YZ, and rYZ120 in primary CK cells. CK cells in 6-well plates were inoculated with the recombinants, the supernatant was harvested at 18, 24, 30, 36, 42, 48, 66, and 90 h post-infection. Viral RNA copies were quantified by real-time RT-PCR. Y axis indicates log10 viral RNA copies/μL culture supernatant. Error bars indicate the standard deviation.

The ectodomain of all CoV spike proteins share the same organization in two domains: an NTD named S1 is responsible for binding and a C-terminal S2 domain is responsible for fusion (Belouzard et al., 2012). For some coronaviruses, including IBV, the two domains are cleaved into S1 and S2 subunits. The S1 subunit, which contains RBDs, is essential for the initial attachment of the virus (Heald-Sargent and Gallagher, 2012). Changes in the amino-terminal domain of the S1 subunit of some CoVs have been related to changes in tissue tropism *in vivo*, which might be attributed to the loss of the ability of the receptor binding (Krempl et al., 1997; Desmarets et al., 2014; Pedersen, 2014). In IBVs, α-2,3 linked sialic acids have been identified to be a receptor bound by the S1 subunit and essential for attachment of avian cells (Winter et al., 2006); however, the molecular mechanism of cell and tissue tropism is yet to be understood. Although virus binding to host cells is the first step in tropism determination that is affected by the S1 subunit (Wickramasinghe et al., 2011), S2 is responsible for membrane fusion. Sometimes, only exchanging the S1 subunit of some IBV strains may not change the tropism; the S2 subunit also affects the cell tropism of the virus. The recombinant that replaced the Beaudette strain S1 subunit with M41-CK S1 corresponding sequence replicated similarly to that of Beau-R in Vero cells, while replacing the S2 subunit of Beaudette strain with S2 subunit of M41-CK strain

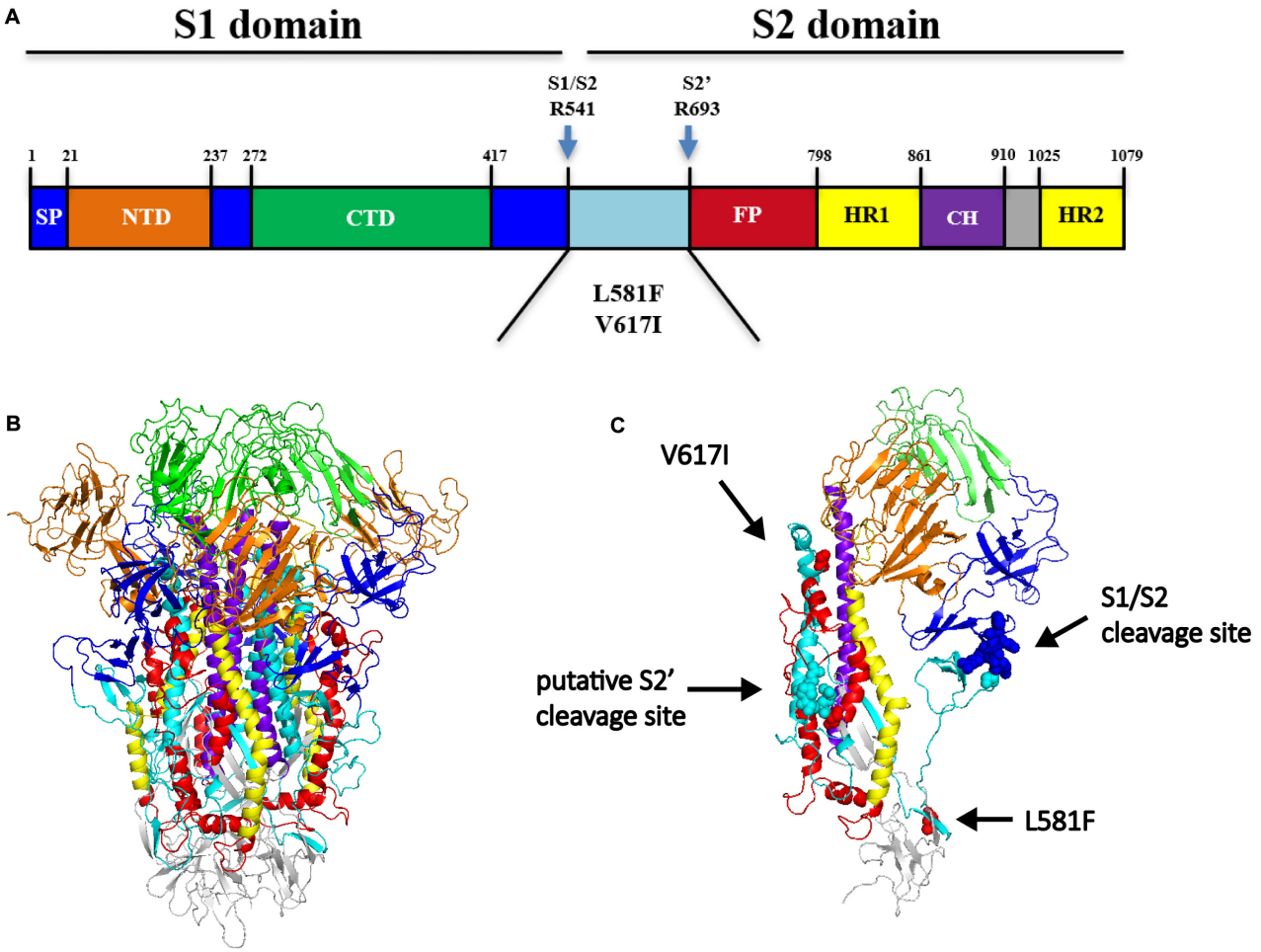


FIGURE 10 | (A) Schematic illustration of IBYZ spike ectodomain drawn by comparing the corresponding sequence of M41 strain. NTD, N-terminal domain. CTD, C-terminal domain. FP, fusion peptide. HR1, heptad repeats 1. CH, central helix. HR2, heptad repeats 2. S2' is a putative cleavage site, which refers to the second fusion cleavage site of the Beaudette strain. The range of FP is not determined. **(B)** Structure of IBYZ spike ectodomain in the pre-fusion conformation predicted by the Swiss-model based on the structure of the M41 S ectodomain trimer (6cv0.1). **(C)** The structure prediction of the monomer of IBYZ spike ectodomain in the prefusion conformation. The different components of Spike are colored differently, which is consistent with the schematic illustration (A). Amino acid residues at position 581 (red) and 617 (red), S1/S2 cleavage site (blue), and the putative S2' cleavage site (cyan) are marked with spheres of different colors.

resulted in the loss of the ability of Beau-R strain to infect and replicate in Vero cells (Bickerton et al., 2018b). This study also indicates that the S1 subunits from both H120 and IBYZ may be able to bind to the IBV receptor, and the S2 subunit is the key factor leading to the difference of CK cell tropism between the two strains.

Due to the numerous sequence differences in the S2 subunits of H120 and IBYZ strains, we obtained a CK cell-adapted strain YZ120, which is a rIBYZ-derived strain continuously passaged in SPF chicken embryos, to identify the key sites on the S2 subunit related to CK cell adaptation. After sequence alignment of the S gene between YZ120 and the parental strain IBYZ, we found three amino acid substitutions, A138V in the S1 subunit and L581F and V617I in the S2 subunit. We also confirmed that A138V substitution was not related to the CK cell tropism change of the YZ120 strain, which was consistent with the conclusion that S2 determines the CK cell tropism difference between IBYZ and H120 strains. Comparing replication curves of the four mutant strains in primary CK cells, we found that rYZ120-S (138−, 581−, 617+) and rYZ120-S (138−, 581+, 617+) could effectively replicate in primary CK cells. Although the replication curve of rYZ120-S (138−, 581+, 617−) also indicated the ability to replicate in CK cells, a V617I codon mutation was found in the S gene of the virus harvested from the supernatant at 96 h post-infection. In the IFA assay, almost no fluorescence was detected in the rYZ120-S (138−, 581+, 617−)-inoculated cells, suggesting that the amino acid substitution at S position 617 might be critical for the CK cell tropism change of YZ120 strain, and L581F might promote the infection efficiency of IBVs in CK cells. Next, we analyzed the position of these two substitution amino sites, and found their localization in the region between S1/S2 cleavage and putative S2' cleavage (**Figure 10A**). As a class I viral fusion protein, the S2 subunit of CoVs contains a putative FP and two heptad repeats (HR1 and HR2) (Belouzard et al., 2012). After the

S1/S2 proteolytic cleavage site is activated, the conformation of S protein changed and the S2' cleavage site and FP were exposed, which led to virus-cell fusion (Matsuyama and Taguchi, 2009; Walls et al., 2017). The analysis of the sequence of S2' cleavage motifs in YZ120 revealed $_{689}$SPRGR/S$_{694}$, which could not be cleaved by furin. Nevertheless, the possibility of proteolysis by other proteases in the infectious process of susceptible chicken cells is yet to be elucidated. We predicted the structure of the S protein of the IBYZ strain, based on the cryo-electron microscopy structure of the IBV M41 S protein, by homology modeling using the SWISS-MODEL server[1]. A clove-like shape, with three S1 subunits forming a crown-like structure on top of a trimeric S2 stalk was observed in the pre-fusion conformation of IBYZ S protein (**Figure 10B**). The S1/S2 cleavage site and the amino acid residue at position 581 were on the periphery of the clove-like structure. In the prefusion state, the amino acid residue at position 617 of the S2 subunit, along with the putative S2' cleavage site and FP, are located within the S1 crown-like structure (**Figure 10C**). At present, there is no evidence for a second cleavage site in the S protein of IBVs, except the Beaudette strain, and the exact location and size of the putative FP could not be determined. Therefore, according to the position of 617 amino acid residues, we speculate that the substitution V617I might affect the conformation of S protein in the pre-fusion state, which contributes to the exposure of the putative S2 cleavage site or FP, thereby promoting the occurrence of virus-cell fusion.

To demonstrate that the amino acid site at position 617 has the same effect in the H120 strain, we carried out a point mutation in the corresponding codon 614 of H120 S gene, which transformed the isoleucine codon into a valine codon. The mutant strain rH120-S (I614V) could not infect primary CK cells, which indicated that valine in the 614 position of S protein inhibited the invasion of IBVs into primary CK cells. To further explore whether the changes in CK cell tropism of YZ120 are determined by changes at only aa 617 in the S2 subunit, we constructed two V617I substitutions with the genomic backbone of rIBYZ or rH120-S/YZ strain. rIBYZ-S (V617I) did not acquire the ability of efficient proliferation in CK cells, while rH120-S (V617I)/YZ could infect the CK cells, its proliferation efficiency was lower than that of YZ120 strain. Based on these results, we suspect that there are other factors encoded in the backbone of the viral genome besides the S protein that can increase the replication efficiency of the virus in host cells.

We selected some of those recombinant viruses and re-tested their replication curves in primary CK cells to analyze the correlation between different recombinants and CK cell infections from the perspective of S protein and other factors (**Supplementary Figure 1** and **Table 1**). The recombinants with the same backbone of H120 or IBYZ strain but expressing different chimeric S protein showed significant differences in the infection of primary CK cells, i.e., the order of the ability of viruses with different S proteins to infect CK cells is as follows: H120 > IBYZ-S(V617I) > IBYZ. With the same S protein from IBYZ-S (V617I), different backbones showed different virus replication efficiency in the CK cells, which indicated that there might be other factors besides S protein associated with the efficiency of IBV infection and replication in different tissues or cells. With the same S protein from IBYZ-S (V617I), different backbones showed different virus replication efficiency in the CK cells (backbone YZ120 > H120 > IBYZ), which indicated that there might be other factors besides S protein associated with the efficiency of IBV infection and replication in different tissues or cells. Thus, the substitution V617I of S protein can make YZ120 obtain the ability to infect CK cells, but the structure of this S protein has not reached the optimal evolutionary state. The effective infection in primary CK cells of YZ120 depends on the role of other structural proteins or non-structural proteins besides S protein of the genome; however, the specific mechanism is yet unclear. However, the substitution of V617I on the S protein triggered the first step of IBV virus to invade the primary CK

[1]https://swissmodel.expasy.org/interactive

TABLE 1 | Analysis of infection and replication of different recombinant viruses in primary CK cells.

Strain	S protein	Backbone	Infection	Replication	Conclusion
rIBYZ-S(V617I)	IBYZ-S(V617I)	IBYZ	Yes	Low	※ The backbone of the IBVgenome besides S gene may affect proliferation efficiency in CK cells
rH120-S(V617I)/YZ	IBYZ-S(V617I)	H120	Yes	Medium	※ Relevant between backbone and CK cell infection:
rYZ120-S(138−, 581−, 617+)	IBYZ-S(V617I)	YZ120	Yes	High	YZ120 > H120 > IBYZ
rIBYZ	IBYZ	IBYZ	No	No	
rIBYZ-S/H120	H120	IBYZ	Yes	Medium	※ Relevant between S protein and CK cell infection:
rIBYZ-S(V617I)	IBYZ-S(V617I)	IBYZ	Yes	Low	H120 > IBYZ-S(V617I) > IBYZ
rH120	H120	H120	Yes	High	
rH120-S/YZ	IBYZ	H120	No	No	※ Relevant between S protein and CK cell infection:
rH120-S(V617I)/YZ	IBYZ-S(V617I)	H120	Yes	Medium	H120 > IBYZ-S(V617I) > IBYZ
rH120-S/YZ	IBYZ	H120	No	No	※ S protein but not backbone of IBVs is a determinant of CK cell tropism. ※ S protein of IBYZ strain does not provide the ability to infect CK cells.
rIBYZ	IBYZ	IBYZ	No	No	
rYZ120-S(138−, 581−, 617−)	IBYZ	YZ120	No	No	

cells, otherwise YZ120 could not enter the CK cells effectively. Therefore, the isoleucine at position 617 in the S protein from both H120 and YZ120 strains are related to CK cell tropism of IBVs, albeit the specific principle remains to be studied further.

CONCLUSION

The S2 subunit is the determinant factor of the difference in CK cell tropism between H120 and IBYZ strains. The adaptation of the YZ120 strain to CK cells is independent of the A138V substitution, while V617I substitution leads to the CK cell tropism changes in the YZ120 strain, and L581F promotes the infectivity of the YZ120 in CK cells. In addition, there could also be other factors in the genomic backbone of IBVs associated with the CK cell tropism, which need further investigation.

DATA AVAILABILITY STATEMENT

All data, models generated or used during the study are available from the corresponding author by request.

AUTHOR CONTRIBUTIONS

YJ and SZ designed and conducted the experiments. YJ, SZ, XC, and YY performed the experiments. YJ, MG, XS, and JL analyzed and interpreted the data. YJ and SZ wrote and revised the manuscript. All the authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program (2017YFD0500703), the National Natural Science Foundation of China (31602091 and 31572524), and the Innovation Capacity Development Plan of Jiangsu Province (BM2018026).

ACKNOWLEDGMENTS

We appreciate all members of the Poultry Institute, Chinese Academy of Agricultural Sciences and College of Veterinary Medicine, Ministry of Education Key Lab for Avian Preventive Medicine for their contribution to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.604335/full#supplementary-material

Supplementary Figure 1 | Replication characteristics of rIBYZ-S (V617I), rH120-S (V617I)/YZ, rYZ120-S (138−, 581−, 617+), rIBYZ, rIBYZ-S/H120, rH120, rH120-S/YZ, in CK cells. **(A)** rIBYZ-S(V617I), rH120-S (V617I)/YZ, rYZ120-S (138−, 581−, 617+), **(B)** rIBYZ, rIBYZ-S/H120, rIBYZ-S (V617I), **(C)** rH120, rH120-S/YZ, rH120-S (V617I)/YZ in primary CK cells. CK cells in 6-well plates were inoculated with the recombinants, the supernatant was harvested at 12, 24, 36, 48, 60, 72, 84, and 96 h post-infection. Viral RNA copies were quantified by real-time RT-PCR. Y axis indicates log10 viral RNA copies/μL culture supernatant. Error bars indicate the standard deviation.

REFERENCES

Baric, R. S., Sullivan, E., Hensley, L., Yount, B., and Chen, W. (1999). Persistent infection promotes cross-species transmissibility of mouse hepatitis virus. *J. Virol.* 73, 638–649. doi: 10.1128/jvi.73.1.638-649.1999

Belouzard, S., Millet, J. K., Licitra, B. N., and Whittaker, G. R. (2012). Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* 4, 1011–1033. doi: 10.3390/v4061011

Bickerton, E., Dowgier, G., and Britton, P. (2018a). Recombinant infectious bronchitis viruses expressing heterologous S1 subunits: potential for a new generation of vaccines that replicate in Vero cells. *J. Gen. Virol.* 99, 1681–1685. doi: 10.1099/jgv.0.001167

Bickerton, E., Maier, H. J., Stevenson-Leggett, P., Armesto, M., and Britton, P. (2018b). The S2 subunit of infectious bronchitis virus beaudette is a determinant of cellular tropism. *J. Virol.* 92:e01044-18. doi: 10.1128/JVI.01044-18

Bosch, B. J., van der Zee, R., de Haan, C. A. M., and Rottier, P. J. M. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* 77, 8801–8811. doi: 10.1128/jvi.77.16.8801-8811.2003

Bouwman, K. M., Parsons, L. M., Berends, A. J., de Vries, R. P., Cipollo, J. F., and Verheije, M. H. (2020). Three amino acid changes in avian coronavirus spike protein allow binding to kidney tissue. *J. Virol.* 94:e01363-19. doi: 10.1128/JVI.01363-19

Casais, R., Dove, B., Cavanagh, D., and Britton, P. (2003). Recombinant avian infectious bronchitis virus expressing a heterologous spike gene demonstrates that the spike protein is a determinant of cell tropism. *J. Virol.* 77, 9084–9089. doi: 10.1128/jvi.77.16.9084-9089.2003

Cavanagh, D. (1983). Coronavirus IBV: structural characterization of the spike protein. *J. Gen. Virol.* 64(Pt 12), 2577–2583. doi: 10.1099/0022-1317-64-12-2577

Cavanagh, D. (2007). Coronavirus avian infectious bronchitis virus. *Vet. Res.* 38, 281–297. doi: 10.1051/vetres:2006055

Cheng, J., Zhao, Y., Xu, G., Zhang, K., Jia, W., Sun, Y., et al. (2019). The S2 subunit of QX-type infectious bronchitis coronavirus spike protein is an essential determinant of neurotropism. *Viruses* 11:972. doi: 10.3390/v11100972

Desmarets, L. M. B., Theuns, S., Roukaerts, I. D. M., Acar, D. D., and Nauwynck, H. J. (2014). Role of sialic acids in feline enteric coronavirus infections. *J. Gen. Virol.* 95(Pt 9), 1911–1918. doi: 10.1099/vir.0.064717-0

Fang, S. G., Shen, S., Tay, F. P., and Liu, D. X. (2005). Selection of and recombination between minor variants lead to the adaptation of an avian coronavirus to primate cells. *Biochem. Biophys. Res. Commun.* 336, 417–423. doi: 10.1016/j.bbrc.2005.08.105

Gombold, J. L., Hingley, S. T., and Weiss, S. R. (1993). Fusion-defective mutants of mouse hepatitis virus A59 contain a mutation in the spike protein cleavage signal. *J. Virol.* 67, 4504–4512. doi: 10.1128/jvi.67.8.4504-4512.1993

Graham, R. L., and Baric, R. S. (2010). Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J. Virol.* 84, 3134–3146. doi: 10.1128/JVI.01394-09 doi: 10.1128/jvi.01394-09

Heald-Sargent, T., and Gallagher, T. (2012). Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses* 4, 557–580. doi: 10.3390/v4040557

Hulswit, R. J., de Haan, C. A., and Bosch, B. J. (2016). Coronavirus spike protein and tropism changes. *Adv. Virus Res.* 96, 29–57. doi: 10.1016/bs.aivir.2016.08.004

Jiang, Y., Cheng, X., Zhao, X., Yu, Y., Gao, M., and Zhou, S. (2020). Recombinant infectious bronchitis coronavirus H120 with the spike protein S1 gene of the nephropathogenic IBYZ strain remains attenuated but induces protective immunity. *Vaccine* 38, 3157–3168. doi: 10.1016/j.vaccine.2020.01.001

Jiang, Y., Zhou, S., Yu, Y., Tang, M. J., Cheng, X., and Zhao, X. M. (2017). Cell tropism of recombinant chicken infectious bronchitis virus expressing heterologous S gene. *Chinese J. Virol.* 33, 732–744.

Kleine-Weber, H., Elzayat, M. T., Hoffmann, M., and Pohlmann, S. (2018). Functional analysis of potential cleavage sites in the MERS-coronavirus spike protein. *Sci. Rep.* 8:16597. doi: 10.1038/s41598-018-34859-w

Krempl, C., Schultze, B., Laude, H., and Herrler, G. (1997). Point mutations in the S protein connect the sialic acid binding activity with the enteropathogenicity of transmissible gastroenteritis coronavirus. *J. Virol.* 71, 3285–3287. doi: 10.1128/jvi.71.4.3285-3287.1997

Kuo, L., Godeke, G. J., Raamsman, M. J., Masters, P. S., and Rottier, P. J. (2000). Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J. Virol.* 74, 1393–1406. doi: 10.1128/jvi.74.3.1393-1406.2000

Li, F. (2012). Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J. Virol.* 86, 2856–2858. doi: 10.1128/JVI.06882-11

Li, F. (2015). Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J. Virol.* 89, 1954–1964. doi: 10.1128/JVI.02615-14

Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., et al. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24, 1634–1643. doi: 10.1038/sj.emboj.7600640

Matsuyama, S., and Taguchi, F. (2009). Two-step conformational changes in a coronavirus envelope glycoprotein mediated by receptor binding and proteolysis. *J. Virol.* 83, 11133–11141. doi: 10.1128/JVI.00959-09

McRoy, W. C., and Baric, R. S. (2008). Amino acid substitutions in the S2 subunit of mouse hepatitis virus variant V51 encode determinants of host range expansion. *J. Virol.* 82, 1414–1424. doi: 10.1128/JVI.01674-07

Millet, J. K., and Whittaker, G. R. (2015). Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* 202, 120–134. doi: 10.1016/j.virusres.2014.11.021

Pedersen, N. C. (2014). An update on feline infectious peritonitis: virology and immunopathogenesis. *Vet. J.* 201, 123–132. doi: 10.1016/j.tvjl.2014.04.017

Qu, X. X., Hao, P., Song, X. J., Jiang, S. M., Liu, Y. X., Wang, P. G., et al. (2005). Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *J. Biol. Chem.* 280, 29588–29595. doi: 10.1074/jbc.M500662200

Schultze, B., Cavanagh, D., and Herrler, G. (1992). Neuraminidase treatment of avian infectious bronchitis coronavirus reveals a hemagglutinating activity that is dependent on sialic acid-containing receptors on erythrocytes. *Virology* 189, 792–794. doi: 10.1016/0042-6822(92)90608-r

Tian, S., Huang, Q., Fang, Y., and Wu, J. (2011). FurinDB: a database of 20-residue furin cleavage site motifs, substrates and their associated drugs. *Int. J. Mol. Sci.* 12, 1060–1065. doi: 10.3390/ijms12021060

Walls, A. C., Tortorici, M. A., Snijder, J., Xiong, X., Bosch, B. J., Rey, F. A., et al. (2017). Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* 114, 11157–11162. doi: 10.1073/pnas.1708727114

Wei, Y. Q., Guo, H. C., Dong, H., Wang, H. M., Xu, J., Sun, D. H., et al. (2014). Development and characterization of a recombinant infectious bronchitis virus expressing the ectodomain region of S1 gene of H120 strain. *Appl. Microbiol. Biotechnol.* 98, 1727–1735. doi: 10.1007/s00253-013-5352-5

Wickramasinghe, I. N., de Vries, R. P., Grone, A., de Haan, C. A., and Verheije, M. H. (2011). Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J. Virol.* 85, 8903–8912. doi: 10.1128/JVI.05112-11

Winter, C., Herrler, G., and Neumann, U. (2008). Infection of the tracheal epithelium by infectious bronchitis virus is sialic acid dependent. *Microbes Infect.* 10, 367–373. doi: 10.1016/j.micinf.2007.12.009

Winter, C., Schwegmann-Wessels, C., Cavanagh, D., Neumann, U., and Herrler, G. (2006). Sialic acid is a receptor determinant for infection of cells by avian Infectious bronchitis virus. *J. Gen. Virol.* 87(Pt 5), 1209–1216. doi: 10.1099/vir.0.81651-0 doi: 10.1099/vir.0.81651-0

Xia, S., Xu, W., Wang, Q., Wang, C., Hua, C., Li, W., et al. (2018). Peptide-based membrane fusion inhibitors targeting HCoV-229E spike protein HR1 and HR2 domains. *Int. J. Mol. Sci.* 19:487. doi: 10.3390/ijms19020487

Yamada, Y., Liu, X. B., Fang, S. G., Tay, F. P., and Liu, D. X. (2009). Acquisition of cell-cell fusion activity by amino acid substitutions in spike protein determines the infectivity of a coronavirus in cultured cells. *PLoS One* 4:e6130. doi: 10.1371/journal.pone.0006130

Yamada, Y. K., Takimoto, K., Yabe, M., and Taguchi, F. (1997). Acquired fusion activity of a murine coronavirus MHV-2 variant with mutations in the proteolytic cleavage site and the signal sequence of the S protein. *Virology* 227, 215–219. doi: 10.1006/viro.1996.8313

Yang, Y., Liu, C., Du, L., Jiang, S., Shi, Z., Baric, R. S., et al. (2015). Two mutations were critical for bat-to-human transmission of middle east respiratory syndrome coronavirus. *J. Virol.* 89, 9119–9123. doi: 10.1128/JVI.01279-15

Zhou, S., Dai, Y. B., Liu, M., Zhao, B. H., Cheng, X., and Lv, X. J. (2011). Expression of green fluorescent protein using an infectious cDNA clone of infectious bronchitis virus. *Chinese J. Virol.* 27, 11–17.

Zhou, S., Tang, M. J., Liu, M., Cheng, X., and Lv, X. J. (2010). Construction of an infectious full-length cDNA clone of infectious bronchitis virus H120 vaccine strain. *China Poultry* 32, 22–26.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fmicb.2020.600254

Addressing Non-linear System Dynamics of Single-Strand RNA Virus–Host Interaction

Alessandra Romano[1,2], Marco Casazza[2] and Francesco Gonella[3]*

[1] Sezione di Ematologia, Dipartimento di Chirurgia Generale e Specialità Medico Chirurgiche (CHIRMED), Università degli Studi di Catania, Catania, Italy, [2] Division of Hematology, U.O.C di Ematologia, Azienda Ospedaliero Universitaria Policlinico "G.Rodolico - San Marco", Catania, Italy, [3] Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari Venezia, Venezia, Italy

Positive single-strand ribonucleic acid [(+)ssRNA] viruses can cause multiple outbreaks, for which comprehensive tailored therapeutic strategies are still missing. Virus and host cell dynamics are tightly connected, generating a complex dynamics that conveys in virion assembly to ensure virus spread in the body. Starting from the knowledge of relevant processes in (+ss)RNA virus replication, transcription, translation, virions budding and shedding, and their respective energy costs, we built up a systems thinking (ST)–based diagram of the virus–host interaction, comprehensive of stocks, flows, and processes as well-described in literature. In ST approach, stocks and flows are expressed by a proxy of the energy embedded and transmitted, respectively, whereas processes are referred to the energy required for the system functioning. In this perspective, healthiness is just a particular configuration, in which stocks relevant for the system (equivalent but not limited to proteins, RNA, DNA, and all metabolites required for the survival) are constant, and the system behavior is stationary. At time of infection, the presence of additional stocks (e.g., viral protein and RNA and all metabolites required for virion assembly and spread) confers a complex network of feedbacks leading to new configurations, which can evolve to maximize the virions stock, thus changing the system structure, output, and purpose. The dynamic trajectories will evolve to achieve a new stationary status, a phenomenon described in microbiology as integration and symbiosis when the system is resilient enough to the changes, or the system may stop functioning and die. Application of external driving forces, acting on processes, can affect the dynamic trajectories adding a further degree of complexity, which can be captured by ST approach, used to address these new configurations. Investigation of system configurations in response to external driving forces acting is developed by computational analysis based on ST diagrams, with the aim at designing novel therapeutic approaches.

Keywords: systems thinking (ST), RNA-virus, virus–host interaction, dynamics, modeling, simulation – computers, evolution trajectories

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Israel Pagan,
Polytechnic University of Madrid, Spain
Caio C. M. Freire,
Federal University of São Carlos, Brazil

****Correspondence:***
Marco Casazza
casazzamarco@gmail.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *29 August 2020*
Accepted: *09 December 2020*
Published: *15 January 2021*

Citation:
Romano A, Casazza M and Gonella F (2021) Addressing Non-linear System Dynamics of Single-Strand RNA Virus–Host Interaction. Front. Microbiol. 11:600254. doi: 10.3389/fmicb.2020.600254

INTRODUCTION

Positive single-stranded ribonucleic acid [(+)ssRNA] viruses, including picornaviruses, flaviviruses, Togaviridae, and human coronaviruses (CoVs) (Ahlquist et al., 2003; Lam et al., 2016; Scutigliani and Kikkert, 2017; Primadharsini et al., 2019), cause multiple outbreaks, for which tailored antiviral strategies are still missing (Zumla et al., 2016; Dinesh et al., 2020; Gordon et al., 2020b). (+)ssRNA viruses package their genomes as messenger sense, single-stranded RNA and replicate those genomes solely through RNA intermediates in the cytosol of the host cells (Den Boon et al., 2010). RNA-dependent RNA polymerases lack coreplicative and postreplicative fidelity-enhancing pathways; this final RNA genome copies incorporate mutations at a high rate (Lauring et al., 2013; Acevedo et al., 2014), providing the viral quasi-species with a higher probability to evolve and adapt to new environments and challenges during infection (Burch and Chao, 2000; Vignuzzi et al., 2006). The diversity is essential for both viral fitness (Wargo and Kurath, 2012) and pathogenesis because of the complex relationships among virus replication (VR), host cells, and immune system, as almost all (+)ssRNA viruses can delay antiviral innate immune response (Kobasa et al., 2007) in multiple ways (Diebold et al., 2003; Hogan et al., 2004; Meylan et al., 2005; Saito et al., 2008; Yang et al., 2015; Beachboard and Horner, 2016; Nelemans and Kikkert, 2019). Host immunogenetic factors can be sensitive to a variation in the viral load, leading to a defective response of the innate immunity, that could explain the variable clinical course of infection (Fanning et al., 2001; Nelemans and Kikkert, 2019).

Recent studies confirmed the complexity of viral dynamics, whose fitness is improved by the complex interactions with the host proteins (Bosl et al., 2019; Sruthi and Prakash, 2019), as previously described in modeling the virus–host interactions at subcell and cell levels (Dapat and Oshitani, 2016; Jonsdottir and Dijkman, 2016; Gao et al., 2017; Patzina et al., 2017; Gordon et al., 2020b). However, models that address a specific aspect of the virus–host interaction do not capture the wide range of intertwined spatial and temporal (hours to days) dynamic scales (Apweiler et al., 2018), which are related to the interactions of different concurrent hierarchical levels. For this reason, we aimed at describing the virus–host cell interaction as a dynamic system by a systems thinking (ST) approach (Northridge and Metcalf, 2016).

In ST, the behavior of a dynamic system can be described and predicted by the temporal evolution of its configurations, given by hierarchical feedback loops and self-organization. The configuration evolution then can be analytically computed by proper simulators (Odum and Odum, 2000; Tegner et al., 2009; Hassmiller Lich et al., 2017; Spill et al., 2018) to further address suitable leverage points for intervening (Meadows and Wright, 2008). In particular, dynamic models, where the temporal evolution of extensive variables (stocks) is simulated in the form of trajectories, derive their initial conditions from available and assessed evidence. Varying the key system parameters, trajectories represent the possible evolutive (structural) patterns of the system at issue, becoming abstracted with respect to local specific attributes related to single case studies. However, being stocks associated to system observables, the relation with those attributes is maintained, making it possible to compare predicted trajectories with observed data and constituting suitable counterfactuals with respect to the laboratory measurements results.

System dynamics (SD) approach is mostly used for strategic modeling, typically for ecological and socioeconomic systems, to understand the supply chain performance. In particular, results of SD modeling provide a set of alternative evolutive patterns in form of graphs, capturing the internal dynamics of a system even in lack of some experimental data to fit. These can provide alternative scenarios that, when fitting experimental evidences, indicate the most effective leverage points to control the system evolution.

In this article, we show that, by approaching the host–virus interaction as a dynamic systemic problem (Sterman, 2002; Meadows and Wright, 2008), it is possible to identify potential systemic leverage points to minimize the release of virions, so addressing effective *systemic* intervention strategies.

MATERIALS AND METHODS

Development of a Stock-Flow Diagram

The basic SD element is the stock and flow diagram. Stocks are countable extensive variables Q_i, $i = 1, 2, \ldots, n$, relevant to the study at issue, that constitute an n-ple of numbers (possibly derived from experimental measurements), which at any time represents the state of the system. A stock may change its value only upon its inflows and/or its outflows, represented by arrows entering or exiting the stock. Processes are any occurrence capable to alter—either quantitatively or qualitatively—a flow, by the action of one or more of the system elements. In a stationary state of the system, stock values are either constant or regularly oscillating. Processes, which cause the stationarity or perturbation of a system, must be activated by a driver, acting on the flows where the process is located. The pattern of the feedbacks acting in the system configurations is the feature that ultimately defines the systems dynamics. Each flow depends on the state variables Q_i by relationships of the kind $dQ_i/dt = kf(Q_j)$, $i, j = 1, \ldots, n$, where n is the number of stocks in the system.

The stocks and flows inventory reported in **Table 1** was based on information from existing knowledge on the biological mechanisms at issue, listing the variables and parameters necessary to set up the equations describing the system dynamics. Turnover times of stocks included in the RNA-virus–host interaction ST diagram have been reported in **Table 2**, derived from the available literature (Konig et al., 2008; Friedel and Haas, 2011; Pfefferle et al., 2011; Jourdan et al., 2012; Munday et al., 2012; Naji et al., 2012; Wu et al., 2012; Emmott et al., 2013; Garcia-Dorival et al., 2014; Verchot, 2014; Watanabe et al., 2014; York et al., 2014; Zheng et al., 2014; Dong et al., 2016; Kuo et al., 2016, 2018; Wang et al., 2016, 2017a,b; Gao et al., 2017; Hafirassou et al., 2017; Khamina et al., 2017; King et al., 2017; Martinez-Gil et al., 2017; Patzina et al., 2017; Coyaud et al., 2018; Iwasaki et al., 2018; Lescar et al., 2018; Ziegler et al., 2018; Bosl et al., 2019; Chakravorty et al., 2019; Garcia-Moreno et al., 2019;

TABLE 1 | Inventory of stocks and flows depicted in the diagram of **Figure 1**.

Stock	Biological meaning	Dynamic equation*	Calibration value	References
Q_1	Resources available for protein synthesis	$dQ_1/dt = J_0 + J_{21A} + J_{21B} - J_1 - J_{13} - J_{15} - J_{17}$	3.9	Odum, 1996, 2002; Odum and Odum, 2000
Q_{2A}	Short-half-life proteins	$dQ_{2A}/dt = J_{2A} - J_{21A} - J_{20A}$	13	Wheatley et al., 1980; Eden et al., 2011; Siwiak and Zielenkiewicz, 2013
Q_{2B}	Long-half-life proteins	$dQ_{2B}/dt = J_{2B} - J_{21B} - J_{23} - J_{25} - J_{27} - J_{20B}$	13	Wheatley et al., 1980; Eden et al., 2011; Siwiak and Zielenkiewicz, 2013
Q_3	Viral ss + RNA	$dQ_3/dt = J_3 - J_4$	3	Baccam et al., 2006
Q_4	Viral proteins	$dQ_4/dt = J_4 - J_5$	0.024	Kummer et al., 2014
Q_5	Virions	$dQ_5/dt = J_6 - J_7$	0	NA
Flow		**Biological role**	**Phenomenological coefficients (*k*)**	**References**
$J_0 = k_0 \times R \times (1 + Q_{2A})$		Enter of resources allocated for protein synthesis	3.9E-06	Odum, 1996
$J_1 = k_1 \times Q_1$		Host-cell RNA transcription and translation	6.9E-05	Eden et al., 2011; Kummer et al., 2014
$J_{2A} = k_{2A} \times Q_1$		Short-half-life protein synthesis	2.6E-05	Adelman et al., 2002; Klumpp and Hwa, 2008; Eden et al., 2011
$J_{2B} = k_{2B} \times Q_1 \times (1 + Q_4)$		Long-half-life protein synthesis	1.6E-05	Proshkin et al., 2010; Eden et al., 2011
$J_3 = k_3 \times Q_1 \times Q_{2B} \times Q_3 \times Q_5$		Virus-RNA replication	1.0E-01	Mahmoudabadi et al., 2017
$J_4 = k_4 \times Q_3$		Viral RNA translation	6.9E-05	Kummer et al., 2014
$J_5 = k_5 \times Q_1 \times Q_{2B} \times Q_4$		Recruitment of resources and host-cell protein machinery for virion assembly	1.7E-03	Mahmoudabadi et al., 2017
$J_6 = k_6 \times Q_1 \times Q_{2B} \times Q_4$		Virion assembly	3.0E-03	Mahmoudabadi et al., 2017
$J_7 = k_7 \times Q_1 \times Q_{2B} \times Q_5$		Virion budding	8.3E-04	Mahmoudabadi et al., 2017
$J_{13} = k_{13} \times Q_{2B} \times Q_1 \times Q_3 \times Q_5$		Flow of host-cell resources diverted to let virus enter	8.3E-02	Mahmoudabadi et al., 2017
$J_{15} = k_{15} \times Q_1 \times Q_4 \times Q_{2B}$		Flow of host-cell resources diverted to let virion assembly	1.7E-03	Mahmoudabadi et al., 2017
$J_{17} = k_{17} \times Q_1 \times Q_5 \times Q_{2B}$		Flow of host-cell resources diverted to let virion shedding	8.3E-04	Mahmoudabadi et al., 2017
$J_{20A} = k_{20A} \times Q_{2A}$		Flow of host-cell short-half-life proteins addressed to degradation	3.9E-06	Eden et al., 2011; Boisvert et al., 2012
$J_{20B} = k_{20B} \times Q_{2B}$		Flow of host-cell long-half-life proteins addressed to degradation	1.6E-06	Eden et al., 2011; Boisvert et al., 2012
$J_{21A} = k_{21A} \times Q_{2A}$		Proteostasis mechanisms, including proteasome degradation and autophagy to re-cycle unfolded, old or not functional host-cell short-half-life proteins	3.9E-06	Eden et al., 2011; Boisvert et al., 2012
$J_{21B} = k_{21B} \times Q_{2B}$		Proteostasis mechanisms, including proteasome degradation and autophagy of to re-cycle unfolded, old or not functional host-cell long-half-life proteins	3.9E-06	Eden et al., 2011; Boisvert et al., 2012
$J_{23} = k_{23} \times Q_1 \times Q_{2B} \times Q_3 \times Q_5$		Flow of host-cell proteins recruited to let virus enter and RNA transcription	8.3E-02	Mahmoudabadi et al., 2017
$J_{25} = k_{25} \times Q_1 \times Q_{2B} \times Q_4$		Flow of host-cell proteins recruited to let virion assembly	1.7E-03	Mahmoudabadi et al., 2017
$J_{27} = k_{27} \times Q_1 \times Q_{2B} \times Q_5$		Flow of host-cell proteins recruited to let virion budding	1.7E-03	Mahmoudabadi et al., 2017
$J_{35} = k_{35} \times Q_4$		Flow of viral RNA to embed in the virion	4.6E-05	Mahmoudabadi et al., 2017
$J_{50} = k_{50} \times Q_1 \times Q_{2B} \times Q_5$		Virion shedding	4.6E-05	Mahmoudabadi et al., 2017

**Equations representing the dynamics of the stocks and calibration values are also shown. Calibration value is expressed in $\times 10^{12}$ ATP-eq.*

TABLE 2 | Relevant flows included for the diagram and simulator development.

Flow	Description	Value	Stock turnover time	References
J_1	Host protein translation		4 h	Kummer et al., 2014
$J_1 + J_{2A}$	Protein synthesis rate	10–20 aa/s		Kudva et al., 2013
$J_1 + J_{2B}$	Protein synthesis rate	10–20 aa/s		Kudva et al., 2013
J_{2A}	Host protein transcription and translation (long-half-life proteins)	13 aa/s		Adelman et al., 2002; Klumpp and Hwa, 2008
J_{2A}	Host protein transcription and translation (long-half-life proteins)	2–500 (140) mRNA/h 1,000/proteins/mRNA/h 660 mRNA/h/cell	1 h	Howard-Ashby et al., 2006; Ben-Tabou De-Leon and Davidson, 2009; Schwanhausser et al., 2011
J_{2A}	Host protein transcription and translation (long-half-life proteins)	60,200 mRNA/h to convert in ATP eq		Pelechano et al., 2010
J_{2B}	Host protein transcription and translation (short half-life proteins)	42 nt/s + 14 aa/s		Proshkin et al., 2010
$J_{2A} + J_{2B}$		15% 20% respiration rate	0.99 μmol/*kg*/day	Buttgereit and Brand, 1995; Waterlow, 2006
J_{20}	Protein degradation	0.08/h		Doherty et al., 2009
J_{20}	Turnover rate of protein	20/hs		Boisvert et al., 2012
J_{13}	Virus transcription and replication	2,500 nt/s	12 s	Mahmoudabadi et al., 2017
J_3	Virus transcription and replication	2,500 nt/s	12 s	Mahmoudabadi et al., 2017
J_4	Viral protein translation		4 h	Kummer et al., 2014
J_5	Virion assembly		10 min	Mahmoudabadi et al., 2017
J_6	Virion assembly		10 min	Mahmoudabadi et al., 2017
J_7	Virion budding		20 min	Mahmoudabadi et al., 2017
J_{50}	Virion shedding	24% in 21.6 h	21.6 h	Murphy et al., 1980
J_{15}	Virion assembly		10 min	Mahmoudabadi et al., 2017
J_{17}	Virion budding		20 min	Mahmoudabadi et al., 2017
J_{23}	Virus transcription and replication	2,500 nt/s	12 s	Mahmoudabadi et al., 2017
J_{25}	Virion assembly			Mahmoudabadi et al., 2017
J_{27}	Virion budding		20 min	Mahmoudabadi et al., 2017
J_{21}	Protein recycling of short and long half unfolded-defective proteins	0.08/h		Doherty et al., 2009
$J_3 + J_4 + J_5 + J_6 + J_7 + J_{50}$	vRNA replication and virion shedding, virus production after infection		6 h	Sedmak and Grossberg, 1973; Baccam et al., 2006

Rothenburg and Brennan, 2020). All stocks, flows, and processes were expressed using a common proxy unit, representing the energy embedded, transmitted, and used, respectively, during the system operation. The proxy unit was expressed as the number of ATP (and ATP-equivalent) hydrolysis events (Mahmoudabadi et al., 2017). This choice allowed calculating each parameter of the system on the basis of stocks and characteristic times of the flows derived from the literature, without further need for experimental data.

Development of the Virus–Host Interaction Systemic Simulator

After setting the initial conditions at time 0 for the stocks, system solutions were obtained using recursive computation for a relative short period of time (identified with the median life of an epithelial cell, 7 days), in order to appreciate the model dynamic behavior. The computational model based on a set of differential equations that describe the rates of change of all stocks in the ST diagram (Odum and Odum, 2000; Bossel, 2007) was developed using the open-source software package SCILAB[1], which uses approximation techniques to evaluate stocks.

Given a set of initial conditions for the stocks (i.e., the initial state of the system) and a set of phenomenological coefficients k associated to flows, the set of interconnected equations was treated by a standard finite-different method, taking care of choosing a time step short enough to evidence the dynamics of

[1]https://www.scilab.org

any of the studied processes. The coefficients k_i were calculated, on the basis of literature data, considering the dynamics of each single stock, by quantifying flows and stocks during the time interval set as simulation step, as shown in **Table 1**. When different flows coparticipate in a process, each coefficient gathers all the actions that concur to the intensity of the outcoming flow(s). In detail, the parameters used to run the model (i.e., the set of values for the k_i coefficients), describing the reaction of each system component to a change in any other one, were derived from the stocks turnover times (Odum and Odum, 2000; Bossel, 2007). Therefore, the host–virus interaction computational model, built on experimental evidences as listed in **Tables 1**, **2**, is not specific for a unique virus, but may represent the patterns of any virus–host interaction, in which stocks, flows, and processes are those relevant for the operation of the system at issue.

The reliability of both available data and modeling was tested by evaluating the effect of the variation of each of the most relevant input data (stocks and processes) on the system trajectories. Unfortunately, here is not a single comprehensive procedure suitable for the validation of all dynamic models, being dependent on their usefulness, in turn referred to the very purpose of the model itself (Grüne-Yanoff and Weirich, 2010). We chose the sensitivity analysis approach (Qudrat-Ullah, 2012; Hekimoğlu and Barlas, 2016), which allows to see to what extent a variation on these values can lead to alternative evaluations of the system dynamics. In particular, we applied a 50% variation (either positive or negative) to those parameters that the results were more sensitive to. As expected, while the corresponding simulations varied as well, the general patterns presented in the following remained the same, especially concerning the overall trends shown by comparing the groups of simulations, providing a model validation.

RESULTS

Stock-Flow Diagram of (+)ssRNA Virus–Host Interaction

First, we identified the important structures in the system and then used to build up the stock-flow diagram of the virus–host interaction system. In **Figure 1**, symbols were borrowed from the energy language (Odum and Odum, 2000; Brown, 2004): shields indicate the stocks; big solid arrows indicate the processes; line arrows indicate the flows; dashed lines show the controls exerted by the stocks on the processes.

The dynamics of energy allocation for protein synthesis contained in the stock Q_1 depended on the cell bioenergetics, e.g., the number of mitochondria, OX-PHOS activity levels, and cell cycle phase (Murayama et al., 2008; Canto et al., 2009; Li et al., 2020). In the absence of virus, the stationary configuration was given by energy required to flow from stock Q_1 (via J_1, J_{2A}, and J_{2B} flows) to stocks Q_{2A} and Q_{2B} to produce, respectively, short- and long-half-life proteins, which could, in turn, be recruited by VR machinery. J_{20A} and J_{20B}, grouped into the flow J_{20}, represented the outflow of folded, fully functional proteins addressed to secretion or surface exposure. Based on basal proteostasis of host cell, recovery of energetic sources from proteins not addressed to leave the system could be possible via several complex processes (e.g., proteasomal degradation, and autophagy), identified by flows of materials J_{21A} and J_{21B}, respectively, from Q_{2A} and Q_{2B} back to Q_1.

The viral load in the system, expressed by the stocks Q_3 (identified as viral RNA content to be used for viral transcription and translation), Q_4 (translated viral proteins content), and Q_5 (full assembled virions to shed virus outside), diverted, at the time of infection, resources directly from Q_1 (through flows J_{13}, J_{15}, and J_{17}) and Q_{2B} (through flows J_{23}, J_{25}, and J_{27}). Virions shedding was represented by the flows J_7 and J_{50} through the contribution of the host flows J_{17} and J_{27}. The output flows J_4 and J_5 were set to be effective only if the value of the respective stock Q_3 and Q_5 was higher than a threshold, as represented by the two switch symbols in the diagram.

We identified four feedback loops (represented by dot lines in **Figure 1**): (i) the positive control of Q_{2A} stock on the energy supply process (occurring when more structural host proteins operate to maintain the energetics homeostasis of the host cell); (ii) the positive control of Q_3 stock on the VR process (highlighting that the more viral RNA is in the system, the more intensive replication can occur if host sources are available); (iii) the positive control of Q_4 stock on the processes of synthesis and maturation of host proteins (highlighting that the more viral proteins are made, the more host proteins are synthesized to be recruited in the virion assembly machinery, increasing J_{2B}); (iv) the positive control of Q_5 stock on the VR process (highlighting that the more virions are produced, the more resources are diverted from the host cell to viral replication).

System Dynamics of (+)ssRNA Virus–Host Interaction

First, a computational model was derived from the stock-flow diagram shown in **Figure 1** using the standardized workflow of systemic modeling (Odum and Odum, 2000; Xue et al., 2018). **Figure 2** shows two different system self-organized patterns (configurations) to guide reader in the overall comprehension of the proposed approach. The virus–host interaction was represented as an evolving set of simulated trajectories, to which the positive value of Q_3 stock had given access, using preexisting stocks, processes, and flows of the host cell, followed overtime by progressive filling of Q_4 and Q_5 stocks. In (**Figure 2A**) configuration, the viral load is null (the stocks Q_3, Q_4, and Q_5 are empty), and the values of stocks Q_1, Q_{2A}, and Q_{2B} are constant; thus, the system behavior is stationary (**Figure 2B**). At time of infection, the Q_3 stock was fed, and its proteins could interact with the host proteome to sustain RNA replication. Based on previous works in the field (Wei et al., 1995; Adelman et al., 2002; Mohler et al., 2005; Regoes et al., 2005; De Boer et al., 2010), we identified a time delay of 2–6 h required to record changes in the Q_5 stock.

Moreover, the value of Q_5 varied over time due to changes that occurred at different timepoints in the stocks Q_{2B}, Q_3, and Q_4. Thus, the network of flows and feedbacks could identify a new configuration (**Figure 2C**), to generate a non-stationary

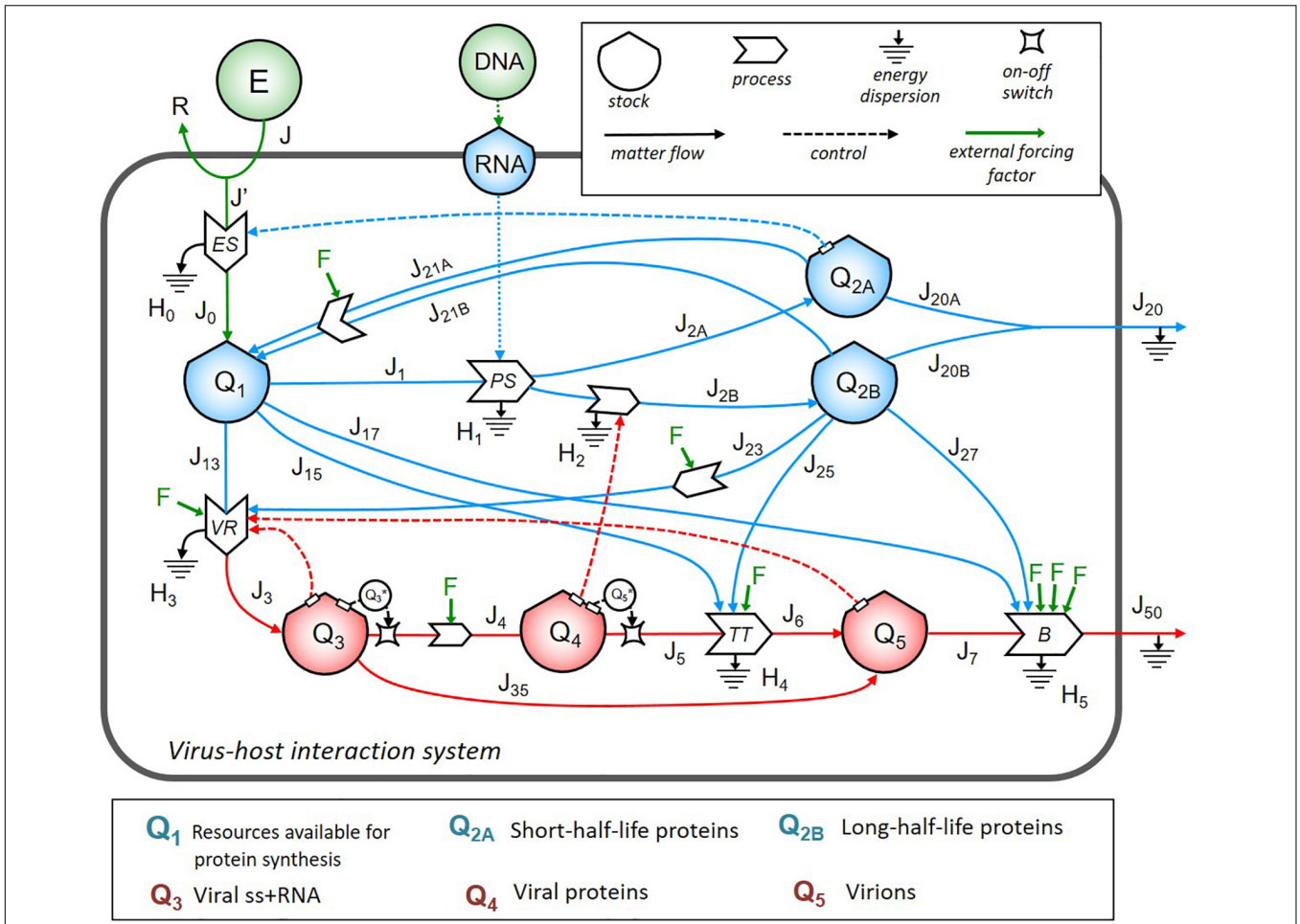


FIGURE 1 | The energy systemic diagram of a cell infected by ss+ RNA virus. Stock-flow diagram of the virus–host interaction system. In the upper right box are the meaning of symbols. The color code is as follows: blue for host cell energy stocks and relative inflows and outflows; red for virus energy stocks and relative inflows and outflows; green for external energy inputs and external driving forces *F* corresponding to different therapeutic strategies. The lower box lists the biological contents of the stocks, all expressed in terms of energy (ATP-equivalent units).

behavior (**Figure 2D**), where the values of Q_3, Q_4, and Q_5 stocks evolved in a non-linear way (**Supplementary Figure 1**), to maximize the value of virions stock in the configuration (**Figure 2E**). We define the (**Figure 2A**) configuration as healthy, the (**Figure 2C**) configuration as early infection associated to asymptomatic disease, the (**Figure 2E**) configuration as late infection associated to symptomatic disease, and the (**Figure 2F**) configuration as symbiotic infection, consequent to any approach derived from the application of external driving forces at any time able to maintain configuration (**Figure 2C**) without crashing the system. The goal for any curative approach should be to recover the (**Figure 2A**) configuration when an infection occurs.

Second, we investigated the system dynamics under different initial conditions, exploring the possible role of different initial viral loads (**Figure 3**). Assuming different initial viral loads (10–10,000 RNA copies range), we found a threshold (at about 5,000 RNA copies) for triggering the progressive reduction of Q_1 (**Figure 3**). Indeed, for low initial viral load (10–1,000 RNA copies), the system perturbation could be absorbed by the configuration itself (**Supplementary Figure 1**), without affecting the overtime stock value of Q_1 and Q_{2B} but maintaining constant Q_3, Q_4, and Q_5.

The behavior of stock values Q_1, Q_{2A}, and Q_{2B} diverged non-linearly at the threshold value, with a progressive decrease, starting at Day 3 from infection (cyan and red lines, respectively). We found a non-linear evolution of the system output (the Q_5 stock) depending on the initial conditions: for Q_3 stock in the range (10–1,000 RNA copies), the Q_5 was linear, whereas for higher initial viral load, the growth of Q_5 was linear in the first day and non-linear in the further timeframe, associated to a progressive, unpredictable reduction of Q_3–Q_4 stocks, reflecting in biological terms the turnover of viral proteins required for virions budding.

The non-linear behavior of the (+)ssRNA virus–host interaction was due to the control of Q_4 stock on processes of the host cell, forcing host proteins Q_{2B} to favor the production of virions (to feed the Q_5 stock value) through the increased J_{23}, J_{25}, and J_{27} flows. In biological terms, the

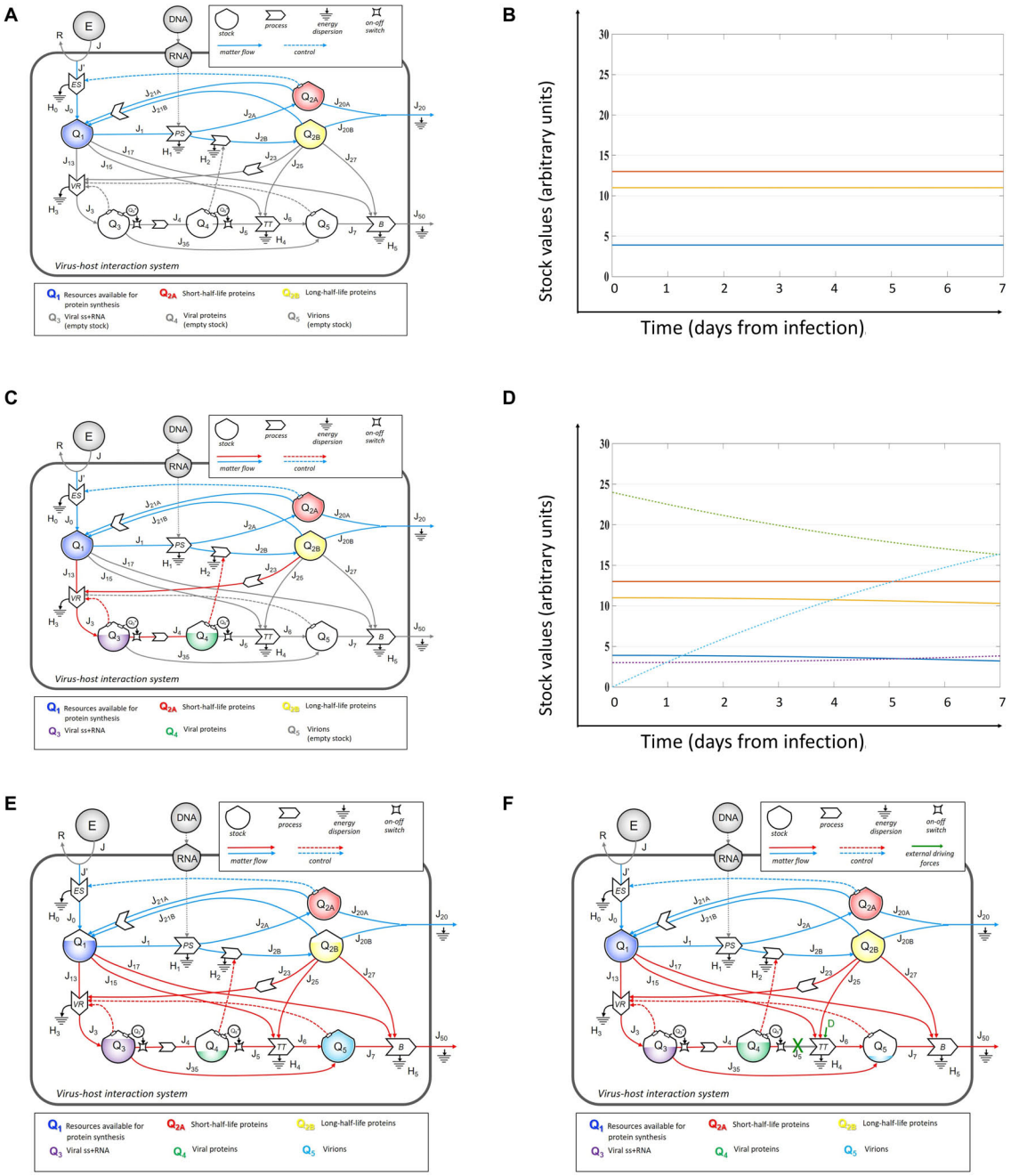


FIGURE 2 | Systems configurations based on initial conditions and effects of external driver forces. In the configuration of initial null viral load **(A)**, the value of stocks Q_1, Q_{2A}, and Q_{2B} were constant, and the system behavior was stationary **(B)**, with constant values of all stocks overtime. At time of infection, the network of flows and feedbacks identified a new configuration **(C)**, to generate a not stationary pattern **(D)**, in which stock values change overtime in response to the other elements of the system, which can evolve to maximize the virions' stock **(E)**. Application of external driving forces, acting on processes (identified by red cross on J_5), can reduce the flows and address new configurations **(F)**, identifying leverage points that can be explored at different magnitude and timepoints with a computational simulator.

progressive Q_1 reduction reflects the metabolic rewiring of infected host cells (Chen et al., 2016), with progressive reduction of resources allocated for the maintenance of host processes, requiring a metabolic shift to less efficient but more rapid source of the available energy required for downstream processes (Thaker et al., 2019). The described system dynamics was experimentally validated for influenza virus (Mahmoudabadi et al., 2017). Results confirmed a previous theoretical assumption, showing a Gibbs free energy for virus lower than its host (Popovic and Minceva, 2020a,b), when virus and host cells are evaluated separately and not as a unique system as herein proposed.

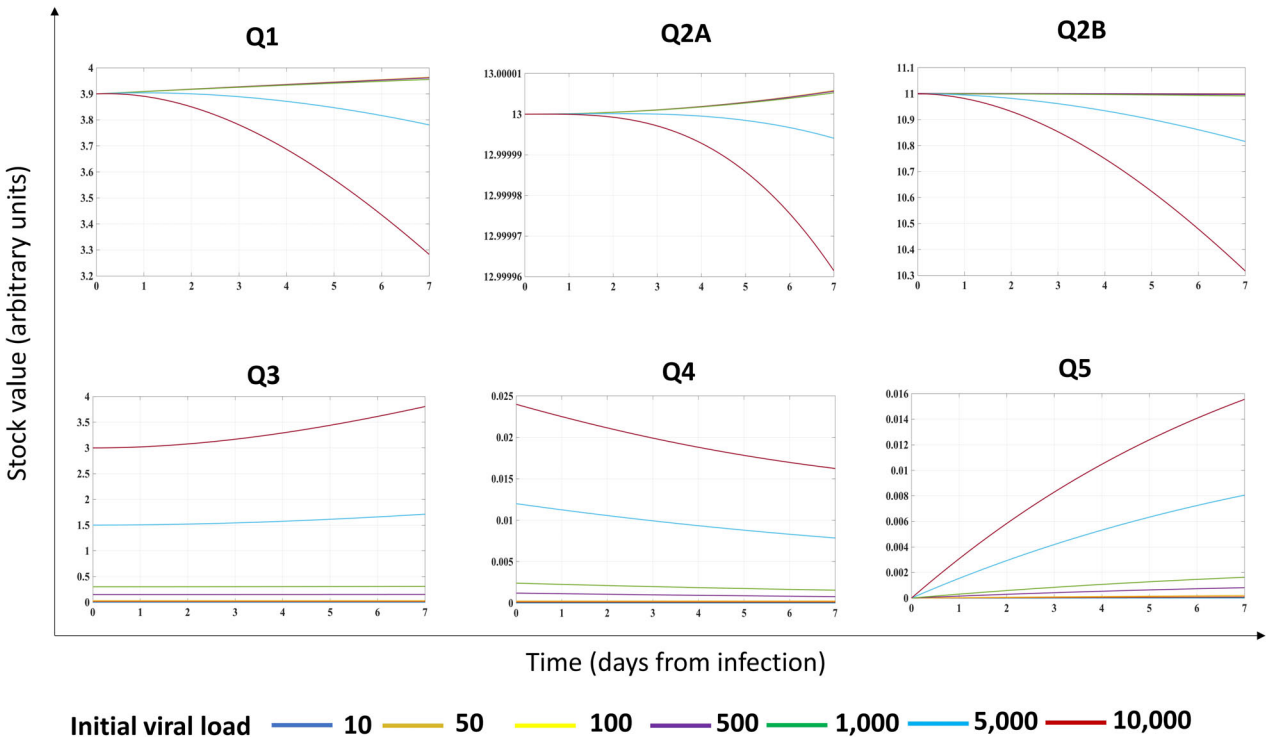


FIGURE 3 | Effects of initial viral load on the energy dynamics of a cell infected by a (+)ssRNA virus. Stock values, expressed in ATP-eq (arbitrary units chosen as proxy), are shown from Day 0 though Day 7 as a function of different initial viral loads (indicated at the bottom with different color codes). Evolutionary pattern for each stock was not linear.

The behavior of Q_1 following different initial values of Q_3 could explain both the variable incubation time in each individual subject and why infections due to (+)ssRNA viruses can occur asymptomatically in most cases. The resilience of virus–host system for a specific range of Q_3 amount could, in turn, depend on intrinsic and extrinsic factors. Indeed, in response to manipulation of stocks and flows, and based on timeframe of observation, the system could evolve along different, non-linear trajectories, requiring early intervention upon infection to make the system resilient to growth of viral stocks.

System Dynamics of (+)ssRNA Virus–Host Interaction in Response to External Driving Forces Applied to Reduce Virions Outflow

Currently, the search for a therapeutic strategy is based on single target-related parameters, while we propose to identify systemic targets (i.e., polytarget) to improve host response to host–virus interaction. Starting from the pharmacodynamics of compounds currently under investigation for a typical (+)ssRNA virus (Thorlund et al., 2020), we could reclassify them, based on their systemic mechanisms of action as listed in **Table 3**. Their effects may be potentially simulated to establish the single-cell effect, the best time, and/or schedule of administration, as shortcut of *in vitro* studies, with a detail level established on the basis of the purpose of the study design. To this end, we applied the search of systemic leverage points by simulating the dynamics of multiple scenarios, upon the action of a generic external driving force (D), assuming that the minimization of the value of Q_5 stock over time should limit the propagation of virions outside the cells.

First, we explored the system configurations upon reduction of the outflows from the Q_5 virions stock, via manipulation of J_7 and/or J_{50}. However, minimization of J_7 was counter-effective, due to the increase of Q_5 as consequence of the feedback action in the VR process (data not shown). The effects of full (100%) or partial (50%) reduction of J_{50} (flow of energy required for virions budding) could prevent the outflow from Q_5 without stopping its growth, diverting resources from Q_1 and Q_{2B} to Q_5, so further supporting viral hijacking of cellular metabolism and impairing host cell homeostasis. As shown in **Figure 4**, the effect of driving forces acting to modulate J_{50} was different based on application time, Day 0 (**Figures 4A–D**) versus Day 1 (**Figures 4E,F**), and initial viral load, as the early abrogation of J_{50} when the amount of Q_3 was 5,000 RNA copies could restore the stationary status (**Figure 4A**), while halving J_{50} maintained homeostasis for host-cell stock, but could not prevent the growth of Q_5 (**Figure 4B**). At increasing initial viral load, reduction of J_{50} applied at Day 0 (**Figures 4C,D**) or at Day 1 (**Figures 4E,F**) could not prevent the growth of Q_5 and the progressive decrease of Q_1 and Q_{2B}. This systemic dynamics can explain the relationship between the time of initiation of neuraminidase inhibitors and their efficacy (Moscona, 2005), as treatment starting within the first 12 h after

TABLE 3 | Examples of drugs that could act as external forcing factors on identified systemic flow targets.

Flow target(s)	Compound	Mechanism of action	References
J_0, J_3, J_4, J_{21A}, J_{21B}	FK506 (tacrolimus)	FKBP15 inhibitor. ER protein quality control regulators. Bioenergetics regulators. mRNA translation inhibitor.	Wiederrecht et al., 1993; Gordon et al., 2020b
J_0, J_4, J_{21A}, J_{21B}	Rapamycin		Jefferies et al., 1997; Wang et al., 2014; Gordon et al., 2020b
J_1, J_3, J_{21B}	Valproic acid	HDAC2 inhibitor	Cong and Bacchetti, 2000; Phiel et al., 2001; Nagesh et al., 2017; Gordon et al., 2020b
J_1, J_4	SAHA	pan HDAC inhibitors	Saha and Parks, 2019
J_1, J_{35}, J_{2B}	Selinexor	mRNA nuclear export complex inhibitor	Bardina et al., 2009; Castelló et al., 2009; Schmidt et al., 2013; Walker et al., 2013; Watters et al., 2017; Gordon et al., 2020b
J_4, J_{2A}, J_{2B}	Dabrafenib	Kinase inhibitor, protein synthesis inhibitor	Burkard et al., 2015; Kindrachuk et al., 2015; Gordon et al., 2020b
J_0, J_{13}	Metformin	Inhibitor of respiratory electron transport, glycolysis regulation	Fontaine et al., 2015; Gordon et al., 2020b
J_3, J_{13}	Camostat, nafamostat	TMPRSS inhibitors proteolytic cleavage of viral spi*k*e protein priming to the receptor ACE2 present in human cell	Gordon et al., 2020b; Kailas et al., 2020
J_4, J_{2A}, J_{2B}	Ponatinib	Kinase inhibitor, protein synthesis inhibitor	Burkard et al., 2015; Gordon et al., 2020b
J_3	Ribavirin	Nucleoside inhibitor (mutagenic ribonucleoside)	Patterson and Fernandez-Larsson, 1990; Crotty et al., 2000; Gordon et al., 2020b
J_{13}	Chloramphenicol, tigecycline, linezolid	Antibiotics, able to inhibit mitochondrial ribosomes	Colberg-Poley et al., 2000; Gordon et al., 2020b
J_{21A}, J_{21B}	Chloroquine	SIGMAR1/SIGMAR2 inhibitor. Autophagy inhibitor.	Keyaerts et al., 2004; Vincent et al., 2005; Saeed et al., 2011; Cottam et al., 2014; Borba et al., 2020; Gordon et al., 2020b
J_3, J_5, J_{21A}, J_{21B}	Hydroxychloroquine	Autophagy inhibitor, antiviral effect	Cottam et al., 2014; Gautret et al., 2020; Gordon et al., 2020b; Liu et al., 2020
J_1, J_4, J_5	Zotatifin, ternatin 4, tomvosertib	mRNA translation inhibitors	Gordon et al., 2020b
J_1, J_{2B}, J_{23}	Silmitasertib or TMCB	Casein *k*inase II inhibitors, protein synthesis inhibitor	Siddiqui-Jain et al., 2010; Reineke et al., 2017; Gordon et al., 2020b
J_3	Remdesivir	Nucleoside analog, interferes with RNA-dependent RNA polymerase	Sheahan et al., 2017
J_{13}	Umifenovir	Inhibitor of the fusion between the viral envelope (surrounding the viral capsid) and the cell membrane of the target cell	Shi et al., 2007; Pécheur et al., 2016
J_{13}	Lisinopril, losartan	ACE inhibitors, prevent the fusion between the viral envelope and the cell membrane of the target cell	Hoffmann et al., 2020
J_3	Favipiravir	Selective inhibitor of RNA-dependent RNA polymerase	Cai et al., 2020; Shiraki and Daikoku, 2020
J_{2A}, J_{2B}, J_3*, J_4, J_7*, J_{21}	Macrolide antibiotics azithromycin*, clarithromycin	Inhibition of ribosomal translation Autophagy inhibition	Tenson et al., 2003; Stamatiou et al., 2009; Gielen et al., 2010; Petroni et al., 2020

**Close to Js, refers to azithromycin, in the other colum, being the compound impactinng on the identified flows.*

the onset of fever shortened the illness by more than 3 days, as compared with treatment starting at 48 h (Aoki et al., 2003).

System Dynamics of (+)ssRNA Virus–Host Interaction in Response to External Driving Forces Applied to Reduce Virions Assembly

Second, we explored the systemic response to full (100%) or partial (50%) reduction of either J_3 (flow of energy required for RNA replication), or J_4 (flow of energy required for viral RNA translation and viral protein synthesis), or J_5 (flow of energy required for virions assembly), which are involved in virions assembly and are typically dependent on intrinsic viral biological properties.

Full reduction of J_5 at Day 0 could recover the systems dynamics in a stationary status, in all scenarios of initial viral load tested (5,000 RNA copies, **Figure 5A**; 10,000 RNA copies, **Figure 5C**), with constant values for all stocks. A partial reduction of J_5 at Day 0 maintained a stationary status only for lower initial viral load (5,000 RNA copies, **Figure 5B**) associated to a reduced amount of stocks Q_4 and Q_5 (**Figure 5D**). The abrogation of J_5 at Day 0 was still effective to preserve the stationary status (**Figure 5E**), but halving J_5 at Day 1 could reduce but not prevent the growth of Q_5 (**Figure 5F**). We also simulated the effect of applying the same external inputs at different times: after 1 (**Supplementary Figure 2**), 3 (**Supplementary Figure 3**),

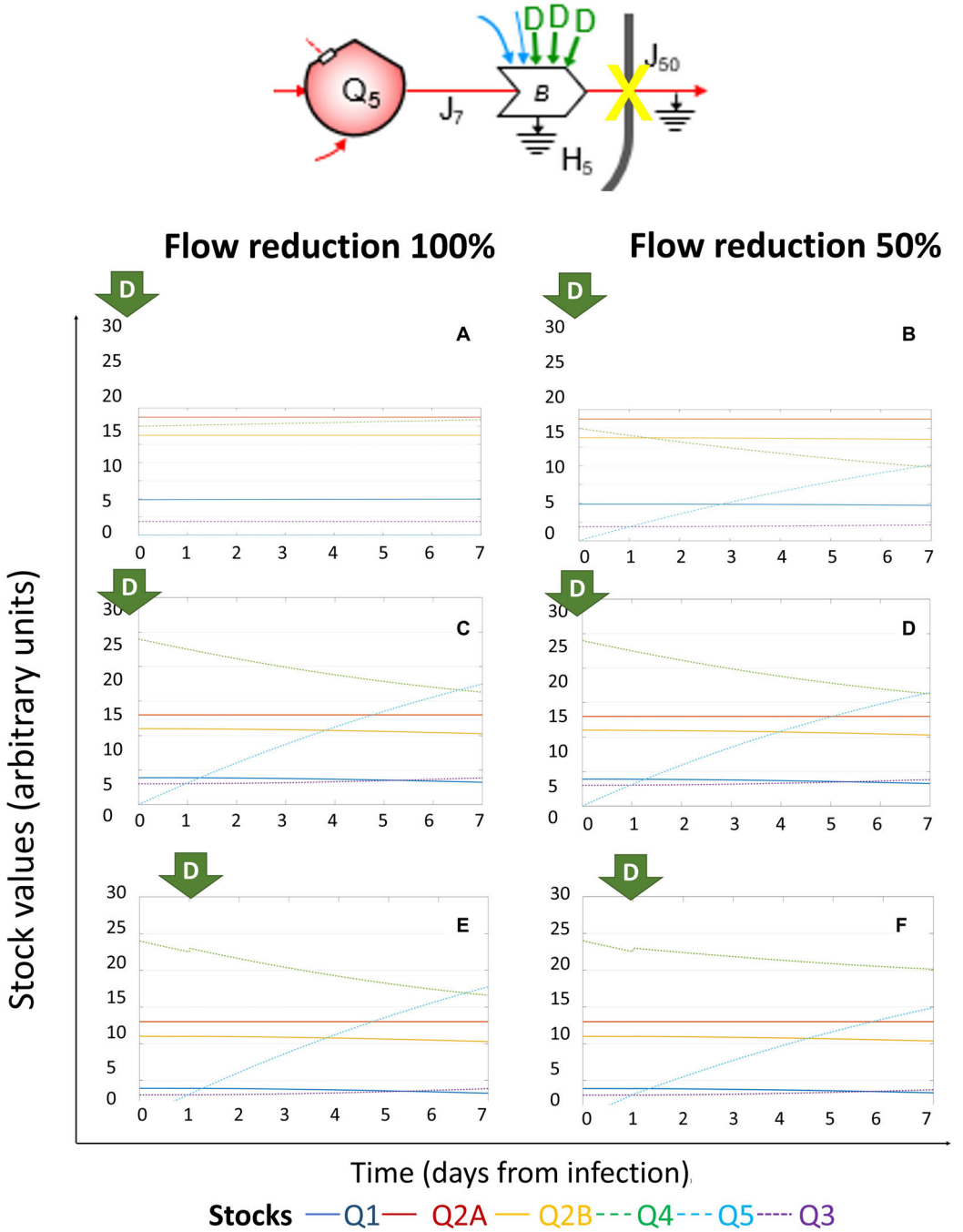


FIGURE 4 | System dynamics of (+)ssRNA virus–host interaction in response to external driving forces applied to reduce virions outflow. Changes over time of the values of each stock of the system diagrammed in **Figure 1** (for the color code, see bottom), expressed in ATP-eq, in response to reduction of J_{50} (flow of energy required for virions budding). Several scenarios are shown: initial viral load 5k and application of full (100%, **A**) or partial (50%, **B**) J_{50} reduction at Day 0; initial viral load 10k and application of full (100%, **C**) or partial (50%, **D**) J_{50} reduction at Day 0; initial viral load 10k and application of full (100%, **E**) or partial (50%, **F**) J_{50} reduction at Day 1.

or 5 days (**Supplementary Figure 4**) from the initial infection, confirming the role of early application of external driving forces to recover the stationary status of the system.

The importance of full abrogation of J_5 flow for a (+)ssRNA virus–host interaction has been indirectly confirmed by the data recently published by Gordon et al., who cloned, tagged, and expressed 26 of 29 severe acute respiratory syndrome (SARS)–CoV-2 proteins individually in HEK293T cells and used mass spectrometry to measure protein–protein interactions (Gordon et al., 2020b), to identify 69 existing drugs, known to target host

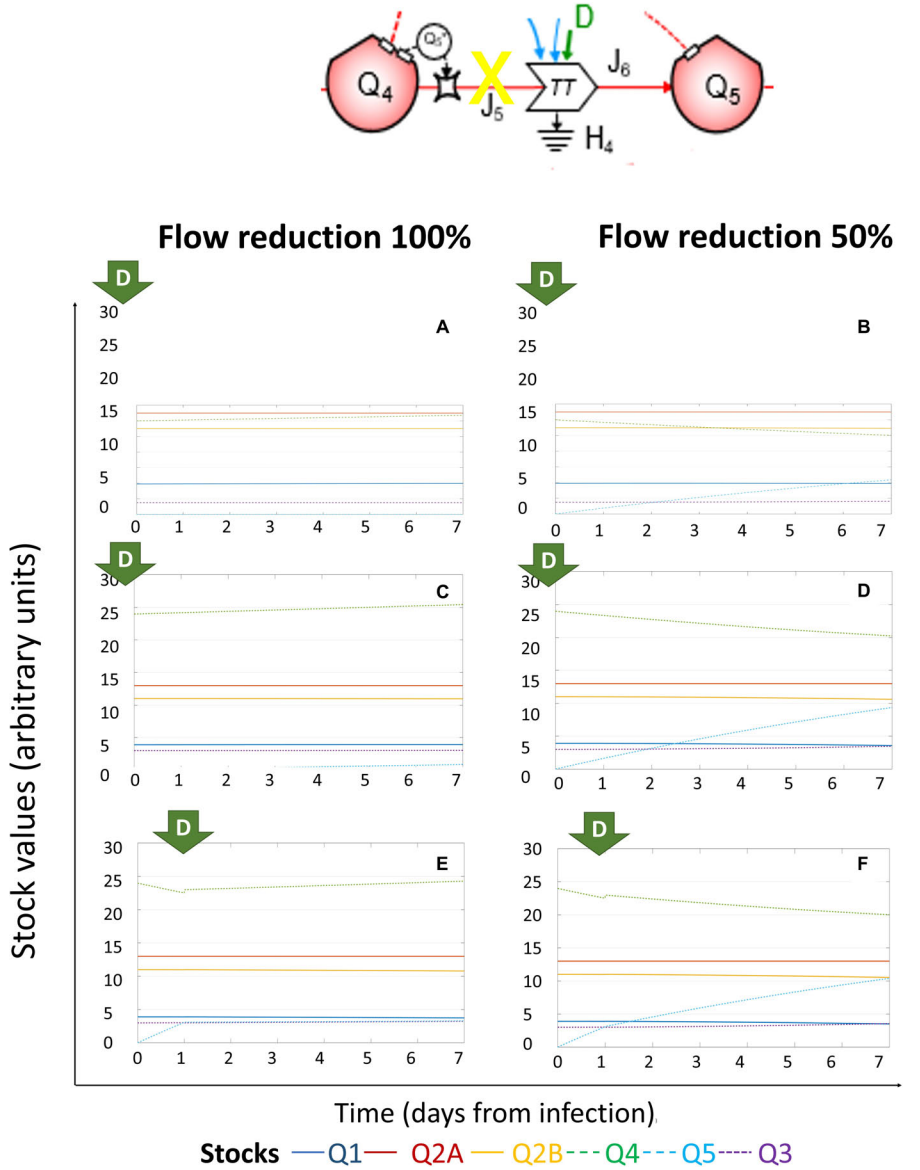


FIGURE 5 | System dynamics of (+)ssRNA virus–host interaction in response to external driving forces applied to reduce virions assembly. Changes over time of the values of each stock of the system diagrammed in **Figure 1** (for the color code, see bottom), expressed in ATP-eq, in response to reduction of J_5 (flow of energy required for virions assembly). Several scenarios are shown: initial viral load 5k and application of full (100%, **A**) or partial (50%, **B**) J_5 reduction at Day 0; initial viral load 10k and application of full (100%, **C**) or partial (50%, **D**) J_5 reduction at Day 0; initial viral load 10k and application of full (100%, **E**) or partial (50%, **F**) J_5 reduction at Day 1.

proteins or associated pathways, which interact with SARS-CoV-2, addressing the importance to target the host–virus interaction at the level of RNA translation.

System Dynamics of (+)ssRNA Virus–Host Interaction in Response to External Driving Forces Applied to Reduce Viral Protein Synthesis

Full (100%) or partial (50%) reduction of J_4 (flow of energy required for viral RNA translation and viral protein synthesis) at Day 0 did not affect the dynamics of the system (**Figure 6**). In particular, for lower initial viral load (5,000 RNA copies, **Figures 6A,B**), the Q_5 stock was always lower than Q_{2A}, thus not affecting Q_1, which remained constant. However, when Q_5 stock was greater than Q_{2A}, Q_1 started to decrease, again suggesting that the effect of viral infection on host metabolism is associated to threshold values specific for each system, and not for each cell type. The observation that—based on initial viral load—the higher value of Q_5 is lower than the stable quantity of Q_{2A} and Q_{2B} for 5k ATP-eq and higher than Q_{2A} and Q_{2B} for 10k ATP-eq can explain the contribution of initial viral load

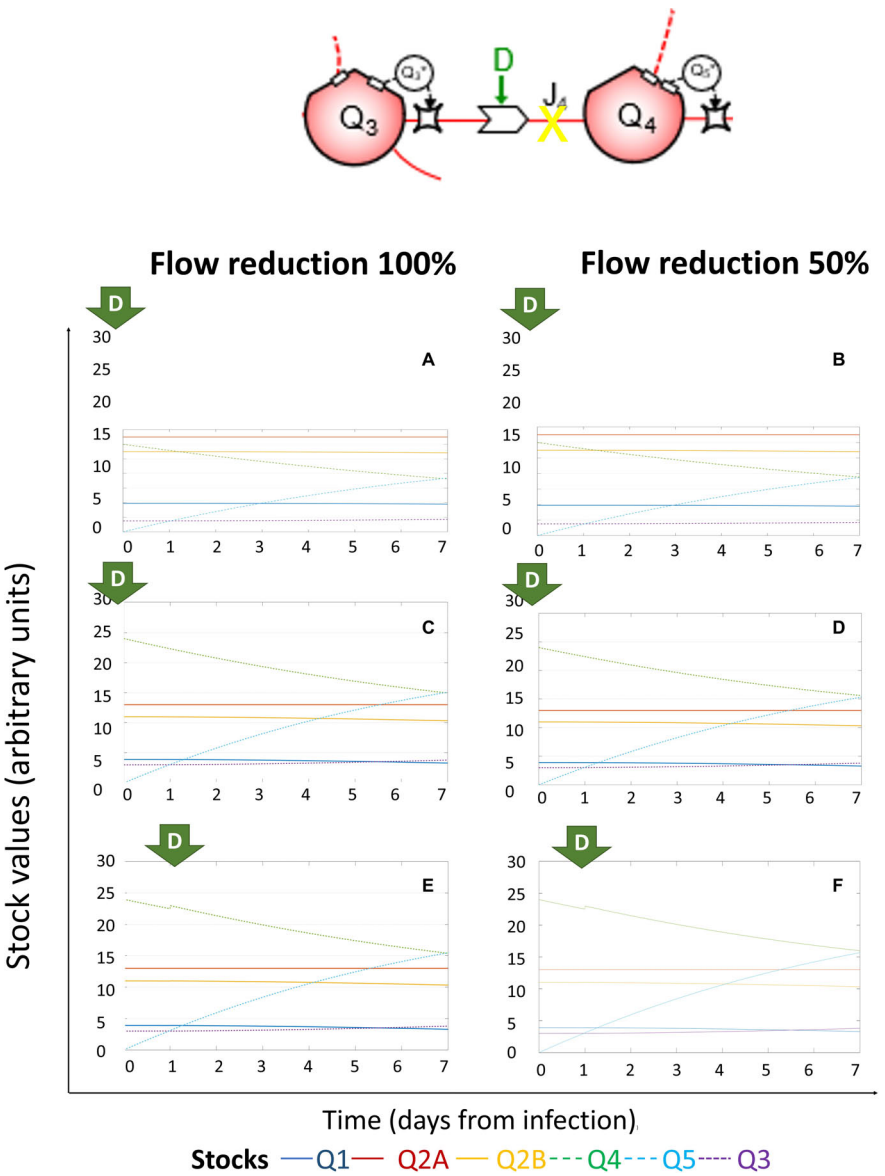


FIGURE 6 | System dynamics of (+)ssRNA virus–host interaction in response to external driving forces applied to reduce viral protein synthesis. Changes over time of the values of each stock of the system diagrammed in **Figure 1** (for the color code, see bottom), expressed in ATP-eq, in response to reduction of J_4 (flow of energy required for viral RNA translation and viral protein synthesis). Several scenarios are shown: initial viral load 5k and application of full (100%, **A**) or partial (50%, **B**) J_4 reduction at Day 0; initial viral load 10k and application of full (100%, **C**) or partial (50%, **D**) J_4 reduction at Day 0; initial viral load 10k and application of full (100%, **E**) or partial (50%, **F**) J_4 reduction at Day 1.

and configuration of host-cell stocks in the viral fitness, which depends on host cell cycle stage, addressed as initial value of Q_{2A} and K_{2A} (Wargo and Kurath, 2012).

System Dynamics of (+)ssRNA Virus–Host Interaction in Response to External Driving Forces Applied to Reduce Viral RNA Replication

Full (100%) or partial (50%) reduction of J_3 (flow of energy required for RNA replication) at Day 0 did not affect the dynamics of the system (**Figure 7**), but—differently from the previous scenario—Q_1 remained constant even if Q_5 was lower than the stable quantity of Q_{2A} and Q_{2B}, even if administered later after infection (on Day 3 or 5, as shown in **Supplementary Figures 2–4**). This suggests that the host cell can preserve its homeostasis upon early exposure to inhibitors of viral replication. Interestingly, the pattern in response to single external driving forces was maintained over time, with the values of each stock just shifted, based on the time of application.

Several strategies can be used to reduce selectively virus RNA, including nucleoside analogs, which are metabolized

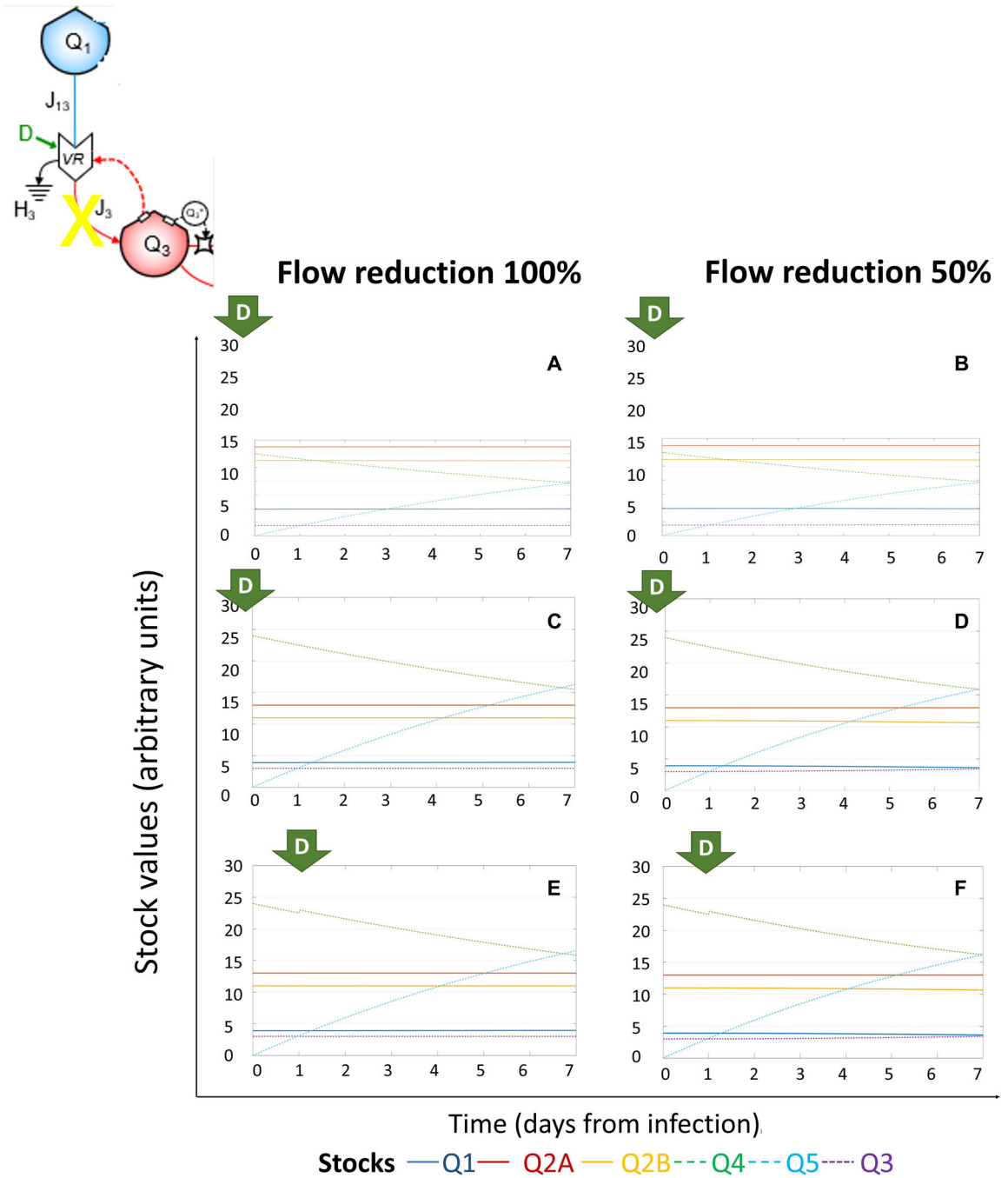


FIGURE 7 | System dynamics of (+)ssRNA virus–host interaction in response to external driving forces applied to reduce viral RNA replication. Changes over time of the values of each stock of the system diagrammed in **Figure 1** (for the color code, see bottom), expressed in ATP-eq, in response to reduction of J_3 (flow of energy required for RNA replication). Several scenarios are shown: initial viral load 5k and application of full (100%, **A**) or partial (50%, **B**) J_3 reduction at Day 0; initial viral load 10k and application of full (100%, **C**) or partial (50%, **D**) J_3 reduction at Day 0; initial viral load 10k and application of full (100%, **E**) or partial (50%, **F**) J_3 reduction at Day 1.

intracellularly into their active ribonucleoside 5′-triphosphate forms and incorporated into the nascent viral RNA by error-prone viral RNA-dependent RNA polymerase (RdRps), to disrupt RNA synthesis directly via chain termination, or accumulation of deleterious mutations in the viral genome. The response to nucleoside analogs could be better tested by our computational model, adding the mutations rate and the DNA/RNA metabolism of host cell, not included in the diagram for lack of experimental data about the turnover of RNA stock, specific for each host cell type of interest.

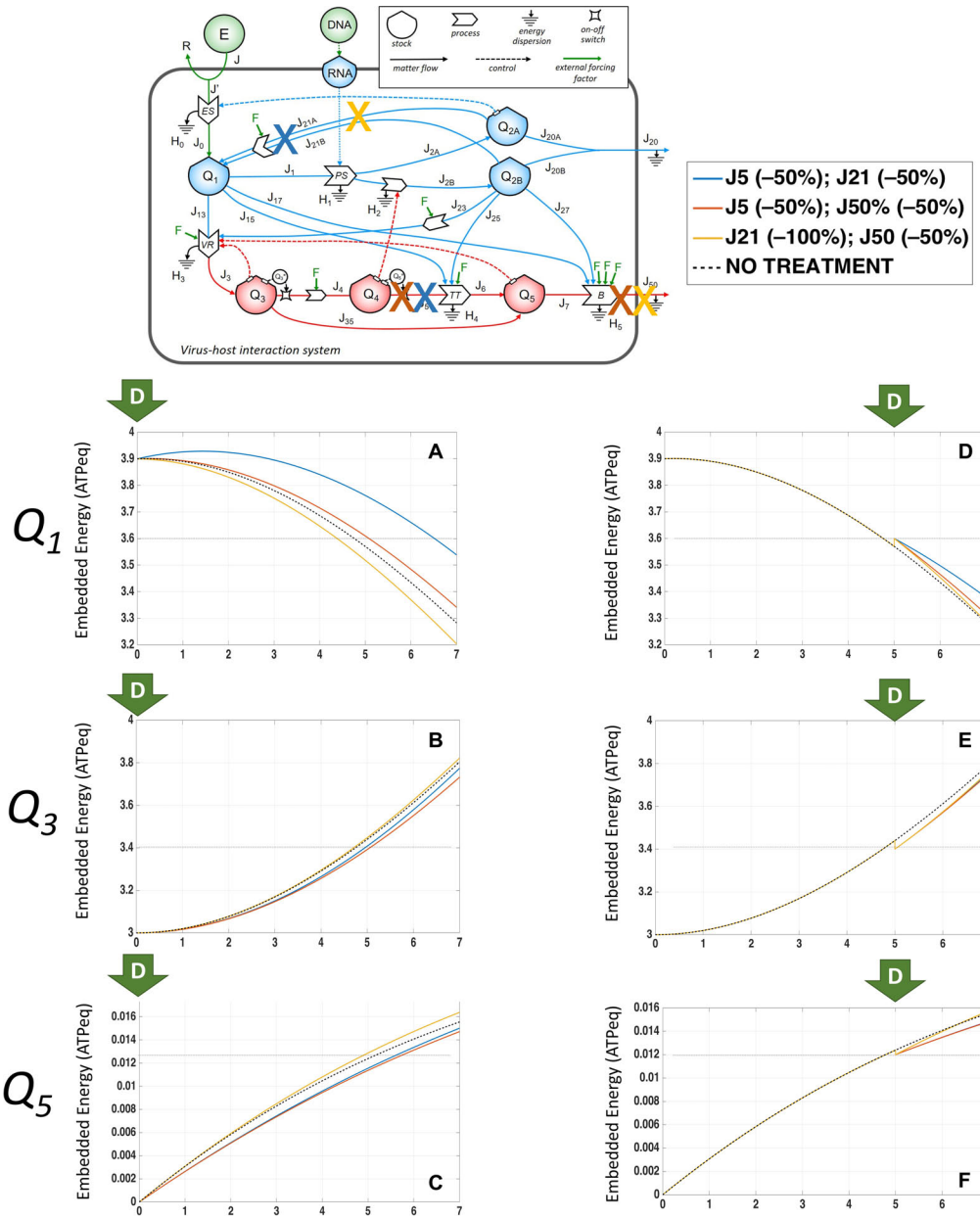


FIGURE 8 | System configurations dynamics in response to multiple external driving forces applied at different timepoints from infection. Changes over time of Q_1 stocks, expressed in ATP-eq, in response to partial (50%) reduction of J_5 and J_{21} (blue line), J_5 and J_{50} (orange line), J_{21} and J_{50} (yellow line), or no application of external driving forces applied at Day 0 **(A)** or at Day 5 **(D)** of infection. Changes over time of Q_3 stocks, expressed in ATP-eq, in response to partial (50%) reduction of J_5 and J_{21} (blue line), J_5 and J_{50} (orange line), J_{21} and J_{50} (yellow line), or no application of external driving forces applied at Day 0 **(B)** or at Day 5 **(E)** of infection. Changes over time of Q_5 stocks, expressed in ATP-eq, in response to partial (50%) reduction of J_5 and J_{21} (blue line), J_5 and J_{50} (orange line), J_{21} and J_{50} (yellow line), or no application of external driving forces applied at Day 0 **(C)** or at Day 5 **(F)** of infection.

System Configurations Dynamics in Response to Multiple External Driving Forces Applied at Different Timepoints From Infection

Third, we next explored the application of multiple external forcing factors, addressing the requirement of a polytarget approach (Bizzarri et al., 2020). The positive effects on Q_1, Q_3, and Q_5 arising from targeting J_5 was mitigated by the combination with reduction of J_{50} or J_{21}, since the combination targeting J_5 and J_{21} (indicated by the blue line in **Figure 8**) applied at Day 0 could preserve the Q_1 amount better than no treatment (**Figure 8A**) associated to slower increase of the growth of Q_3 (**Figure 8B**) and Q_5 (**Figure 8C**), leading us to explore further combinations. Application of external driving forces acting on J_{50} and J_{21} (yellow line in **Figure 8**) increased the

Q_5 stock value instead of the expected decrease (**Figure 8C**); thus, action on a single flow was more efficient than on two of them, given the emergence of compensatory feedbacks and flows. These observations are supported by the controversial findings about the efficacy of macrolides, chloroquine, and their derivatives in the recent COVID-19 pandemic, as their systemic effects include reduction of J_5 and J_{21} (**Table 3**), and could be affected by the viral load and time of application, with weak changes in virions' spread if applied later in the clinical course of disease as shown by preliminary results of randomized trials (Derwand et al., 2020; Kashour et al., 2020; Magagnoli et al., 2020; Fiolet et al., 2021).

When the combination targeting J_5 and J_{50} was applied later (at Day 5 from infection, orange line in **Figure 8**), Q_1 was higher than untreated (**Figure 8D**), associated to slower growth of Q_3 (**Figure 8E**) and Q_5 (**Figure 8F**), but not leading to restore of stationarity. This behavior is a typical systemic feature, where an intervention on a specific local process may lead to counterintuitive rearrangements in the system dynamics. Thus, once the main flow to reduce was found to identify the *n*-ple stock associated with the desired system output, molecular insights should suggest the biological process to assess a tailored treatment.

DISCUSSION

In this work, we approached the host–virus interaction dynamics as a systemic problem, and for the first time in the field, we used combined ST tools as a conceptual framework to build up a systemic description of the viral action and host response, critically depending on the existing metabolic environment.

The basic ST idea is to integrate the traditional bottom-up approach—which describes "local" behaviors through cause–effect chains and functional units—with a top-down approach, which points out at the global behavior of the system in terms of its operational configurations, emerging from the feedback-driven response to different driving forces, like those represented, for example, by the chemistry of new drugs (Odum, 1996). The utility of computational simulations stems from their capacity to identify structural side effects, non-linearities, and time delays, which are left unexplained by other approaches. Systems biology already recognized the relevance of complexity in the study of microbiological systems (Kaneko, 2006; Loscalzo and Barabasi, 2011), but although successfully applied in several fields ranging from hard sciences to ecology and economics (Brown, 2004; Brown and Ulgiati, 2004), the potential of ST in the study of biological systems is still underexploited. Thanks to its abstract nature, stock-flow description can be used in a wide range of different fields, realizing the conceptual bridge that connects the language of biological systems to that of ecology.

In targeting the virus–host interaction, there is an emerging need of tools that could early identify those compounds, not primarily designed for their antiviral action, identifiable by *in silico* approaches (Gordon et al., 2020b), which alone or in combination can provide clinical efficacy (Mina et al., 2016; Cheng et al., 2018; Bogdanow et al., 2019; Panja et al., 2019; Gordon et al., 2020b). There are known advantages of *in silico* modeling of the action of therapeutic agents on known diseases through agent-based modeling (Mao et al., 2018). However, the literature evidenced some intrinsic limitations on the choice of parameters, such as the size of investigated populations (Mina et al., 2016), while major problems are related to model validation (Mina et al., 2016; Donkin et al., 2017), also requiring to supplement the models with adequate formal ontologies (Kalfoglou and Schorlemmer, 2003; Gotts et al., 2019).

The proposed model was developed at the single-cell scale. However, in order to define an overall therapeutic approach, the integration with a multiscale approach would be also desirable. In particular, depending on the availability of appropriate data, a future model could focus on different scales, with a more detailed description of some components at the subcellular level, which were grouped (e.g., short- and long-half-life proteins, lipids, and vesicles trafficking) in the present study. On the other hand, the interaction between different cell populations in the host could be also developed, to represent the interaction between healthy and infected cells, and the contribution of immune system (Tan et al., 2007; Katze et al., 2008; Wu et al., 2020) or the repertoire of receptors on the surface of the host cell (Gordon et al., 2020a; Hoffmann et al., 2020; Kailas et al., 2020) to surveil and limit the size of Q_3 stock at the single-cell level. Other natural system constraints could be also included, like some physiological parameters (e.g., temperature and metabolic rate), whose impact on the human body energy dynamics is already understood. The use of a multiscale hierarchical perspective is in principle already possible, as discussed in previous works adopting the same sort of system representation (Ulgiati and Brown, 2009; Ulgiati et al., 2011).

CONCLUSION

This work highlights the advantages of applying an ST-based approach to the study of virus–host interaction, being reflected in the possibility of extracting systemic dynamic features that would be otherwise counterintuitive. While a traditional single-target approach would address strategies targeting the viral RNA (Q_3) or the replication process (J_3), our results suggest that the virus growth is more vulnerable if the process of virion growth before expulsion (process TT, involving flow J_5) is targeted.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

AR designed the study and collected medical knowledge. FG built up the diagram. MC built up the simulator. All authors performed simulation, analyzed data, prepared the figures, and wrote the manuscript.

FUNDING

This work was supported by SIES, Società Italiana di Ematologia Sperimentale.

ACKNOWLEDGMENTS

The authors thank Luca Naso, Francesco Di Raimondo, and Maria Sanfilippo from SIES, Società Italiana di Ematologia Sperimentale for the support provided.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.600254/full#supplementary-material

Supplementary Figure 1 | System dynamics of (+)ssRNA virus–host interaction in response to initial viral load. Changes overtime of the values of each stock of the system diagrammed in **Figure 1** (for the color code see at the bottom), expressed in ATP-eq: in absence of infection the system status was stationary **(A)**. Upon infection initial values of Q_3 stock identify the response of system, at 50 **(B)**, 100 **(C)**, 1,000 **(D)**, 5,000 **(E)** and 10,000 **(F)** RNA copies, expressed in ATP-eq. For Q3 stock in the range (10–1,000 RNA copies) the trajectory of Q_5 evolution was linear, while for higher initial viral load the growth of Q_5 was linear in the first day and non-linear in the further timeframe.

Supplementary Figure 2 | Effects of targeting leverage point applying external driving force (*D*) at Day 1. Time change of the stocks values (for the color code see at the bottom), expressed in ATP-eq, were depicted over-time, from infection (Day 0, Time 0) through the next 7 days, upon application, at Day 1, of generic external driving forces (*D*) able to reduce the target flows (schemed on the left) of 100% or 50%, respectively.

Supplementary Figure 3 | Effects of targeting leverage point applying external driving force (*D*) at Day 3. Time change of the stocks values (for the color code see at the bottom), expressed in ATP-eq, were depicted over-time, from infection (Day 0, Time 0) through the next 7 days, upon application, at Day 3, of generic external driving forces (*D*) able to reduce the target flows (schemed on the left) of 100% or 50%, respectively.

Supplementary Figure 4 | Effects of targeting leverage point applying external driving force (*D*) at Day 5. Time change of the stocks values (for the color code see at the bottom), expressed in ATP-eq, were depicted over-time, from infection (Day 0, Time 0) through the next 7 days, upon application, at Day 5, of generic external driving forces (*D*) able to reduce the target flows (schemed on the left) of 100% or 50%, respectively.

REFERENCES

Acevedo, A., Brodsky, L., and Andino, R. (2014). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505, 686–690. doi: 10.1038/nature12861

Adelman, K., La Porta, A., Santangelo, T. I. J. I., Lis, J. T., Roberts, J. W., and Wang, M. D. (2002). Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. U S A* 99, 13538–13543. doi: 10.1073/pnas.212358999

Ahlquist, P., Noueiry, A. O., Lee, W.-M., Kushner, D. B., and Dye, B. T. (2003). Host Factors in Positive-Strand RNA Virus Genome Replication. *J. Virol.* 77, 8181–8186. doi: 10.1128/jvi.77.15.8181-8186.2003

Aoki, F. Y., Macleod, M. D., Paggiaro, P., Carewicz, O., El Sawy, A., Wat, C., et al. (2003). Early administration of oral oseltamivir increases the benefits of influenza treatment. *J. Antimicrob. Chemother.* 51, 123–129. doi: 10.1093/jac/dkg007

Apweiler, R., Beissbarth, T., Berthold, M. R., Blüthgen, N., Burmeister, Y., Dammann, O., et al. (2018). Whither systems medicine? *Exp. Mole. Med.* 50, e453–e453.

Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G., and Perelson, A. S. (2006). Kinetics of influenza A virus infection in humans. *J. Virol.* 80, 7590–7599. doi: 10.1128/jvi.01623-05

Bardina, M. V., Lidsky, P. V., Sheval, E. V., Fominykh, K. V., Van Kuppeveld, F. J. M., Polyakov, V. Y., et al. (2009). Mengovirus-Induced Rearrangement of the Nuclear Pore Complex: Hijacking Cellular Phosphorylation Machinery. *J. Virol.* 83, 3150–3161. doi: 10.1128/jvi.01456-08

Beachboard, D. C., and Horner, S. M. (2016). Innate immune evasion strategies of DNA and RNA viruses. *Curr. Opin. Microbiol.* 32, 113–119. doi: 10.1016/j.mib.2016.05.015

Ben-Tabou De-Leon, S., and Davidson, E. H. (2009). Modeling the dynamics of transcriptional gene regulatory networks for animal development. *Dev. Biol.* 325, 317–328. doi: 10.1016/j.ydbio.2008.10.043

Bizzarri, M., Minini, M., and Monti, N. (2020). "Revisiting the Concept of Human Disease," in *Approaching Complex Diseases*, ed. M. Bizzarri (Berlin: Springer Nature), 1–34. doi: 10.1007/978-3-030-32857-3_1

Bogdanow, B., Wang, X., Eichelbaum, K., Sadewasser, A., Husic, I., Paki, K., et al. (2019). The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant. *Nat. Commun.* 10:5518.

Boisvert, F. M., Ahmad, Y., Gierlinski, M., Charriere, F., Lamont, D., Scott, M., et al. (2012). A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol. Cell Proteomics* 11:M111.011429.

Borba, M. G. S., Val, F. F. A., Sampaio, V. S., Alexandre, M. A. A., Melo, G. C., Brito, M., et al. (2020). Effect of High vs Low Doses of Chloroquine Diphosphate as Adjunctive Therapy for Patients Hospitalized With Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection: A Randomized Clinical Trial. *JAMA Netw. Open* 3:e208857.

Bosl, K., Ianevski, A., Than, T. T., Andersen, P. I., Kuivanen, S., Teppor, M., et al. (2019). Common Nodes of Virus-Host Interaction Revealed Through an Integrated Network Analysis. *Front. Immunol.* 10:2186.

Bossel, H. (2007). *Systems and models: complexity, dynamics, evolution, sustainability*. Norderstedt: Books on Demand GmbH.

Brown, M. T. (2004). A picture is worth a thousand words: energy systems language and simulation. *Ecol. Model.* 178, 83–100. doi: 10.1016/j.ecolmodel.2003.12.008

Brown, M. T., and Ulgiati, S. (2004). Energy quality, emergy, and transformity: H.T. Odum's contributions to quantifying and understanding systems. *Ecol. Model.* 178, 201–213. doi: 10.1016/j.ecolmodel.2004.03.002

Burch, C. L., and Chao, L. (2000). Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* 406, 625–628. doi: 10.1038/35020564

Burkard, C., Verheije, M. H., Haagmans, B. L., Van Kuppeveld, F. J., Rottier, P. J., Bosch, B. J. , et al. (2015). ATP1A1-mediated Src signaling inhibits coronavirus entry into host cells. *J. Virol.* 89, 4434–4448. doi: 10.1128/jvi.03274-14

Buttgereit, F., and Brand, M. D. (1995). A hierarchy of ATP-consuming processes in mammalian cells. *Biochem. J.* 312(Pt 1), 163–167. doi: 10.1042/bj3120163

Cai, Q., Yang, M., Liu, D., Chen, J., Shu, D., Xia, J., et al. (2020). Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study. *Engineering* 6, 1192–1198. doi: 10.1016/j.eng.2020.03.007

Canto, C., Gerhart-Hines, Z., Feige, J. N., Lagouge, M., Noriega, L., Milne, J. C., et al. (2009). AMPK regulates energy expenditure by modulating NAD+ metabolism and SIRT1 activity. *Nature* 458, 1056–1060. doi: 10.1038/nature07813

Castelló, A., Izquierdo, J. M., Welnowska, E., and Carrasco, L. (2009). RNA nuclear export is blocked by poliovirus 2A protease and is concomitant with nucleoporin cleavage. *J. Cell Sci.* 122, 3799–3809. doi: 10.1242/jcs.055988

Chakravorty, S., Yan, B., Wang, C., Wang, L., Quaid, J. T., Lin, C. F., et al. (2019). Integrated Pan-Cancer Map of EBV-Associated Neoplasms Reveals Functional Host-Virus Interactions. *Cancer Res.* 79, 6010–6023. doi: 10.1158/0008-5472.can-19-0615

Chen, I. T., Lee, D. Y., Huang, Y. T., Kou, G. H., Wang, H. C., Chang, G. D., et al. (2016). Six Hours after Infection, the Metabolic Changes Induced by WSSV Neutralize the Host's Oxidative Stress Defenses. *Sci. Rep.* 6:27732.

Cheng, F., Desai, R. J., Handy, D. E., Wang, R., Schneeweiss, S., Barabasi, A. L., et al. (2018). Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.* 9:2691.

Colberg-Poley, A. M., Patel, M. B., Erezo, D. P., and Slater, J. E. (2000). Human cytomegalovirus UL37 immediate-early regulatory proteins traffic through the secretory apparatus and to mitochondria. *J. Gen. Virol.* 81, 1779–1789. doi: 10.1099/0022-1317-81-7-1779

Cong, Y. S., and Bacchetti, S. (2000). Histone deacetylation is involved in the transcriptional repression of hTERT in normal human cells. *J. Biol. Chem.* 275, 35665–35668. doi: 10.1074/jbc.c000637200

Cottam, E. M., Whelband, M. C., and Wileman, T. (2014). Coronavirus NSP6 restricts autophagosome expansion. *Autophagy* 10, 1426–1441. doi: 10.4161/auto.29309

Coyaud, E., Ranadheera, C., Cheng, D., Goncalves, J., Dyakov, B. J. A., Laurent, E. M. N., et al. (2018). Global Interactomics Uncovers Extensive Organellar Targeting by Zika Virus. *Mol. Cell Proteom.* 17, 2242–2255. doi: 10.1074/mcp.tir118.000800

Crotty, S., Maag, D., Arnold, J. J., Zhong, W., Lau, J. Y. N., Hong, Z., et al. (2000). The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat. Med.* 6, 1375–1379. doi: 10.1038/82191

Dapat, C., and Oshitani, H. (2016). Novel insights into human respiratory syncytial virus-host factor interactions through integrated proteomics and transcriptomics analysis. *Exp. Rev. Anti. Infect. Ther.* 14, 285–297. doi: 10.1586/14787210.2016.1141676

De Boer, R. J., Ribeiro, R. M., and Perelson, A. S. (2010). Current estimates for HIV-1 production imply rapid viral clearance in lymphoid tissues. *PLoS Comput. Biol.* 6:e1000906. doi: 10.1371/journal.pcbi.1000906

Den Boon, J. A., Diaz, A., and Ahlquist, P. (2010). Cytoplasmic viral replication complexes. *Cell Host Microb.* 8, 77–85. doi: 10.1016/j.chom.2010.06.010

Derwand, R., Scholz, M., and Zelenko, V. (2020). COVID-19 outpatients: early risk-stratified treatment with zinc plus low-dose hydroxychloroquine and azithromycin: a retrospective case series study. *Int. J. Antimicrob. Agents* 56:106214. doi: 10.1016/j.ijantimicag.2020.106214

Diebold, S. S., Montoya, M., Unger, H., Alexopoulou, L., Roy, P., Haswell, L. E., et al. (2003). Viral infection switches non-plasmacytoid dendritic cells into high interferon producers. *Nature* 424, 324–328. doi: 10.1038/nature01783

Dinesh, D. C., Tamilarasan, S., Rajaram, K., and Boura, E. (2020). Antiviral Drug Targets of Single-Stranded RNA Viruses Causing Chronic Human Diseases. *Curr. Drug Targets* 21, 105–124. doi: 10.2174/1389450119666190920153247

Doherty, M. K., Hammond, D. E., Clague, M. J., Gaskell, S. J., and Beynon, R. J. (2009). Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. *J. Proteome. Res.* 8, 104–112. doi: 10.1021/pr800641v

Dong, S., Liu, L., Wu, W., Armstrong, S. D., Xia, D., Nan, H., et al. (2016). Determination of the interactome of non-structural protein12 from highly pathogenic porcine reproductive and respiratory syndrome virus with host cellular proteins using high throughput proteomics and identification of HSP70 as a cellular factor for virus replication. *J. Proteomic.* 146, 58–69. doi: 10.1016/j.jprot.2016.06.019

Donkin, E., Dennis, P., Ustalakov, A., Warren, J., and Clare, A. (2017). Replicating complex agent based models, a formidable task. *Environ. Model. Soft.* 92, 142–151. doi: 10.1016/j.envsoft.2017.01.020

Eden, E., Geva-Zatorsky, N., Issaeva, I., Cohen, A., Dekel, E., Danon, T., et al. (2011). Proteome half-life dynamics in living human cells. *Science* 331, 764–768. doi: 10.1126/science.1199784

Emmott, E., Munday, D., Bickerton, E., Britton, P., Rodgers, M. A., Whitehouse, A., et al. (2013). The cellular interactome of the coronavirus infectious bronchitis virus nucleocapsid protein and functional implications for virus biology. *J. Virol.* 87, 9486–9500. doi: 10.1128/jvi.00321-13

Fanning, L. J., Levis, J., Kenny-Walsh, E., Whelton, M., O'sullivan, K., and Shanahan, F. (2001). HLA class II genes determine the natural variance of hepatitis C viral load. *Hepatology* 33, 224–230. doi: 10.1053/jhep.2001.20642

Fiolet, T., Guihur, A., Rebeaud, M. E., Mulot, M., Peiffer-Smadja, N., Mahamat-Saleh, Y., et al. (2021). Effect of hydroxychloroquine with or without azithromycin on the mortality of coronavirus disease 2019 (COVID-19) patients: a systematic review and meta-analysis. *Clin. Microbiol. Infect.* 27, 19–27. doi: 10.1016/j.cmi.2020.08.022

Fontaine, K. A., Sanchez, E. L., Camarda, R., and Lagunoff, M. (2015). Dengue virus induces and requires glycolysis for optimal replication. *J. Virol.* 89, 2358–2366. doi: 10.1128/jvi.02309-14

Friedel, C. C., and Haas, J. (2011). Virus-host interactomes and global models of virus-infected cells. *Trends Microbiol.* 19, 501–508. doi: 10.1016/j.tim.2011.07.003

Gao, Z., Hu, J., Liang, Y., Yang, Q., Yan, K., Liu, D., et al. (2017). Generation and Comprehensive Analysis of Host Cell Interactome of the PA Protein of the Highly Pathogenic H5N1 Avian Influenza Virus in Mammalian Cells. *Front. Microbiol.* 8:739.

Garcia-Dorival, I., Wu, W., Dowall, S., Armstrong, S., Touzelet, O., Wastling, J., et al. (2014). Elucidation of the Ebola virus VP24 cellular interactome and disruption of virus biology through targeted inhibition of host-cell protein function. *J. Proteome Res.* 13, 5120–5135. doi: 10.1021/pr500556d

Garcia-Moreno, M., Noerenberg, M., Ni, S., Jarvelin, A. I., Gonzalez-Almela, E., Lenz, C. E., et al. (2019). System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. *Mol. Cell* 74, 196.e–211.e.

Gautret, P., Lagier, J. C., Parola, P., Hoang, V. T., Meddeb, L., Mailhe, M., et al. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int. J. Antimicrob. Agents* 56:105949. doi: 10.1016/j.ijantimicag.2020.105949

Gielen, V., Johnston, S. L., and Edwards, M. R. (2010). Azithromycin induces anti-viral responses in bronchial epithelial cells. *Eur. Respir. J.* 36, 646–654. doi: 10.1183/09031936.00095809

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020a). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468.

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020b). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468.

Gotts, N. M., Van Voorn, G. A. K., Polhill, J. G., Jong, E. D., Edmonds, B., Hofstede, G. J. , et al. (2019). Agent-based modelling of socio-ecological systems: Models, projects and ontologies. *Ecol. Compl.* 40:100728. doi: 10.1016/j.ecocom.2018.07.007

Grüne-Yanoff, T., and Weirich, P. (2010). The Philosophy and Epistemology of Simulation: A Review. *Simulation Gaming* 41, 20–50. doi: 10.1177/1046878109353470

Hafirassou, M. L., Meertens, L., Umana-Diaz, C., Labeau, A., Dejarnac, O., Bonnet-Madin, L., et al. (2017). A Global Interactome Map of the Dengue Virus NS1 Identifies Virus Restriction and Dependency Host Factors. *Cell Rep.* 21, 3900–3913. doi: 10.1016/j.celrep.2017.11.094

Hassmiller Lich, K., Urban, J. B., Frerichs, L., and Dave, G. (2017). Extending systems thinking in planning and evaluation using group concept mapping and system dynamics to tackle complex problems. *Eval. Prog. Plann* 60, 254–264. doi: 10.1016/j.evalprogplan.2016.10.008

Hekimoğlu, M., and Barlas, Y. (2016). Sensitivity analysis for models with multiple behavior modes: a method based on behavior pattern measures. *Syst. Dyn. Rev.* 32, 332–362. doi: 10.1002/sdr.1568

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, 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. *Cell* 181, 271.e–280.e.

Hogan, R. J., Gao, G., Rowe, T., Bell, P., Flieder, D., Paragas, J., et al. (2004). Resolution of primary severe acute respiratory syndrome-associated coronavirus infection requires Stat1. *J. Virol.* 78, 11416–11421. doi: 10.1128/jvi.78.20.11416-11421.2004

Howard-Ashby, M., Materna, S. C., Brown, C. T., Chen, L., Cameron, R. A., and Davidson, E. H. (2006). Identification and characterization of homeobox transcription factor genes in Strongylocentrotus purpuratus, and their expression in embryonic development. *Dev. Biol.* 300, 74–89. doi: 10.1016/j.ydbio.2006.08.039

Iwasaki, M., Minder, P., Cai, Y., Kuhn, J. H., Yates, J. R. III, Torbett, B. E., et al. (2018). Interactome analysis of the lymphocytic choriomeningitis virus nucleoprotein in infected cells reveals ATPase Na+/K+ transporting subunit Alpha 1 and prohibitin as host-cell factors involved in the life cycle of

mammarenaviruses. *PLoS Pathog.* 14:e1006892. doi: 10.1371/journal.ppat.1006892

Jefferies, H. B., Fumagalli, S., Dennis, P. B., Reinhard, C., Pearson, R. B., and Thomas, G. (1997). Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70s6k. *Embo J.* 16, 3693–3704. doi: 10.1093/emboj/16.12.3693

Jonsdottir, H. R., and Dijkman, R. (2016). Coronaviruses and the human airway: a universal system for virus-host interaction studies. *Virol J.* 13:24.

Jourdan, S. S., Osorio, F., and Hiscox, J. A. (2012). An interactome map of the nucleocapsid protein from a highly pathogenic North American porcine reproductive and respiratory syndrome virus strain generated using SILAC-based quantitative proteomics. *Proteomics* 12, 1015–1023. doi: 10.1002/pmic.201100469

Kailas, S., Sagar, S. B., Maruti, J. D., Shailesh, R. W., Naiem, H. N., Subodh, A. K., et al. (2020). Homology Modeling and Docking Studies of TMPRSS2 with Experimentally Known Inhibitors Camostat Mesylate, Nafamostat and Bromhexine Hydrochloride to Control SARS-Coronavirus-2. *ChemRxiv. Preprint* doi: 10.26434/chemrxiv.12162360.v1

Kalfoglou, Y., and Schorlemmer, M. (2003). Ontology mapping: the state of the art. *Know. Eng. Rev.* 18, 1–31. doi: 10.1017/s0269888903000651

Kaneko, K. (2006). *Life: An Introduction to Complex Systems Biology*. Berlin: Springer.

Kashour, Z., Riaz, M., Garbati, M. A., Aldosary, O., Tlayjeh, H., Gerberi, D., et al. (2020). Efficacy of chloroquine or hydroxychloroquine in COVID-19 patients: a systematic review and meta-analysis. *J. Antimicrob. Chemother.* 76, 30–42.

Katze, M. G., Fornek, J. L., Palermo, R. E., Walters, K.-A., and Korth, M. J. (2008). Innate immune modulation by RNA viruses: emerging insights from functional genomics. *Nat. Rev. Immunol.* 8, 644–654. doi: 10.1038/nri2377

Keyaerts, E., Vijgen, L., Maes, P., Neyts, J., and Van Ranst, M. (2004). In vitro inhibition of severe acute respiratory syndrome coronavirus by chloroquine. *Biochem. Biophys. Res. Commun.* 323, 264–268. doi: 10.1016/j.bbrc.2004.08.085

Khamina, K., Lercher, A., Caldera, M., Schliehe, C., Vilagos, B., Sahin, M., et al. (2017). Characterization of host proteins interacting with the lymphocytic choriomeningitis virus L protein. *PLoS Pathog.* 13:e1006758. doi: 10.1371/journal.ppat.1006758

Kindrachuk, J., Ork, B., Hart, B. J., Mazur, S., Holbrook, M. R., Frieman, M. B., et al. (2015). Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome analysis. *Antimicrob. Agents Chemother.* 59, 1088–1099. doi: 10.1128/aac.03659-14

King, B. R., Hershkowitz, D., Eisenhauer, P. L., Weir, M. E., Ziegler, C. M., Russo, J., et al. (2017). A Map of the Arenavirus Nucleoprotein-Host Protein Interactome Reveals that Junin Virus Selectively Impairs the Antiviral Activity of Double-Stranded RNA-Activated Protein Kinase (PKR). *J. Virol.* 91, e763–e717.

Klumpp, S., and Hwa, T. (2008). Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. *Proc. Natl. Acad. Sci.* 105, 18159–18164. doi: 10.1073/pnas.0806084105

Kobasa, D., Jones, S. M., Shinya, K., Kash, J. C., Copps, J., Ebihara, H., et al. (2007). Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature* 445, 319–323. doi: 10.1038/nature05495

Konig, R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M., Irelan, J. T., et al. (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 135, 49–60. doi: 10.1016/j.cell.2008.07.032

Kudva, R., Denks, K., Kuhn, P., Vogt, A. (2013). gative bacteria: the Sec and Tat dependent protein transport pathways. *Res. Microbiol.* 164, 505–534. doi: 10.1016/j.resmic.2013.03.016

Kummer, S., Flöttmann, M., Schwanhäusser, B., Sieben, C., Veit, M., Selbach, M., et al. (2014). Alteration of protein levels during influenza virus H1N1 infection in host cells: a proteomic survey of host and virus reveals differential dynamics. *PLoS One* 9:e94257. doi: 10.1371/journal.pone.0094257

Kuo, R. L., Chen, C. J., Tam, E. H., Huang, C. G., Li, L. H., Li, Z. H., et al. (2018). Interactome Analysis of NS1 Protein Encoded by Influenza A H7N9 Virus Reveals an Inhibitory Role of NS1 in Host mRNA Maturation. *J. Proteome Res.* 17, 1474–1484. doi: 10.1021/acs.jproteome.7b00815

Kuo, R. L., Li, Z. H., Li, L. H., Lee, K. M., Tam, E. H., Liu, H. M., et al. (2016). Interactome Analysis of the NS1 Protein Encoded by Influenza A H1N1 Virus Reveals a Positive Regulatory Role of Host Protein PRP19 in Viral Replication. *J. Proteome Res.* 15, 1639–1648. doi: 10.1021/acs.jproteome.6b00103

Lam, T. T., Zhu, H., Guan, Y., and Holmes, E. C. (2016). Genomic Analysis of the Emergence, Evolution, and Spread of Human Respiratory RNA Viruses. *Annu. Rev. Genomics Hum. Genet.* 17, 193–218. doi: 10.1146/annurev-genom-083115-022628

Lauring, A. S., Frydman, J., and Andino, R. (2013). The role of mutational robustness in RNA virus evolution. *Nat. Rev. Microb.* 11, 327–336. doi: 10.1038/nrmicro3003

Lescar, J., Soh, S., Lee, L. T., Vasudevan, S. G., Kang, C., and Lim, S. P. (2018). The Dengue Virus Replication Complex: From RNA Replication to Protein-Protein Interactions to Evasion of Innate Immunity. *Adv. Exp. Med. Biol.* 1062, 115–129. doi: 10.1007/978-981-10-8727-1_9

Li, A., Wu, Q., Warnick, G., Li, S., Libby, E. N., Garcia, D. A., et al. (2020). The incidence of thromboembolism for lenalidomide versus thalidomide in older patients with newly diagnosed multiple myeloma. *Ann. Hematol.* 99, 121–126. doi: 10.1007/s00277-019-03860-2

Liu, J., Cao, R., Xu, M., Wang, X., Zhang, H., Hu, H., et al. (2020). Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. *Cell Discov.* 6:16.

Loscalzo, J., and Barabasi, A. L. (2011). Systems biology and the future of medicine. *Wiley Interdiscip Rev. Syst. Biol. Med.* 3, 619–627. doi: 10.1002/wsbm.144

Magagnoli, J., Narendran, S., Pereira, F., Cummings, T. H., Hardin, J. W., Sutton, S. S., et al. (2020). Outcomes of Hydroxychloroquine Usage in United States Veterans Hospitalized with COVID-19. *medRxiv. Preprint* doi: 10.1101/2020.04.16.20065920

Mahmoudabadi, G., Milo, R., and Phillips, R. (2017). Energetic cost of building a virus. *Proc. Natl. Acad. Sci.* 114, E4324–E4333. doi: 10.1073/pnas.1701670114

Mao, X., Mcmanaway, S., Jaiswal, J. K., Patel, P. B., Wilson, W. R., Hicks, K. O., et al. (2018). An agent-based model for drug-radiation interactions in the tumour microenvironment: Hypoxia-activated prodrug SN30000 in multicellular tumour spheroids. *PLoS Comput. Biol.* 14:e1006469. doi: 10.1371/journal.pcbi.1006469

Martinez-Gil, L., Vera-Velasco, N. M., and Mingarro, I. (2017). Exploring the Human-Nipah Virus Protein-Protein Interactome. *J. Virol.* 91, e1461–e1417. doi: 10.1128/JVI.01461-17

Meadows, D. H., and Wright, D. (2008). Thinking in systems : a primer.

Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R., et al. (2005). Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* 437, 1167–1172. doi: 10.1038/nature04193

Mina, P., Tsaneva-Atanasova, K., and Bernardo, M. D. (2016). Entrainment and Control of Bacterial Populations: An in Silico Study over a Spatially Extended Agent Based Model. *ACS Synthetic Biol.* 5, 639–653. doi: 10.1021/acssynbio.5b00243

Mohler, L., Flockerzi, D., Sann, H., and Reichl, U. (2005). Mathematical model of influenza A virus production in large-scale microcarrier culture. *Biotechnol. Bioeng.* 90, 46–58. doi: 10.1002/bit.20363

Moscona, A. (2005). Neuraminidase inhibitors for influenza. *N. Engl. J. Med.* 353, 1363–1373. doi: 10.1056/NEJMra050740

Munday, D. C., Surtees, R., Emmott, E., Dove, B. K., Digard, P., Barr, J. N., et al. (2012). Using SILAC and quantitative proteomics to investigate the interactions between viral and host proteomes. *Proteomics* 12, 666–672. doi: 10.1002/pmic.201100488

Murayama, A., Ohmori, K., Fujimura, A., Minami, H., Yasuzawa-Tanaka, K., Kuroda, T., et al. (2008). Epigenetic Control of rDNA Loci in Response to Intracellular Energy Status. *Cell* 133, 627–639. doi: 10.1016/j.cell.2008.03.030

Murphy, B. R., Rennels, M. B., Douglas, R. G. Jr., Betts, R. F., Couch, R. B., Cate, T. R. Jr., et al. (1980). Evaluation of influenza A/Hong Kong/123/77 (H1N1) ts-1A2 and cold-adapted recombinant viruses in seronegative adult volunteers. *Infect Immun.* 29, 348–355.

Nagesh, P. T., Hussain, M., Galvin, H. D., and Husain, M. (2017). Histone Deacetylase 2 Is a Component of Influenza A Virus-Induced Host Antiviral Response. *Front. Microbiol.* 8:1315. doi: 10.3389/fmicb.2017.01315

Naji, S., Ambrus, G., Cimermancic, P., Reyes, J. R., Johnson, J. R., Filbrandt, R., et al. (2012). Host cell interactome of HIV-1 Rev includes RNA helicases involved in multiple facets of virus production. *Mol. Cell Proteom.* 11, M111.015313. doi: 10.1074/mcp.M111.015313

Nelemans, T., and Kikkert, M. (2019). Viral Innate Immune Evasion and the Pathogenesis of Emerging RNA Virus Infections. *Viruses* 11:916. doi: 10.20944/preprints201909.0212.v1

Northridge, M. E., and Metcalf, S. S. (2016). Enhancing implementation science by applying best principles of systems science. *Health Res. Policy Syst.* 14:74. doi: 10.1186/s12961-016-0146-8

Odum, H. T. (1996). *Environmental accounting : EMERGY and environmental decision making*. New York,NY: Wiley.

Odum, H. T. (2002). Explanations of ecological relationships with energy systems concepts. *Ecol. Model.* 158, 201–211. doi: 10.1016/s0304-3800(02)00232-6

Odum, H. T., and Odum, E. C. (2000). *Modeling for all scales : an introduction to system simulation*. San Diego: Academic Press.

Panja, A. S., Sarkar, A., Biswas, R., Bandyopadhyay, B., and Bandopadhyay, R. (2019). Modification of drug-binding proteins associated with the efflux pump in MDR-MTB in course of evolution: an unraveled clue based on in silico approach. *J. Antibiot.* 72, 282–290. doi: 10.1038/s41429-019-0146-3

Patterson, J. L., and Fernandez-Larsson, R. (1990). Molecular Mechanisms of Action of Ribavirin. *Rev. Infect. Dis.* 12, 1139–1146. doi: 10.1093/clinids/12.6.1139

Patzina, C., Botting, C. H., Garcia-Sastre, A., Randall, R. E., and Hale, B. G. (2017). Human interactome of the influenza B virus NS1 protein. *J. Gen. Virol.* 98, 2267–2273. doi: 10.1099/jgv.0.000909

Pécheur, E. I., Borisevich, V., Halfmann, P., Morrey, J. D., Smee, D. F., Prichard, M., et al. (2016). The Synthetic Antiviral Drug Arbidol Inhibits Globally Prevalent Pathogenic Viruses. *J. Virol.* 90, 3086–3092. doi: 10.1128/jvi.02077-15

Pelechano, V., Chavez, S., and Perez-Ortin, J. E. (2010). A complete set of nascent transcription rates for yeast genes. *PLoS One* 5:e15442. doi: 10.1371/journal.pone.0015442

Petroni, G., Bagni, G., Iorio, J., Duranti, C., Lottini, T., Stefanini, M., et al. (2020). Clarithromycin inhibits autophagy in colorectal cancer by regulating the hERG1 potassium channel interaction with PI3K. *Cell Death Dis.* 11:161.

Pfefferle, S., Schopf, J., Kogl, M., Friedel, C. C., Muller, M. A., Carbajo-Lozoya, J., et al. (2011). The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS Pathog.* 7:e1002331. doi: 10.1371/journal.ppat.1002331

Phiel, C. J., Zhang, F., Huang, E. Y., Guenther, M. G., Lazar, M. A., and Klein, P. S. (2001). Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. *J. Biol. Chem.* 276, 36734–36741. doi: 10.1074/jbc.m101287200

Popovic, M., and Minceva, M. (2020a). Thermodynamic insight into viral infections 2: empirical formulas, molecular compositions and thermodynamic properties of SARS, MERS and SARS-CoV-2 (COVID-19) viruses. *Heliyon* 6:e04943. doi: 10.1016/j.heliyon.2020.e04943

Popovic, M., and Minceva, M. (2020b). A thermodynamic insight into viral infections: do viruses in a lytic cycle hijack cell metabolism due to their low Gibbs energy? *Heliyon* 6:e03933. doi: 10.1016/j.heliyon.2020.e03933

Primadharsini, P. P., Nagashima, S., and Okamoto, H. (2019). Genetic Variability and Evolution of Hepatitis E Virus. *Viruses* 127, 216–228. doi: 10.1016/j.virusres.2007.02.002

Proshkin, S., Rahmouni, A. R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328, 504–508. doi: 10.1126/science.1184939

Qudrat-Ullah, H. (2012). On the validation of system dynamics type simulation models. *Telecommun. Syst.* 51, 159–166. doi: 10.1007/s11235-011-9425-4

Regoes, R. R., Crotty, S., Antia, R., and Tanaka, M. M. (2005). Optimal replication of poliovirus within cells. *Am. Nat.* 165, 364–373. doi: 10.2307/3473412

Reineke, L. C., Tsai, W. C., Jain, A., Kaelber, J. T., Jung, S. Y., and Lloyd, R. E. (2017). Casein Kinase 2 Is Linked to Stress Granule Dynamics through Phosphorylation of the Stress Granule Nucleating Protein G3BP1. *Mol. Cell Biol.* 37, e596–e516. doi: 10.1128/MCB.00596-16

Rothenburg, S., and Brennan, G. (2020). Species-Specific Host-Virus Interactions: Implications for Viral Host Range and Virulence. *Trends Microbiol.* 28, 46–56. doi: 10.1016/j.tim.2019.08.007

Saeed, M., Suzuki, R., Watanabe, N., Masaki, T., Tomonaga, M., Muhammad, A., et al. (2011). Role of the Endoplasmic Reticulum-associated Degradation (ERAD) Pathway in Degradation of Hepatitis C Virus Envelope Proteins and Production of Virus Particles. *J. Biol. Chem.* 286, 37264–37273. doi: 10.1074/jbc.M111.259085

Saha, B., and Parks, R. J. (2019). Histone Deacetylase Inhibitor Suberoylanilide Hydroxamic Acid Suppresses Human Adenovirus Gene Expression and Replication. *J. Virol.* 93, e88–e19. doi: 10.1128/JVI.00088-19

Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J., and Gale, M. Jr. (2008). Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* 454, 523–527. doi: 10.1038/nature07106

Schmidt, J., Braggio, E., Kortuem, K. M., Egan, J. B., Zhu, Y. X., Xin, C. S., et al. (2013). Genome-wide studies in multiple myeloma identify XPO1/CRM1 as a critical target validated using the selective nuclear export inhibitor KPT-276. *Leukemia* 27, 2357–2365. doi: 10.1038/leu.2013.172

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098

Scutigliani, E. M., and Kikkert, M. (2017). Interaction of the innate immune system with positive-strand RNA virus replication organelles. *Cytokine Growth Fact. Rev.* 37, 17–27. doi: 10.1016/j.cytogfr.2017.05.007

Sedmak, J. J., and Grossberg, S. E. (1973). Interferon bioassay: reduction in yield of myxovirus neuraminidases. *J. Gen. Virol.* 21, 1–7. doi: 10.1099/0022-1317-21-1-1

Sheahan, T. P., Sims, A. C., Graham, R. L., Menachery, V. D., Gralinski, L. E., Case, J. B., et al. (2017). Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci. Transl. Med.* 9:eaal3653. doi: 10.1126/scitranslmed.aal3653

Shi, L., Xiong, H., He, J., Deng, H., Li, Q., Zhong, Q., et al. (2007). Antiviral activity of arbidol against influenza A virus, respiratory syncytial virus, rhinovirus, coxsackie virus and adenovirus in vitro and in vivo. *Arch. Virol.* 152, 1447–1455. doi: 10.1007/s00705-007-0974-5

Shiraki, K., and Daikoku, T. (2020). Favipiravir, an anti-influenza drug against life-threatening RNA virus infections. *Pharm. Therap.* 209:107512. doi: 10.1016/j.pharmthera.2020.107512

Siddiqui-Jain, A., Drygin, D., Streiner, N., Chua, P., Pierre, F., O'brien, S. E., et al. (2010). CX-4945, an orally bioavailable selective inhibitor of protein kinase CK2, inhibits prosurvival and angiogenic signaling and exhibits antitumor efficacy. *Cancer Res.* 70, 10288–10298. doi: 10.1158/0008-5472.CAN-10-1893

Siwiak, M., and Zielenkiewicz, P. (2013). Transimulation - protein biosynthesis web service. *PLoS One* 8:e73943. doi: 10.1371/journal.pone.0073943

Spill, F., Bakal, C., and Mak, M. (2018). Mechanical and Systems Biology of Cancer. *Comput. Struct. Biotechnol. J* 16, 237–245. doi: 10.1016/j.csbj.2018.07.002

Sruthi, C. K., and Prakash, M. K. (2019). Statistical characteristics of amino acid covariance as possible descriptors of viral genomic complexity. *Sci. Rep.* 9:18410. doi: 10.1038/s41598-019-54720-y

Stamatiou, R., Paraskeva, E., Boukas, K., Gourgoulianis, K. I., Molyvdas, P.-A., and Hatziefthimiou, A. A. (2009). Azithromycin has an antiproliferative and autophagic effect on airway smooth muscle cells. *Eur. Respir. J.* 34, 721–730. doi: 10.1183/09031936.00089407

Sterman, J. D. (2002). All models are wrong: reflections on becoming a systems scientist. *Syst. Dyn. Rev.* 18, 501–531. doi: 10.1002/sdr.261

Tan, S.-L., Ganji, G., Paeper, B., Proll, S., and Katze, M. G. (2007). Systems biology and the host response to viral infection. *Nat. Biotechnol.* 25, 1383–1389. doi: 10.1038/nbt1207-1383

Tegner, J. N., Compte, A., Auffray, C., An, G., Cedersund, G., Clermont, G., et al. (2009). Computational disease modeling - fact or fiction? *BMC Syst. Biol.* 3:56. doi: 10.1186/1752-0509-3-56

Tenson, T., Lovmar, M., and Ehrenberg, M. (2003). The mechanism of action of macrolides, lincosamides and streptogramin B reveals the nascent peptide exit path in the ribosome. *J. Mol. Biol.* 330, 1005–1014. doi: 10.1016/S0022-2836(03)00662-4

Thaker, S. K., Ch'ng, J., and Christofk, H. R. (2019). Viral hijacking of cellular metabolism. *BMC Biol.* 17:59. doi: 10.1186/s12915-019-0678-9

Thorlund, K., Dron, L., Park, J., Hsu, G., Forrest, J. I, and Mills, E. J. (2020). A real-time dashboard of clinical trials for COVID-19. *Lancet Digit. Health* 2, e286–e287. doi: 10.1016/S2589-7500(20)30086-8

Ulgiati, S., and Brown, M. T. (2009). Emergy and ecosystem complexity. *Commun. Nonlinear Sci. Numerical. Simul.* 14, 310–321. doi: 10.1016/j.cnsns.2007.05.028

Ulgiati, S., Ascione, M., Zucaro, A., and Campanella, L. (2011). Emergy-based complexity measures in natural and social systems. *Ecol. Indic.* 11, 1185–1190. doi: 10.1016/j.ecolind.2010.12.021

Verchot, J. (2014). The ER quality control and ER associated degradation machineries are vital for viral pathogenesis. *Front. Plant Sci.* 5:66. doi: 10.3389/fpls.2014.00066

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., and Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439, 344–348. doi: 10.1038/nature04388

Vincent, M. J., Bergeron, E., Benjannet, S., Erickson, B. R., Rollin, P. E., Ksiazek, T. G., et al. (2005). Chloroquine is a potent inhibitor of SARS coronavirus infection and spread. *Virol J.* 2:69. doi: 10.1186/1743-422X-2-69

Walker, E. J., Younessi, P., Fulcher, A. J., Mccuaig, R., Thomas, B. J., Bardin, P. G., et al. (2013). Rhinovirus 3C protease facilitates specific nucleoporin cleavage and mislocalisation of nuclear proteins in infected host cells. *PLoS One* 8:e71316. doi: 10.1371/journal.pone.0071316

Wang, C. H., Chung, F. T., Lin, S. M., Huang, S. Y., Chou, C. L., Lee, K. Y., et al. (2014). Adjuvant treatment with a mammalian target of rapamycin inhibitor, sirolimus, and steroids improves outcomes in patients with severe H1N1 pneumonia and acute respiratory failure. *Crit. Care Med.* 42, 313–321. doi: 10.1097/CCM.0b013e3182a2727d

Wang, L., Fu, B., Li, W., Patil, G., Liu, L., Dorf, M. E., et al. (2017a). Comparative influenza protein interactomes identify the role of plakophilin 2 in virus restriction. *Nat. Commun.* 8:13876. doi: 10.1038/ncomms13876

Wang, Q., Li, Q., Liu, R., Zheng, M., Wen, J., and Zhao, G. (2016). Host cell interactome of PA protein of H5N1 influenza A virus in chicken cells. *J. Proteomics* 136, 48–54. doi: 10.1016/j.jprot.2016.01.018

Wang, Q., Li, Y., Dong, H., Wang, L., Peng, J., An, T., et al. (2017b). Identification of host cellular proteins that interact with the M protein of a highly pathogenic porcine reproductive and respiratory syndrome virus vaccine strain. *Virol J.* 14:39. doi: 10.1186/s12985-017-0700-1

Wargo, A. R., and Kurath, G. (2012). Viral fitness: definitions, measurement, and current insights. *Curr. Opin. Virol.* 2, 538–545.

Watanabe, T., Kawakami, E., Shoemaker, J. E., Lopes, T. J., Matsuoka, Y., Tomita, Y., et al. (2014). Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host Microb.* 16, 795–805.

Waterlow, J. C. (2006). *Protein turnover*. Cambridge, MA: CABI Pub.

Watters, K., Inankur, B., Gardiner, J. C., Warrick, J., Sherer, N. M., Yin, J., et al. (2017). Differential Disruption of Nucleocytoplasmic Trafficking Pathways by Rhinovirus 2A Proteases. *J. Virol.* 91, e2472–e2416. doi: 10.1128/JVI.02472-16

Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., et al. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373, 117–122. doi: 10.1038/373117a0

Wheatley, D. N., Giddings, M. R., and Inglis, M. S. (1980). Kinetics of degradation of 'short-' and 'long-lived' proteins in cultured mammalian cells. *Cell Biol. Int. Rep.* 4, 1081–1090. doi: 10.1016/0309-1651(80)90045-4

Wiederrecht, G., Lam, E., Hung, S., Martin, M., and Sigal, N. (1993). The mechanism of action of FK-506 and cyclosporin A. *Ann. N. Y. Acad. Sci.* 696, 9–19. doi: 10.1111/j.1749-6632.1993.tb17137.x

Wu, N. C., Otwinowski, J., Thompson, A. J., Nycholat, C. M., Nourmohammad, A., and Wilson, I. A. (2020). Major antigenic site B of human influenza H3N2 viruses has an evolving local fitness landscape. *Nat. Commun.* 11:1233. doi: 10.1038/s41467-020-15102-5

Wu, W., Tran, K. C., Teng, M. N., Heesom, K. J., Matthews, D. A., Barr, J. N., et al. (2012). The interactome of the human respiratory syncytial virus NS1 protein highlights multiple effects on host cell biology. *J. Virol.* 86, 7777–7789. doi: 10.1128/JVI.00460-12

Xue, J., Liu, G., Casazza, M., and Ulgiati, S. (2018). Development of an urban FEW nexus online analyzer to support urban circular economy strategy planning. *Energy* 164, 475–495. doi: 10.1016/j.energy.2018.08.198

Yang, Y., Ye, F., Zhu, N., Wang, W., Deng, Y., Zhao, Z., et al. (2015). Middle East respiratory syndrome coronavirus ORF4b protein inhibits type I interferon production through both cytoplasmic and nuclear targets. *Sci. Rep.* 5:17554. doi: 10.1038/srep17554

York, A., Hutchinson, E. C., and Fodor, E. (2014). Interactome analysis of the influenza A virus transcription/replication machinery identifies protein phosphatase 6 as a cellular factor required for efficient virus replication. *J. Virol.* 88, 13284–13299. doi: 10.1128/JVI.01813-14

Zheng, L. L., Li, C., Ping, J., Zhou, Y., Li, Y., and Hao, P. (2014). The domain landscape of virus-host interactomes. *Biomed. Res. Int.* 2014:867235. doi: 10.1155/2014/867235

Ziegler, C. M., Eisenhauer, P., Kelly, J. A., Dang, L. N., Beganovic, V., Bruce, E. A., et al. (2018). A Proteomics Survey of Junin Virus Interactions with Human Proteins Reveals Host Factors Required for Arenavirus Replication. *J. Virol.* 92, e1565–e1517.

Zumla, A., Chan, J. F. W., Azhar, E. I., Hui, D. S. C., and Yuen, K.-Y. (2016). Coronaviruses — drug discovery and therapeutic options. *Nat. Rev. Drug Discov.* 15, 327–347. doi: 10.1038/nrd.2015.37

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 17 February 2021
doi: 10.3389/fmicb.2020.615280

Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil

Paola Cristina Resende[1†], Edson Delatorre[2†], Tiago Gräf[3†], Daiana Mir[4], Fernando Couto Motta[1], Luciana Reis Appolinario[1], Anna Carolina Dias da Paixão[1], Ana Carolina da Fonseca Mendonça[1], Maria Ogrzewalska[1], Braulia Caetano[1], Gabriel Luz Wallau[5], Cássia Docena[5], Mirleide Cordeiro dos Santos[6], Jessylene de Almeida Ferreira[6], Edivaldo Costa Sousa Junior[6], Sandro Patroca da Silva[6], Sandra Bianchini Fernandes[7], Lucas Alves Vianna[8], Larissa da Costa Souza[9], Jean F. G. Ferro[10], Vanessa B. Nardy[11], Cliomar A. Santos[12], Irina Riediger[13], Maria do Carmo Debur[13], Júlio Croda[14,15], Wanderson K. Oliveira[16], André Abreu[17], Gonzalo Bello[18*‡] and Marilda M. Siqueira[1‡]*

[1] Laboratory of Respiratory Viruses and Measles, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (FIOCRUZ), SARS-CoV-2 National Reference Laboratory for the Brazilian Ministry of Health (MoH) and Regional Reference Laboratory in Americas for the Pan-American Health Organization (PAHO/WHO), Rio de Janeiro, Brazil, [2] Departamento de Biologia, Centro de Ciencias Exatas, Naturais e da Saude, Universidade Federal do Espirito Santo, Alegre, Brazil, [3] Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil, [4] Unidad de Genomica y Bioinformatica, Centro Universitario Regional del Litoral Norte, Universidad de la Republica, Salto, Uruguay, [5] Instituto Aggeu Magalhaes, Fundação Oswaldo Cruz, Recife, Brazil, [6] Instituto Evandro Chagas, Belem, Para, [7] Laboratorio Central de Saude Publica do Estado de Santa Catarina (LACEN-SC), Florianopolis, Brazil, [8] Laboratorio Central de Saude Publica do Estado Espirito Santo (LACEN-ES), Vitoria, Brazil, [9] Laboratorio Central de Saude Publica do Distrito Federal (LACEN-DF), Brazilia, Brazil, [10] Laboratorio Central de Saude Publica de Alagoas (LACEN-AL), Maceio, Brazil, [11] Laboratorio Central de Saude Publica da Bahia (LACEN-BA), Salvador, Brazil, [12] Laboratorio Central de Saude Publica de Sergipe (LACEN-SE), Aracaju, Brazil, [13] Laboratorio Central de Saude Publica de Parana (LACEN-PR), Curitiba, Brazil, [14] Fiocruz Mato Grosso do Sul, Campo Grande, Brazil, [15] Universidade Federal de Mato Grosso do Sul – UFMS, Campo Grande, Brazil, [16] Hospital das Forças Armadas, Ministério da Defesa, Brasília, Brazil, [17] Coordenadoria Geral de Laboratorios – Ministério da Saude, Brazilia, Brazil, [18] Laboratorio de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil

A previous study demonstrates that most of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) Brazilian strains fell in three local clades that were introduced from Europe around late February 2020. Here we investigated in more detail the origin of the major and most widely disseminated SARS-CoV-2 Brazilian lineage B.1.1.33. We recovered 190 whole viral genomes collected from 13 Brazilian states from February 29 to April 31, 2020 and combined them with other B.1.1 genomes collected globally. Our genomic survey confirms that lineage B.1.1.33 is responsible for a variable fraction of the community viral transmissions in Brazilian states, ranging from 2% of all SARS-CoV-2 genomes from Pernambuco to 80% of those from Rio de Janeiro. We detected a moderate prevalence (5–18%) of lineage B.1.1.33 in some South American countries and a very low prevalence (<1%) in North America, Europe, and Oceania. Our study reveals that lineage B.1.1.33 evolved from an ancestral clade, here designated B.1.1.33-like, that carries one of the two B.1.1.33 synapomorphic mutations. The B.1.1.33-like lineage may have been introduced from Europe or arose in

OPEN ACCESS

Edited by:
Josanne Hinke Verhagen,
Erasmus Medical Center, Netherlands

Reviewed by:
Luiz Carlos Junior Alcantara,
Oswaldo Cruz Foundation (Fiocruz), Brazil
Junki Maruyama,
University of Texas Medical Branch at Galveston, United States

***Correspondence:**
Paola Cristina Resende
paola@ioc.fiocruz.br
Gonzalo Bello
gbellobr@gmail.com

[†]These authors share first authorship

[‡]These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 08 October 2020
Accepted: 21 December 2020
Published: 17 February 2021

Citation:
Resende PC, Delatorre E, Gräf T, Mir D, Motta FC, Appolinario LR, Paixão ACD, Mendonça ACF, Ogrzewalska M, Caetano B, Wallau GL, Docena C, Santos MC, de Almeida Ferreira J, Sousa Junior EC, Silva SP, Fernandes SB, Vianna LA, Souza LC, Ferro JFG, Nardy VB, Santos CA, Riediger I, do Carmo Debur M, Croda J, Oliveira WK, Abreu A, Bello G and Siqueira MM (2021) Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. Front. Microbiol. 11:615280. doi: 10.3389/fmicb.2020.615280

Brazil in early February 2020 and a few weeks later gave origin to the lineage B.1.1.33. These SARS-CoV-2 lineages probably circulated during February 2020 and reached all Brazilian regions and multiple countries around the world by mid-March, before the implementation of air travel restrictions in Brazil. Our phylodynamic analysis also indicates that public health interventions were partially effective to control the expansion of lineage B.1.1.33 in Rio de Janeiro because its median effective reproductive number (R_e) was drastically reduced by about 66% during March 2020, but failed to bring it to below one. Continuous genomic surveillance of lineage B.1.1.33 might provide valuable information about epidemic dynamics and the effectiveness of public health interventions in some Brazilian states.

Keywords: coronavirus disease 2019, severe acute respiratory syndrome coronavirus-2, coronavirus, Brazil, genetic lineages, community transmission

INTRODUCTION

Coronavirus disease 2019 (COVID-19), the disease caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has gone pandemic, leading to high rates of acute respiratory syndrome, hospitalization, and death worldwide (Zhu et al., 2020). Brazil, the third most affected country in the world so far, has reported over 5.7 million cases and over 163,000 deaths (last update November 12, 2020) (Dong et al., 2020). The first positive case of SARS-CoV-2 infection in Brazil was reported on February 25, 2020 in an individual traveling from Europe to Sao Paulo metropolitan region (de Souza et al., 2020), and during the following 2 weeks, the virus was detected in all country regions (Jesus et al., 2020).

The rapid worldwide genomic surveillance of SARS-CoV-2, mainly shared via the EpiCoV database in the GISAID initiative[1] coupled with the development of real-time interactive platforms for pathogen genomic data[2], has been crucial for managing this public health emergency, enabling the tracking of viral transmission patterns as the epidemic unfolds. According to the nomenclature systems for designation of SARS-CoV-2 lineages[3], there are currently two major groups, designated as lineages A and B, that have diversified while the virus spread across the world, in several phylogenetic subclades characterized by the presence of particular sets of shared mutations (Shu and McCauley, 2017; Hadfield et al., 2018; Rambaut et al., 2020). The nomenclature implemented in the Pangolin software defined further SARS-CoV-2 active lineages by assigning a numerical value to descendants from either lineage A or B that fulfill a set of conditions. A maximum of three sublevels is allowed (e.g., A.1.1.1); after that, new descendent lineages are given a letter. The SARS-CoV-2 lineage B.1 was initially identified as the most common variant in Europe and is currently also the predominant viral lineage in the Americas (Candido et al., 2020; Deng et al., 2020; Fauver et al., 2020; Franco et al., 2020; Gonzalez-Reiche et al., 2020; Laiton-Donato et al., 2020; Paiva et al., 2020; Paniz-Mondolfi et al., 2020; Ramirez et al., 2020; Salazar et al., 2020; Taboada et al., 2020; Xavier et al., 2020).

[1]https://www.gisaid.org/

[2]https://nextstrain.org/ncov/global?c=region

[3]https://pangolin.cog-uk.io/

Genomic epidemiology has been a useful tool to track the community transmission of SARS-CoV-2, revealing that its epidemics in Oceania (Rockett et al., 2020; Seemann et al., 2020), Europe (Bluhm et al., 2020; Dellicour et al., 2020; Díez-Fuertes et al., 2020; Gambaro et al., 2020; Gudbjartsson et al., 2020; Miller et al., 2020; Oude Munnink et al., 2020), and the Americas (Deng et al., 2020; Gonzalez-Reiche et al., 2020; Worobey et al., 2020) resulted from multiple independent introductions followed by local dissemination of some viral strains. One previous study estimated that most (76%, n = 370/490) of the Brazilian SARS-CoV-2 strains recovered from different country regions group into three main clades designated with numbers (Candido et al., 2020). The clade 1, characterized by one non-synonymous mutation in the spike protein (G25088T), was subsequently classified by the Pangolin nomenclature as lineage B.1.1.28, and the clade 2, defined by two non-synonymous synapomorphies in ORF6 (T27299C/I33T) and the nucleocapsid protein (T29148C/I292T), was classified as lineage B.1.1.33. According to that study, the lineage B.1.1.28 circulates predominantly in the state of Sao Paulo and its origin was dated to February 28 (February 21 to March 4), while lineage B.1.1.33 circulates in different Brazilian states, and its origin was traced to February 22 (February 17–24), 2020. The precise dispersion pattern, demographical dynamics, and global prevalence of these major Brazilian clades are currently unclear.

Here we perform an in-depth analysis of the origin and spatiotemporal dissemination dynamics of B.1.1.33 lineage in Brazil and globally. To this end, we generated 190 SARS-CoV-2 whole-genomes sampled in 13 different Brazilian states during the first 2 months of the COVID-19 epidemic. New B.1.1.33 viral sequences identified here were combined with other Brazilian and global SARS-CoV-2 genomes containing the same synapomorphic mutations and then subjected to maximum likelihood (ML) and Bayesian phylogenetic analyses.

MATERIALS AND METHODS

Sampling and Ethical Aspects

We recovered viral whole-genomes (>99% coverage) from nasopharyngeal–throat combined swabs collected from 190

individuals with confirmed SARS-CoV-2 infection, who underwent testing and/or genomic sequencing at the Laboratory of Respiratory Viruses and Measles-Oswaldo Cruz Institute-FIOCRUZ, in Rio de Janeiro, the Technological Platform and Bioinformatic Core-Aggeu Magalhaes Institute-FIOCRUZ, in Pernambuco, and the Evandro Chagas Institute, in Para Brazil between February 29 and April 30, 2020. Samples were collected from clinically ill individuals (between the first and the 11th day after their first symptoms) or from asymptomatic individuals suspicious of SARS-CoV-2 infection that reside in 13 different Brazilian states from the Southeastern (Rio de Janeiro and Espirito Santo), Central-Western (Distrito Federal), Northern (Acre, Amapa, and Para), Northeastern (Alagoas, Bahia Maranhao, Pernambuco, and Sergipe), and Southern (Parana and Santa Catarina) regions (**Supplementary Table 1**). Samples were conserved in the viral transport medium at 4–8°C up to processing. This study was approved by the FIOCRUZ-IOC Ethics Committee (68118417.6.0000.5248 and CAAE 32333120.4.0000.5190) and the Brazilian Ministry of Health SISGEN (A1767C3).

Nucleic Acid Isolation and RT–qPCR

The viral RNA was extracted manually from 140 μl of clinical samples using QIAamp Viral RNA Mini kit (QIAGEN, Hilden, Germany) or automatedly using 300 μl of the sample and Perkin-Elmer Chemagic machine/chemistry, according to the manufacturer's instructions. SARS-CoV-2-positive cases were confirmed by real-time RT-PCR assays using the SARS-CoV-2 Molecular E/RP kit (Biomanguinhos, Rio de Janeiro, Brazil) based on the protocol previously designed by Corman et al. (2020). Amplifications were conducted in the ABI7500 platform using the following conditions: reverse transcription (50°C, 15 min), reverse transcriptase inactivation and DNA polymerase activation (95°C, 2 min), followed by 45 cycles of DNA denaturation (95°C, 20 s) and annealing-extension (58°C, 30 s). The fluorescence data was collected in the annealing-extension step, and all samples with sigmoid curves crossing the threshold line up to cycle 40 were named positive. Negative and positive controls were included in each extraction and real-time RT-PCR batch.

SARS-CoV-2 Whole-Genome Amplification and Sequencing

Total RNA from positive samples presenting Ct values up to 300 for gene E was reverse transcribed using SuperScript™ IV First Strand Synthesis System (Invitrogen). Two multiplex PCR reactions using the primer scheme previously described (Resende et al., 2020b) (Pool A = nine amplicons and Pool B = eight amplicons) were performed using the Q5® High-Fidelity DNA Polymerase (NEB). Amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter™), and the DNA quantified with Qubit 4 Fluorometer (Invitrogen) using the Qubit dsDNA HS Assay kit (Invitrogen) and sequenced using Illumina MiSeq or NextSeq (San Diego, CA, United States) and Nanopore (Oxford, United Kingdom). Illumina short reads DNA libraries were generated from the pooled amplicons using Nextera XT DNA Sample Preparation kit (Illumina, San Diego, CA, United States) according to the manufacturer specifications. The size distribution of these libraries was evaluated using a 2100 Bioanalyzer (Agilent, Santa Clara, United States), and the samples were pair-end sequenced (Micro V2, 300 cycles) on a MiSeq/NextSeq equipment (Illumina, San Diego, United States) in around 18 h. The Nanopore library protocol is optimized for long reads (2 kb amplicons) (Resende et al., 2020b). Library preparation was conducted using Ligation Sequencing 1D (SQK-LSK109 Oxford Nanopore Technologies (ONT) and Native Barcoding kit 1–24 (ONT), according to the manufacturer's instructions. After end repair using the NEBNext® Ultra™ II End Repair/dA-Tailing Module (New England Biolabs, NEB), the native barcodes were attached using a NEBNext® Ultra™ II Ligation Module (NEB). Up to 23 samples were pooled for sequencing in the same flow cell (FLOMIN106 flow cell R9.4.1), and a negative mock sample was loaded in each run for validation. The sequencing was performed for 12 h using the high-accuracy base calling in the MinKNOW software; however, the run was monitored by RAMPART[4] allowing us to stop the assay after 2 h, when $\geq$20 × depth for all amplicons was achieved.

Data Analysis to Recover the SARS-CoV-2 Whole-Genome Consensus Sequences

Demultiplexed fastq files generated by Illumina sequencing were used as the input for the analysis. Reads were trimmed based on quality scores with a cutoff of Q30, in order to remove low-quality regions, and adapter sequences were filtered. Following standard pre-processing steps, reads were mapped to the hCoV-19/Wuhan/Hu-1/2019 strain (GISAID accession number EPI_ISL_402125). Duplicate reads were removed from the alignment and the consensus sequence called at a threshold of 10×. The entire workflow was carried out in CLC Genomics Workbench software version 20.0[5]. For the ONT sequencing data, the pipeline used was an adaptation of the nCoV-2019 novel coronavirus bioinformatics protocol developed by Artic[6]. We used an earlier version of the workflow, which used Porechop[7] to demultiplex the reads and trim off the adaptors. The mapping to the hCoV-19/Wuhan/Hu-1/2019 reference sequence was done using Minimap2 (Li, 2018) and Medaka[8]. This was all carried out within the artic-ncov2019 conda environment[9]. All genomes produced in this study can be accessed in https://www.gisaid.org/ under the accession number listed in **Supplementary Table 1**.

Maximum Likelihood Phylogenetic Analyses

New Brazilian SARS-CoV-2 genome sequences were assigned to viral lineages according to the nomenclature proposed in

[4]https://github.com/articnetwork/rampart
[5]https://digitalinsights.qiagen.com
[6]https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html
[7]https://github.com/rrwick/Porechop
[8]https://github.com/nanoporetech/medaka
[9]https://github.com/artic-network/artic-ncov2019

Rambaut et al. (2020), using the Pangolin web application[10]. Subsequently, an international dataset composed by all genomes available at GISAID[11] as of July 31 was screened for sequences harboring the B.1.1.33 characteristic mutations ORF6:T27299C and/or N:T29148C. As a reference dataset, we also download from GISAID high quality (<10% of ambiguous positions) complete (>29 kilobases) genome sequences belonging to the global lineage B.1.1 (n = 7,766), which also prevails in Brazil. To reduce the size of the global B.1.1 dataset while preserving the full genetic diversity, we construct a "non-redundant" subset (n = 3,053) by grouping sequences by similarity (sequence identity cut-off = 1.0) with the CD-HIT program (Li, 2018) and then selecting one representative sequence from each group. SARS-CoV-2 B.1.1 complete genome sequences were aligned using MAFFT v7.467 (Katoh and Standley, 2013) and subject to maximum likelihood (ML) phylogenetic analysis. The ML phylogenetic tree was inferred using IQ-TREE v1.6.12 (Nguyen et al., 2015) under the GTR + F + I + G4 nucleotide substitution model, as selected by the ModelFinder application (Kalyaanamoorthy et al., 2017), and the branch support was assessed by the approximate likelihood-ratio test based on the Shimodaira–Hasegawa-like procedure (SH-aLRT) with 1,000 replicates. The ML phylogenetic tree was visualized using the software FigTree v1.4[12].

Analysis of Temporal Signal and Phylogeographic Structure

All genomes identified as B.1.1.33 and closely related basal sequences were submitted to the Bayesian phylogenetic analysis in MrBayes program (Ronquist et al., 2012), under the GTR + I + G substitution model. Two Markov Chain Monte Carlo (MCMC) were run for 10 million generations and stationarity (constant mean and variance of trace plots) and good mixing (effective sample size > 200) were assessed using TRACER v1.7 (Rambaut et al., 2018). The temporal signal was assessed from the maximum clade credibility (MCC) tree by performing a regression analysis of the root-to-tip divergence against sampling time using TempEst (Rambaut et al., 2016). The degree of phylogeographic structure (phylogenetic clustering by sampling location) was investigated using the BaTS program (Parker et al., 2008), which assesses phylogeny-trait association in a posterior sampling of trees. This analysis included metrics as association index (AI), the parsimony score (PS), and the maximum clade (MC) and compared to a null hypothesis generated by tip randomization. Results were considered significant for $P < 0.01$.

Bayesian Phylogeographic Analyses

The age of the most recent common ancestor (T_{MRCA}) and the spatial diffusion pattern of the B.1.1.33 lineage and closely related basal sequences were jointly estimated using a Bayesian MCMC approach implemented in BEAST 1.10 (Suchard et al., 2018), using the BEAGLE library v3 (Ayres et al., 2019) to improve computational time. Time-scaled Bayesian trees were estimated under GTR + I + G nucleotide substitution model and a strict molecular clock with a uniform substitution rate prior ($8–10 \times 10^{-4}$ substitutions/site/year) based on previous estimates (Duchene et al., 2020; Ghafari et al., 2020). The non-parametric Bayesian skyline (BSKL) (Drummond et al., 2005) and Bayesian skygrid (BSKG) (Gill et al., 2013) coalescent models were applied to adjust the demographic signal contained in the dataset. Viral migrations across time were reconstructed using a reversible discrete phylogeographic model (Lemey et al., 2009) with a CTMC rate reference prior (Ferreira and Suchard, 2008). For each analysis, two MCMC chains were run for 100–250 million generations and then combined. Stationarity and good mixing for all parameter estimates were assessed using TRACER v1.7 (Rambaut et al., 2018) as explained above. The maximum clade credibility (MCC) tree was summarized with TreeAnnotator v1.10 and visualized using the FigTree v1.4 program.

Effective Reproductive Number (R_e) Estimation

To estimate the R_e of the B.1.1.33 lineage in Rio de Janeiro state through time, we used the birth-death skyline (BDSKY) model (Stadler et al., 2013) implemented within BEAST2 v2.6.2 (Bouckaert et al., 2019). The sampling rate (δ) was set to zero for the period prior to the oldest sample and estimated from the data afterward. The BDSKY prior settings for the model parameters are shown in **Supplementary Table 2**. Origin parameter was conditioned to the root height, and the R_e was estimated in a piece-wise manner over seven time-intervals defined from the date of the most recent sample (April 30, 2020) up to the root of the tree. One MCMC chain was run for 50 million generations and then checked for stationarity and mixing as explained above.

RESULTS

Prevalence of SARS-CoV-2 Lineage B.1.1.33 in Brazil and Worldwide

To estimate the prevalence of lineage B.1.1.33 across different Brazilian states over the first 2 months of the epidemic, we generated 190 high-quality viral whole-genomes (>99% coverage) from individuals with confirmed SARS-CoV-2 infection between February 29 and April 30, 2020 that reside in 13 different Brazilian states from the Southeast (n = 94, 49%), Northeast (n = 66, 35%), North (n = 14, 7%), South (n = 11, 6%), and Central-West (n = 5, 3%) regions (**Supplementary Table 1**). These SARS-CoV-2 Brazilian sequences here obtained (n = 190, 26%) were then combined with Brazilian sequences available in GISAID on July 31 (n = 737, 100%) (**Supplementary Table 3**). The overall prevalence of the lineage B.1.1.33 among Brazilian samples was 33% (n = 244/737), but with high variability across regions and time. The highest prevalence of this lineage was observed in the Southeastern state of Rio de Janeiro (80%, n = 122/153), while the lowest prevalence was observed in the Northeastern states of Ceara (3%, n = 1/31)

[10]https://pangolin.cog-uk.io

[11]https://www.gisaid.org/

[12]http://tree.bio.ed.ac.uk/software/figtree/

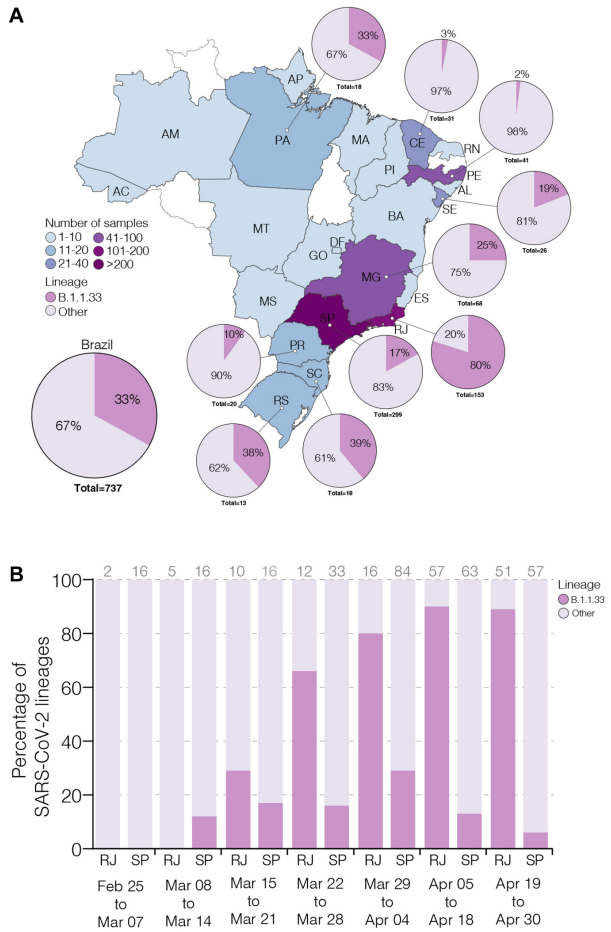


FIGURE 1 | Prevalence of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) B.1.1.33 lineage in Brazil. **(A)** Map showing the prevalence of SARS-CoV-2 B.1.1.33 across Brazilian states with more than 10 viral genomes available. The color of each state represents the number of sequences used in this study. The pie charts show the proportion of sequences that belong to the B.1.1.33 lineage in each Brazilian state with >10 sequences and in the whole country. **(B)** Dynamics of B.1.1.33 lineage prevalence in Rio de Janeiro (RJ) and Sao Paulo (SP) states over time. The total of sequences analyzed in each period is indicated above each column. The two letter codes defining the Brazilian states follow the ISO 3166-2:BR standard.

and Pernambuco (2%, n = 1/41) (**Figure 1A**). Longitudinal analyses in the most heavily affected states, Rio de Janeiro and Sao Paulo, also revealed very different temporal dynamics of lineage B.1.1.33 prevalence (**Figure 1B**). In Rio de Janeiro, there was a fast increase in the relative proportion of B.1.1.33 up to 90% (n = 110/122) in early April and a sustained high prevalence during the following weeks. While in Sao Paulo, there was a more gradual increase in prevalence of lineage B.1.1.33 up to 30% (n = 90/299) in early April and a subsequent decline toward the end of the month. None of the 23 early SARS-CoV-2 Brazilian sequences sampled between February 25 and March 7, 2020, mostly derived from imported cases, belong to lineage B.1.1.33. The lineage B.1.1.33 was first detected in Brazil on March 9 in Sao Paulo state and between March 12 and 19 in different states from all country regions (**Supplementary Table 3**).

To estimate the prevalence of lineage B.1.1.33 outside Brazil, we analyzed almost 75,000 SARS-CoV-2 complete genome sequences sampled worldwide available in GISAID on July 31. We estimated a moderate prevalence (5–18%) of lineage B.1.1.33 across countries from the South American cone (Argentina, Chile and Uruguay), and a very low prevalence (<1%) in some countries from North America (Canada and United States), Europe (England, Portugal, and Scotland), and Oceania (Australia) (**Supplementary Table 3**). This finding could not be explained by sampling bias as North America, Europe, and Oceania comprise the most densely sampled countries worldwide and clearly suggests that lineage B.1.1.33 was much prevalent in Brazil and other countries from the Southern cone than in any other geographic region. One interesting finding was the detection of some B.1.1.33 strains outside Brazil at very early times, including one virus sampled in Argentina on March 1 (the oldest B.1.1.33 virus detected so far) and viruses sampled in the United States and Canada on March 7, in England on March 12, in Chile and Australia on March 17, and in Portugal on March 18. Thus, the lineage B.1.1.33 was first detected in Argentina, Canada, and the United States at the first week of March 2020, in Brazil and England at the second week of March, and in Australia, Chile, and Portugal at the third week of March. The nearly simultaneous detection of lineage B.1.1.33 in different countries from the Americas and Europe as well as in different Brazilian regions is consistent with a period of cryptic circulation and dissemination of this lineage before its detection in early March 2020.

Origin and Worldwide Dissemination of SARS-CoV-2 Lineages B.1.1.33

To investigate in more detail the origin of the lineage B.1.1.33, we thoroughly searched all B.1.1 sequences available in GISAID on July 31 (n = 7,766) for the presence of the B.1.1.33 lineage-defining non-synonymous synapomorphies (ORF6:T27299C and/or N:T29148C). That search retrieved all B.1.1.33 sequences and a group of 28 B.1.1 sequences (19 from Europe, eight from Brazil, and one from Australia) that only harbor the mutation N:T29148C (**Supplementary Table 4**). These sequences, here named B.1.1.33-like, were aligned with all B.1.1.33 sequences described above and with a reference dataset of 3,053 B.1.1 sequences representing the global diversity of that lineage. The ML phylogenetic analysis revealed that the B.1.1.33 and B.1.1.33-like sequences branched together in a highly supported (SH-*aLRT* = 91%) monophyletic clade, with the B.1.1.33 sequences grouping in a monophyletic subcluster (SH-*aLRT* = 85%) nested among the basal B.1.1.33-like sequences (**Figure 2**). Most (85%, n = 19/22) B.1.1.33-like sequences from Europe were sampled between February 2 and April 6, 2020, and all Brazilian B.1.1.33-like sequences were sampled between March 13 and 19 in the states of Minas Gerais and Federal District. The overall prevalence of lineage B.1.1.33-like was low in Brazil (1.1%, n = 8/737) and outside (<1%, n = 20/3,926) but was moderately high (10%, n = 7/68) in the Brazilian state of Minas

Gerais (**Supplementary Table 4**). None of the individuals from Minas Gerais infected with the B.1.1.33-like viruses reported international travel, supporting that they correspond to local transmission (Xavier et al., 2020). Interestingly, the earliest B.1.1.33-like strain detected in Brazil was identical to some of those detected in Europe (**Supplementary Table 5**).

The B.1.1.33 and B.1.1.33-like sequences sampled worldwide were next combined and analyzed using Bayesian phylogeographic approaches. The worldwide B.1.1.33/B.1.1.33-like dataset displayed a significant geographic structure (**Supplementary Table 6**), but a weak temporal structure ($R^2 = 0.07$) (**Supplementary Figure 1**) and time-scaled Bayesian trees were thus reconstructed using a stringent molecular clock prior. The BSKG analysis traced the origin of the B.1.1.33-like lineage most probably to Europe [Posterior state probability (PSP) = 0.89] on February 4 [95% high posterior density (HPD): January 18 to February 17], its dissemination to Brazil around February 16 (95% HPD: February 1–27), and the origin of lineage B.1.1.33 to Brazil (*PSP* = 0.86) on February 22 (95% HPD: February 9–29) (**Figure 3**). The paucity of SARS-CoV-2 Brazilian sequences sampled in February and early March (**Figure 3**, inset), however, might represent a critical constraint for accurate phylogeographic reconstructions. To test this hypothesis, we conducted a new phylogeographic analysis with a subset of 16 B.1.1.33-like sequences (eight from Brazil, seven from Europe, and one from Australia) and 98 B.1.1.33 sequences (72 from Brazil) sampled between March 5 and 31. The new sampling scheme had a tremendous impact on the inferred transmission history as it traced the most probable origin of both B.1.1.33-like (*PSP* = 0.93) and B.1.1.33 (*PSP* = 0.99) lineages to Brazil on February 21 (95% HPD: February 3 to March 3) and February 29 (95% HPD: February 21 to March 5), respectively (**Figure 4** and **Supplementary Table 7**). Thus, any conclusion about the spatial and temporal origin of lineages B.1.1.33-like and B.1.1.33 should be interpreted with caution.

Dissemination Dynamics of SARS-CoV-2 Lineage B.1.1.33 in Rio de Janeiro

The phylogeographic structure of the B.1.1.33 lineage circulating in different Brazilian states did not differ from the null hypotheses of panmixis (**Supplementary Table 8**), demonstrating that the low genetic diversity and rapid spread of the B.1.1.33 lineage in Brazil imposes an important constraint to reconstruct the early within-country geographic dispersion pattern of this viral lineage. Phylogenomic modeling of the B.1.1.33 lineage, however, could offer important clues about the epidemic dynamics and effectiveness of control measures in those Brazilian states where the COVID-19 epidemic was dominated by this viral variant, like the Rio de Janeiro state. Indeed, the analysis of B.1.1.33 sequences from Rio de Janeiro using the BDSKY model (**Figure 5A**) supports an epidemic pattern that is fully consistent with changes in human mobility according to the Google Community Mobility Reports (**Figure 5B**) and with temporal trend of new diagnoses of SARS-CoV-2 (**Figure 5C**). The BDSKY model suggests a steep (66%) decrease in the R_e from 3.8 (95% HPD: 1.6–7.0) in late February and beginning in March, to 1.3 (95% HPD: 0.7–2.1) by the end of March, which coincides with the implementation of different control measures in this Brazilian state between March 10 and 20 that drastically reduce both human mobility, and the rate of incidence increase. Our phylogenomic analysis also suggests that the median *Re* stabilizes just above 1.0 during April, with the exception of a short interval between the 6th and 12th when median *Re* was about one, consistent with a transient stabilization of incidence trend during the second week of April. Thus, although early social distancing measures were able to drastically decrease the *Re* of the lineage B.1.1.33 in the Rio de Janeiro state, they probably failed to bring its spread under control ($Re < 1$).

DISCUSSION

Our genomic survey confirms that SARS-CoV-2 lineage B.1.1.33 is responsible for a substantial fraction of the early community viral transmissions in Brazil (∼33%, n = 244/737), but it is not homogenously distributed across the country (Candido et al., 2020). The estimated prevalence of lineage B.1.1.33 ranges from 2% (n = 1/41) of the SARS-CoV-2 genomes sampled in Pernambuco state to 80% (n = 122/153) in Rio de Janeiro between February and April 2020. Our study also suggests that lineage B.1.1.33 followed different temporal trends across Brazilian states. The prevalence of lineage B.1.1.33 rapidly increased during March in Rio de Janeiro, reaching up to 90% (n = 110/122) of SARS-CoV-2 genomes sampled by early April. In Sao Paulo, by contrast, this lineage displayed a more gradual increase in prevalence up to 30% (n = 90/299) of viral genomes sampled in early April and then declined toward the end of the month. These findings support that SARS-CoV-2 genetic diversity exhibits great variation across different Brazilian states and that in-depth molecular epidemiologic studies at a local scale should be necessary to fully understand the complex dynamics of the COVID-19 epidemic in different locations in the country during the early phase.

Our study shows for the first time that lineage B.1.1.33 was not unique to Brazil, but also circulated in other countries from the Americas, Europe, and Australia. The lineage B.1.1.33 exhibits a medium prevalence (5–20%) in South American countries from the Southern Cone and a very low prevalence (<1%) in countries from North America, Europe, and Oceania. Notably, we found that lineage B.1.1.33 strain was first detected in Argentina (March 1) and North America (March 7) before Brazil (Sao Paulo state, March 9). This may reflect the origin and earliest circulation of lineage B.1.1.33 outside Brazil or may be the consequence of limitations of public health systems to detect the early introduction and local circulation of SARS-CoV-2. Indeed, most SARS-CoV-2 cases detected in the very early phase of Brazilian epidemic (<March 5, 2020) correspond to imported cases, and community viral transmission was only officially recognized on March 20 (Melo et al., 2020). The paucity of diagnosis of local SARS-CoV-2 cases in Brazil during February and early March certainly might have limited the ability to detect early circulation of the autochthonous SARS-CoV-2 Brazilian lineage B.1.1.33.

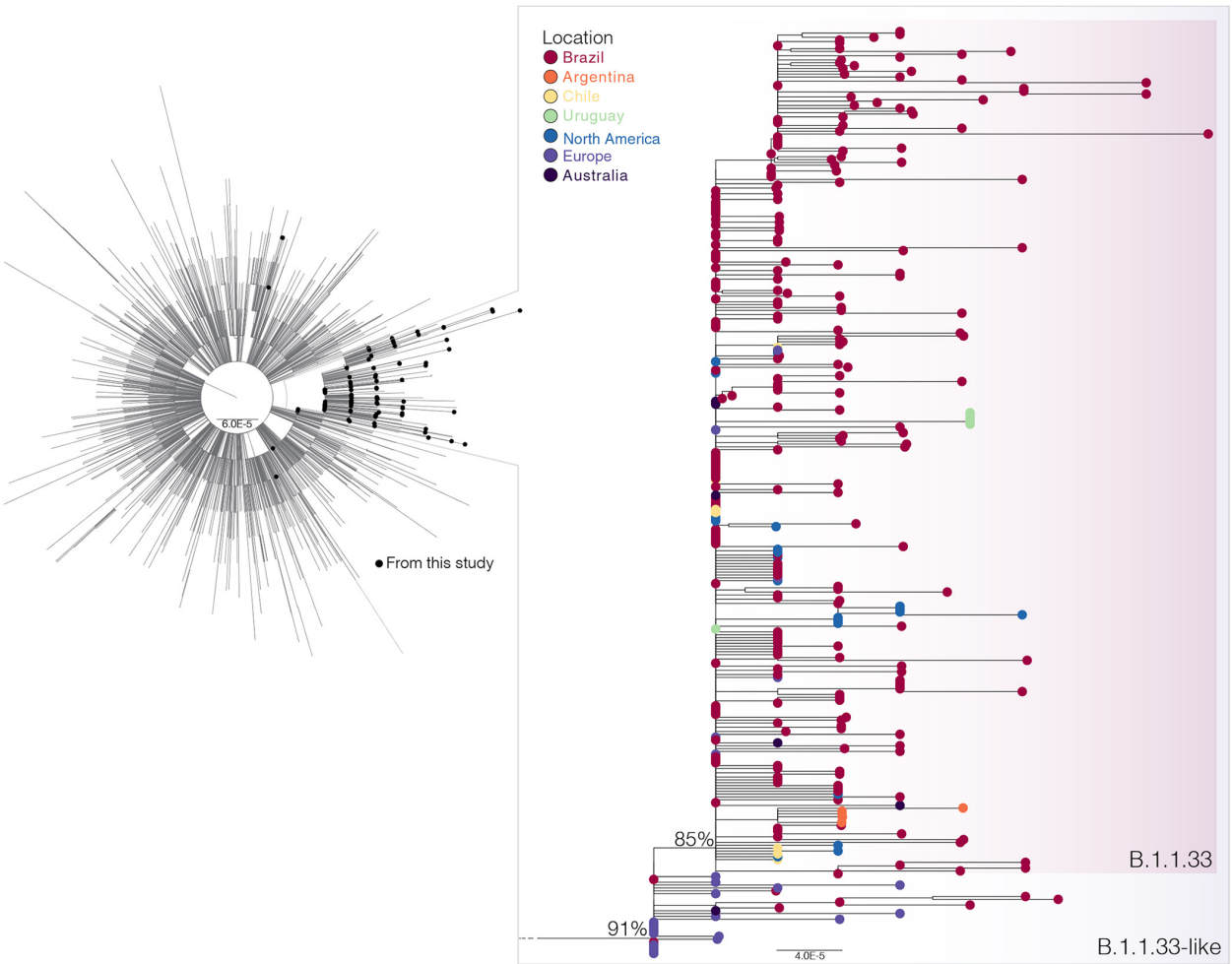


FIGURE 2 | Phylogenetic relationships of SARS-CoV-2 B.1.1 Brazilian and global strains. Maximum likelihood (ML) phylogenetic tree of 190 B.1.1 Brazilian genomes obtained in this survey (black circles) along with 3,053 B.1.1 international reference sequences from the GISAID database. The zoomed view shows the clusters B.1.1.33-like and B.1.1.33. Tip circles are colored according to the sampling location. Only node supports (aLRT) above 70% are shown. Shaded boxes highlight the position of clusters B.1.1.33-like and B.1.1.33. Tree was rooted on midpoint, and branch lengths are drawn to scale with the bars at the bottom indicating nucleotide substitutions per site.

The lineage B.1.1.33 was originally defined by the presence of two non-synonymous synapomorphies in ORF6 (T27299C/I33T) and the nucleocapsid protein (T29148C/I292T) (Candido et al., 2020). Our analysis of over 7,000 B.1.1 sequences available on GISAID initiative, however, revealed a group of 28 sequences (19 from Europe, eight from Brazil, and one from Australia) that also harbor mutation N:T29148C. These sequences, designated here as B.1.1.33-like, branched basal to lineage B.1.1.33 and represents an evolutionary intermediate between lineages B.1.1 and B.1.1.33. Thus, mutation N:T29148C is a synapomorphy common to lineages B.1.1.33-like and B.1.1.33. Lineage B.1.1.33-like was first detected in Netherlands and Switzerland (February 2–29) and only later in Brazil (March 13–19, 2020) and displayed a very low prevalence in Europe (<1%) and Brazil (1%). Although this viral lineage was only detected in a few early samples from Brazil, epidemiological data supports local transmission of this viral lineage in the Southeastern state of Minas Gerais (Xavier et al., 2020). It is currently unclear whether the lineage B.1.1.33-like represents an extinct viral variant or continues to be transmitted among Brazilian and/or European populations.

Several lines of evidence support that clades B.1.1.33 and B.1.1.33-like circulated for several weeks before being detected in symptomatic carriers in Brazil between March 9 and 13, 2020. Our phylogeographic reconstructions with the complete dataset traced the introduction of lineage B.1.1.33-like to Brazil on February 16 (95% HPD: February 1 to February 27) and the subsequent local origin of lineage B.1.1.33 on February 22 (95% HPD: February 9–29), consistent with a previous estimate (Candido et al., 2020). The nearly simultaneous detection of clade B.1.1.33 in distant states from all Brazilian regions between March 9 and 19 is also compatible with a period of untracked local circulation of this variant. Our findings are also supported by epidemiological surveillance data (available from: https://covid.saude.gov.br/) that reveals the

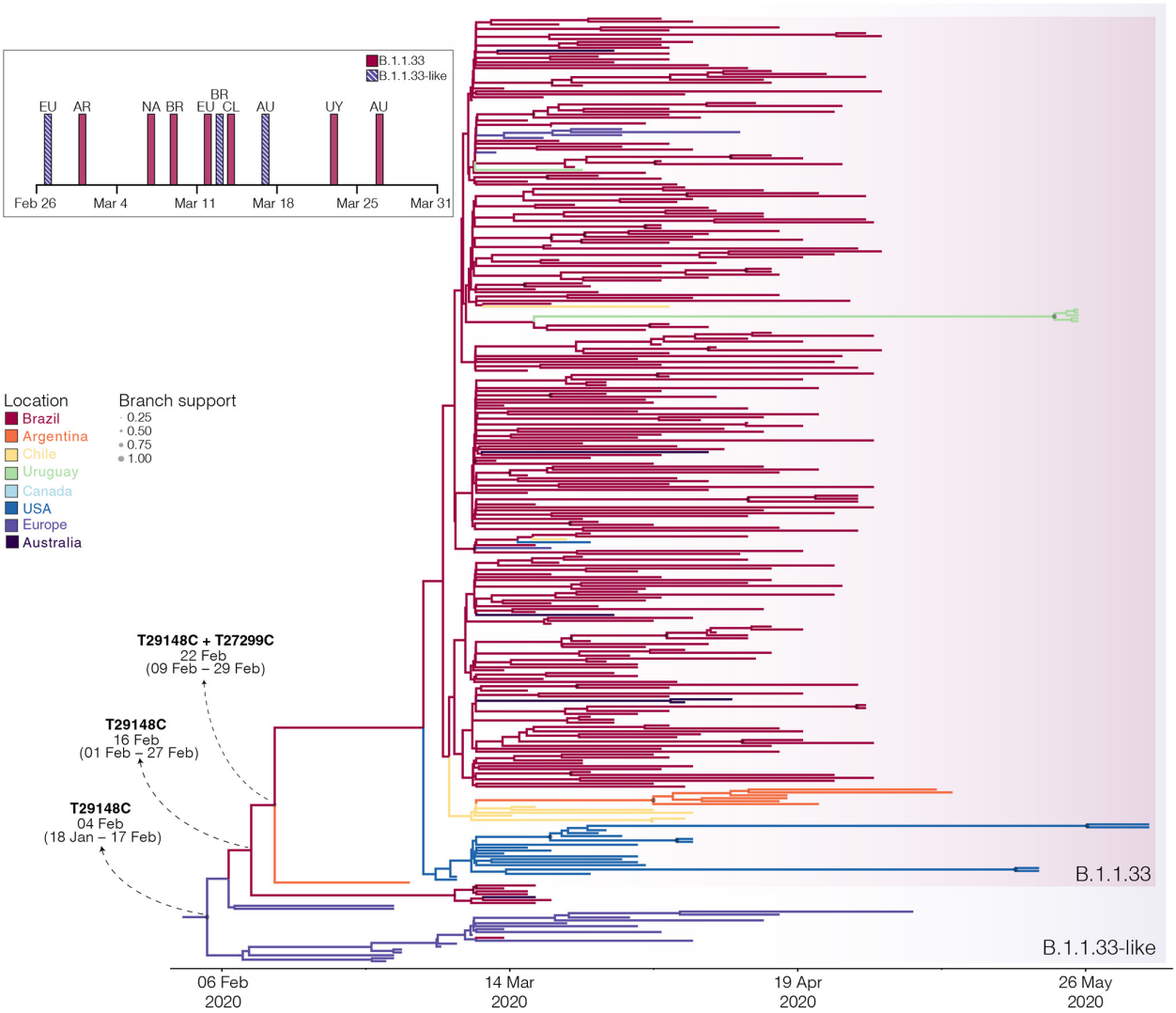


FIGURE 3 | Spatiotemporal dissemination of the SARS-CoV-2 clades B.1.1.33-like and B.1.1.33. Time-scaled Bayesian phylogeographic MCC tree of the major B.1.1 lineages circulating in Brazil. Branches are colored according to the most probable location state of their descendent nodes as indicated at the legend. Circle sizes at internal nodes are proportional to the corresponding posterior probability support as indicated at the legend. The inferred T_{MRCA} (based on the median of the posterior heights) and nucleotide substitutions fixed at ancestral key nodes are shown. Shaded boxes highlight the position of the clades B.1.1.33-like and B.1.1.33. The tree is automatically rooted under the assumption of a strict molecular clock, and all horizontal branch lengths are drawn to a scale of years. The inset figure depicts the timeline of the earliest detection of clades B.1.1.33-like (blue dashed bars) and B.1.1.33 (red bars) in Europe (EU), North America (NA), Australia (AU), Argentina (AR), Brazil (BR), Chile (CL), and Uruguay (UY).

onset symptoms of the first SARS-CoV-2-positive cases with severe acute respiratory illness in Brazil since February 16–22. Also, in agreement with our results, a recent study based on projections from the cumulative number of deaths during the early phase estimated the establishment of local transmission of SARS-CoV-2 in Brazil between early and mid-February 2020 (Delatorre et al., 2020).

Reconstructions of the spatial and temporal origin of SARS-CoV-2 lineages B.1.1.33-like and B.1.1.33 could be biased due the uneven spatial and temporal sampling during early phase as was described for other lineages (Lemey et al., 2020; Villabona-Arenas et al., 2020; Worobey et al., 2020). Indeed, our phylogeographic analyses supports two different viral transmission histories according to the dataset analyzed. Analysis of the complete dataset supports that lineage B.1.1.33-like most probably arose in Europe and was later disseminated to Brazil, where it spread and gave origin to lineage B.1.1.33 (**Figure 6A**). When early B.1.1.33-like sequences sampled in Europe at late February were removed, however, phylogeographic analysis indicates that both lineages originated at sequential steps during viral local spread in Brazil (**Figure 6B**). This second phylogenetic hypothesis is consistent with some epidemiological data. First, the extremely low prevalence of the clade B.1.1.33-like in Europe makes its transmission to Brazil a highly unlikely epidemiological scenario.

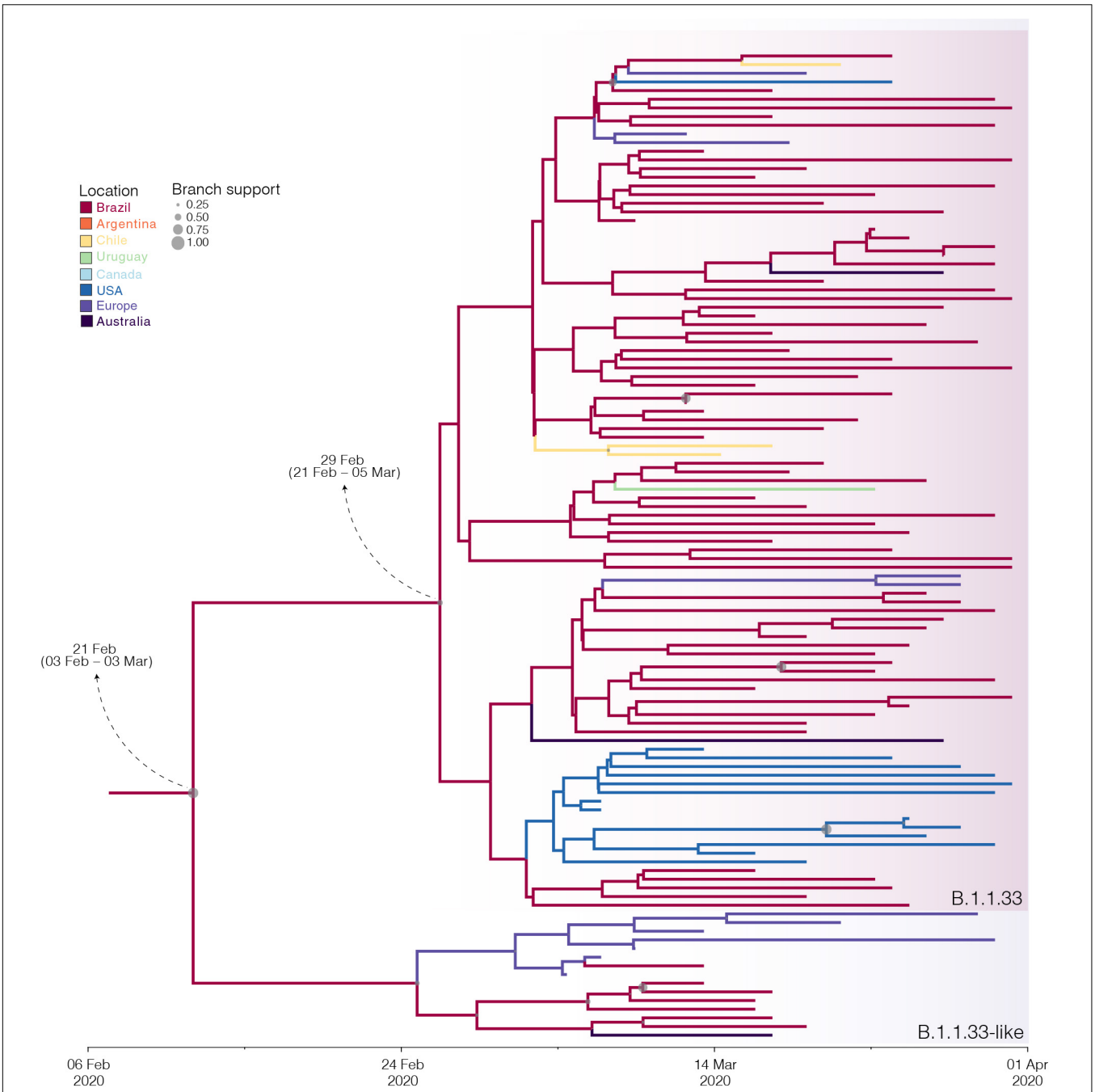


FIGURE 4 | Influence of sampling on the phylogeographic reconstruction of SARS-CoV-2 clade B.1.1.33. Time-scaled Bayesian phylogeographic maximum clade credibility (MCC) tree of a subset of sequences comprising the oldest B.1.1.33-like and B.1.1.33 strains. Branches are colored according to the most probable location state of their descendent nodes as indicated at the legend. Circle size at internal nodes are proportional to the corresponding posterior probability support as indicated at the legend. The inferred T_{MRCA} (based on the median of the posterior heights) and nucleotide substitutions fixed at ancestral key nodes are shown. Shaded boxes highlight the position of the clades B.1.1.33-like and B.1.1.33. The tree is automatically rooted under the assumption of a strict molecular clock, and all horizontal branch lengths are drawn to a scale of years.

Second, the paucity of B.1.1.33-like and B.1.1.33 sequences detected in Europe after early April 2020 coincides with a sharp decrease in the flow of international air travels to/from Brazil (**Figure 6C**), suggesting that those infections might be associated to imported cases from Brazil.

Our analyses support that lineage B.1.1.33 was exported from Brazil to neighboring South American countries (Argentina, Chile, and Uruguay) and also to more distant countries (i.e., Canada, United States, Portugal, United Kingdom, and Australia) at multiple times. It is interesting to note that most (>80%,

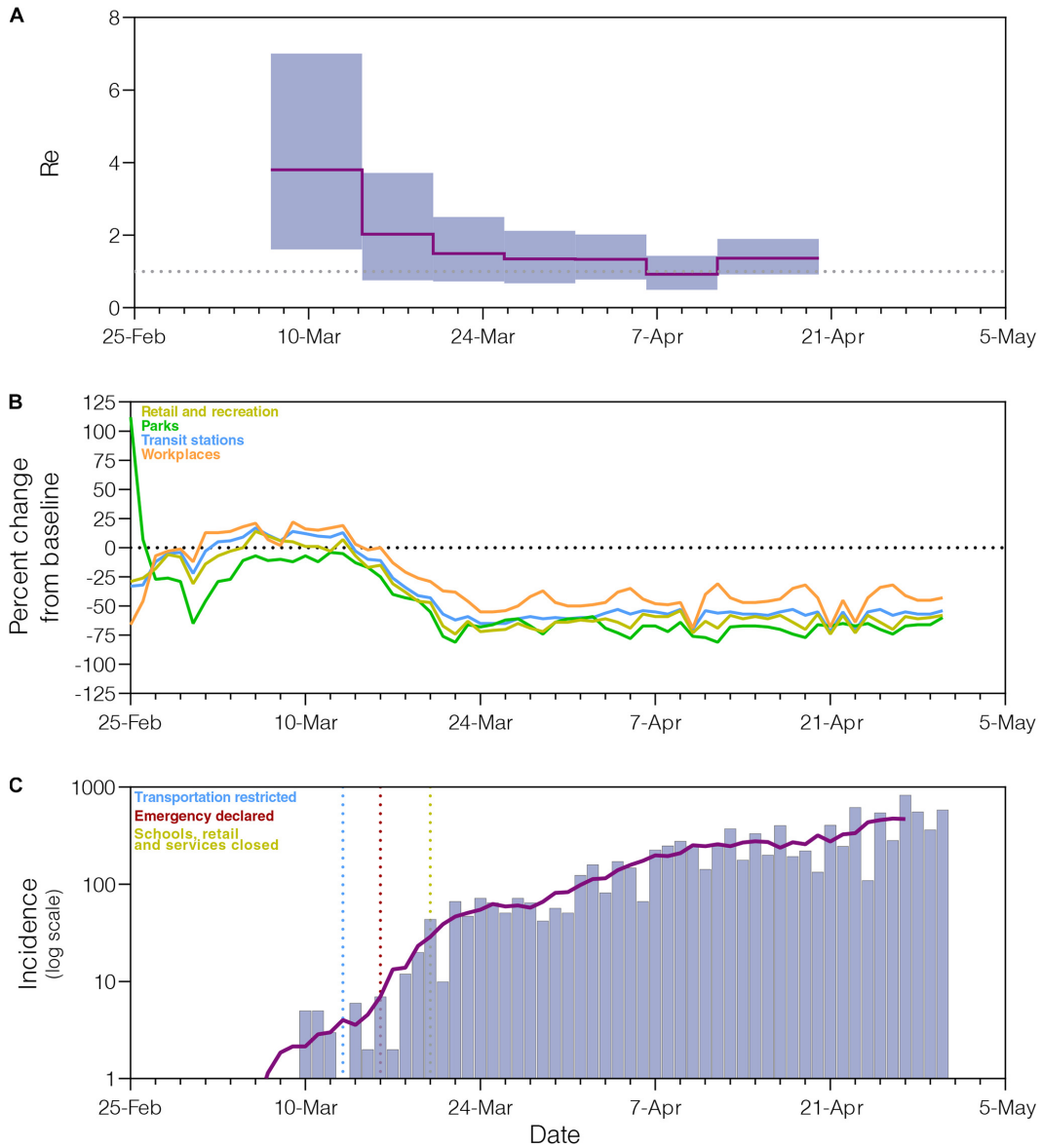


FIGURE 5 | Epidemiological and mobility indicators of the SARS-CoV-2 epidemic in Rio de Janeiro. **(A)** Temporal variation of the effective reproductive number (Re) of the B.1.1.33 lineage in Rio de Janeiro estimated using the Bayesian birth–death approach. **(B)** Mobility data trends reported as percentage change measured against baseline. Each parameter is colored following the legend. **(C)** Progress of incidence of SARS-CoV-2 in Rio de Janeiro. The line represents the weekly average. The date of the main epidemic control measures is indicated by the vertical dotted lines.

$n = 40/53$) B.1.1.33 sequences detected in Europe, North America, Australia, and Chile were sampled before implementation of international air travel restrictions to/from Brazil around early April. By contrast, nearly all B.1.1.33 sequences detected in South American countries of Argentina and Uruguay were sampled between mid-April and late May, thus long after Brazil closed its land borders and limited international air travels. These findings support that air travel restrictions were able to limit the long-distance dispersion of lineage B.1.1.33 from Brazil, but closing of land borders were not so efficient to limit the short-distance dissemination to neighboring South American countries. Our phylogeographic reconstruction also suggests that lineage B.1.1.33 might have seeded secondary outbreaks in Argentina and Uruguay, but those findings should be interpreted with caution because of the relative low number of B.1.1.33 sequences available from those countries.

Our study shows a random phylogenetic clustering of B.1.1.33 strains from most Brazilian states, supporting the notion that SARS-CoV-2 transmission clusters do not contain sufficient phylogenetic information to allow reliable within-country phylogeographic inferences (Mavian et al., 2020; Villabona-Arenas et al., 2020). Despite this limitation, our results highlight

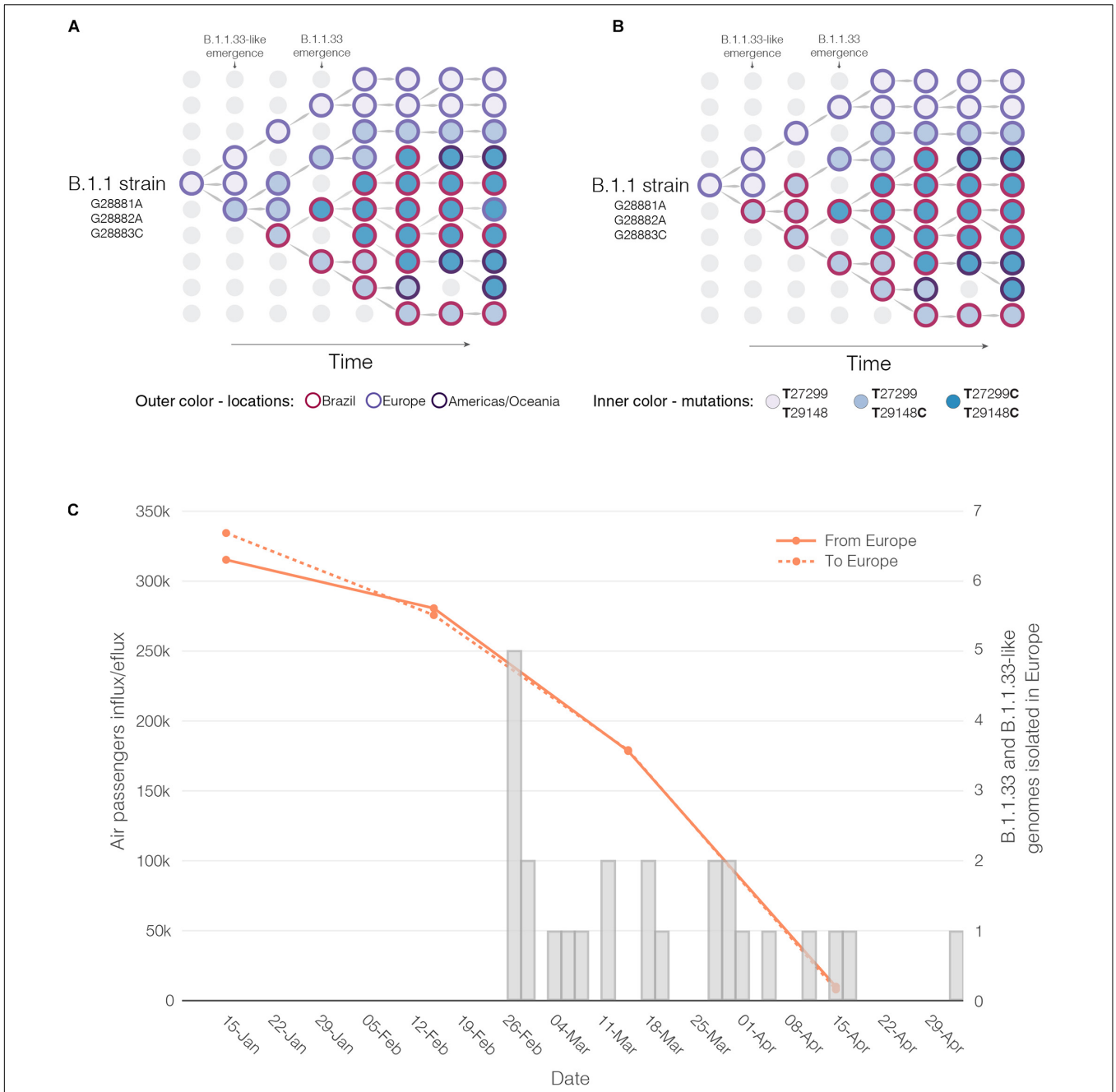


FIGURE 6 | Putative origin and transmission history of the SARS-CoV-2 clades B.1.1.33-like and B.1.1.33. **(A)** Diagrams showing two alternative scenarios for the origin and dissemination of clades B.1.1.33-like and B.1.1.33. The left panel depicts the hypothetical scenario where a B.1.1.33-like strain carrying the mutation T29148C was introduced into Brazil from Europe (marked with an asterisk) and after a period of local transmission in Brazil arose the B.1.1.33 variant carrying the mutation T27299C, which dispersed all over the country and from Brazil to other countries in the Americas and Oceania. **(B)** This panel depicts the hypothetical scenario where a B.1.1 strain was introduced from Europe to Brazil, and mutations T29148C and T27299C arose at sequential steps during local transmission. According to this second scenario, Brazil was the epicenter of dissemination of both clades B.1.1.33-like and B.1.1.33 to other countries in Europe, the Americas, and Oceania. **(C)** Graphic showing the monthly number of international air passengers from South America, North America, and Europe that arrived in Brazil during 2020 (available at: https://www.anac.gov.br) (left-hand axis) along with probability density of T_{MRCA} estimates for clades B.1.1.33.like (gray) and B.1.1.33 (red).

the potential of genomic data to inform about the disease dynamics and the effect of public health interventions in specific geographic settings in Brazil. The fast-increasing prevalence of lineage B.1.1.33 in Rio de Janeiro supports that this clade was a major driver of community transmission in this Brazilian state. Phylodynamic modeling indicates a significant reduction of 66% in the median R_e (from 3.8 to 1.3) of lineage B.1.1.33 following non-pharmaceutical interventions in Rio de Janeiro, but also

points that such interventions probably failed to bring R_e below 1.0 during the early phase. These findings are fully consistent with the pattern inferred from epidemiological modeling in Rio de Janeiro (Candido et al., 2020; de Souza et al., 2020; Mellan et al., 2020) and contrast with data from Europe and Oceania that points to stronger reductions in R_e (71–82%) after public health interventions, sufficient to bring the epidemic under control (Binny et al., 2020; Flaxman et al., 2020; Jarvis et al., 2020; Seemann et al., 2020).

In summary, this study confirms that SARS-CoV-2 B.1.1.33 is a widespread lineage associated with community transmission in Brazil but also reveals a high geographic compartmentalization of SARS-CoV-2 genetic diversity across states. While lineage B.1.1.33 was a major driver of community transmission in Rio de Janeiro, it displayed a very low prevalence in other heavily affected Brazilian sates. Our study demonstrates that Brazilian lineage B.1.1.33 probably evolved from a B.1.1.33-like basal lineage that also carries the synapomorphic mutation at the nucleocapsid protein. Both viral clades probably circulated in Brazil since the early or mid-February 2020 and reached all Brazilian regions and other countries around the world by mid-March 2020. Finally, our data supports that public health interventions at the early epidemic phase were successful in slowing down the transmission of B.1.1.33 lineage in Rio de Janeiro state, but failed to fully control its dissemination, which broadly reflects the past and ongoing COVID-19 dynamics in the state.

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 below: GISAID (https://www.gisaid.org/), accession numbers in **Supplementary Table 1**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by FIOCRUZ-IOC Ethics Committee (68118417.6.0000.5248 and CAAE 32333120.4.0000.5190) and the Brazilian Ministry of Health SISGEN (A1767C3). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

PR, ED, TG, DM, FM, GB, and MS conceived, designed the analysis, and wrote the manuscript. PR, ED, TG, DM, FM, LA, AP, AM, MO, BC, GW, CD, MS, JA, ES, SS, SF, LV, LC, JF, VN, CS, IR, MC, and GB collected the data and performed experiments. PR, ED, TG, DM, FM, and GB conceived, designed, and performed the analysis. All authors reviewed and agreed with the content of the manuscript.

FUNDING

We would like to thank the funding support from CGLab/MoH (General Laboratories Coordination of Brazilian Ministry of Health), CVSLR/FIOCRUZ (Coordination of Health Surveillance and Reference Laboratories of Oswaldo Cruz Foundation), CNPq COVID-19 MCTI 402457/2020-0, and INOVA VPPCB-005-FIO-20-2.

ACKNOWLEDGMENTS

We wish to thank all the health care workers and scientists, who have worked hard to deal with this pandemic threat, the GISAID team, and all the submitters of the database. GISAID acknowledgment tables containing sequences used in this study are in **Supplementary Table 10** (Global SARS-CoV-2 genomes B.1.1 lineage). Locally, we acknowledge the Respiratory Viruses Genomic Surveillance Network of the General Laboratory Coordination (CGLab) of the Brazilian Ministry of Health (MoH), Brazilian Central Laboratory States (LACENs), and local surveillance teams for the partnership in the viral surveillance in Brazil. This manuscript has been released as a pre-print at bioRxiv, the preprint server for biology https://www.biorxiv.org/content/10.1101/2020.06.17.158006v1 (Resende et al., 2020a).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.615280/full#supplementary-material

REFERENCES

Ayres, D. L., Cummings, M. P., Baele, G., Darling, A. E., Lewis, P. O., Swofford, D. L., et al. (2019). BEAGLE 3: improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *Syst. Biol.* 68, 1052–1061. doi: 10.1093/sysbio/syz020

Binny, R. N., Lustig, A., Brower, A., Hendy, S. C., James, A., Parry, M., et al. (2020). Effective reproduction number for COVID-19 in Aotearoa New Zealand. *medRxiv* [Preprint], doi: 10.1101/2020.08.10.20172320

Bluhm, A., Christandl, M., Gesmundo, F., Klausen, F. R., Mańcinska, L., Steffan, V., et al. (2020). SARS-CoV-2 transmission chains from genetic data: a Danish case study. *bioRxiv* [Preprint], doi: 10.1101/2020.05.29.123612

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. doi: 10.1371/journal.pcbi.1006650

Candido, D. S., Claro, I. M., de Jesus, J. G., Souza, W. M., Moreira, F. R. R., Dellicour, S., et al. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 369, 1255–1260. doi: 10.1126/science.abd2161

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., et al. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* 25:45. doi: 10.2807/1560-7917.ES.2020.25.3.2000045

de Souza, W. M., Buss, L. F., Candido, D. D. S., Carrera, J. P., Li, S., Zarebski, A. E., et al. (2020). Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nat. Hum. Behav.* 4, 856–865. doi: 10.1038/s41562-020-0928-4

Delatorre, E., Mir, D., Graf, T., and Bello, G. (2020). Tracking the onset date of the community spread of SARS-CoV-2 in western countries. *Mem. Inst. Oswaldo Cruz.* 115:e200183. doi: 10.1590/0074-02760200183

Dellicour, S., Durkin, K., Hong, S. L., Vanmechelen, B., Martí-Carreras, J., Gill, M. S., et al. (2020). A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. *bioRxiv* [Preprint], doi: 10.1101/2020.05.05.078758

Deng, X., Gu, W., Federman, S., du Plessis, L., Pybus, O. G., Faria, N. R., et al. (2020). Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* 369, 582–587. doi: 10.1126/science.abb9263

Díez-Fuertes, F., Iglesias-Caballero, M., Monzón, S., Jiménez, P., Varona, S., Cuesta, I., et al. (2020). Phylodynamics of SARS-CoV-2 transmission in Spain. *bioRxiv* [Preprint], doi: 10.1101/2020.04.20.050039

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 20, 533–534. doi: 10.1016/S1473-3099(20)30120-1

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192. doi: 10.1093/molbev/msi103

Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., and Baele, G. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* 6:veaa061.

Fauver, J. R., Petrone, M. E., Hodcroft, E. B., Shioda, K., Ehrlich, H. Y., Watts, A. G., et al. (2020). Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell* 181, 990–996.e5. doi: 10.1016/j.cell.2020.04.021

Ferreira, M. A. R., and Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Statist.* 36, 355–368. doi: 10.1002/cjs.5550360302

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 584, 257–261. doi: 10.1038/s41586-020-2405-7

Franco, D., Gonzalez, C., Abrego, L. E., Carrera, J. P., Diaz, Y., Caisedo, Y., et al. (2020). Early transmission dynamics, spread, and genomic characterization of SARS-CoV-2 in Panama. *medRxiv* [Preprint], doi: 10.1101/2020.07.31.20160929

Gambaro, F., Behillil, S., Baidaliuk, A., Donati, F., Albert, M., Alexandru, A., et al. (2020). Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March 2020. *Euro Surveill.* 25:2001200. doi: 10.2807/1560-7917.ES.2020.25.26.2001200

Ghafari, M., Plessis, L. D., Pybus, O., and Katzourakis, A. (2020). *Time Dependence of SARS-CoV-2 Substitution Rates*. Available from: https://virological.org/t/time-dependence-of-sars-cov-2-substitution-rates/542 (accessed September 29, 2020).

Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B., and Suchard, M. A. (2013). Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* 30, 713–724. doi: 10.1093/molbev/mss265

Gonzalez-Reiche, A. S., Hernandez, M. M., Sullivan, M. J., Ciferri, B., Alshammary, H., Obla, A., et al. (2020). Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* 369, 297–301. doi: 10.1126/science.abc1917

Gudbjartsson, D. F., Helgason, A., Jonsson, H., Magnusson, O. T., Melsted, P., Norddahl, G. L., et al. (2020). Spread of SARS-CoV-2 in the Icelandic population. *N. Engl. J. Med.* 382, 2302–2315. doi: 10.1056/NEJMoa2006100

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

Jarvis, C. I., Van Zandvoort, K., Gimma, A., Prem, K., CMMID Covid-19 Working Group, Klepac, P., et al. (2020). Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med.* 18:124. doi: 10.1186/s12916-020-01597-8

Jesus, J. G., Sacchi, C., Candido, D. D. S., Claro, I. M., Sales, F. C. S., Manuli, E. R., et al. (2020). Importation and early local transmission of COVID-19 in Brazil, 2020. *Rev. Inst. Med. Trop. Sao Paulo* 62:e30. doi: 10.1590/s1678-9946202062030

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

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

Laiton-Donato, K., Villabona Arenas, C. J., Usme Ciro, J. A., Franco Munoz, C., Alvarez-Diaz, D. A., Villabona-Arenas, L. S., et al. (2020). Genomic epidemiology of SARS-CoV-2 in Colombia. *medRxiv* [Preprint], doi: 10.1101/2020.06.26.20135715

Lemey, P., Hong, S., Hill, V., Baele, G., Poletto, C., Colizza, V., et al. (2020). Accommodating individual travel history, global mobility, and unsampled diversity in phylogeography: a SARS-CoV-2 case study. *bioRxiv* [Preprint], doi: 10.1101/2020.06.22.165464

Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5:e1000520. doi: 10.1371/journal.pcbi.1000520

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Mavian, C., Marini, S., Prosperi, M., and Salemi, M. A. (2020). Snapshot of SARS-CoV-2 genome availability up to April 2020 and its implications: data analysis. *JMIR Public Health Surveill.* 6:e19170. doi: 10.2196/19170

Mellan, T. A., Hoeltgebaum, H. H., Mishra, S., Whittaker, C., Schnekenberg, R. P., Gandy, A., et al. (2020). Subnational analysis of the COVID-19 epidemic in Brazil. *medRxiv* [Preprint], doi: 10.1101/2020.05.09.20096701

Melo, C. M. L., Silva, G. A. S., Melo, A. R. S., and Freitas, A. C. (2020). COVID-19 pandemic outbreak: the Brazilian reality from the first case to the collapse of health services. *Ann. Acad. Bras. Cienc.* 92:e20200709. doi: 10.1590/0001-3765202020200709

Miller, D., Martin, M. A., Harel, N., Kustin, T., Tirosh, O., Meir, M., et al. (2020). Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *medRxiv* [Preprint], doi: 10.1101/2020.05.21.20104521

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Oude Munnink, B. B., Nieuwenhuijse, D. F., Stein, M., O'Toole, A., Haverkate, M., Mollers, M., et al. (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* 26, 1405–1410. doi: 10.1038/s41591-020-0997-y

Paiva, M. H. S., Guedes, D. R. D., Docena, C., Bezerra, M. F., Dezordi, F. Z., Machado, L. C., et al. (2020). Multiple introductions followed by ongoing community spread of SARS-CoV-2 at one of the largest metropolitan areas in the Northeast of Brazil. *medRxiv* [Preprint], doi: 10.1101/2020.08.25.20171595

Paniz-Mondolfi, A., Munoz, M., Florez, C., Gomez, S., Rico, A., Pardo, L., et al. (2020). SARS-CoV-2 spread across the Colombian-venezuelan border. *medRxiv* [Preprint], doi: 10.1101/2020.07.09.20149856

Parker, J., Rambaut, A., and Pybus, O. G. (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* 8, 239–246. doi: 10.1016/j.meegid.2007.08.001

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

Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407. doi: 10.1038/s41564-020-0770-5

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

Ramirez, J. D., Florez, C., Munoz, M., Hernandez, C., Castillo, A., Gomez, S., et al. (2020). The arrival and spread of SARS-CoV-2 in Colombia. *J. Med. Virol.* 93, 1158–1163. doi: 10.1002/jmv.26393

Resende, P. C., Delatorre, E., Gräf, T., Mir, D., Motta, F. D. C., Appolinario, L. R., et al. (2020a). Genomic surveillance of SARS-CoV-2 reveals community transmission of a major lineage during the early pandemic phase in Brazil. *bioRxiv* [Preprint], doi: 10.1101/2020.06.17.158006

Resende, P. C., Motta, F. C., Roy, S., Appolinario, L., Fabri, A., Xavier, J., et al. (2020b). SARS-CoV-2 genomes recovered by long amplicon tiling multiplex

approach using nanopore sequencing and applicable to other sequencing platforms. *bioRxiv* [Preprint], doi: 10.1101/2020.04.30.069039

Rockett, R. J., Arnott, A., Lam, C., Sadsad, R., Timms, V., Gray, K. A., et al. (2020). Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat. Med.* 26, 1398–1404. doi: 10.1038/s41591-020-1000-7

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Salazar, C., Díaz-Viraqué, F., Pereira-Gómez, M., Ferrés, I., Moreno, P., Moratorio, G., et al. (2020). Multiple introductions, regional spread and local differentiation during the first week of COVID-19 epidemic in Montevideo, Uruguay. *bioRxiv* [Preprint], doi: 10.1101/2020.05.09.086223

Seemann, T., Lane, C. R., Sherry, N. L., Duchene, S., Goncalves da Silva, A., Caly, L., et al. (2020). Tracking the COVID-19 pandemic in Australia using genomics. *Nat. Commun.* 11:4376. doi: 10.1038/s41467-020-18314-x

Shu, Y., and McCauley, J. (2017). GISAID: global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* 22:30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494

Stadler, T., Kuhnert, D., Bonhoeffer, S., and Drummond, A. J. (2013). Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. U.S.A.* 110, 228–233. doi: 10.1073/pnas.1207965110

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

Taboada, B., Vazquez-Perez, J. A., Munoz-Medina, J. E., Ramos-Cervantes, P., Escalera-Zamudio, M., Boukadida, C., et al. (2020). Genomic Analysis of Early SARS-CoV-2 Variants Introduced in Mexico. *J. Virol.* 94:e01056-20. doi: 10.1128/JVI.01056-20

Villabona-Arenas, C. J., Hanage, W. P., and Tully, D. C. (2020). Phylogenetic interpretation during outbreaks requires caution. *Nat. Microbiol.* 5, 876–877. doi: 10.1038/s41564-020-0738-5

Worobey, M., Pekar, J., Larsen, B. B., Nelson, M. I., Hill, V., Joy, J. B., et al. (2020). The emergence of SARS-CoV-2 in Europe and North America. *Science* 370, 564–570. doi: 10.1126/science.abc8169

Xavier, J., Giovanetti, M., Adelino, T., Fonseca, V., Barbosa da Costa, A. V., Ribeiro, A. A., et al. (2020). The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. *Emerg. Microb. Infect.* 9, 1824–1834. doi: 10.1080/22221751.2020.1803146

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 March 2021
doi: 10.3389/fmicb.2021.655567

Molecular Evolution of Human Norovirus GII.2 Clusters

Xingguang Li[1], Haizhou Liu[2], Brittany Rife Magalis[3], Sergei L. Kosakovsky Pond[3] and Erik M. Volz[4]*

[1] Department of Hospital Office, The First People's Hospital of Fangchenggang, Fangchenggang, China, [2] Centre for Emerging Infectious Diseases, The State Key Laboratory of Virology, Wuhan Institute of Virology, University of Chinese Academy of Sciences, Wuhan, China, [3] Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, United States, [4] MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, United Kingdom

Background: The human norovirus GII.2 outbreak during the 2016–2017 winter season was of unprecedented scale and geographic distribution.

Methods: We analyzed 519 complete *VP1* gene sequences of the human norovirus GII.2 genotype sampled during the 2016–2017 winter season, as well as prior (dating back to 1976) from 7 countries. Phylodynamic analyses of these sequences were performed using maximum likelihood and Bayesian statistical frameworks in order to estimate viral evolutionary and population dynamics associated with the outbreak.

Results: Our results revealed an increase in the genetic diversity of human norovirus GII.2 during the recent Asian outbreak and diversification was characterized by at least eight distinct clusters. Bayesian estimation of viral population dynamics revealed a highly fluctuating effective population size, increasing in frequency during the past 15 years.

Conclusion: Despite an increasing viral diversity, we found no evidence of an elevated evolutionary rate or significant selection pressure in human norovirus GII.2, indicating viral evolutionary adaptation was not responsible for the volatility of or spread of the virus during this time.

Keywords: human norovirus, genetic diversity, positive selection, virus evolution, phylogenetic analyses

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Silva Luciana Damascena,
Evandro Chagas Institute, Brazil
Tao Jin,
Guangdong Meige Gene Technology Co., Ltd., China

****Correspondence:***
Xingguang Li
xingguanglee@hotmail.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *19 January 2021*
Accepted: *15 February 2021*
Published: *22 March 2021*

Citation:
Li X, Liu H, Rife Magalis B, Kosakovsky Pond SL and Volz EM (2021) Molecular Evolution of Human Norovirus GII.2 Clusters. Front. Microbiol. 12:655567. doi: 10.3389/fmicb.2021.655567

INTRODUCTION

Human norovirus (HuNoV) is a pathogenic agent that contributes substantially to the global burden of sporadic cases and outbreaks of acute gastroenteritis across all settings and age groups in humans (Ahmed et al., 2014). HuNoV is a non-enveloped, single-stranded, positive-sense RNA virus from the genus *Norovirus* in the family *Caliciviridae* (Glass et al., 2009). HuNoV is classified into at least 10 genogroups (GI–GX) and 48 genotypes, based on phylogenetic analyses of the capsid gene (Zheng et al., 2006; Vinje, 2015; Chhabra et al., 2019, 2020). Among these different genogroups, genotypes belonging to the GI, GII, and GIV are primarily responsible for the acute gastroenteritis cases in humans (Kroneman et al., 2013; Vinje, 2015). GI and GII strains can be further stratified into nine and twenty-six genotypes, respectively (Kroneman et al., 2013; Chhabra et al., 2019, 2020).

Genotype GII.2 strains recently emerged, causing large outbreaks of acute gastroenteritis in Japan and China during the 2016–2017 winter season, mainly in childcare centers

(Ao et al., 2017, 2018; Hata et al., 2018; Nagasawa et al., 2018). Many other countries also reported GII.2-dominated HuNoV outbreaks during the 2016–2017 winter season (Bidalot et al., 2017; Niendorf et al., 2017; Tohma et al., 2017; Medici et al., 2018; Thanusuwannasak et al., 2018).

To gain a better understanding of genetic variation during the recent HuNoV epidemic, it is essential to analyze the *VP1* gene, which is crucial for viral attachment and entry, as well as the production of neutralizing antibodies (Prasad et al., 1999; Tan et al., 2004; Chakravarty et al., 2005). Therefore, in order to gain a more comprehensive understanding of the viral evolutionary factors responsible for the emergence and spread accompanying the recent HuNoV epidemic, we performed a detailed analysis of the complete *VP1* gene of the GII.2 strains with known sampling dates and geographic locations. Our findings suggest that close monitoring of the global spread of this emergent GII.2 strains is necessary for the prevention and mitigation of HuNoV acute gastroenteritis outbreaks.

MATERIALS AND METHODS

Sequence Data Set

All available complete ORF2 region sequences (at least 1,626 bp in length) of GII.2 strains with known sampling date and geographic location were retrieved from GenBank[1] on 28 April 2018. Sequences were aligned using MAFFT v7.409 (Katoh and Standley, 2013) and then adjusted manually in BioEdit v7.2.5 (Hall, 1999). The best-fit model of nucleotide substitution for this data set was identified according to the Bayesian information criterion in jModelTest v2.1.10 (Darriba et al., 2012). Given the phylogenetic scope of our analysis, we chose to exclude models that allowed a proportion of invariable sites (Jia et al., 2014). For this data set, the SYM + Γ_4 model provided the best fit and was used in genetic distance calculations and tree reconstruction.

Phylogenetic, Likelihood-Mapping, and Genetic Distance Analyses

To evaluate the phylogenetic signal for this data set, we performed a likelihood mapping analysis (Strimmer and von Haeseler, 1997) using TREE-PUZZLE v5.3.rc16 (Schmidt et al., 2002). We then inferred the phylogeny using maximum likelihood (ML) in PhyML v3.1 (Guindon and Gascuel, 2003). Bootstrap support values were calculated using 1,000 pseudoreplicates. Methods for identifying and defining phylogenetic clusters differ between studies. In the present study, we identify them on the basis of bootstrap support (cut-off of 99%) for groupings with more than two sequences. Genetic distances between and within clusters were calculated in MEGA v5.05 (Tamura et al., 2011) using the maximum composite likelihood model (Tamura et al., 2004) with 1,000 bootstrap replicates.

[1]http://www.ncbi.nlm.nih.gov/

Evolutionary Rate, Time Origin, and Past Population Dynamic Inferences

To evaluate the temporal structure of this data set, we plotted root-to-tip genetic distances against date of sampling within the ML tree using TempEst v1.5 (Rambaut et al., 2016). We also estimated the evolutionary rate and the time to the most recent common ancestor for this data set using least-squares dating in LSD v0.3beta (To et al., 2016), a Gamma-Poisson mixture model in treedater package (Volz and Frost, 2017), and ML dating in TreeTime package (Sagulenko et al., 2018), respectively. We then employed a Bayesian phylogenetic approach using Markov chain Monte Carlo (MCMC) sampling to estimate the rate of evolution and the time to the most recent common ancestor for all sequences in BEAST v1.8.2 (Drummond et al., 2012). Evolutionary rates were estimated using a strict molecular clock model. A non-parametric coalescent Bayesian skygrid tree prior model was employed for demographic inference. The overall evolutionary rate was given an uninformative continuous-time Markov chain (CTMC) reference prior. The MCMC analysis was run for 500 million steps, with samples drawn every 50,000 steps. The first 1,000 samples in each chain were removed as burn-in. All parameters had effective sample sizes greater than 200, which is indicative of sufficient sampling. We ran multiple chains to check for convergence to the target distribution using Tracer v1.7.1 (Rambaut et al., 2018). Trees were summarized as maximum clade credibility (MCC) trees using TreeAnnotator v1.8.2 after discarding the first 10% as burn-in, and then visualized and annotated in FigTree v1.4.3[2]. We also employed the skygrowth package in R to estimate the effective population size for the *VP1* gene of the HuNoV GII.2 strains (Volz and Didelot, 2018).

Analysis of Selection

The evidence of gene-wide episodic positive selection along the backbone of the phylogenetic tree was performed using

[2]http://tree.bio.ed.ac.uk/software/figtree

TABLE 1 | Geographic source and sampling year of human norovirus GII.2 sequences used in the present study.

	n	Sampling year				
Geographic source		**1970–1979**	**1980–1989**	**1990–1999**	**2000–2009**	**2010–2017**
Australia	3					3
China	302					302
Hong Kong	22					22
Japan	166			1	52	113
Malaysia	1	1				
Netherlands	10			1	9	
Taiwan	9					9
United Kingdom	1		1			
United States	5	1		1		3
Total	**519**	**2**	**1**	**3**	**61**	**452**

BUSTED, a test for episodic diversifying selection (Murrell et al., 2015), implemented in HyPhy software package (Pond et al., 2005). Initial branch lengths within the fixed MCC tree topology were estimated using the GTR (general time reversible) nucleotide model. A likelihood ratio test (LRT) was used to compare optimized likelihoods under the unconstrained, alternative model wherein the ratio of non-synonymous to synonymous rates (*dN/dS*) is allowed to exceed 1 (positive selection) and under the constrained (*dN/dS* = 1), null model (neutrality). A related test, RELAX (Wertheim et al., 2015) was used to investigate relaxation or intensification of selective regimes within each of the eight highly supported clades as compared to the persisting viral lineages. External branches were placed in the nuisance category to avoid *dN/dS* estimate inflation owing to transient polymorphisms. The selection intensity parameter (k) was estimated for each clade and the LRT used to compare the null model in which k is constrained to 1 (i.e., the same *dN/dS* distribution on test and reference branches) to an alternative model in which k is a free parameter for each clade. $p \leq 0.05$ were considered significant evidence in favor of the alternative selection model for both tests.

RESULTS

Sequence Data Set Information

This data set included 519 complete ORF2 region sequences of GII.2 strains from Australia (n = 3); the Mainland of China (n = 302); Hong Kong (n = 22); Japan (n = 166); Malaysia (n = 1); Netherlands (n = 10); Taiwan (n = 9); United Kingdom (n = 1); and United States (n = 5), with sampling dates between 1976 and 2017 (**Table 1** and **Supplementary Table S1**).

Phylogenetic Analysis

The phylogeny of this data set, inferred using maximum likelihood, indicated the presence of eight transmission clusters (**Figures 1**, **2**, **Supplementary Figure S1**, **Table 2**, and **Supplementary Table S2**). Sequences from Cluster I (n = 4) were found in Netherlands and Japan between 1999 and 2002. Cluster II sequences (n = 6) were found in Japan between 2000 and 2006. Sequences from Cluster III (n = 21) were collected from Japan between 2004 and 2006. Cluster IV sequences (n = 25) were found in Japan between 2007 and 2015, whereas Cluster V sequences (n = 64) were collected from

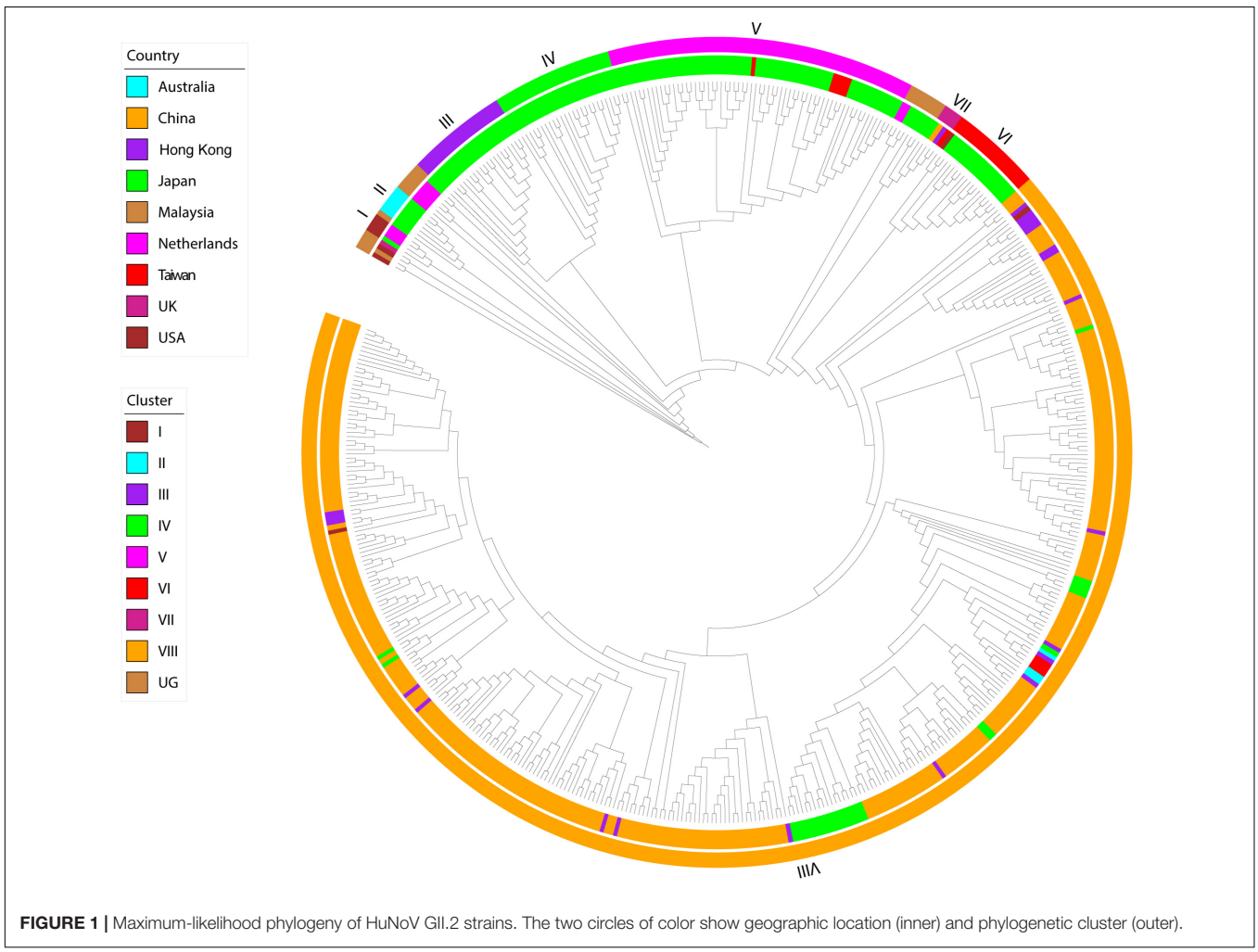


FIGURE 1 | Maximum-likelihood phylogeny of HuNoV GII.2 strains. The two circles of color show geographic location (inner) and phylogenetic cluster (outer).

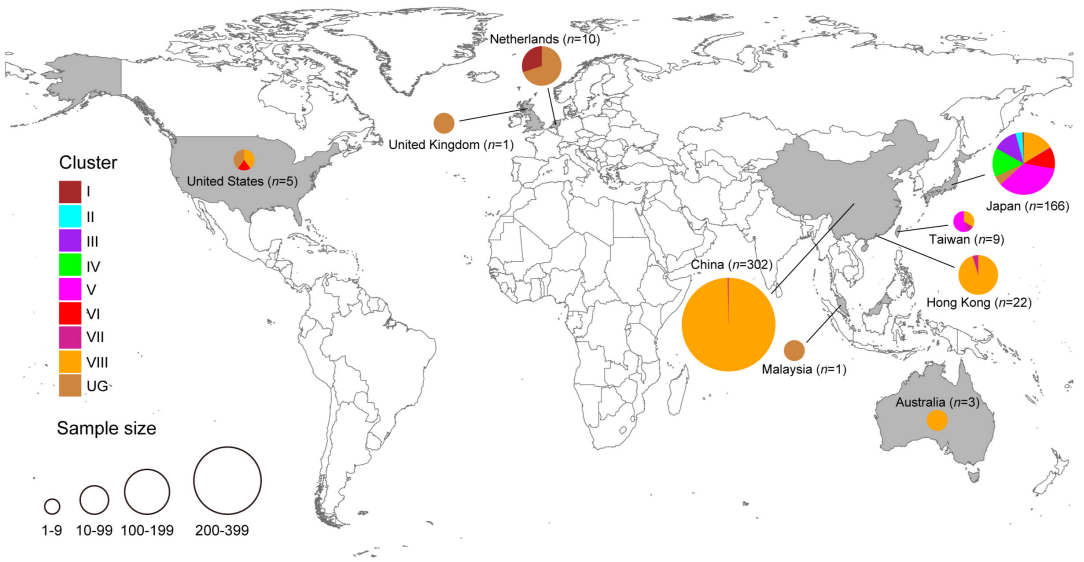


FIGURE 2 | Geographic distribution of HuNoV GII.2 clusters identified in the present study. Each GII.2 cluster identified in this study is color-coded, as shown on the left. The figure was created using Adobe Illustrator CS5 version 15.0.0 software, based on the maps obtained from Craft MAP website (craftmap.box-i.net/).

TABLE 2 | Classification and geographic source of human norovirus GII.2 sequences used in the present.

Cluster	*n*	Geographic source								
		Australia	China	Hong Kong	Japan	Malaysia	Netherlands	Taiwan	United Kingdom	United States
I	4				1		3			
II	6				6					
III	21				21					
IV	25				25					
V	64				59			5		
VI	19				18					1
VII	4		1	1	1			1		
VIII	357	3	301	21	27			3		2
Ungrouped	19				8	1	7		1	2
Total	**519**	**3**	**302**	**22**	**166**	**1**	**10**	**9**	**1**	**5**

Japan and Taiwan between 2008 and 2015. Cluster VI included eighteen sequences collected from Japan and one sequence collected from United States between 2010 and 2012. The four sequences from Cluster VII were collected from Japan, Taiwan, Hong Kong, and the Mainland of China between 2015 and 2017. Cluster VIII sequences (n = 357) were found in Australia, the Mainland of China, Hong Kong, Japan, Taiwan, and United States between 2016 and 2017.

The remaining 19 sequences (designated as Ungrouped) were scattered throughout the main GII.2 genotype and had been sampled in Japan (n = 8), Malaysia (n = 1), Netherlands (n = 7), United Kingdom (n = 1), and United States (n = 2) between 1976 and 2014.

Of the 302 Chinese strains in our analysis, all were found within either cluster VII (n = 1) or cluster VIII (n = 301) and all were sampled within either 2016 (n = 133) or 2017 (n = 169). However, the 166 Japanese strains were interspersed throughout the phylogeny and were sampled within a wide range (1997, 2000, 2002, and 2004–2016.

Likelihood-Mapping and Evolutionary Divergence Analysis

Our likelihood-mapping analysis revealed significant differences in phylogenetic signal and the amount of evolutionary information between each cluster in our data set (**Supplementary Figure S2**). The most notable result of the likelihood-mapping analysis revealed that 45.4% of the quartets from cluster III were distributed in the center of the triangle, indicating a relative strong star-like phylogenetic signal reflecting a new cluster, which might be due to exponential epidemic spread. Intriguingly, cluster III was characterized by the lowest genetic divergence among the eight distinct HuNoV GII.2 clusters (**Supplementary Figure S3**), which further confirmed that cluster III was most likely a new cluster, in accordance with

likelihood-mapping analysis. Likewise, 38.5% of the quartets from cluster VI were distributed in the center of the triangle and showed lower genetic divergence. Conversely, 12.3% of the quartets from cluster VIII were distributed in the center of the triangle and showed higher genetic divergence than cluster VI. Notably, 43.5% of the quartets from cluster IV

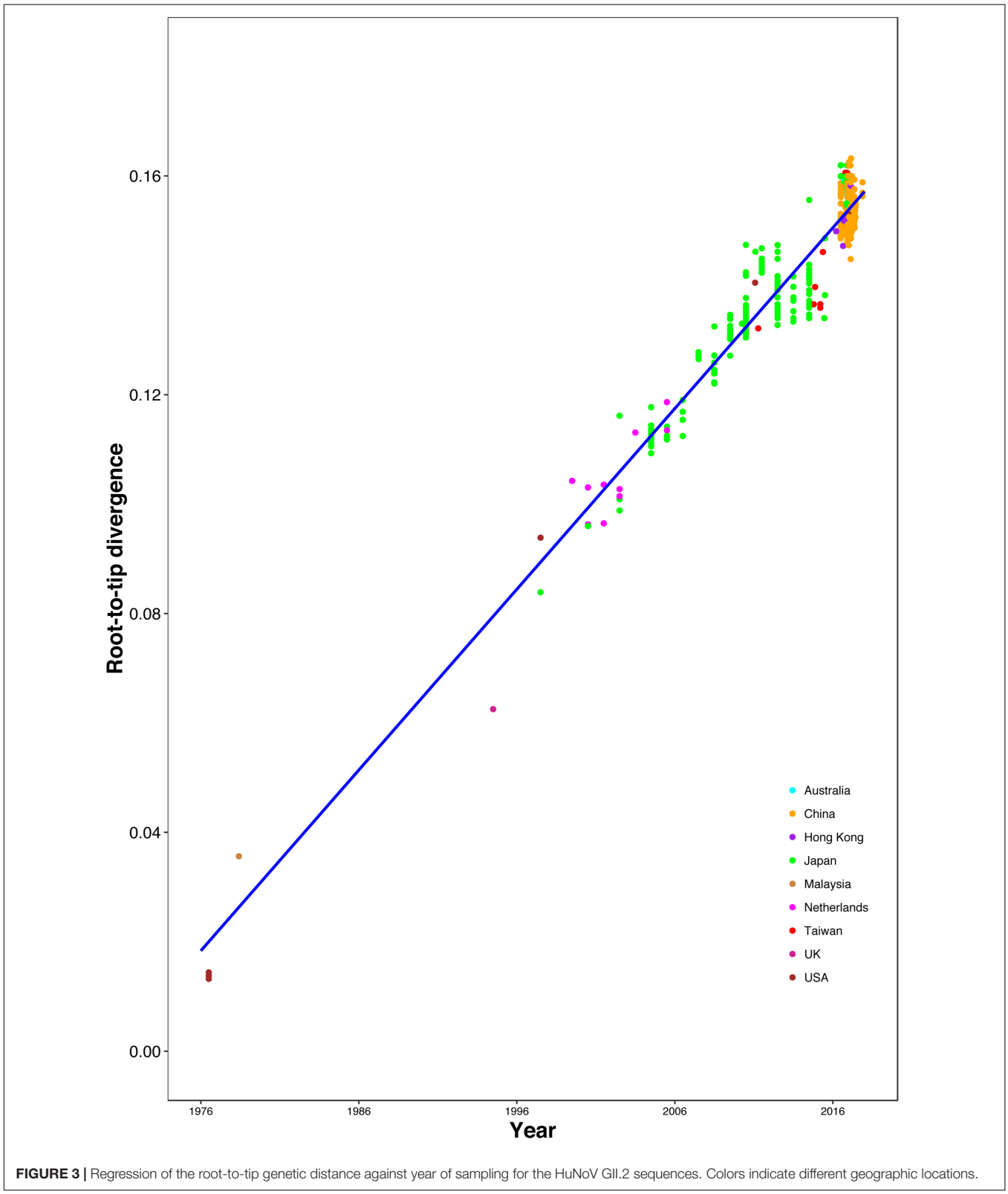


FIGURE 3 | Regression of the root-to-tip genetic distance against year of sampling for the HuNoV GII.2 sequences. Colors indicate different geographic locations.

were distributed in the center of the triangle, but with higher genetic divergence than cluster VI. In contrast, 0, 0, 3.9, and 0% of the quartets from clusters I, II, V, and VII were distributed in the center of the triangle, but with higher genetic divergence, indicating the presence of a strong phylogenetic signal in the four clusters. Also of note, 3.4% of the quartets from the overall epidemics were distributed in the center of the triangle, but with lower genetic divergence (4.1%). We also observed that the smallest genetic distance separated clusters III and IV (3.9%), whereas the largest was between clusters I and VIII (10.1%). Notably, all the genetic distances between each of the eight clusters are higher than within each of the eight clusters, except for the genetic distance between clusters III and IV. We can also observe mixed genetic distances between and within each of the eight clusters (**Supplementary Figure S3**).

Evolutionary Rate and Time Origin Analyses

The correlation between root-to-tip distances and sampling time indicated a strong temporal signal (R^2 = 0.9359), without any clear outlier sequences (**Figure 3**). This result suggests a relatively clocklike pattern of molecular evolution, with an estimated evolutionary rate of 3.29×10^{-3} substitutions per site per year and the most recent common ancestor occurring in 1970.37. We also analyzed the correlation between genetic distances and sampling time intervals, which indicated a strong linear regression signal for sampling time intervals > 12 years (**Supplementary Figure S4**). The estimated evolutionary rate and the time to the most recent common ancestor using a Gamma-Poisson mixture model were 2.76×10^{-3} substitutions per site per year and 1968.60, respectively. The estimated evolutionary rate and the time to the most recent common ancestor using maximum likelihood dating were 2.48×10^{-3} substitutions per site per year (95% credibility interval: 2.18×10^{-3}–2.78×10^{-3}) and 1969.0 (95% credibility interval: 1966.35–1971.65), respectively. The estimated evolutionary rate and the time to the most recent common ancestor using least-squares dating were 2.63×10^{-3} substitutions per site per year (95% credibility interval: 2.30×10^{-3}–2.82×10^{-3}) and 1968.61 (95% credibility interval: 1965.27–1970.84), respectively. In our Bayesian phylogenetic analysis, we estimated a evolutionary rate of 3.00×10^{-3} substitutions per site per year (95% credibility interval: 2.67×10^{-3}–3.35×10^{-3}). And the time to the most recent common ancestor of 1968.89 (95% credibility interval: 1965.49–1972.06). The age of each GII.2 cluster was also estimated in the analysis (**Supplementary Table S3** and **Supplementary Figure S5**). The divergences between the sequences from clusters VI and VIII were estimated to have occurred in 2007.96 (95% credibility interval: 2007.10–2008.75) and 2012.39 (95% credibility interval: 2011.42–2013.44), respectively. And the time to the most recent common ancestor of 2004.61 (95% credibility interval: 2003.22–2005.83) for clusters VI and VIII.

Analysis of Past Population Dynamics

We further investigated the past population dynamics of this data set using a Bayesian skygrid plot. The effective population size of the *VP1* gene in the HuNoV GII.2 strains experienced very high volatility of population size, especially for the past 15 years (**Figure 4**). The effective population size of the *VP1* gene in the HuNoV GII.2 strains was also estimated using MCMC method in skygrowth package (**Supplementary Figure S6**), the estimates of the effective population size were consistent with the Bayesian skygrid model. The estimates of the phylogenetic relationships among the HuNoV GII.2 sequences using Bayesian coalescent framework were consistent with maximum likelihood tree reconstruction (**Figure 5**).

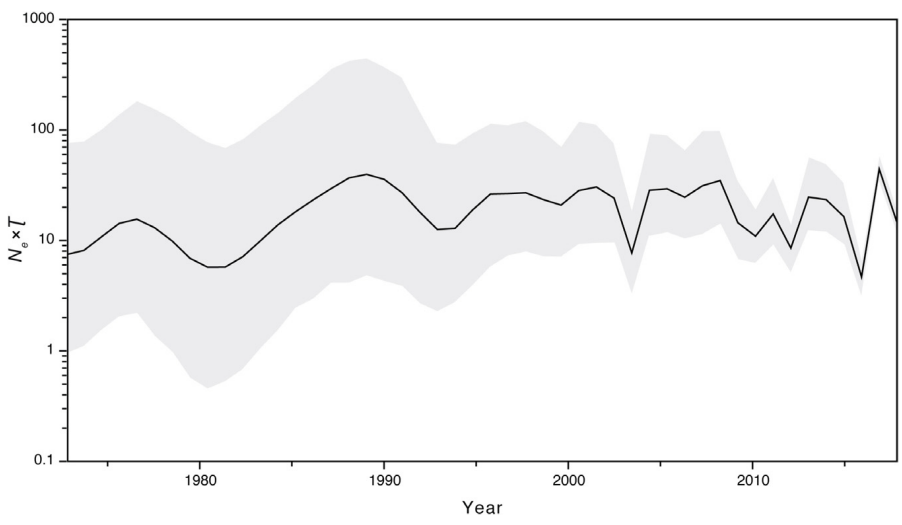


FIGURE 4 | Bayesian skygrid demographic reconstruction of HuNoV GII.2. The vertical axis shows the effective number of infections (N_e) multiplied by mean viral generation time (τ). The solid line and shaded region represent the median and 95% credibility interval, respectively, of the inferred $N_e\tau$ through time.

Analysis of Selection

Given the observed variations in evolutionary dynamics for this data set, we sought to investigate the selective pressures driving the persistence of HuNoV GII.2. No evidence of episodic positive selection was detected among the persisting viral lineages (**Supplementary Table S4**, $p = 0.18$); however, significant differences in selective strength among outbreaks when compared to these persisting lineages were found ($p = 0.04$). Clusters II, III, V, and VII exhibited intensified selection ($k > 2$), whereas sister clusters VI and VIII were considered to have relaxed in selective strength ($k < 0.5$). Clusters I and IV were characterized by relatively similar selective regimes in comparison to the backbone ($k = 0.72$ and 0.6, respectively). The results suggested that the evolutionary

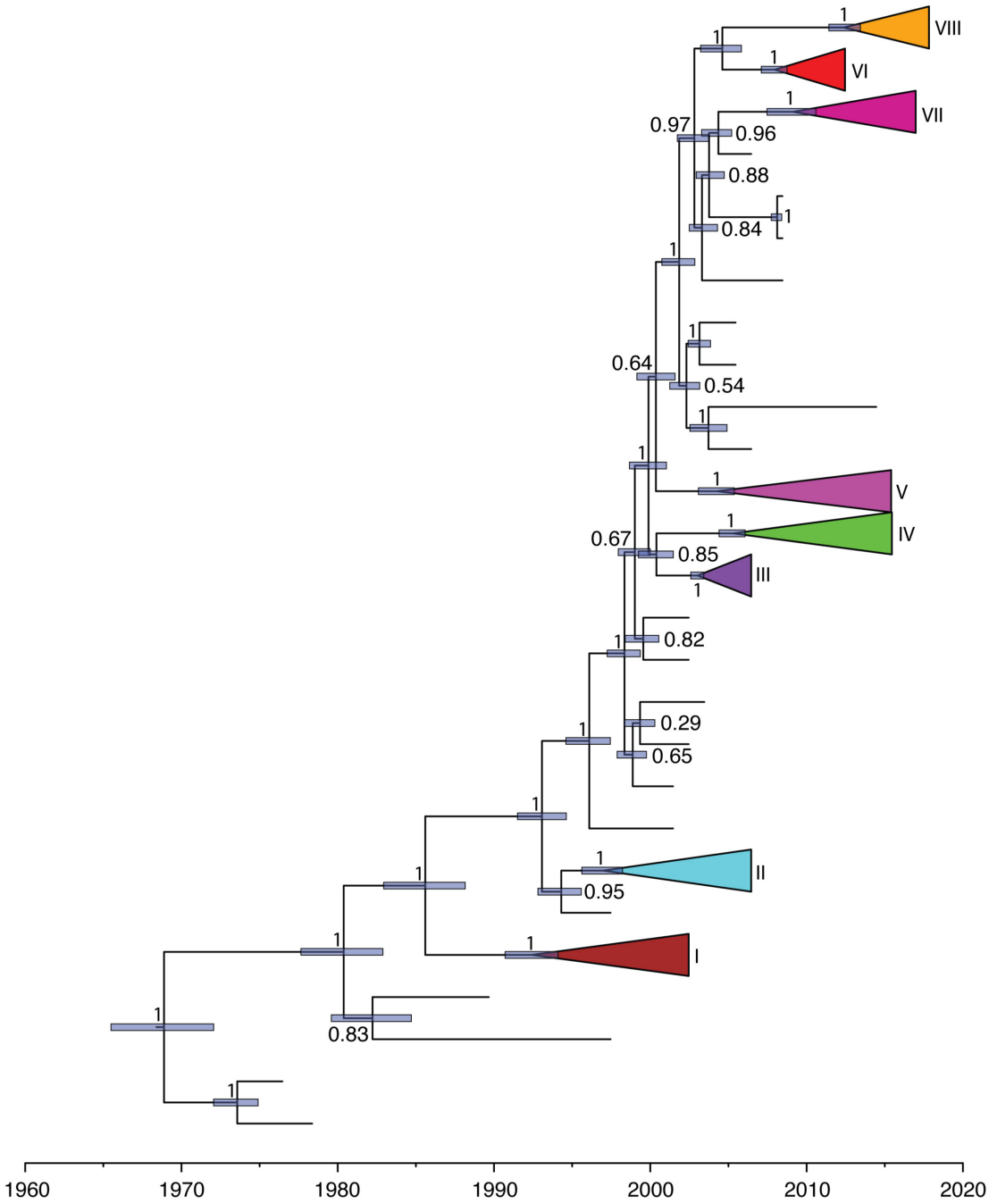


FIGURE 5 | Maximum-clade-credibility tree estimated from complete *VP1* gene sequences of HuNoV GII.2. Light blue horizontal bars represent 95% credibility intervals for estimates of node times. Posterior probabilities are shown at the node. Each cluster is shown with different colors.

dynamics of the *VP1* gene of HuNoV GII.2 strains in this data set were influenced largely by natural evolutionary forces, rather than adaptation to selective pressure from the host immune system, which was consistent with the previous study (Tohma et al., 2017).

DISCUSSION

Our evolutionary analyses, based on all of the available complete *VP1* sequences of HuNoV GII.2 genotype that included country of origin and year of sampling, revealed the presence of at least eight independent clusters. Cluster VII was restricted to countries within Asia, whereas cluster VIII had dispersed to Asia and North America. The recent sequences from cluster VIII outbreak formed a highly distinct sister group to cluster VII. As such, the putative HuNoV GII.2 reservoirs for cluster VIII outbreak sequences might be distinct from that of cluster VII.

Our estimates of the evolutionary rate for this data set were reassuringly consistent across different methods and models. The mean estimated rates of HuNoV GII.2 molecular evolution using TempEst v1.5 and BEAST v1.8.2 were 3.00×10^{-3} and 3.29×10^{-3} substitutions per site per year, respectively. These were also similar to the mean estimates previously reported for HuNoV GII.2 using BEAST v2.4.5, which was 3.26×10^{-3} substitutions per site per year. However, the mean estimated rates of HuNoV GII.2 molecular evolution using LSD v0.3beta, TreeTime package, and treedater package were 2.63×10^{-3}, 2.48×10^{-3}, and 2.76×10^{-3} substitutions per site per year, respectively, which were lower than the results using TempEst v1.5 and BEAST v1.8.2. In summary, despite the extensive and prolonged human-to-human transmission in the recent HuNoV GII.2 outbreak, HuNoV GII.2 was not mutating at an unusually elevated rate. Thus, we found little evolutionary signal that the virus has enhanced its virulence and/or transmissibility in humans during the recent HuNoV GII.2 outbreak. Coalescent-based demographic inference revealed a rapidly increasing population size for HuNoV GII.2 from late 2015 to early 2017 (**Figure 4**), consistent with the outbreak during this period of time.

Our analyses of selection did not reveal episodic positive selection representative of continual adaptation, nor was the more recent human norovirus GII.2 outbreak characterized by intensified selective pressure. Thus, the results from the present study seem to be consistent with earlier reports (Kobayashi et al., 2016; Parra et al., 2017; Nagasawa et al., 2018) and suggest that the recent evolution of HuNoV GII.2 has been dominated by genetic drift. Furthermore, the results also suggested a relationship between selection strength variation and geographical spread, as Clusters VI and VIII were the only clusters to contain sequences outside of Asia after 2010.

Our study has several limitations. First, our analysis was based on all of the available complete *VP1* sequences of HuNoV GII.2 genotype that included country of origin and year of sampling, and we did not randomly collected sequences, we may therefore have underestimated the total number of infections. Second, the number of sample size is limited, which may provide an incomplete picture of phylogenetic relationship of the virus, therefore, which may have underestimated the total number of phylogenetic clusters. Additional sequences are critically needed to investigate the phylogenetic cluster changes over time, estimate its date of emergence, monitor local transmission, and infer past population dynamics of the virus. Additional sequences will also help to reveal the presence of additional phylogenetic clusters. Third, there are probably biases in geographic and temporal distributions of the *VP1* sequences presented in the present study, given that our study relies on data from different countries or stages that is varied significantly, which may introduce bias our estimates. Despite the aforementioned limitations, we believe our findings are reliable and robust.

Taken together, we analyzed 519 complete *VP1* gene sequences of the human norovirus GII.2 genotype. Most of the samples were collected during the outbreak during the 2016–2017 winter season from the Mainland of China. We explored several aspects about geographic distribution, demographic analysis and selective pressure of the GII.2 norovirus strains. This information is important for surveillance of viral strains with pandemic potential. Based on the phylogenetic analysis of these sequences, we showed an increase in the genetic diversity of human norovirus GII.2 for at least eight distinct clusters, and the geographic and gene divergence varied among clusters. Our results sufficiently demonstrated an increase in the genetic diversity of human norovirus GII.2 during the recent Asian outbreak and a highly fluctuating effective population size. We also indicated that viral evolutionary adaptation was not responsible for the volatility of or spread of the virus during this time. Our results also emphasize the importance of ongoing genetic and demographic surveillance of the epidemic, especially, sampled during the epidemic and across all the geographical range of the virus. These efforts, combined with epidemiological investigation, clinical recognition, and population movement data, are required to trace the history of recent epidemic waves and to assess adaptability and selection in HuNoV GII.2 strains. This is increasingly important for surveillance of viral strains with pandemic potential, and will further guide research on vaccines and therapeutic targets.

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 author/s.

AUTHOR CONTRIBUTIONS

XL conceived and designed the study, performed the experiments, and drafted the manuscript. XL, BR, and EV analyzed the data. XL, HL, BR, EV, and SK interpreted data and provided critical comments. All authors reviewed and approved the final manuscript.

FUNDING

This study was supported by a grant from the National Key Research and Development Program of China

(2018YFC1603803) to HL, and this study was also supported by grants from U01 GM110749 (NIH/NIGMS) and R01 GM093939 (NIH/NIGMS) to SK.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.655567/full#supplementary-material

Supplementary Figure 1 | Phylogenetic relationships of complete *VP1* gene sequences from HuNoV GII.2. The tree was inferred using a maximum-likelihood analysis of 519 complete *VP1* gene sequences. Bootstrap values were shown at the nodes. Clusters were shown with different colors. The scale bar showed nucleotide substitutions per site.

Supplementary Figure 2 | Likelihood-mapping analyses of complete *VP1* gene sequences from HuNoV GII.2. Likelihood-mapping analyses were performed for this data set, and also each of eight clusters defined by the maximum-likelihood phylogenetic analyses of complete *VP1* gene sequences from HuNoV GII.2.

Supplementary Figure 3 | Average pairwise genetic distances within and between HuNoV GII.2 clusters using ggplot2 package. Green and orange colors indicated within and between human norovirus GII.2 clusters.

Supplementary Figure 4 | Sampling time interval with respective genetic distance between pairs of HuNoV GII.2 clusters. Green and orange colors indicated within and between HuNoV GII.2 clusters.

Supplementary Figure 5 | Estimated time to the most recent common ancestor of each HuNoV GII.2 cluster. Clusters were shown with different colors.

Supplementary Figure 6 | Estimated effective population size through time using Bayesian Gibbs-within-Metropolis MCMC and fast maximum *a posteriori* algorithm. Estimated effective population size through time was based on the tree in **Supplementary Figure 1**.

REFERENCES

Ahmed, S. M., Hall, A. J., Robinson, A. E., Verhoef, L., Premkumar, P., Parashar, U. D., et al. (2014). Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect. Dis.* 14, 725–730. doi: 10.1016/s1473-3099(14)70767-4

Ao, Y., Cong, X., Jin, M., Sun, X., Wei, X., Wang, J., et al. (2018). Genetic analysis of reemerging GII.P16-GII.2 Noroviruses in 2016-2017 in China. *J. Infect. Dis.* 218, 133–143. doi: 10.1093/infdis/jiy182

Ao, Y., Wang, J., Ling, H., He, Y., Dong, X., Wang, X., et al. (2017). Norovirus GII.P16/GII.2-associated gastroenteritis, China, 2016. *Emerg. Infect. Dis.* 23, 1172–1175.

Bidalot, M., Thery, L., Kaplon, J., De Rougemont, A., and Ambert-Balay, K. (2017). Emergence of new recombinant noroviruses GII.p16-GII.4 and GII.p16-GII.2, France, winter 2016 to 2017. *Euro. Surveill.* 22:30508.

Chakravarty, S., Hutson, A. M., Estes, M. K., and Prasad, B. V. (2005). Evolutionary trace residues in noroviruses: importance in receptor binding, antigenicity, virion assembly, and strain diversity. *J. Virol.* 79, 554–568. doi: 10.1128/jvi.79.1.554-568.2005

Chhabra, P., De Graaf, M., Parra, G. I., Chan, M. C., Green, K., Martella, V., et al. (2019). Updated classification of norovirus genogroups and genotypes. *J. Gen. Virol.* 100, 1393–1406. doi: 10.1099/jgv.0.001318

Chhabra, P., Graaf, M., Parra, G. I., Chan, M. C., Green, K., Martella, V., et al. (2020). Corrigendum: updated classification of norovirus genogroups and genotypes. *J. Gen. Virol.* 101:893. doi: 10.1099/jgv.0.001475

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Glass, R. I., Parashar, U. D., and Estes, M. K. (2009). Norovirus gastroenteritis. *N. Engl. J. Med.* 361, 1776–1785.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Hata, M., Nakamura, N., Kobayashi, S., Onouchi, A., Saito, T., Hirose, E., et al. (2018). Emergence of new recombinant noroviruses GII.P16-GII.2 and GII.P16-GII.4 in Aichi, Japan, during the 2016/17 season. *Jpn. J. Infect. Dis.* 71, 319–322. doi: 10.7883/yoken.jjid.2017.520

Jia, F., Lo, N., and Ho, S. Y. (2014). The impact of modelling rate heterogeneity among sites on phylogenetic estimates of intraspecific evolutionary rates and timescales. *PLoS One* 9:e95722. doi: 10.1371/journal.pone.0095722

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

Kobayashi, M., Matsushima, Y., Motoya, T., Sakon, N., Shigemoto, N., Okamoto-Nakagawa, R., et al. (2016). Molecular evolution of the capsid gene in human norovirus genogroup II. *Sci. Rep.* 6:29400.

Kroneman, A., Vega, E., Vennema, H., Vinje, J., White, P. A., Hansman, G., et al. (2013). Proposal for a unified norovirus nomenclature and genotyping. *Arch. Virol.* 158, 2059–2068. doi: 10.1007/s00705-013-1708-5

Medici, M. C., Tummolo, F., Martella, V., De Conto, F., Arcangeletti, M. C., Pinardi, F., et al. (2018). Emergence of novel recombinant GII.P16_GII.2 and GII. P16_GII.4 Sydney 2012 norovirus strains in Italy, winter 2016/2017. *New Microbiol.* 41, 71–72.

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., et al. (2015). Gene-wide identification of episodic selection. *Mol. Biol. Evol.* 32, 1365–1371. doi: 10.1093/molbev/msv035

Nagasawa, K., Matsushima, Y., Motoya, T., Mizukoshi, F., Ueki, Y., Sakon, N., et al. (2018). Genetic Analysis of human Norovirus strains in Japan in 2016-2017. *Front. Microbiol.* 9:1. doi: 10.3389/fmicb.2018.00001

Niendorf, S., Jacobsen, S., Faber, M., Eis-Hubinger, A. M., Hofmann, J., Zimmermann, O., et al. (2017). Steep rise in norovirus cases and emergence of a new recombinant strain GII.P16-GII.2, Germany, winter 2016. *Euro. Surveill.* 22:30447.

Parra, G. I., Squires, R. B., Karangwa, C. K., Johnson, J. A., Lepore, C. J., Sosnovtsev, S. V., et al. (2017). Static and evolving Norovirus genotypes: implications for epidemiology and immunity. *PLoS Pathog.* 13:e1006136. doi: 10.1371/journal.ppat.1006136

Pond, S. L., Frost, S. D., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079

Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G., and Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* 286, 287–290. doi: 10.1126/science.286.5438.287

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarisation 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

Sagulenko, P., Puller, V., and Neher, R. A. (2018). TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol.* 4:vex042.

Schmidt, H. A., Strimmer, K., Vingron, M., and Von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504. doi: 10.1093/bioinformatics/18.3.502

Strimmer, K., and von Haeseler, A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment.

Proc. Natl. Acad. Sci. U.S.A. 94, 6815–6819. doi: 10.1073/pnas.94.13.6815

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11030–11035. doi: 10.1073/pnas.0404206101

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Tan, M., Hegde, R. S., and Jiang, X. (2004). The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. *J. Virol.* 78, 6233–6242. doi: 10.1128/jvi.78.12.6233-6242.2004

Thanusuwannasak, T., Puenpa, J., Chuchaona, W., Vongpunsawad, S., and Poovorawan, Y. (2018). Emergence of multiple norovirus strains in Thailand, 2015-2017. *Infect. Genet. Evol.* 61, 108–112. doi: 10.1016/j.meegid.2018.03.021

To, T. H., Jung, M., Lycett, S., and Gascuel, O. (2016). Fast dating using least-squares criteria and algorithms. *Syst. Biol.* 65, 82–97. doi: 10.1093/sysbio/syv068

Tohma, K., Lepore, C. J., Ford-Siltz, L. A., and Parra, G. I. (2017). Phylogenetic analyses suggest that factors other than the capsid protein play a role in the epidemic potential of GII.2 Norovirus. *mSphere* 2:e00187-17.

Vinje, J. (2015). Advances in laboratory methods for detection and typing of norovirus. *J. Clin. Microbiol.* 53, 373–381. doi: 10.1128/jcm.01535-14

Volz, E. M., and Didelot, X. (2018). Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. *Syst. Biol.* 67, 719–728. doi: 10.1093/sysbio/syy007

Volz, E. M., and Frost, S. D. W. (2017). Scalable relaxed clock phylogenetic dating. *Virus Evol.* 3:vex025.

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., and Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32, 820–832. doi: 10.1093/molbev/msu400

Zheng, D. P., Ando, T., Fankhauser, R. L., Beard, R. S., Glass, R. I., and Monroe, S. S. (2006). Norovirus classification and proposed strain nomenclature. *Virology* 346, 312–323. doi: 10.1016/j.virol.2005.11.015

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 April 2021
doi: 10.3389/fmicb.2021.653719

Evolutionary Mechanism of Immunological Cross-Reactivity Between Different GII.17 Variants

Yueting Zuo[1,2†], Liang Xue[2†], Junshan Gao[2], Yingyin Liao[2], Yanhui Liang[2], Yueting Jiang[3], Weicheng Cai[2], Zhiwei Qin[2], Jiale Yang[2], Jumei Zhang[2], Juan Wang[2], Moutong Chen[2], Yu Ding[2] and Qingping Wu[2]**

[1]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China, [2]Guangdong Provincial Key Laboratory of Microbial Safety and Health, State Key Laboratory of Applied Microbiology Southern China, Institute of Microbiology, Guangdong Academy of Sciences, Guangzhou, China, [3]Department of Laboratory Medicine, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Janet Mans,
University of Pretoria, South Africa
Luciana Damascena Silva,
Evandro Chagas Institute, Brazil

***Correspondence:**
Qingping Wu
wuqp203@163.com
Yu Ding
dingyu@jnu.edu.cn

†*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *15 January 2021*
Accepted: *09 March 2021*
Published: *06 April 2021*

Citation:
Zuo Y, Xue L, Gao J, Liao Y, Liang Y, Jiang Y, Cai W, Qin Z, Yang J, Zhang J, Wang J, Chen M, Ding Y and Wu Q (2021) Evolutionary Mechanism of Immunological Cross-Reactivity Between Different GII.17 Variants. Front. Microbiol. 12:653719. doi: 10.3389/fmicb.2021.653719

Human norovirus is regarded as the leading cause of epidemic acute gastroenteritis with GII.4 being the predominant genotype during the past decades. In the winter of 2014/2015, the GII.17 Kawasaki 2014 emerged as the predominant genotype, surpassing GII.4 in several East Asian countries. Hence, the influence of host immunity response on the continuous evolution of different GII.17 variants needs to be studied in depth. Here, we relate the inferences of evolutionary mechanisms of different GII.17 variants with the investigation of cross-reactivity and cross-protection of their respective antisera using the expression of norovirus P particles in *Escherichia coli*. The cross-reactivity assay showed that the antisera of previous strains (GII.17 A and GII.17 B) reacted with recent variants (GII.17 C and GII.17 D) at high OD values from 0.8 to 1.16, while recent variant antisera cross-reacting with previous strains were weak with OD values between 0.26 and 0.56. The cross-protection assay indicated that the antisera of previous strains had no inhibitory effect on recent variants. Finally, mutations at amino acids 353–363, 373–384, 394–404, and 444–454 had the greatest impact on cross-reactivity. These data indicate that the recent pandemic variants GII.17 C and GII.17 D avoided the herd immunity effect of previous GII.17 A and GII.17 B strains through antigenic variation.

Keywords: norovirus, GII.17, capsid protein, immunogenicity, evolutionary mechanism

INTRODUCTION

Norovirus (NoV) is one of the most leading causes of epidemic acute gastroenteritis (AGE) worldwide, affecting people of all ages and asymptomatic transmission is common (Quee et al., 2020). It is estimated that there are approximately 699 million norovirus illnesses and 219,000 deaths each year, resulting in $4.2 billion in health system costs and $60.3 billion in societal costs (Bartsch et al., 2016). In recent years, the NoV infection rate of sporadic diarrhea in China has ranged from 8.2 to 28.57% (Xue et al., 2019). The main routes of NoV transmission are foodborne, waterborne, airborne, and contaminated surfaces or infected persons (Alsved et al., 2020). The vomitus and feces of infected subjects contain a large number of virions,

whereas as low as 10 infectious particles are required to cause AGE (Teunis et al., 2010). It can be seen that NoV is a threat to public health.

NoV in the family Caliciviridae is nonenveloped single-stranded positive-sense RNA viruses. The genome is 7.5–7.7 kb in length, which contains three open reading frames (ORF1, ORF2, and ORF3). ORF2 encodes the major structural protein VP1 and ORF3 encodes the minor structural protein VP2. X-ray crystallographic structure studies show that the capsid protein VP1 has a shell (S) and a protruding (P) domain (Prasad et al., 1999). The S domain surrounds the viral RNA, while the P domain is linked to the S domain *via* a flexible hinge (Vinjé, 2015). P domain is further divided into the central P1 subdomain and a highly variable P2 subdomain, with the latter, which is essential for virion binding with potential neutralizing antibodies and histo-blood group antigens (HBGAs; Tan et al., 2003). Structural changes of the capsid may lead to the possible emergence of a new phenotype (Mikael et al., 2003).

In 2014/2015, the sudden outbreaks of GII.17 Kawasaki variant have raised widespread concern (Xue et al., 2016b; da Silva Ribeiro de Andrade et al., 2018). GII.17 genotype was grouped into four different clusters (clusters A–D) based on the maximum likelihood analysis (Xue et al., 2016c). Strains in clusters A ranged from 1978 to 2015 and in cluster B emerged from 2005 to 2009. Strains detected after 2012 were divided into cluster C and cluster D (Sang and Yang, 2018). Distances of strains in cluster A ranged from 2.1 to 2.6% indicated the stability of the *VP1* gene over a long period of time. However, the distances between cluster B and cluster A were 7.0–8.0%. The distances between recent strains (cluster C and D) to earlier strains (cluster A and B) increased by 10.0–15.0%, indicating that the recently emerged GII.17 strains were new variants (Chen et al., 2015). GII.17 genotype evolved at a rate of 1.68×10^{-3} (95% HPD: 0.79×10^{-3} – 2.65×10^{-3}) nucleotide substitutions/site/year (Sang and Yang, 2018).

The lack of cell culture systems or small animal models for human diseases severely hinders the development of vaccines and therapeutic agents. Moreover, these viruses are highly heterogeneous and undergoing antigenic variation in response to human immunity, further complicating our understanding of the complex immune interactions that regulate susceptibility and disease (Debbink et al., 2012b). Recent studies have shown that HBGAs are the attachment factors of human norovirus, which are critical in determining NoV host susceptibility and epidemiology (Van Trang et al., 2014; Nordgren and Svensson, 2019). In addition, the blockade assay has been used as a surrogate neutralization assay, and pig gastric mucin (PGM) type III is used as a substrate (Harrington et al., 2002; Lindesmith et al., 2011, 2012). NoV specific antibodies that block the binding of P particles to HBGAs are considered as the best correlate of protection against NoV.

On the capsid surface, amino acids surrounding the binding pocket evolve rapidly, altering HBGA and antibody binding affinity (Debbink et al., 2012a). On one hand, these viruses may evolve to bind with different HBGAs, which indicate that individuals resistant to one pandemic strain may be susceptible to the other. On the other hand, minor sequence and structural differences allow these viruses to escape pre-existing herd immunity, ultimately leading to the emergence of new strains, which are competent to infect the same individuals who had been infected before (Donaldson et al., 2008; **Figure 1**). While the epidemiology and genetic analyses of new GII.17 variants were reported (Xue et al., 2016a,c), and further study to probe the emergence and its high prevalence is significant. Extensive antigenic relationships among NoV GII.17 variants are still

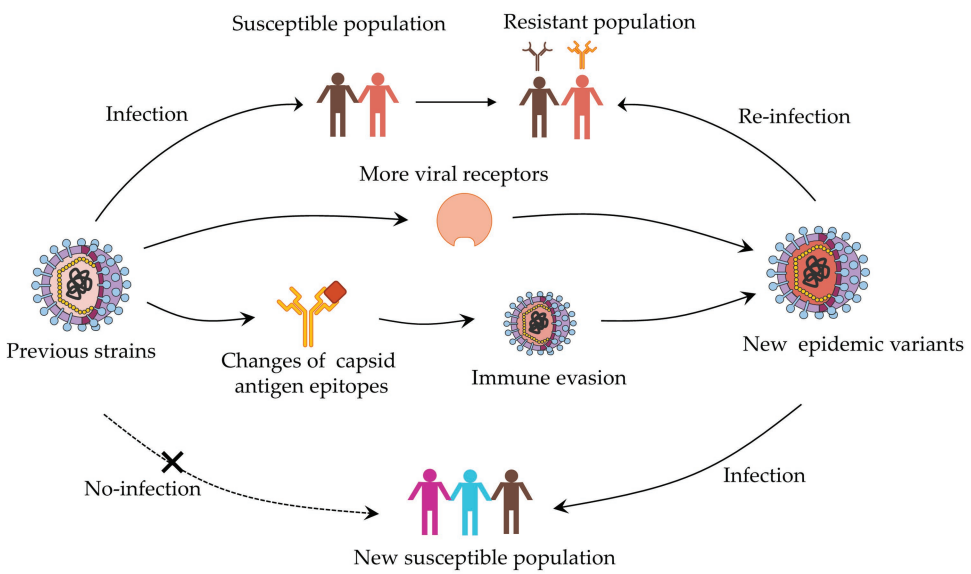


FIGURE 1 | Evolution mechanism of norovirus based on capsid protein variation.

unknown and the complex relationship between host protective immunity and virus antigenic heterogeneity are the primary obstacles to vaccine development. In this study, we systematically investigated the antigenic and ligand-binding characteristics of the four time-ordered GII.17 strains and found potentially critical amino acids that contribute to antigenic variation.

MATERIALS AND METHODS

Ethical Statement

All the animal protocols performed in this study were approved by the Guangdong Institute of Microbiology Laboratory Animal Ethics Committee (permission number GT-IACUC202004271).

Expression and Purification of P Particles in *Escherichia coli*

Four clusters of representative capsid P proteins of NoV GII.17 strains were prepared by *E. coli* prokaryotic expression system (Tan and Jiang, 2005). A cysteine-containing peptide was linked to the N or C terminus of the P domains to promote P-particle formation by forming intermolecular disulfide bridges (Tan and Jiang, 2005). Briefly, the sequences encoding the P domains of NoV GII.17 A-D were amplified by PCR and cloned into the expression vector pGEX-4T-1. After sequence confirmation by sequencing, the bacterial cultures were expressed in *E. coli* BL21 (DE3) with an induction of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 20°C for 20 h. Purification of P protein-glutathione S-transferase (GST) fusion proteins was performed using GSTrap FF. The GST tag was removed by thrombin cleavage for 16 h at 22°C. SDS-PAGE verified the purity of P-protein. Finally, we used a BCA assay kit to measure the concentration of P particle.

Production of Capsid P Protein of GII.17 D Mutants

Capsid P protein of NoV GII.17 D mutants were produced by our laboratory (unpublished data). In brief, after multiple alignments of GII.17 NoVs capsid protein sequences, six mutations were designed on hyper-variable regions by changing these amino acids to alanine, which located in 293-300aa, 342-349aa, 353-363aa, 373-384aa, 394-404aa, and 444-454aa. All mutants were expressed and purified as described above.

Mice Immunization and Sample Collection

Female BALB/c mice purchased from the Laboratory Animal Center of Southern Medical University were divided into four groups of three mice. At 7 weeks old, the mice were immunized four times by subcutaneous injection of 30 μg of GST-P domain fusion protein in Freund's adjuvant at weeks 0, 2, 4, and 5. PBS of the equivalent volume was injected as a control. P particles were emulsified in Freund's complete adjuvant in the first injection and Freund's incomplete adjuvant in the other three injections. Blood samples were collected at weeks 0 (pre-immune serum), 2, 4, 5, and 6. Centrifuging the microtainer tubes in a microcentrifuge to separate serum from the whole blood. Serum samples were transferred to a 1.5-ml microcentrifuge tube and stored at −20°C for later use.

Enzyme Linked Immunosorbent Assay

Antigen-specific and cross-reactive antibody responses were determined by indirect enzyme-linked immunosorbent assay (ELISA). Ninety-six-well microtiter plates were coated with P particle 2 μg/ml at 4°C overnight and blocked with 5% skimmed milk at 37°C for 2 h. Serum samples were added as 2-fold dilution series starting from 1:200 dilution, followed by incubation for 1 h at 37°C. Subsequently, the plates were incubated with the secondary HRP-conjugated goat anti-mouse IgG antibody (diluted 1:3,000) for 30 min at 37°C. Tetramethylbenzidine (TMB) was added as the peroxidase chromogenic substrate to react for 7 min at 37°C. After stopping the substrate reaction with 2M H_2SO_4 and the optical density (OD) values were measured at 450 nm. A sample was considered as positive if the net OD value was above the set cut-off and at least 0.2 OD (Uusi-Kerttula et al., 2014), and the highest reciprocal serum dilution that yielded absorbance >2-fold over the background value was taken as the end-point titer. Geometric mean titers (GMTs) with 95% confidence intervals (CIs) for each immunization group were counted from individual mice end-point titers. Sera dilution with OD450 value of 0.8–1.2 was used for sera antibody cross-reactivity.

HBGAs Binding Assay

The HBGAs binding assay was carried out with the P particle and two different sources of HBGAs: human saliva (type A, B, O, and non-secretor) and PGM type III. Microtiter plates were coated with saliva samples (1:1,000 dilution) and PGM at 4°C overnight. After blocking with 5% skimmed milk, the coated saliva samples and PGM were incubated with P particle at concentrations of 1 μg/ml for 1 h at 37°C. Subsequently, the plates were incubated with antisera from mice immunized with P particles for 1 h at 37°C, followed by incubation with a secondary HRP-conjugated goat anti-mouse IgG antibody (1:3,000 dilution) for 30 min at 37°C. Signal was detected with a TMB Liquid Substrate System. Samples were assayed in triplicate. A sample was considered as positive if the net OD value was above the set cut-off and at least 0.2 OD (Uusi-Kerttula et al., 2014).

HBGAs Blockade Assay

Blockade assay was used to test the ability of antibodies blocking the binding of P particles to ligands. Briefly, microtiter plates were coated with PGM at 4°C overnight, followed by blocking with 5% skimmed milk. Two-fold diluted mouse sera (starting dilution 1:100) were pre-incubated with GII.17 P particle (1 μg/ml) for 1 h at 37°C. After that, adding the pre-incubated mixtures to the plates and then incubated for 1 h at 37°C. The plate were further incubated with immunized rabbit sera at the dilution of 1:2,000 for 1 h at 37°C, followed by a secondary HRP-conjugated goat anti-rabbit IgG antibody (1:3,000 dilution) at 37°C for 30 min. Signal was detected with the TMB Liquid Substrate System. Samples were assayed in triplicate.

The blocking index (%) was calculated as 100% – [(OD wells with serum/OD wells without serum) × 100%; Tamminen et al., 2012]. Sera that did not block 50% of P particle binding to PGM at the lowest serum dilution tested were assigned a titer equal to 0.5 times the limit of detection for statistical analysis (Lindesmith et al., 2017).

Statistical Analysis

All serological data are expressed as the geometric mean titers (GMT). Statistical analysis was performed with Graphpad prism8.0. Schapiro-Wilk test and Mann-Whitney U test were performed to assess the antibody binding properties. Significant differences from different groups were determined using One-way ANOVA. Value of $p < 0.05$ was statistically significant.

RESULTS

Selection of Representative Strains of GII.17 Four Clusters

Phylogenetic relationships of the GII.17 strains based on their capsid protein sequences have been analyzed by our previous studies (Xue et al., 2016b,c). As a result, GII.17 NoV was divided into four clusters (GII.17-A, GII.17-B, GII.17-C, and GII.17-D) during the evolution process. Multiple amino acid sequence alignments showed that a total of 87 differences were found in the capsid protein sequences of four clusters, of which 21 sites were specific to the newly emerged GII.17 Kawasaki 2014 variant. In addition to amino acid changes, six deletion/insertion mutations were identified at site 296, 297, 344, 380, 386, and 400 (corresponding to the consensus capsid protein sequence of all GII.17 strains).

In this study, four representative strains of NoV GII.17 were selected as reference strains to explore the evolutionary mechanism: GII.17-A|2002|CS-E1 (GenBank accession number AY502009.1), GII.17-B|2005|Katrina-17 (GenBank accession number DQ438972.1), GII.17-C|2013|Saitama5203 (GenBank accession number LC043167.1), and GII.17-D|2015|GZ2015-L343 (GenBank accession number KT970376.1).

Preparation of Capsid P Protein and Their Antisera Titer

The P domain of representative strain clusters A (GII.172002), B (GII.172005), C (GII.172013), and D (GII.172014) were expressed for P particle production. As shown in **Figure 2A**, four GII.17 GST-fused capsid proteins eluted from the Glutathione Sepharose beads were 63 kDa. ELISA results showed that after each subsequent immunization, the titer of specific antibodies increased (**Figure 2B**). Immunization with GII.17 A, GII.17 B, GII.17 C, and GII.17 D GST-P fusion protein resulted in equally high titer of specific sera antibodies, with GMT values of 20, 000 (95% CI = 20, 000 – 20, 000); 100, 000 (95% CI = 37, 321 – 278, 576); 640, 000 (95% CI = 640, 000 – 640, 000); and 158, 489 (95% CI = 37, 321 – 278, 576), respectively.

HBGA Binding Profile of Four GII.17 P Proteins

GII.17 P particles were assessed for their capacity to bind to cellular ligands, that is, HBGAs present in human saliva (types A, B, O, and non-secretor) and PGM (**Figure 3**). The results showed that limited binding or no binding signals were observed for the binding of previous strains (GII.17 A and GII.17 B) to HBGAs in the saliva samples of types A, B, and O, and to PGM. However, newly emerged GII.17 variants (GII.17 C and GII.17 D) showed significantly higher binding signals, and both previous and recent strains hardly binding to non-secretors.

Cross-Reactivity of Four GII.17 Antisera

Immunized mouse sera were used to quantify type-specific and cross-reactive sera antibody responses. To determine the cross-reactivity of antibodies induced by the four clusters, sera were analyzed by indirect ELISA (**Figure 4**). Mice immunized

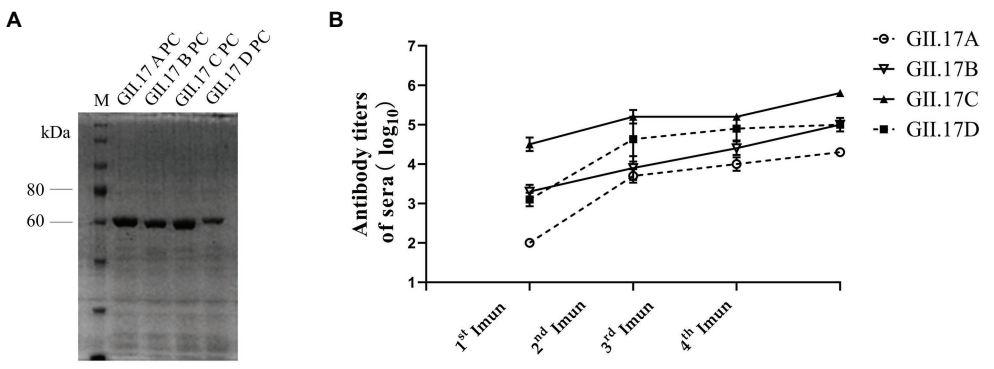


FIGURE 2 | (A) SDS-PAGE analysis of capsid P protein of NoV GII.17 A, GII,17 B, GII.17C, and GII.17 D. The gel was stained with Brilliant Blue (G-250). The M lane contains protein standards. **(B)** Titers of NoV GII.17 sera antibodies. Sera samples were collected from mice immunized with GII.17 A, GII.17 B, GII.17 C, and GII.17 D fusion protein at weeks 2, 4, 5, and 6. Antibody titers were measured by indirect ELISA using a 1:200 dilution of sera samples, and the results are represented in log10. 1st Imun, the first immunization; 2nd Imun, the second immunization; 3rd Imun, the third immunization; and 4th Imun, the fourth immunization.

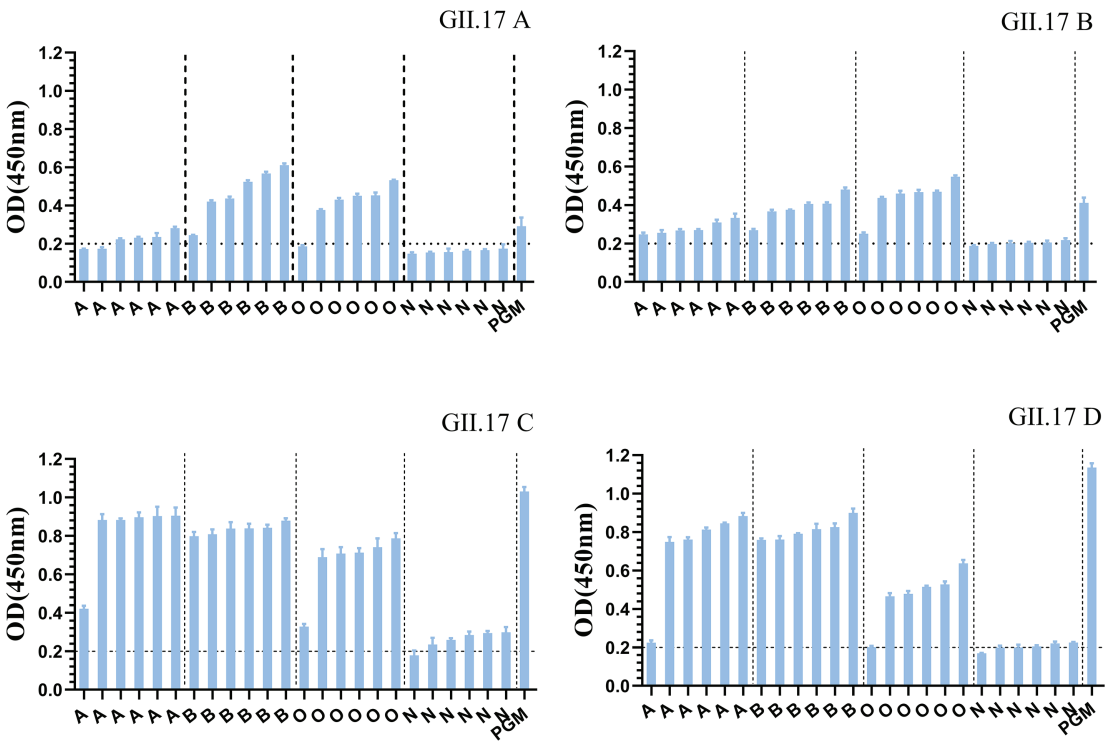


FIGURE 3 | Binding of four clusters of GII.17 NoV P particles to saliva samples. GII.17 P particles were assayed for the binding property to saliva samples phenotyped by ELISA. Saliva samples from six individuals of each carbohydrate type (type A, B, O, and N), pig gastric mucin (PGM) were coated onto 96-well microplates, and then add P protein from four clusters of GII.17 viruses (GII.17 A, GII.17 B, GII.17 C, and GII.17 D). Shown are mean OD values with standard errors of P particles binding to saliva type A, B, O, non-secretor, and pig gastric mucin (PGM). Each group was sorted from the lowest to the highest in terms of the intensity of binding to saliva according to the OD 450 nm readings. Dotted line represents the cut-off value for samples considered positive. O, A, B, and N represent the type O (H antigen), A, B, and non-secretor saliva, respectively.

with capsid P proteins from previous strains (GII.17 A and GII.17 B) induced relatively high cross-reactive responses against the other two cluster variants, with a minimum OD value of 0.8; however, the cross-reactivity of the antisera against the recent variants (GII.17-C and GII.17-D) with that against previous strains (GII.17-A and GII.17-B) showed a significant decrease ($p < 0.0001$), with a maximum OD value of 0.56. These results indicate that the major capsid proteins of the previous and recent GII.17 strains diverged from each other at the antigenic level.

Preparation of GII.17 Mutants and Cross-Reactivity With Four Antisera

In addition to the changes in receptor binding function, spontaneous mutations in the capsid gene are also responsible for the evolution of GII.17 strains. This study compared nucleotide and amino acid sequences differences between the four representative GII.17 strains (**Supplementary Table S1**; **Supplementary Figure S1**). The identity of selected genomes at the nucleotide level was 85.4%. As a result of multiple amino acid sequence alignments, a total of 84 concrete differences in the capsid protein sequences of four strains were found. Based on the changes in amino acid sequences during the evolution of GII.17 NoV, we designed six mutants: 293-300A, 342-349A, 353-363A, 373-384A, 394-404A, and 444-454A.

In our previous study about the variation of capsid proteins, the stability and binding ability of six GII.17 mutants (293-300A, 342-349A, 353-363A, 373-384A, 394-404A, and 444-454A) were verified by GII.17 monoclonal antibody. ELISA results showed that the six GII.17 mutants reacted with GII.17-2A111C1 monoclonal antibodies at OD value of 0.45 ± 0.07, 0.51 ± 0.08, 0.53 ± 0.02, 0.42 ± 0.06, 0.60 ± 0.17, and 0.48 ± 0.04, respectively, while the binding signal of positive control (GII.17 wild-type) and negative control were 0.46 ± 0.06 ($p > 0.05$) and 0.10 ± 0.04 (Unpublished data). The data indicated that mutation has no significant effect on the stability and binding ability of the P particles. ELISA results (**Figure 5**) showed that the cross-reactivity of all mutants tested with GII.17 A immunized sera decreased obviously, as is the case with GII.17 C immunized sera ($p < 0.0001$). Both 353-363A and 373-384A mutants had a robust decreased ($p < 0.05$ and $p < 0.01$, respectively) reaction with GII.17 B immunized sera, the other mutants have no obvious changes. As to GII.17 D immunized sera, the reaction of 293-300A mutant hardly changed; 353-363A, 373-384A, 394-404A, and 444-454A mutants had lost the binding property; 343-349A mutant had a significantly decreased reaction ($p < 0.0001$). In general, replacing residues 353-363aa, 373-384aa,

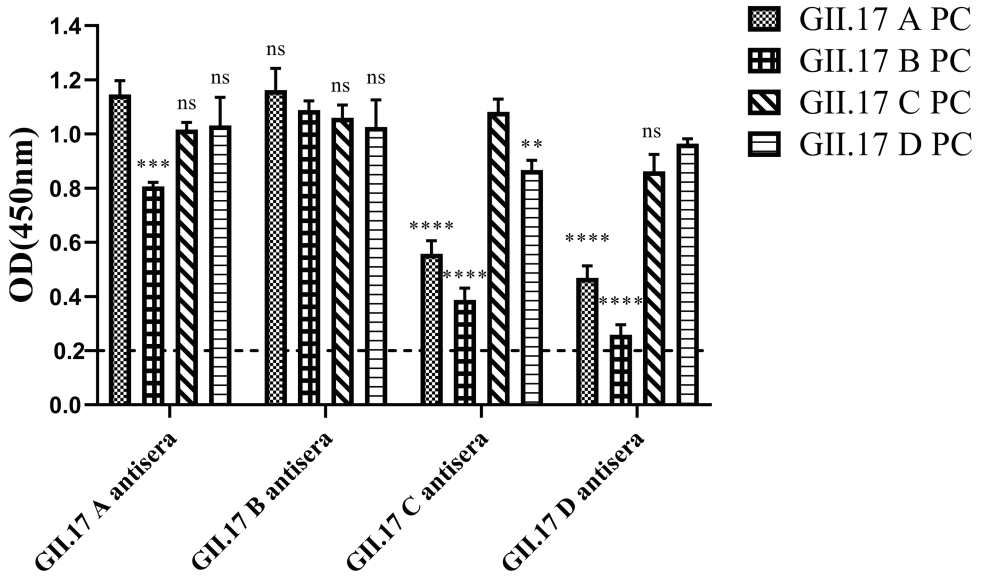


FIGURE 4 | GII.17 antibody responses in homologous and heterologous mouse immune antisera. Antisera of GII.17 A, GII.17 B, GII.17 C, and GII.17 D were analyzed for cross-reactive responses. OD values were expressed as the mean ± SD. Dotted line represents the cut-off value for samples was considered as positive. *Significantly different from homotypic sera (by ANOVA). $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; and $^{****}p < 0.0001$. ns, not significant.

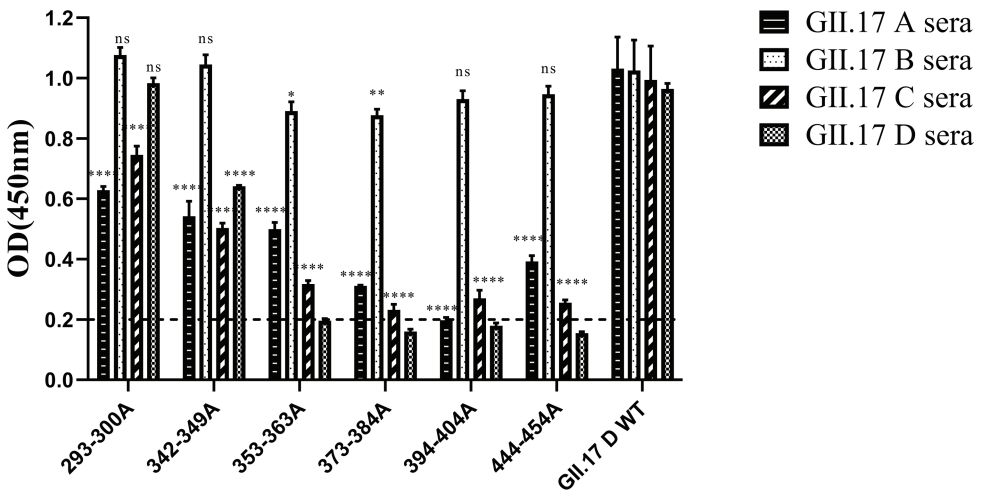


FIGURE 5 | Cross-reactivity of six NoV GII.17 D mutants with GII.17 (A–D) antisera. OD values were expressed as the mean ± SD. Horizontal line illustrates the cut-off value for samples considered positive. *Significantly different from wild-type GII.17 D variant sera (by analysis of variance). $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; and $^{****}p < 0.0001$. ns, not significant.

394-404aa, and 444-454aa had the greatest impact on cross-reactivity, followed by 293-300aa and 342-349aa.

Cross-Clusters Blockade Antibody Responses of GII.17 Antisera

To measure the potential for cross-clusters protection as the viruses evolve, four GII.17 clusters and their antisera were tested for blockade activity. Since HGBAs binding ligands for GII.17 A and GII.17 B have not been identified, blockade potential for them could not be determined (**Figure 6**). As a result, GII.17 C and GII.17 D variants induced considerable homologous blocking activity as mean serum titers blocking at least 50% of P particle binding, with GMT of 3,162 (95% CI, 2,736–3,215) and 1,995 (95% CI, 160.3–3,942). GII.17 C antisera (GMT, 2,512; 95% CI, 253.2–4,793) were able to block GII.17 D P particle binding to PGM. Serum from mice immunized with GII.17 A and GII.17 B P particle was not able to block GII.17 C and GII.17 D variants. The result further

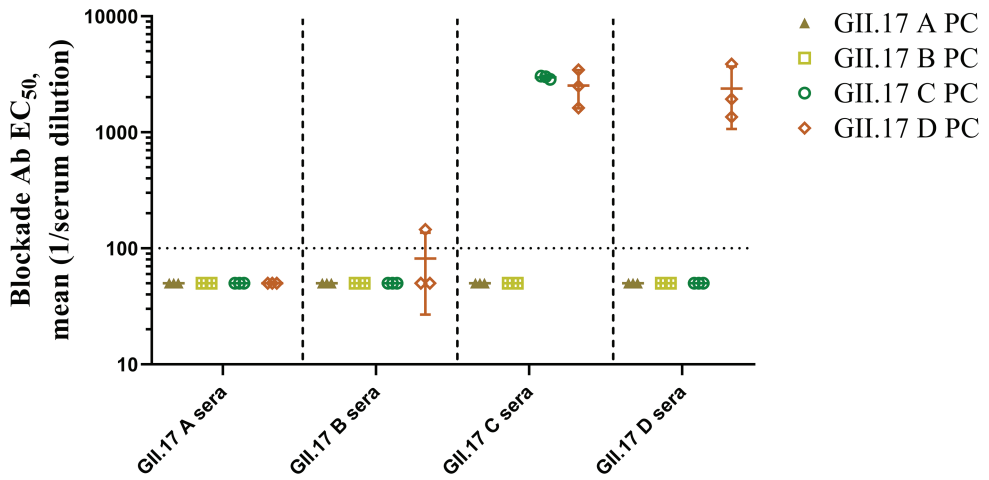


FIGURE 6 | GII.17 type-specific and cross-reactive blockade antibodies. Sera from mice immunized with GII.17 P particles were evaluated for the ability to block P particle binding with carbohydrate ligand pig gastric mucin type III (PGM). The results are represented in log10 antibody titers. EC_{50}, half-maximum effective concentration. Bars represent the geometric mean titer (GMT). Dotted line equals the assay limit of detection.

suggested that major capsid protein of previous and recent GII.17 strains are undergoing antigenic variation over time.

DISCUSSION

NoV is the most common cause of acute viral gastroenteritis worldwide and GII.4 genotype variants have been predominant during the past two decades (Siebenga et al., 2009). However, GII.17 Kawasaki 2014 pandemic variants were found to cause an alarming number of outbreaks in certain parts of East Asian countries in 2014–2015 (Lu et al., 2015; Matsushima et al., 2015; Shi et al., 2016). The first GII.17 strain in the National Center for Biotechnology Information (NCBI) databank is from 1978 (Rackoff et al., 2013). Since then viruses with a GII.17 capsid genotype have sporadically been detected in Africa, Asia, Europe, North America, and South America (de Graaf et al., 2015).

When the novel GII.17 Kawasaki 2014 variant emerged and surpassed the predominating GII.4 viruses, there has been a concern of global spreading of GII.17 NoV (da Silva Ribeiro de Andrade et al., 2018). A previous study estimated that GII.17 VP1 evolves one order magnitude faster than GII.4 (Chan et al., 2015), which may allow their rapid emergence and cause a spread of a pandemic. Therefore, it is very important and necessary to evaluate the impact of sequence changes on antigenicity. This study further explored the mechanism behind the emergence and the high prevalence of the new variants.

The first GII.4 NoV emerged and caused a pandemic in the mid-1990s (Noel et al., 1999). Since 2002, new variants emerging every 2–4 years, resulting in epidemics and sometimes global pandemics (Siebenga et al., 2007; Huhti et al., 2014). A molecular evolution model has been proposed, that is, NoV GII.4 strains persist by evolving novel carbohydrate-binding domains in response to immune-driven selection and by antigenic drift in the receptor-binding regions of the P2 subdomain (Lindesmith et al., 2008). Monoclonal antibodies targeting distinct GII.4 strains show that antigenic variation is high and these strains are evolving in response to human herd immunity (Debbink et al., 2012b).

Multiple studies have shown that GII.4 NoV exhibited the broadest binding spectrum, attaching efficiently to saliva from secretor A, B, and O individuals, which may have contributed to the emergence of GII.4 viruses in human populations (de Rougemont et al., 2011; Dai et al., 2015). In this study, we characterized the binding spectrum of NoV GII.17 clusters to HBGAs in human saliva. Concordantly with the previous findings (Qian et al., 2019), recent GII.17 variants bound strongly to types A, B, O saliva and PGM, while previous GII.17 strains did not exhibit such strong binding. Strong binding of recent GII.17 variants to HBGAs in saliva samples of types A and B was observed, while binding to type O saliva were somewhat lower, especially GII.17 D (Malm et al., 2018). The HBGAs binding spectrum seemed to be consistent with the prevalence of the variants, in which the strains with strong binding capacity are more prevalent than those with weak binding, explaining the predominance of the new GII.17 variants. It further showed that the high prevalence of GII.17 D variant is the result of long-term evolution under the selection of HGBAs.

Antigenic change is another important factor that drives the evolution and the emergence of new variants that were often associated with new pandemics (Debbink et al., 2012a). Mice immunized with GST-P domain capsid proteins of the four clusters variants obtained high titer sera. Whether GST was removed from the immunogens or not did not significantly impact the immune responses (Wang et al., 2013). Previous strains antisera had high cross-reactivity against recent variants, with the minimum OD value of 0.8. However, sera from mice immunized with recent variants had a relatively weak reaction with previous strains, with the maximum OD value of 0.56. These results revealed that previous strains and recent variants

were diverging from each other at the antigenic level, the significant antigenic epitopes of previous strains were still preserved and recent variants probably contain evolving antigenic epitopes and innovative structural changes to the capsid protein (Nilsson et al., 2003), such that changes probably lead to the evasion of the predominant protective immune response.

Receptor binding blockade assay with the results of sera from mice immunized with previous GII.17 strains had no blockade activity against recent GII.17 variants, combined with cross-reactivity data demonstrated that previous strains antisera may recognize highly conserved linear epitopes. Five blockade epitopes on VP1 in GII.17 NoV have been hypothesized, these are Epitope A (294–301aa, 374aa, and 378aa), Epitope B (336aa and 395aa), Epitope C (343aa and 382aa), Epitope D (residues 400–402aa), and Epitope E (residues 410aa, 414–416aa; Chen et al., 2015). Most variation we observed in the VP1 sequences of recent GII.17 strains mapped in or near these epitopes, which may lead to herd immune escape or alter human susceptibility to GII.17 NoV, thus partially explaining the prevalence of this novel variant (Chen et al., 2015).

To identify amino acids regulating antigenic differences over time, six GII.17 D mutants reacted with four GII.17 wild-type strains immunized serum. Mutational analysis showed that the mutations at amino acids 353–363, 373–384, 394–404, and 444–454 had the most significant impact on cross-reactivity, which further supported that those antigenic sites played an important role in promoting antigen evolution. In addition, we found that all wild-type variants and mutants had high cross-reactivity with GII.17 B antisera, and there is no significant difference in type-specific sera antibody response, which indicated that GII.17 B strain has strong potential for vaccine design.

This study explored the connection between antisera and different NoV GII.17 variants by immunizing mice. Although the mouse model is very extensive in immunological research, it will be more meaningful for using human antisera.

It is conceivable to anticipate cross-protection among GII.17 viruses since our data showed that GII.17 D variant immunized sera had a strong cross-reactivity response against the other clusters. Dai et al. (2017) found GII.4 antisera from patients had limited cross-reactivity or even no cross-blockade activity against GII.17, providing new evidence explaining the emergence and rapid expansion and indicating GII.17 variants need to be considered in future vaccine strategies. Important issues in NoV vaccine design are the genetic diversity and evolution resulting in escaping herd immunity, and our study provides a new perspective of them to produce broad protection against many GII.17 NoV variants. Understanding the molecular mechanism governing GII.17 antigenic variation is a key to design vaccines.

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 at: https://www.ncbi.nlm.nih.gov/genbank/, KT970376.1.

ETHICS STATEMENT

The animal study was reviewed and approved by the Guangdong Institute of Microbiology Laboratory Animal Ethics Committee.

AUTHOR CONTRIBUTIONS

YZ and LX designed the experiments and wrote the manuscript. YJ and WC contributed reagents, materials, and analysis tools. JG, YLo, and YLn performed the experiments. ZQ and JY analyzed the data. MC and JW revised the manuscript. YD, JZ, and QW provided the project funding. All authors have read and agreed to the published version of the manuscript.

FUNDING

This project was supported by the National Natural Science Foundation of China (31872912 and 31701717), the Natural Science Foundation of Guangdong Province for Distinguished Young Scholars (2019B151502065), the Key Research and Development Program of Guangdong Province (2019B020209001), the Key Research and Development Program of Jiangxi Province (20202BBGL73053), and the GDAS' Project of Science and Technology Development (2020GDASYL-20200104008).

ACKNOWLEDGMENTS

We are grateful to Dr. Yuqi Huo (the sixth people's hospital of Zhengzhou) for providing rabbit antisera. The personnel of Institutional Animal Care are thanked for technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.653719/full#supplementary-material

REFERENCES

Alsved, M., Fraenkel, C.-J., Bohgard, M., Widell, A., Söderlund-Strand, A., Lanbeck, P., et al. (2020). Sources of airborne norovirus in hospital outbreaks. *Clin. Infect. Dis.* 70, 2023–2028. doi: 10.1093/cid/ciz584

Bartsch, S. M., Lopman, B. A., Ozawa, S., Hall, A. J., and Lee, B. Y. (2016). Global Economic Burden of Norovirus Gastroenteritis. *PLoS One* 11:e0151219. doi: 10.1371/journal.pone.0151219

Chan, M. C. W., Lee, N., Hung, T.-N., Kwok, K., Cheung, K., Tin, E. K. Y., et al. (2015). Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. *Nat. Commun.* 6:10061. doi: 10.1038/ncomms10061

Chen, H., Qian, F., Xu, J., Chan, M., Shen, Z., Zai, S., et al. (2015). A novel norovirus GII.17 lineage contributed to adult gastroenteritis in Shanghai, China, during the winter of 2014–2015. *Emerg. Microbes Infect.* 4:e67. doi: 10.1038/emi.2015.67

da Silva Ribeiro de Andrade, J., Fumian, T. M., Leite, J. P. G., de Assis, M. R., Fialho, A. M., Mouta, S., et al. (2018). Norovirus GII.17 Associated with a Foodborne Acute Gastroenteritis Outbreak in Brazil, 2016. *Food Environ. Virol.* 10, 212–216. doi: 10.1007/s12560-017-9326-0

Dai, Y. C., Xia, M., Huang, Q., Tan, M., Qin, L., Zhuang, Y. L., et al. (2017). Characterization of Antigenic Relatedness between GII.4 and GII.17 Noroviruses by Use of Serum Samples from Norovirus-Infected Patients. *J. Clin. Microbiol.* 55, 3366–3373. doi: 10.1128/JCM.00865-17

Dai, Y. C., Zhang, X. F., Xia, M., Tan, M., Quigley, C., Lei, W., et al. (2015). Antigenic relatedness of norovirus GII.4 variants determined by human challenge sera. *PLoS One* 10:e0124945. doi: 10.1371/journal.pone.0124945

de Graaf, M., Van Beek, J., Vennema, H., Podkolzin, A. T., Hewitt, J., Bucardo, F., et al. (2015). Emergence of a novel GII.17 norovirus – End of the GII.4 era? *Euro Surveill.* 20:21178. doi: 10.2807/1560-7917.es2015.20.26.21178

de Rougemont, A., Ruvoen-Clouet, N., Simon, B., Estienney, M., Elie-Caille, C., Aho, S., et al. (2011). Qualitative and quantitative analysis of the binding of GII.4 norovirus variants onto human blood group antigens. *J. Virol.* 85, 4057–4070. doi: 10.1128/JVI.02077-10

Debbink, K., Donaldson, E. F., Lindesmith, L. C., and Baric, R. S. (2012a). Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. *J. Virol.* 86, 1214–1226. doi: 10.1128/JVI.06189-11

Debbink, K., Lindesmith, L. C., Donaldson, E. F., and Baric, R. S. (2012b). Norovirus immunity and the great escape. *PLoS Pathog.* 8:e1002921. doi: 10.1371/journal.ppat.1002921

Donaldson, E. F., Lindesmith, L. C., Lobue, A. D., and Baric, R. S. (2008). Norovirus pathogenesis: mechanisms of persistence and immune evasion in human populations. *Immunol. Rev.* 225, 190–211. doi: 10.1111/j.1600-065X.2008.00680.x

Harrington, P. R., Lisa, L., Boyd, Y., Moe, C. L., and Baric, R. S. (2002). Binding of Norwalk virus-like particles to ABH histo-blood group antigens is blocked by antisera from infected human volunteers or experimentally vaccinated mice. *J. Virol.* 76, 12335–12343. doi: 10.1128/JVI.76.23.12335-12343.2002

Huhti, L., Blazevic, V., Puustinen, L., Hemming, M., Salminen, M., and Vesikari, T. (2014). Genetic analyses of norovirus GII.4 variants in Finnish children from 1998 to 2013. *Infect. Genet. Evol.* 26, 65–71. doi: 10.1016/j.meegid.2014.05.003

Lindesmith, L. C., Debbink, K., Swanstrom, J., Vinje, J., Costantini, V., Baric, R. S., et al. (2012). Monoclonal antibody-based antigenic mapping of norovirus GII.4-2002. *J. Virol.* 86, 873–883. doi: 10.1128/JVI.06200-11

Lindesmith, L. C., Donaldson, E. F., and Baric, R. S. (2011). Norovirus GII.4 strain antigenic variation. *J. Virol.* 85, 231–242. doi: 10.1128/JVI.01364-10

Lindesmith, L. C., Donaldson, E. F., Lobue, A. D., Cannon, J. L., Zheng, D. -P., Vinje, J., et al. (2008). Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med.* 5:e31. doi: 10.1371/journal.pmed.0050031

Lindesmith, L. C., Kocher, J. F., Donaldson, E. F., Debbink, K., Mallory, M. L., Swann, E. W., et al. (2017). Emergence of novel human norovirus GII.17 strains correlates with changes in blockade antibody epitopes. *J. Infect. Dis.* 216, 1227–1234. doi: 10.1093/infdis/jix385

Lu, J., Sun, L., Fang, L., Yang, F., Mo, Y., Lao, J., et al. (2015). Gastroenteritis outbreaks caused by norovirus GII.17, Guangdong Province, China, 2014-2015. *Emerg. Infect. Dis.* 21, 1240–1242. doi: 10.3201/eid2107.150226

Malm, M., Tamminen, K., Vesikari, T., and Blazevic, V. (2018). Norovirus GII.17 virus-like particles bind to different histo-blood group antigens and cross-react with genogroup II-specific mouse sera. *Viral Immunol.* 31, 649–657. doi: 10.1089/vim.2018.0115

Matsushima, Y., Ishikawa, M., Shimizu, T., Komane, A., Kasuo, S., Shinohara, M., et al. (2015). Genetic analyses of GII.17 norovirus strains in diarrheal disease outbreaks from December 2014 to March 2015 in Japan reveal a novel polymerase sequence and amino acid substitutions in the capsid region. *Euro Surveill.* 20:21173. doi: 10.2807/1560-7917.es2015.20.26.21173

Mikael, N., Kjell-Olof, H., Margareta, T., GöRan, L., Kari, J., Anders, E., et al. (2003). Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. *J. Virol.* 77:13117. doi: 10.1128/jvi.77.24.13117-13124.2003

Nilsson, M., Hedlund, K. O., Thorhagen, M., Larson, G., Johansen, K., Ekspong, A., et al. (2003). Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. *J. Virol.* 77, 13117–13124. doi: 10.1128/jvi.77.24.13117-13124.2003

Noel, J. S., Fankhauser, R. L., Ando, T., Monroe, S. S., and Glass, R. I. (1999). Identification of a distinct common strain of "Norwalk-like viruses" having a global distribution. *J. Infect. Dis.* 179, 1334–1344. doi: 10.1086/314783

Nordgren, J., and Svensson, L. (2019). Genetic susceptibility to human norovirus infection: an update. *Viruses* 11:226. doi: 10.3390/v11030226

Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G., and Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* 286, 287–290. doi: 10.1126/science.286.5438.287

Qian, Y., Song, M., Jiang, X., Xia, M., Meller, J., Tan, M., et al. (2019). Structural adaptations of norovirus GII.17/13/21 lineage through two distinct evolutionary paths. *J. Virol.* 93:e01655–18. doi: 10.1128/jvi.01655-18

Quee, F. A., De Hoog, M. L. A., Schuurman, R., and Bruijning-Verhagen, P. (2020). Community burden and transmission of acute gastroenteritis caused by norovirus and rotavirus in the Netherlands (RotaFam): a prospective household-based cohort study. *Lancet Infect. Dis.* 20, 598–606. doi: 10.1016/S1473-3099(20)30058-X

Rackoff, L. A., Bok, K., Green, K. Y., and Kapikian, A. Z. (2013). Epidemiology and evolution of rotaviruses and noroviruses from an archival WHO Global Study in Children (1976-79) with implications for vaccine design. *PLoS One* 8:e59394. doi: 10.1371/journal.pone.0059394

Sang, S., and Yang, X. (2018). Evolutionary dynamics of GII.17 norovirus. *PeerJ* 6:e4333. doi: 10.7717/peerj.4333

Shi, C., Feng, W. H., Shi, P., Ai, J., Guan, H. X., Sha, D., et al. (2016). An acute gastroenteritis outbreak caused by GII.17 norovirus in Jiangsu Province, China. *Int. J. Infect. Dis.* 49, 30–32. doi: 10.1016/j.ijid.2016.05.004

Siebenga, J. J., Vennema, H., Renckens, B., De Bruin, E., Van Der Veer, B., Siezen, R. J., et al. (2007). Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *J. Virol.* 81, 9932–9941. doi: 10.1128/JVI.00674-07

Siebenga, J. J., Vennema, H., Zheng, D. P., Vinjé, J., Lee, B. E., Pang, X. L., et al. (2009). Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001-2007. *J. Infect. Dis.* 200, 802–812. doi: 10.1086/605127

Tamminen, K., Huhti, L., Koho, T., Lappalainen, S., Hytönen, V. P., Vesikari, T., et al. (2012). A comparison of immunogenicity of norovirus GII-4 virus-like particles and P-particles. *Immunology* 135, 89–99. doi: 10.1111/j.1365-2567.2011.03516.x

Tan, M., Huang, P., Meller, J., Zhong, W., Farkas, T., and Jiang, X. (2003). Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. *J. Virol.* 77, 12562–12571. doi: 10.1128/jvi.77.23.12562-12571.2003

Tan, M., and Jiang, X. (2005). The p domain of norovirus capsid protein forms a subviral particle that binds to histo-blood group antigen receptors. *J. Virol.* 79, 14017–14030. doi: 10.1128/JVI.79.22.14017-14030.2005

Teunis, P. F. M., Moe, C. L., Liu, P., Miller, S. E., Lindesmith, L., Baric, R. S., et al. (2010). Norwalk virus: how infectious is it? *J. Med. Virol.* 80, 1468–1476. doi: 10.1002/jmv.21237

Uusi-Kerttula, H., Tamminen, K., Malm, M., Vesikari, T., and Blazevic, V. (2014). Comparison of human saliva and synthetic histo-blood group antigens usage as ligands in norovirus-like particle binding and blocking assays. *Microbes Infect.* 16, 472–480. doi: 10.1016/j.micinf.2014.02.010

Van Trang, N., Vu, H. T., Le, N. T., Huang, P., Jiang, X., and Anh, D. D. (2014). Association between norovirus and rotavirus infection and histo-blood group antigen types in Vietnamese children. *J. Clin. Microbiol.* 52, 1366–1374. doi: 10.1128/jcm.02927-13

Vinjé, J. (2015). Advances in laboratory methods for detection and typing of norovirus. *J. Clin. Microbiol.* 53, 373–381. doi: 10.1128/jcm.01535-14

Wang, L., Huang, P., Fang, H., Xia, M., Zhong, W., Mcneal, M. M., et al. (2013). Polyvalent complexes for vaccine development. *Biomaterials* 34, 4480–4492. doi: 10.1016/j.biomaterials.2013.02.041

Xue, L., Cai, W., Gao, J., Zhang, L., Dong, R., Li, Y., et al. (2019). The resurgence of the norovirus GII.4 variant associated with sporadic gastroenteritis in the post-GII.17 period in South China, 2015 to 2017. *BMC Infect. Dis.* 19:696. doi: 10.1186/s12879-019-4331-6

Xue, L., Cai, W., Wu, Q., Zhang, J., and Guo, W. (2016a). Direct sequencing and analysis of the genomes of newly emerging GII.17 norovirus strains in South China. *J. Appl. Microbiol.* 120, 1130–1135. doi: 10.1111/jam.13052

Xue, L., Dong, R., Wu, Q., Li, Y., Cai, W., Kou, X., et al. (2016b). Molecular epidemiology of noroviruses associated with sporadic gastroenteritis in Guangzhou, China, 2013-2015. *Arch. Virol.* 161, 1377–1384. doi: 10.1007/s00705-016-2784-0

Xue, L., Wu, Q., Cai, W., Zhang, J., and Guo, W. (2016c). Molecular characterization of new emerging GII.17 norovirus strains from South China. *Infect. Genet. Evol.* 40, 1–7. doi: 10.1016/j.meegid.2016.02.026

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 April 2021
doi: 10.3389/fmicb.2021.667084

Isolation and Identification of a Recombinant Porcine Epidemic Diarrhea Virus With a Novel Insertion in S1 Domain

Dongliang Li[1,2†], Yongtao Li[1†], Yunchao Liu[2], Yumei Chen[3], Wenqiang Jiao[2], Hua Feng[2], Qiang Wei[2], Jucai Wang[1], Yuhang Zhang[1] and Gaiping Zhang[1,2,3]*

[1] College of Veterinary Medicine, Henan Agricultural University, Zhengzhou, China, [2] Henan Provincial Key Laboratory of Animal immunology, Henan Academy of Agricultural Sciences, Zhengzhou, China, [3] School of Life Sciences, Zhengzhou University, Zhengzhou, China

Porcine epidemic diarrhea virus (PEDV) is the major pathogen that causes diarrhea and high mortality in newborn piglets with devastating impact to the pig industry. Recombination and mutation are the main driving forces of viral evolution and genetic diversity of PEDV. In 2016, an outbreak of diarrhea in piglets occurred in an intensive pig farm in Central China. A novel PEDV isolate (called HNAY) was successfully isolated from clinical samples. Sequence analysis and alignment showed that HNAY possessed 21-nucleotide (nt) insertion in its S1 gene, which has never been reported in other PEDV isolates. Moreover, the sequence of the insertion was identical with the sequence fragment in PEDV N gene. Notably, the HNAY strain exhibited two unique mutations (T500A and L521Y) in the neutralizing epitopes of the S1 protein that were different from those of other PEDV variant strains and CV777-based vaccine strains. Additionally, PEDV HNAY might be derived from a natural recombination between two Chinese variant PEDV strains. Animal experiments demonstrated that HNAY displayed higher pathogenicity compared with two other clinical isolates. This study lays the foundation for better understanding of the genetic evolution and molecular pathogenesis of PEDV.

Keywords: porcine epidemic diarrhea virus, spike protein, novel insertion, recombination, pathogenicity

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Baochao Fan,
Jiangsu Academy of Agricultural Sciences (JAAS), China
Deping Song,
Jiangxi Agricultural University, China

****Correspondence:***
Gaiping Zhang
zhanggaip@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *12 February 2021*
Accepted: *29 March 2021*
Published: *20 April 2021*

Citation:
Li D, Li Y, Liu Y, Chen Y, Jiao W, Feng H, Wei Q, Wang J, Zhang Y and Zhang G (2021) Isolation and Identification of a Recombinant Porcine Epidemic Diarrhea Virus With a Novel Insertion in S1 Domain. Front. Microbiol. 12:667084. doi: 10.3389/fmicb.2021.667084

INTRODUCTION

Coronaviruses (CoVs) belong to RNA viruses and infect different vertebrates, such as mammalian and avian species (Durham et al., 1979; Law et al., 2005). As an important member of CoVs, porcine epidemic diarrhea virus (PEDV) can cause an acute enteritis disease, which is characterized by vomiting, diarrhea and dehydration, leading to high mortality in newborn piglets (Song and Park, 2012). The genome of PEDV is about 28 kb in size and composed of at least six open reading frames (ORFs) including ORF1a, ORF1b, spike (S), envelope (E), membrane (M), and nucleocapsid (N) (Wang et al., 2019). Among them, S protein is one of the most important structural protein due to its involvement in multiple functions, such as virus attachment, receptor binding, virus-cell membrane fusion, and inducing the production of virus neutralizing antibodies in the host (Vaughn et al., 1995; Vennema et al., 1998; Vijgen et al., 2005; Hasoksuz et al., 2007;

Lorusso et al., 2008). Therefore, it is very necessary to continuously monitor the genetic alterations of PEDV S gene in the field (Park et al., 2014; Zhao et al., 2017).

Since its first appearance in Europe in 1971, PEDV has caused considerable economic losses, especially in Asia (Pensaert and de Bouck, 1978; Sun et al., 2016). Since 2010, PEDV variant strain has emerged in Asia, which can infect the pigs of all ages but most severely in suckling pigs, reaching up to 95% mortality (Bi et al., 2012; Li et al., 2012; Pan et al., 2012). In 2013, the first outbreak of PEDV in the United States (US) was reported, and then spread to Canada and Mexico (Stevenson et al., 2013; Chen et al., 2014; Pasick et al., 2014). Phylogenetic analyses showed the emergent US PEDV strains are most closely related to a Chinese strain detected in 2012 (Huang et al., 2013). From a broad perspective, phylogenetic analysis based on S gene can divide PEDVs into two major genotypes in the US: the original non-S-INDEL (S-INDEL standing for insertions/deletions in the S gene) and the variant S-INDEL (Vlasova et al., 2014; Lin et al., 2016). PEDV strains which were first identified in the US in 2013 with high virulence are classified as the non-S INDEL strains (Stevenson et al., 2013). The new variant strains which were reported in the US in 2014 with relatively milder virulence are termed as S-INDEL strains (Wang et al., 2014). In addition, PEDVs can be divided into genome 1 (G1) and genome 2 (G2) clades and further divided into four subgroups: G1a, G1b, G2a, and G2b (Wang et al., 2016). Despite they have relatively conserved neutralizing epitopes, commercial vaccines based on CV777 strains cannot provide complete protection in piglets against variant PEDV infections which are currently circulating in China and the US (Zhu et al., 2019). Monitoring the genetic alterations in the S gene of PEDV could provide more clues for understanding the epidemic characteristics of PEDV and adjusting the immunization program of PED in pig farms.

Henan province, located in the center of China, is the largest pig-producing province in China. Since 2010, severe diarrhea outbreaks caused by variant PEDV frequently occurs in this area. In our routine epidemiological investigation, 52 sequences of the full-length S genes and 9 complete genomes of PEDV have been determined, indicating the complexity of PEDV in Henan (Su et al., 2016). In this study, a PEDV strain with novel insertion in its S gene (called HNAY) was successfully isolated from an intensive farm in Henan in 2016. We determined the genetic relationships, variant characteristics, potential recombination features, and the clinical characterization of HNAY. Our findings could provide valuable information for PED outbreaks in Central china and contribute to prevent and control of PED.

MATERIALS AND METHODS

Sample Collection, Reverse Transcription (RT)-PCR, and Sequencing

Intestine and fecal samples were collected from pig farms in Henan, China during 2016–2018 and stored at −80°C. These samples came from piglets with clinical signs including vomiting, and severe diarrhea. Total RNA was extracted from the small intestine samples using TaKaRa MiniBEST Viral RNA/DNA Extraction Kit Ver.5.0 (TaKaRa, Dalian, China). Viral cDNA was generated via reverse transcription using PrimeScript reverse transcriptase (TaKaRa, Dalian, China) according to the manufacturer's instructions. The samples were examined for PEDV by RT-PCR (Li et al., 2018). Three samples tested PEDV positive were sequenced for complete genome as described in previous study (Wang et al., 2013). Purified PCR products were cloned into PMD-19T vector and recombinant DNA clones were sequenced by the BigDye Terminator v3.1 Cycle Sequencing kit (Biotech Company, China).

Phylogenetic and Recombination Analysis

Complete genome sequence of PEDV and their S gene were aligned using ClustalW in DNASTAR7.1. Phylogenetic tree of PEDV was constructed using neighbor-joining (NJ) method with 1,000 bootstrap replicates in MEGA7. The sequences of reference strains are downloaded from GenBank database (**Table 1**). Potential recombination in the genome of PEDV HNAY were assessed by the Recombination Detection Program v4 (RDP4) which was supported by ≥6 programs.

Protein Structure Prediction

According to the crystal structure of S protein of PEDV USA/Colorado/2013 available in the PDB database (accession code 6VV5) as the input model, the three-dimensional structure of the S protein of HNAY was generated via SWISS-MODEL Homology Modeling server and figures were made with PymoL (Version 1.0 Schrödinger, LLC; Wrapp and McLellan, 2019).

Virus Isolation and Immunofluorescence Assay

Vero cells were used for viral isolation as previously described (Oka et al., 2014). Cells were inoculated with sterile supernatants of prepared samples, maintained in Dulbecco's modified Eagle's medium (DMEM; HyClone) with 10 μg/mL trypsin (Life Technologies) and monitored daily for cytopathic effect (CPE). All the PEDV isolates were passed in Vero cells for 10 generations, and the 10th-generation of plaque purified viruses were chosen for the following experiments.Then, the cells were fixed with anhydrous ethanol when 70% of cells showed CPE. Finally, an immunofluorescence assay (IFA) was performed with a mouse anti-N protein monoclonal antibody diluted 1:1,000, and with a fluorescence isothiocyanate (FITC)-conjugated goat anti-mouse secondary antibody (1:200 dilution), 4',6-diamidino-2-phenylindole (DAPI) was used for nucleic acids staining. The plates were examined using a fluorescence microscope.

Determination for the Viral Growth Curve

To determine the growth curve of PEDV strains of HNAY, HB, and HNXX, viruses were inoculated onto cell monolayers at MOI of 0.1. After adsorption at 37°C for 1 h, the cells were maintained with DMEM containing 10 μg/mL trypsin at 37°C with 5% CO_2. Cells were collected for virus titration every 12 h post infection. After centrifugation, the supernatants were collected and the median tissue culture infective dose per milliliter ($TCID_{50}$/mL)

TABLE 1 | PEDV strains used in this study.

Strain name	Countries	Time	Genotype	Accession no.	Strain name	Countries	Time	Genotype	Accession no.
AH2012	China	2012	G2a	KC210145	CHSD2014	China	2014	G2b	KX791060.1
CH ZMDZY 11	China	2011	G2a	KC196276.1	AH2012-12	China	2012	G2b	KU64683.1
Kansas125	USA	2014	G2a	KJ645701.1	ZL29	China	2015	S-INDEL	KU847996
KNU-1305	Korae	2013	G2a	KJ662670.1	KNU-1406-1	Korea	2014	S-INDEL	KM403155.1
KGS-1	Japan	2013	G2a	LC063814.1	KCH-2	Japan	2014	S-INDEL	LC063845.1
NIG-1	Japan	2014	G2a	LC063830.1	Indiana12.83	USA	2013	S-INDEL	KJ645635.1
MN	USA	2013	G2a	KF468752.1	Iow106	USA	2013	S-INDEL	KJ645695.1
MEX124	MEX	2014	G2a	KR206669	OH851	USA	2013	S-INDEL	KJ399978.1
EHM-1	Japan	2014	G2a	LC063812	Ohio126	USA	2014	S-INDEL	KJ645702.1
PEDV-Hjms	China	2016	G2a	KY007139.1	MYZ1	USA	2013	S-INDEL	LC063846
TC PC177-P2	USA	2013	G2a	KM392229.1	aDR13	Korea	2003	G1	JQ023162.1
TTR-2	Japan	2014	G2a	LC063828.1	CV777	Belgium	1978	G1	AF353511.1
CH JX-2013	China	2015	G2a	KJ526096.1	JS2008	China	2008	G1	KC109141.1
GDS28	China	None	G2a	MH726372.1	SM98	Korea	1998	G1	GU937797.1
PC22A	USA	2013	G2a	KY499262.1	vDR13	Korea	1999	G1	JQ023161.1
Colorado	USA	2013	G2a	KF272920.1	SD-M	China	2012	G1	JX560761
HNAY 2015	China	2015	G2a	KR809885.1	LZC	China	2006	G1	EF185992
HNZZ47	China	2016	G2a	KX981440.1	CHM	China	2013	G1	KM88144
Missouri270	USA	2015	G2a	KR265846.1	KPEDV9	Korea	2013	G1	KF898124.1
LC	China	2012	G2b	JX489155.1	HNXX[a]	China	2016	G2a	MT338517.1
AJ1102	China	2011	G2b	JX188454.1	HNAY[a]	China	2016	G2	MT338518.1
CH SCZG	China	2017	G2b	MH061337	HB[a]	China	2016	G2b	KY928065.1

[a] *The PEDV strains isolated in current study.*

was determined with a microtitration infection assay described by Reed and Muench (1938). Virus titration at different time points was performed in triplicates.

Piglet Challenge Experiment

To determine the virulence of PEDV isolates, 42 two-day-old piglets, which were confirmed negative for PEDV and other common viral pathogens in pigs by PCR or RT-PCR, were chosen for challenge experiment. Then they were randomized into four groups by weight with 10 pigs in each group, and housed in separate rooms. The infection groups were orally challenged with 0.5 mL of the 10th passage of plaque purified PEDV HNAY, HB, and HNXX, respectively, at 10^5 $TCID_{50}$/mL. The 10 piglets of control group were inoculated with equal volume of medium. All the animals were monitored daily for clinical signs. Rectal swabs were collected for scoring fecal denseness according to previous studies (Lin et al., 2015). The fecal viral RNA loads were quantified by real-time RT-qPCR (Su et al., 2018). The animal experiment was approved by Institutional Animal Care and Use Committee of Henan Academy of Agriculture Science (LLSC100085).

Histopathological Examination and Immunohistochemistry

Intestine and other major organs were collected and fixed for pathological examination. Samples of duodenum, jejunum, ileum, caecum, colon, rectum, and mesenteric lymph nodes were collected, and fixed in formalin. After 48 h, the jejunum was stained with haematoxylin and eosin (H&E) and the villous height-to-crypt depth (VH:CD) ratios were calculated according to previous study (Jung et al., 2014). Serial sections of ileums were evaluated for viral antigen by immunohistochemistry (IHC) using a mouse anti-N protein monoclonal antibody and biotinylated goat anti-mouse antibody as the secondary antibody with 3,3′-diaminobenzidine (DAB) as the chromogen (Lin et al., 2011).

Statistical Analysis

All data are expressed as means ± standard deviations (SD) and were analyzed with the GraphPad Prism software (GraphPad Software Inc.). Statistical analysis was performed using one-way ANOVA or two-way ANOVA followed by *t*-test. Differences were considered statistically significant at $P < 0.05$.

RESULTS

Viral Isolation and Identification

To determine causative agent associated with diarrhea in piglets, total RNA of clinical samples were extracted and subjected to RT-PCR. PEDV positive samples were processed and used for virus isolation on Vero cell as described previously (Chen et al., 2014). CPEs were observed on cells inoculated with PEDV positive samples, characterized by the rounding up, enlarging and detachment of cells as well as the formation of syncytia. In the end, three PEDV strains were successfully isolated, named HNAY, HNXX, and HB, respectively. The IFA results showed that PEDV specific fluorescence was observed in the inoculated cells,

but not in the non-infected cells (**Figure 1A**). The viral titres of three PEDV isolates were determined in Vero cells. The titers of HNXX and HB reached the maximum ($10^{4.6}$ $TCID_{50}$/mL and $10^{5.3}$ $TCID_{50}$/mL) at 24 hpi, whereas the peak virus titer of HNAY only reached $10^{3.8}TCID_{50}$/mL at 36 hpi (**Figure 1B**).

Sequence Analysis of the PEDVs

The genome of the isolates were sequenced and submitted to GenBank (accession numbers MT338517.1, MT338518.1, and KY928065.1). From nt identity analysis of the PEDV genome, HNAY exhibited 98.0–98.5% identity with the reference S-INDEL strains, 97.9–98.9% identity with the reference G2a strains, 98.0–98.5% identity with the reference G2b strains, and 96.4–97.6% identity with the reference G1 strains. HNAY shared 97.9 and 97.6% identity with HB (G2b), and HNXX (G2a), respectively (**Table 2**). From the nt identity analysis of PEDV S genes, HNAY exhibited 94–96.9% identity with the S-INDEL strains, 98–99.1% identity with the G2a strains, 97.5–98.5% identity with the G2b strains, and 93.7–94.3% identity with the G1 strains. However, HNAY shared 98.5 and 98.6% with HNXX, and HB, respectively (**Table 3**). From the amino acid (aa) identity analysis of PEDV S proteins, HNAY exhibited 92.9–96.2% identity with the S-INDEL

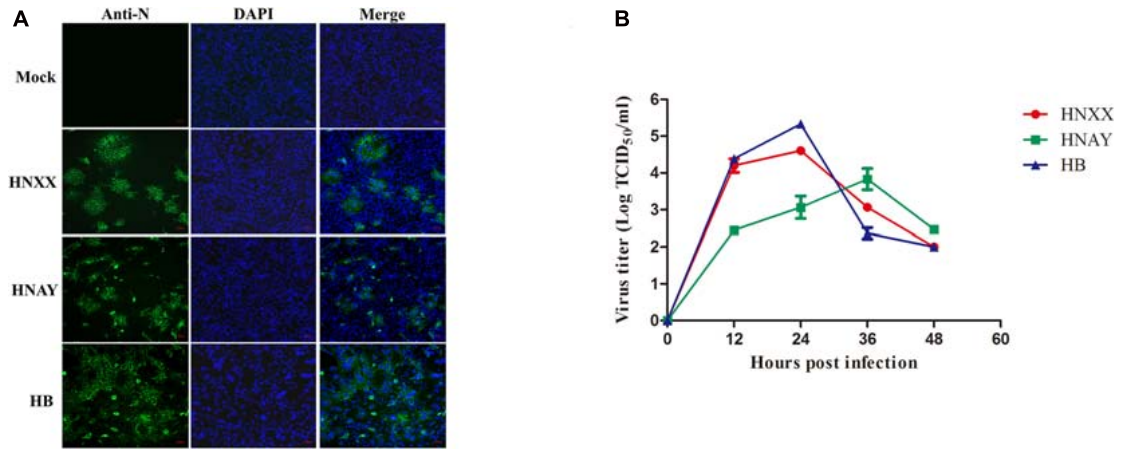


FIGURE 1 | PEDV isolates replicated in Vero cells. **(A)** Detection of PEDV infection in Vero cells by IFA. **(B)** Growth curves of three PEDV isolates at passage 5 in Vero cells. Data is presented as mean ± SD by triplicates.

TABLE 2 | Nucleotide sequence identity (%) of complete genome of different PEDV strains.

Strain name	HNXX	HNAY	HB	Genotype	Strain name	HNXX	HNAY	HB	Genotype
AH2012	98.6	98.8	98.0	G2a	AJ1102	98.2	98.4	98.2	G2b
CH ZMDZY11	98.8	98.8	97.9	G2a	LC	98.2	98.4	98.2	G2b
Kanasas125	98.5	98.6	97.8	G2a	ZL29	98.2	98.1	97.2	S-INDEL
KNU-1305	98.6	98.8	98.0	G2a	ohio126	98.1	98.3	97.4	S-INDEL
KGS-1	98.2	98.4	97.5	G2a	Indiana12.83	98.3	98.3	97.4	S-INDEL
NIG-1	98.2	98.5	97.5	G2a	low106	98.2	98.2	97.4	S-INDEL
MN	98.6	98.8	98.0	G2a	KNU-01406-1	98.3	98.2	97.5	S-INDEL
MEX124	98.5	98.6	97.8	G2a	KCH-2	97.8	98.1	97.1	S-INDEL
EHM-1	98.2	98.4	97.5	G2a	OH851	98.4	98.3	97.5	S-INDEL
PEDV-Hjms	98.3	98.4	97.7	G2a	MYZ1	98.7	98.4	97.7	S-INDEL
TC PC177-P2	98.6	98.9	98.0	G2a	CV777	96.4	96.5	96.4	G1
TTR-2	98.5	98.9	97.9	G2a	LZC	96.1	96.2	96.1	G1
CH JX-2013	98.5	98.8	98.0	G2a	VSM98	96.0	96.0	96.0	G1
CH SCZG	98.4	98.7	98.2	G2a	CHM	96.1	96.1	96.0	G1
HNZZ47	98.9	99.2	98.0	G2a	JS2008	96.6	96.7	96.7	G1
HNAY 2015	98.7	99.4	98.0	G2a	SD-M	96.5	96.6	96.6	G1
GDS28	98.7	99.1	98.1	G2a	vDR13	97.2	97.4	97.1	G1
PC22A	98.9	98.7	98.0	G2a	aDR13	96.5	96.5	96.6	G1
Colorado	98.8	98.6	98.0	G2a	HNXX	100	98.5	97.6	G2a
Missouri270	98.8	98.6	97.9	G2a	HNAY	98.5	100	97.9	G2
AN2012-12	97.7	97.9	98.2	G2b	HB	97.6	97.9	100	G2b
CHSD2014	98.3	98.5	98.2	G2b					

strains, 98.4–99% identity with the G2a strains, 97.7–98.1% identity with the G2b strains, and 92.6–94.4% identity with the G1 strains. However, HNAY shared 97.7 and 98.1% with HNXX, and HB, respectively (**Table 4**).

Interestingly, sequence analysis showed that HNAY strain had a novel 21-nt insertion in its S1 gene at 1,063–1,083 nt (**Figure 2A**). Besides, the inserted sequence (AGAAGAACAAAUCCAGAGCCA) were the same as the

TABLE 3 | Nucleotide sequence identity (%) of S genes of different PEDV strains.

Strain name	HNXX	HNAY	HB	Genotype	Strain name	HNXX	HNAY	HB	Genotype
KGS-1	98.8	98.9	98.1	G2a	CH SCZG	97.9	98.5	98.3	G2b
KNU-1305	98.8	98.9	98.1	G2a	AH2012-12	97.0	97.5	98.3	G2b
MIN	98.8	98.9	98.1	G2a	ZL29	96.6	96.9	95.5	S-INDEL
NIG-1	98.8	99.0	98.1	G2a	KNU-1406-1	96.2	96.7	95.8	S-INDEL
Kansas 125	98.8	98.9	98.0	G2a	KCH-2	96.2	96.8	95.8	S-INDEL
CHZMDZY11	98.6	98.8	98.0	G2a	OH851	96.2	96.8	95.8	S-INDEL
AH2012	98.3	98.5	97.7	G2a	Ohio126	96.2	96.8	95.8	S-INDEL
MEX 124	98.7	98.8	97.7	G2a	Indiana12.83	96.3	96.8	95.8	S-INDEL
EHM-1	98.8	98.9	97.8	G2a	Iowa 106	96.2	96.8	95.8	S-INDEL
PEDV-Hjms	97.9	98.4	97.5	G2a	SD-M	93.4	94.0	93.9	S-INDEL
CH JX	98.2	98.4	97.6	G2a	MYZ1	96.3	96.8	95.6	S-INDEL
TC PC177-P2	97.5	98.0	96.6	G2a	CHM	93.2	93.9	93.6	G1
TTR-2	97.8	98.2	96.8	G2a	SM98	93.2	93.8	93.8	G1
GDS28	98.5	99.3	98.1	G2a	vDR13	95.0	94.1	93.5	G1
PC22A	98.8	98.9	97.8	G2a	aDR13	93.5	94.1	94.0	G1
Colorado	98.7	98.8	97.7	G2a	CV777	93.5	94.2	94.1	G1
HNAY 2015	98.3	99.1	98.0	G2a	JS2008	93.7	94.3	94.2	G1
HNZZ47	98.5	98.7	97.5	G2a	LZC	93.0	93.7	93.4	G1
Missouri270	98.6	98.8	97.9	G2a	KPEDV9	93.3	93.9	93.6	G1
CHSD2014	97.4	97.6	98.0	G2b	HNXX	100	98.5	97.6	G2a
LC	97.3	97.7	98.1	G2b	HNAY	98.5	100	98.6	G2
AJ1102	97.3	97.6	98.1	G2b	HB	97.6	98.6	100	G2b

TABLE 4 | Amino acid identity (%) of S proteins of different PEDV strains.

Strain name	HNXX	HNAY	HB	Genotype	Strain name	HNXX	HNAY	HB	Genotype
KGS-1	98.1	98.8	97.9	G2a	CH SCZG	97.2	98.1	98.5	G2b
KNU-1305	98.1	98.8	97.9	G2a	AH2012-12	96.8	97.7	98.5	G2b
MIN	98.1	98.8	97.9	G2a	ZL29	95.5	96.2	95.3	S-INDEL
NIG-1	98.3	99.0	98.1	G2a	KNU-1406-1	95.2	95.9	95.1	S-INDEL
Kansas 125	98.1	98.6	97.7	G2a	KCH-2	95.3	96.0	95.2	S-INDEL
CHZMDZY11	97.6	98.4	97.4	G2a	OH851	95.2	96.0	95.1	S-INDEL
AH2012	98.1	98.5	97.5	G2a	Ohio126	95.4	96.1	95.3	S-INDEL
MEX 124	98.1	98.8	97.8	G2a	Indiana12.83	95.4	96.1	95.3	S-INDEL
EHM-1	98.2	98.9	98.0	G2a	Iowa 106	95.3	96.1	95.2	S-INDEL
PEDV-Hjms	97.5	98.3	97.3	G2a	SD-M	92.3	92.9	92.8	S-INDEL
CH JX	97.8	98.3	97.4	G2a	MYZ1	95.4	96.1	95.3	S-INDEL
TC PC177-P2	98.2	99.1	97.9	G2a	CHM	91.9	92.7	92.5	G1
TTR-2	97.9	98.7	97.6	G2a	SM98	91.9	92.7	92.5	G1
GDS28	97.6	98.7	97.9	G2a	vDR13	93.6	94.4	94.4	G1
PC22A	98.1	98.8	97.9	G2a	aDR13	92.6	93.3	93.2	G1
Colorado	98.1	98.8	98.0	G2a	CV777	92.8	93.6	93.3	G1
HNAY 2015	97.7	99.0	98.1	G2a	JS2008	92.5	93.5	93.1	G1
HNZZ47	97.3	98.0	97.2	G2a	LZC	91.8	92.6	92.3	G1
Missouri270	98.1	98.8	98.0	G2a	KPEDV9	92.9	93.5	93.1	G1
CHSD2014	97.0	97.8	98.3	G2b	HNXX	100	97.7	96.9	G2a
LC	97.3	98.1	98.6	G2b	HNAY	97.7	100	98.1	G2
AJ1102	97.3	98.1	98.6	G2b	HB	96.9	98.1	100	G2b

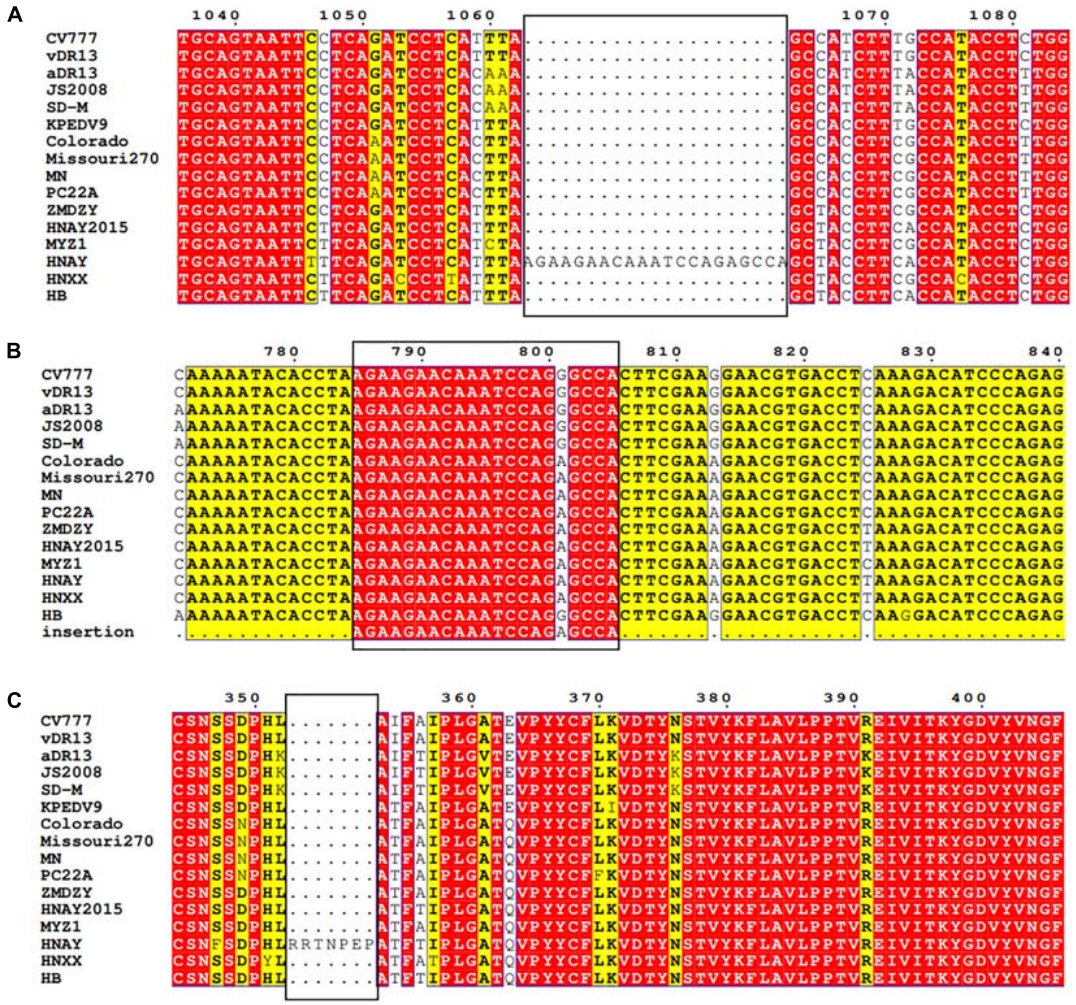


FIGURE 2 | Sequence analysis of S gene of PEDV HNAY with representative PEDV strains. **(A)** Nucleotide alignment of partial S gene of PEDV HNAY with different types of reference strains. The novel insertion (AGAAGAACAAATCCAGAGCCA) in HNAY S gene is depicted as a black box. **(B)** Nucleotide alignment of partial N gene of PEDV HNAY with different types of reference strains. The black box indicates the position of 21 nucleotides inserted into the S gene in N gene. **(C)** Amino acid alignments of the partial S1 protein, insertions of seven amino acids are depicted as black boxes.

sequence of its N gene at 785–805 nt (**Figure 2B**). Insertion of the deduced 7 aa was located at the 358–364 aa of PEDV S1 protein (**Figure 2C**). Besides, HNAY S protein structure was predicted by SWISS-MODEL according to the structure of PEDV USA/Colorado/2013 strain in PDB database (accession code 6VV5; **Figure 3**). Structure prediction showed that the 7-aa insertion might be located at the flexible loop (Leu354 to Ala363) at the apex of the trimer of PEDV S1A domain (Wrapp and McLellan, 2019).

To date, three neutralizing epitopes of the PEDV S protein have been reported, including core neutralizing epitope (COE), SS2 and SS6 (Chiou et al., 2017). Multiple alignments revealed that similar to other variant strains, compared to CV777 strain, HNAY S protein possessed nine aa substitutions (A517S, S523G, V527I, T549S, G594S, A605E, L612F, I635V, and Y766S) in the COE and one amino acid substitution (Y766S) in the SS6 (**Figure 4**). Compared with CH/HNAY/2015 strain which was isolated in the same region as HNAY, two additional aa substitutions (D520G and H521Y) in the COE were observed in the HNAY strain. In addition to these shared mutations, HNAY strain have two unique mutations (T500A and H521Y) in the COE that were absent in other PEDV variant strains and CV777-based vaccine strains.

Phylogenetic and Recombination Analysis

The phylogenetic tree based on the sequences of 3 PEDV isolates along with 33 PEDV reference strains revealed that they can be clustered into two groups, classical G1 and variant G2 (**Figure 5A**). G1 contains eight isolates: CV777, SM98, and DR13 isolated from Korea, and SD-M, LZC, CHM, and JS2008 from China. G2 includes many virulent strains from Japan, Korea, China, and the US. HNAY, HNXX, and HB were grouped

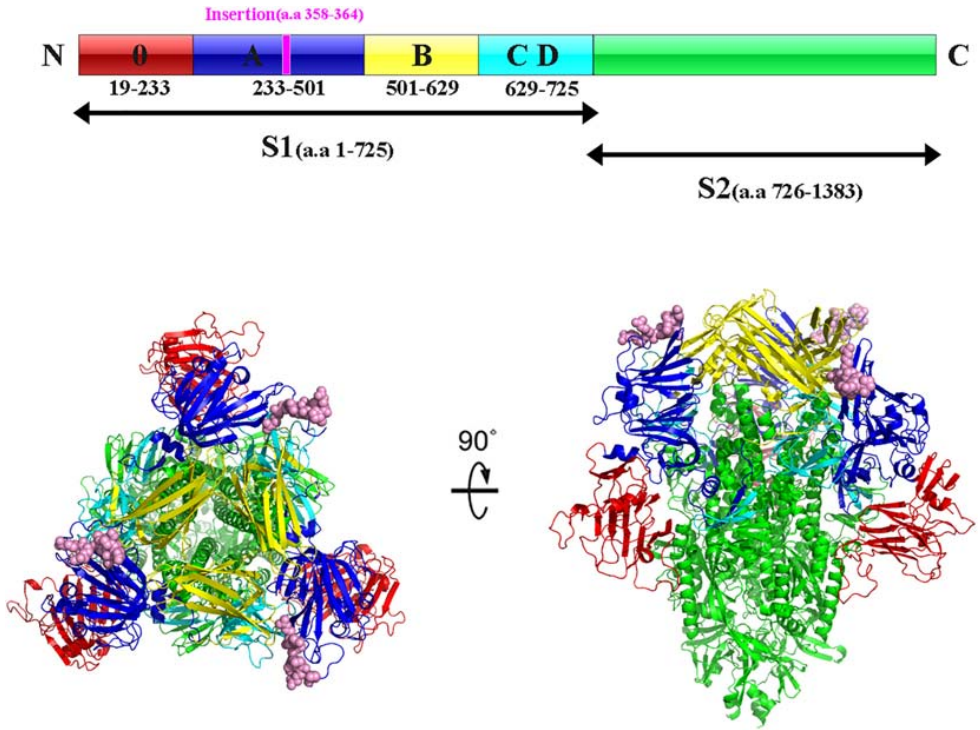


FIGURE 3 | Structure prediction of S protein of PEDV HNAY in the prefusion conformation. The primary sequence of PEDV S protein can be divided into S1 and S2. The S1 region can be further divided into S10, S1A, S1B, and S1CD (in red, admiral blue, yellow, and arctic blue, respectively). The 7-aa insertion was shown in purple within S1A domain. Viewing the trimeric PEDV HNAY S protein from the membrane distal apex and a 90° view based on the crystal structure of S protein of USA/Colorado/2013. S10, S1A, S1B, S1CD, and S2 were shown in red, admiral blue, yellow, and arctic blue, respectively, and the 7-AA insertion was in pink in the crystal structure model.

in subgroup G2 as most of other isolates from China in the past 10 years. Furthermore, phylogenetic analysis of complete S genes delineated G2 group into 2a, 2b, and S-INDEL subgroups. HNXX and HB belongs to G2a and G2b, respectively, which was consistent with the phylogenetic analysis based on their complete genomes. Interestingly, HNAY strain belongs to a separate branch other than G2a and G2b (**Figure 5B**).

To further analyze the association between PEDV HNAY and reference strains, a recombination analysis was performed by using RDP4 software (Martin et al., 2015). As shown in **Figure 6**, PEDV HNAY might arise by the recombination of HNZZ47 and GDS28 strains, which was supported by 6 programs (**Figure 6**). The major parent strain of the recombination might be HNZZ47 isolated from Henan, May 2016; the minor parent strain was GDS28 which was isolated in Guangdong, December 2012. Furthermore, we also identified potential breakpoints for recombination in the ORF1b and S region (nt 17,205–21,832). This suggests that HNAY might be derived from a natural recombination between two Chinese variant PEDV strains.

Pathogenicity of PEDV Isolates in New-Born Piglets

To investigate the virulence of PEDV HNAY in piglets in comparison with that of HNXX and HB, suckling piglets were orally infected with the three PEDV strains. Clinical observation showed that piglets in challenge groups exhibited lethargy, loss of appetite, diarrhea, dehydration and vomiting at 12–20 hpi. The control pigs remained healthy. The challenged piglets began to die at 1 day post infection (dpi), and all piglets in the HNAY-, HB- and HNXX-challenged groups died within 2, 3, and 4 dpi, respectively (**Figure 7A**). From fecal score analysis, there was no significant difference between piglets in three groups (**Figure 7B**). Necropsy examinations were performed immediately after the death of the infected piglets. All the infected piglets displayed typical lesions of PED. The wall of the small intestine was thin and transparent. The small intestine and stomach were, respectively, distended and filled with curdled and undigested milk. By contrast, the intestinal organs of control piglets appeared grossly normal. Furthermore, viral RNA in fecal samples were tested positive for PEDV in challenged groups (2–6$\log_{10}$ genomic equivalents/mL). HNXX and HB reached peak at 1 dpi, but HNAY reached peak at 2 dpi (**Figure 7C**).

Microscopic examination of jejunum revealed moderate to severe, extensive, atrophic enteritis in three PEDV isolates infected piglets (**Figure 7D**). The jejunum VH:CD ratios ranged from 1.14 to 2.49 in PEDVs infected piglets. The VH:CD ratios in the three PEDVs infected piglets were lower when compared with that of non-infected pigs, but there was no obvious difference between three PEDVs infected groups (**Figure 7E**).

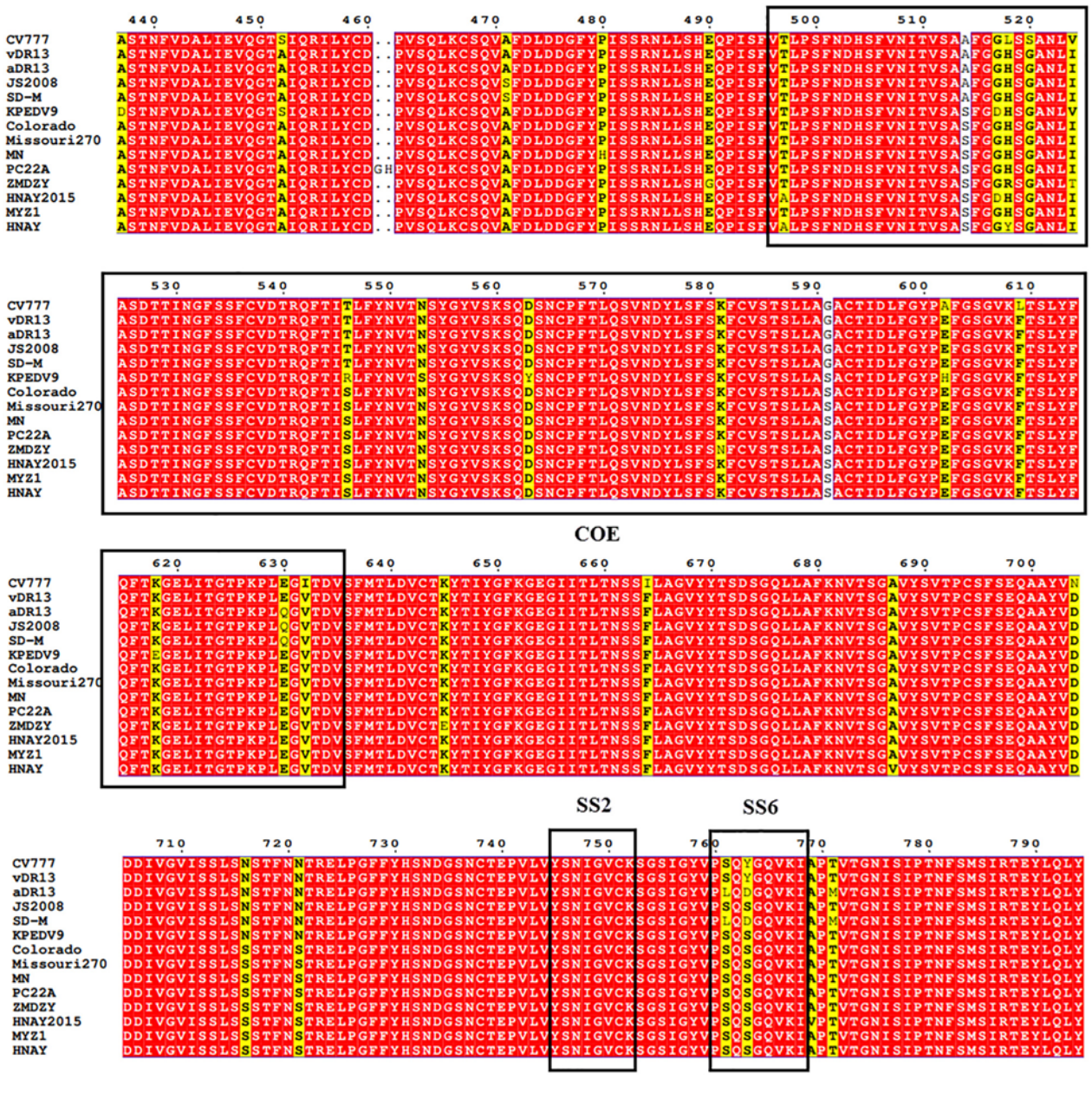


FIGURE 4 | The neutralizing epitopes analysis of S protein of PEDV HNAY with reference strains. The positions of COE, SS2, and SS6 are shown in black box.

In addition, PEDV antigens were detected in the duodenum, jejunum, ileum, caecum, colon, rectum and MLNs of HNXX-, HB-, and HNAY-infected piglets. IHC results showed that PEDV antigens were mainly distributed in the villus epithelial cells. Specifically, viral antigen was expressed in the colon and cecum but mainly distributed in the mucosa and submucosa of the HNXX-infected piglets; viral antigen of HB-infected piglets was expressed in the mucosal epithelium of each segment of the intestine. While the tissues from the piglets in the HNAY-infected group showed remarkable levels of viral antigen in the intestinal glands and mucosal epithelium of the jejunum, and the mucosal layer of the colon and rectum (**Figure 8**). The results of PEDV challenge test indicate that HNAY displayed higher pathogenicity compared with two other clinical isolates.

DISCUSSION

PED is one of the most critical diarrheal diseases threatening the global pig industry. In 2010, a new variant PEDV (G2 genotype) first emerged in Central China and spread rapidly throughout the country, resulting in big economic losses (Li et al., 2012). Subsequently, this type of virus spread to many countries in Asia, America and Europe (Peng et al., 2017). In 2014, another

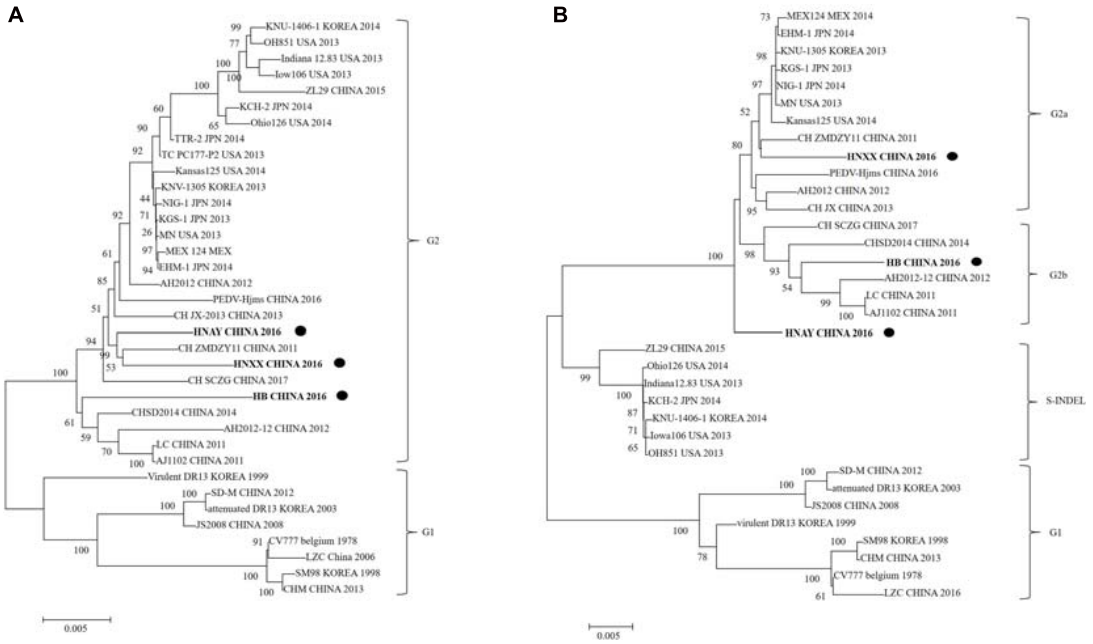


FIGURE 5 | Phylogenetic analysis of PEDV complete genome and S genes. **(A)** Multiple sequence alignments were generated with Clustal X and complete genome phylogenetic tree was constructed using neighbor joining (NJ) method and supported with a bootstrap test of 1,000 replicates in MEGA7. "●" indicates the PEDV strains isolated in this study. **(B)** S gene phylogenetic tree was constructed by using the same method as above. "●" indicates the PEDV strains isolated in this study. Scale bars indicate nucleotide substitutions per site.

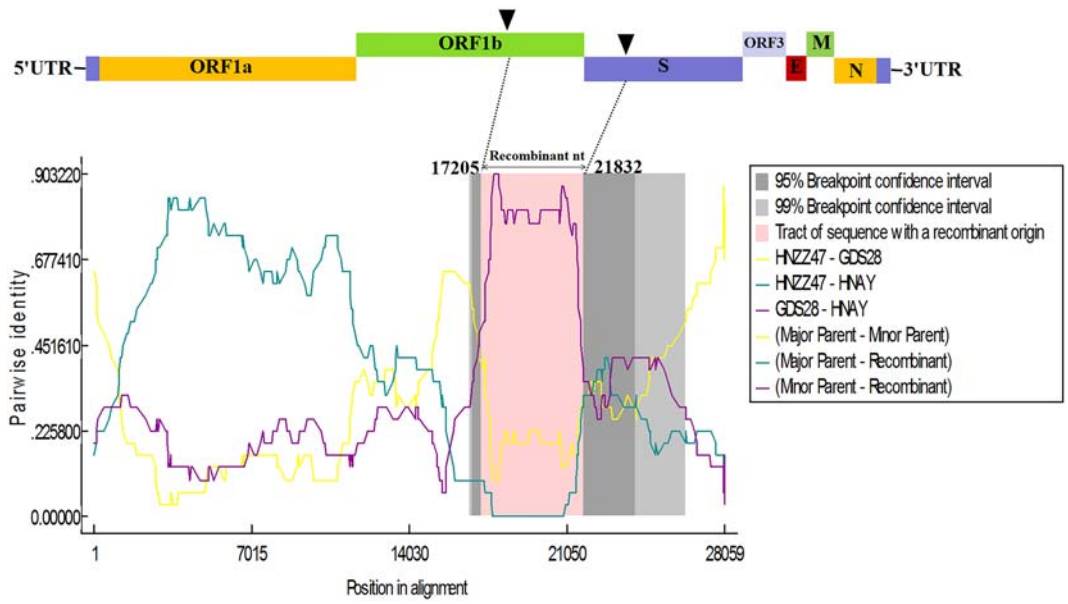


FIGURE 6 | Recombination analysis of PEDV HNAY with indicated PEDV strains. The result was described using RDP method which was supported by ≥6 programs to further characterize the potential recombination events. The black arrow indicates the regions where recombination event may occur.

genotype of PEDV, designated as S-INDEL, was first detected in the USA (Wang et al., 2014). Since then, more and more variant PEDV strains, such as those with a large deletion in the S gene and deletions in the ORF1a gene, have been discovered worldwide (Lin et al., 2016). The S glycoprotein of coronaviruses contains two domains called S1 and S2. S1 domain contributes to receptor binding and S2 domain is thought to be associated with direct fusion between the viral and cellular membranes (Walls et al., 2016). Similar to other coronavirus S proteins, the PEDV S protein plays a critical role in viral entry and production of

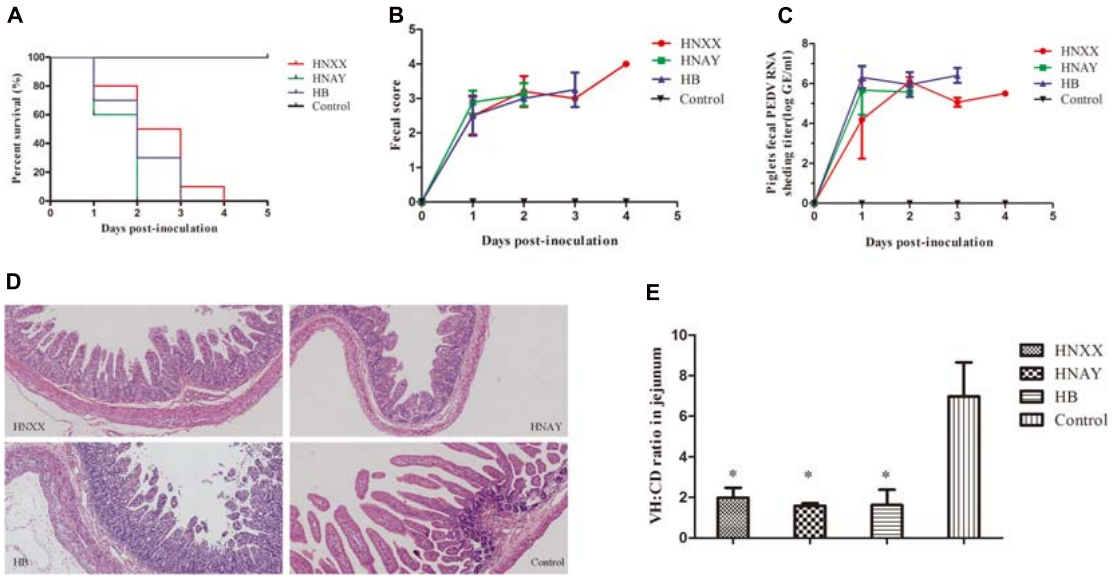


FIGURE 7 | Pathogenicity analysis of PEDV isolates in new-born piglets. **(A)** Survival rate of piglets in each group. **(B)** The severity of diarrhea was scored based on clinical examination; 0 = normal and no diarrhea; 1, mild and fluidic diarrhea; 2, severe watery diarrhea; with scores of 1 or more considered diarrheic. **(C)** Fecal virus shedding in PEDV-challenged piglets by RT-qPCR. **(D)** Microscopic damage in Jejunum in piglets (Original magnifications: ×20). **(E)** Further analysis of jejunum villous high: crypt depth ratios (VH:CD).

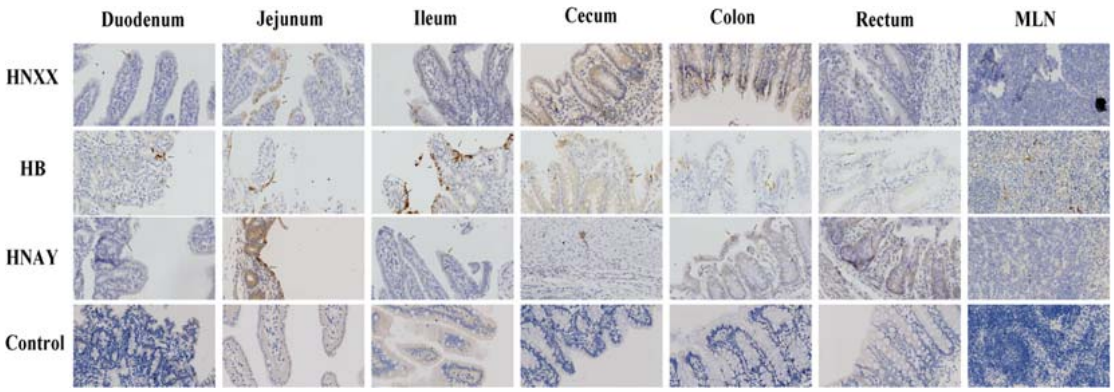


FIGURE 8 | PEDV antigens detection in different intestinal tissues of piglets infected with PEDV isolates by immunohistochemistry. The representative pictures of viral antigens in duodenum, jejunum, ileum, caecum, colon, rectum, mesenteric lymph nodes of intestines from piglets of each group are shown, and antigens are depicted as arrows. Original magnifications: ×40.

neutralization antibodies in the host. Most of the variant PEDV strains possessed insertion and deletions in the S0 domain (19–233 aa; Oka et al., 2014; Vlasova et al., 2014). Interestingly, HNAY was identified as a novel PEDV strain with 7-aa insertion (358–364 aa) in S1A domain (233–501 aa), and the inserted sequence was identical with the sequence of 785–805 nt in N gene. Although HNAY had higher sequence identity with G2a and G2b strains, phylogenetic analysis revealed that HNAY belonged to a single branch instead of G1, G2 or S-INDEL. In addition, a recombination event was identified in the genome of HNAY at 27,105–28,059 nt, where PEDV HNZZ47 and GDS28 were major and minor parents, respectively. Currently, none of clinically isolated PEDV strain has been reported to have multiple amino acid insertions in the S1A region. The prevalence of this type of variants in pig farms needs further epidemiological investigation.

Mutations in the S1 region of PEDV have been shown to be responsible for the alternation of viral tropism and pathogenicity in pigs (Gallagher and Buchmeier, 2001; Sato et al., 2011; Lee, 2015). In this study, HNAY-infected Vero cells were obviously characterized by cell fusion and syncytial formation, but viral titre was lower compared to that of the other two PEDV strains. Animal experiments showed that pigs that infected by HNAY showed most severe disease compared to those infected by HNXX and HB, as all of HNAY-infected piglets died within 2 dpi. Moreover, a partial of the PEDV N antigen was also found in caecum muscularis of HNAY-infected pigs, but not in other

virus strains infected groups. Considering the particularity of this insertion position, further studies are required to decipher its role in receptor binding and virulence in the PEDV strain.

Recently, the atomic-resolution structure of prefusion PEDV spike protein has been resolved (Wrapp and McLellan, 2019). Similar to other class I fusion proteins in the prefusion conformation, PEDV S1 subunit is made up of a series of β-sheets, whereas the S2 subunit is almost composed of a series of discontinuous α-helices (Wrapp and McLellan, 2019). Interestingly, the 7-aa insertion happens to be in the flexible loop at the apex of the trimer (354–363 aa). However, this region was not clearly resolved in structure model of PEDV S protein (Wrapp and McLellan, 2019). Therefore, we predicted that this region may be an ideal insertion site for foreign proteins or peptides, which could provide clues for constructing recombinant PEDV strains with molecular markers in S1 domain by reverse genetics technology. Whether the insertion in this region affects the infectious characteristics and pathogenicity of PEDV needs further study.

CONCLUSION

In this study, a novel 7-aa insertion in the S1 and a recombination event were detected in PEDV HNAY strain.This insertion might affect the structure of S protein, thereby directly or indirectly altering viral tropism and pathogenicity, resulting in consistent infection in pigs. Therefore, further research is needed to determine the impact of this insertion on the pathogenicity of HNAY strain. In addition, how such PEDV variants emerged and evolved in the field needs further investigations, which would significantly contribute to the prevention and control of PED outbreaks worldwide.

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 Animal Experiment Committee of Henan Academy of Agricultural Sciences.

AUTHOR CONTRIBUTIONS

DL, YL, and GZ wrote the manuscript and conceived and initiated the study. DL, YL, and YC devised the experimental methods. WJ and HF curated the data. JW and YZ performed animal experiments. DL prepared the original manuscript draft. YL and GZ reviewed the manuscript and edited it. All authors read and approved the final manuscript.

FUNDING

This work was supported by Grants from the National Key R&D Program (2016YFD0500704 and 2017YFD0501103), the National Natural Science Funds of China (31902281), the Science-Technology Foundation for Outstanding Young Scientists of Henan Academy of Agricultural Sciences (2020YQ25), and the State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, and the Chinese Academy of Agricultural Sciences (SKLVEB2020KFKT019).

ACKNOWLEDGMENTS

We thank to Dr. Wentao Li of Huazhong Agricultural University for the revision of this manuscript.

REFERENCES

Bi, J., Zeng, S., Xiao, S., Chen, H., and Fang, L. (2012). Complete genome sequence of porcine epidemic diarrhea virus strain AJ1102 isolated from a suckling piglet with acute diarrhea in China. *J. Virol.* 86, 10910–10911. doi: 10.1128/JVI.01919-12

Chen, Q., Li, G., Stasko, J., Thomas, J. T., Stensland, W. R., Pillatzki, A. E., et al. (2014). Isolation and characterization of porcine epidemic diarrhea viruses associated with the 2013 disease outbreak among swine in the United States. *J. Clin. Microbiol.* 52, 234–243. doi: 10.1128/JCM.02820-13

Chiou, H. Y., Huang, Y. L., Deng, M. C., Chang, C. Y., Jeng, C. R., Tsai, P. S., et al. (2017). Phylogenetic analysis of the spike (S) gene of the new variants of porcine epidemic diarrhoea virus in taiwan. *Transbound Emerg. Dis.* 64, 157–166. doi: 10.1111/tbed.12357

Durham, P. J., Stevenson, B. J., and Farquharson, B. C. (1979). Rotavirus and coronavirus associated diarrhoea in domestic animals. *N. Z. Vet. J.* 27, 30–32. doi: 10.1080/00480169.1979.34595

Gallagher, T. M., and Buchmeier, M. J. (2001). Coronavirus spike proteins in viral entry and pathogenesis. *Virology* 279, 371–374. doi: 10.1006/viro.2000.0757

Hasoksuz, M., Alekseev, K., Vlasova, A., Zhang, X., Spiro, D., Halpin, R., et al. (2007). Biologic, antigenic, and full-length genomic characterization of a bovine-like coronavirus isolated from a giraffe. *J. Virol.* 81, 4981–4990. doi: 10.1128/JVI.02361-06

Huang, Y. W., Dickerman, A. W., Piñeyro, P., Li, L., Fang, L., Kiehne, R., et al. (2013). Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. *mBio* 4, e00737–13. doi: 10.1128/mBio.00737-13

Jung, K., Wang, Q., Scheuer, K. A., Lu, Z., Zhang, Y., and Saif, L. J. (2014). Pathology of US *porcine epidemic diarrhea virus* strain PC21A in gnotobiotic pigs. *Emerg. Infect. Dis.* 20, 662–665. doi: 10.3201/eid2004.131685

Law, H. K., Cheung, C. Y., Ng, H. Y., Sia, S. F., Chan, Y. O., Luk, W., et al. (2005). Chemokine up-regulation in SARS-coronavirus-infected, monocyte-derived human dendritic cells. *Blood* 106, 2366–2374. doi: 10.1182/blood-2004-10-4166

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, D., Feng, H., Liu, Y., Chen, Y., Wei, Q., Wang, J., et al. (2018). Molecular evolution of porcine epidemic diarrhea virus and porcine deltacoronavirus strains in Central China. *Res. Vet. Sci.* 120, 63–69. doi: 10.1016/j.rvsc.2018.06.001

Li, W., Li, H., Liu, Y., Pan, Y., Deng, F., Song, Y., et al. (2012). New variants of porcine epidemic diarrhea virus, China, 2011. *Emerg. Infect. Dis.* 18, 1350–1353. doi: 10.3201/eid1808.120002

Lin, C. M., Annamalai, T., Liu, X., Gao, X., Lu, Z., El-Tholoth, M., et al. (2015). Experimental infection of a US spike-insertion deletion porcine epidemic diarrhea virus in conventional nursing piglets and cross-protection to the original US PEDV infection. *Vet. Res.* 46:134. doi: 10.1186/s13567-015-0278-9

Lin, C. M., Jeng, C. R., Hsiao, S. H., Liu, J. P., Chang, C. C., Chiou, M. T., et al. (2011). Immunopathological characterization of porcine circovirus type 2 infection-associated follicular changes in inguinal lymph nodes using high-throughput tissue microarray. *Vet. Microbiol.* 149, 72–84. doi: 10.1016/j.vetmic.2010.10.018

Lin, C. M., Saif, L. J., Marthaler, D., and Marthaler, D. (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

Lorusso, A., Decaro, N., Schellen, P., Rottier, P. J. M., Buonavoglia, C., Haijema, B. J., et al. (2008). Gain, preservation, and loss of a group 1a coronavirus accessory glycoprotein. *J. Virol.* 82, 10312–10317. doi: 10.1128/JVI.01031-08

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003

Oka, T., Saif, L. J., Marthaler, D., Esseili, M. A., Meulia, T., Lin, C. M., et al. (2014). Cell culture isolation and sequence analysis of genetically diverse US porcine epidemic diarrhea virus strains including a novel strain with a large deletion in the spike gene. *Vet. Microbiol.* 173, 258–269. doi: 10.1016/j.vetmic.2014.08.012

Pan, Y., Saif, L. J., Marthaler, D., Esseili, M. A., Meulia, T., Lin, C. M., 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

Park, S., Kim, S., Song, D., and Park, B. (2014). Novel porcine epidemic diarrhea virus variant with large genomic deletion, South Korea. *Emerg. Infect Dis.* 20, 2089–2092. doi: 10.3201/eid2012.131642

Pasick, J., Berhane, Y., Ojkic, D., Maxie, G., Embury−Hyatt, C., Swekla, K., et al. (2014). Investigation into the role of potentially contaminated feed as a source of the first-detected outbreaks of porcine epidemic diarrhea in Canada. *Transbound Emerg. Dis.* 61, 397–410. doi: 10.1111/tbed.12269

Peng, J. Y., Jian, C. Z., Chang, C. Y., and Chang, H. W. (2017). "Porcine epidemic diarrhea," in *Emerging and Re-Emerging Infectious Diseases of Livestock*, ed. J. Bayry (Cham: Springer), doi: 10.1007/978-3-319-47426-7

Pensaert, M. B., and de Bouck, P. (1978). A new coronavirus-like particle associated with diarrhea in swine. *Arch. Virol.* 58, 243–247. doi: 10.1007/BF01317606

Reed, L. J., and Muench, H. (1938). A simple method of estimating fifty percent endpoints. *Am. J. Hyg.* 1938, 493–497. doi: 10.1093/oxfordjournals.aje.a118408

Sato, T., Takeyama, N., Katsumata, A., Tuchiya, K., Kodama, T., and Kusanagi, K. (2011). Mutations in the spike gene of porcine epidemic diarrhea virus associated with growth adaptation in vitro and attenuation of virulence in vivo. *Virus Genes* 43, 72–78. doi: 10.1007/s11262-011-0617-5

Song, D., and Park, B. (2012). Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus Genes* 44, 167–175. doi: 10.1007/s11262-012-0713-1

Stevenson, G. W., Hoang, H., Schwartz, K. J., Burrough, E. R., Sun, D., Madson, D., et al. (2013). Emergence of Porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. *J. Vet. Diagn. Invest.* 25, 649–654. doi: 10.1177/1040638713501675

Su, Y., Liu, Y., Chen, Y., Xing, G., Hao, H., Wei, Q., et al. (2018). A novel duplex TaqMan probe-based real-time RT-qPCR for detecting and differentiating classical and variant porcine epidemic diarrhea viruses. *Mol. Cell Probes.* 37, 6–11. doi: 10.1016/j.mcp.2017.10.003

Su, Y., Liu, Y., Chen, Y., Zhao, B., Ji, P., Xing, G., et al. (2016). Detection and phylogenetic analysis of porcine epidemic diarrhea virus in central China based on the ORF3 gene and the S1 gene. *Virol. J.* 13:192. doi: 10.1186/s12985-016-0646-8

Sun, D., Wang, X., Wei, S., Chen, J., and Feng, L. (2016). Epidemiology and vaccine of porcine epidemic diarrhea virus in China: a mini-review. *J. Vet. Med. Sci.* 78, 355–363. doi: 10.1292/jvms.15-0446

Vaughn, E. M., Halbur, P. G., and Paul, P. S. (1995). Sequence comparison of porcine respiratory coronavirus isolates reveals heterogeneity in the S, 3, and 3-1 genes. *J. Virol.* 69, 3176–3184. doi: 10.1128/JVI.69.5.3176-3184.1995

Vennema, H., Poland, A., Foley, J., and Pedersen, N. C. (1998). Feline infectious peritonitis viruses arise by mutation from endemic feline enteric coronaviruses. *Virology* 243, 150–157. doi: 10.1006/viro.1998.9045

Vijgen, L., Keyaerts, E., Moës, E., Thoelen, I., Wollants, E., Lemey, P., et al. (2005). Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J. Virol.* 79, 1595–1604. doi: 10.1128/JVI.79.3.1595-1604.2005

Vlasova, A. N., Marthaler, D., Wang, Q., Culhane, M. R., Rossow, K., Rovira, A., et al. (2014). Distinct characteristics and complex evolution of PEDV strains, North America, May 2013-February 2014. *Emerg. Infect. Dis.* 20, 1620–1628. doi: 10.3201/eid2010.140491

Walls, A. C., Tortorici, M. A., Jan Bosch, B., Frenz, B., Rottier, P. J. M., DiMaio, F., et al. (2016). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* 531, 114–117. doi: 10.1038/nature16988

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, L., Byrum, B., and Zhang, Y. (2014). New variant of porcine epidemic diarrhea virus, United States, 2014. *Emerg. Infect. Dis.* 20, 917–919. doi: 10.3201/eid2005.140195

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

Wang, X. M., Niu, B. B., Yan, H., Gao, D. S., Yang, X., Chen, L., et al. (2013). Genetic properties of endemic Chinese porcine epidemic diarrhea virus strains isolated since 2010. *Arch. Virol.* 158, 2487–2494. doi: 10.1007/s00705-013-1767-7

Wrapp, D., and McLellan, J. S. (2019). The 3.1-angstrom cryo-electron microscopy structure of the porcine epidemic diarrhea virus spike protein in the prefusion conformation. *J. Virol.* 93:e00923-19. doi: 10.1128/JVI.00923-19

Zhao, X., Li, Z., Zeng, X., Zhang, G., Niu, J., Sun, B., et al. (2017). Sequence analysis of the spike gene of *Porcine epidemic diarrhea virus* isolated from South China during 2011-2015. *J. Vet. Sci.* 18, 237–243. doi: 10.4142/jvs.2017.18.2.237

Zhu, T., Du, S., Cao, D., Pei, Z., Guo, Y., Shao, H., et al. (2019). Isolation and identification of a variant subtype G 2b porcine epidemic diarrhea virus and S gene sequence characteristic. *Infect. Genet. Evol.* 71, 82–90. doi: 10.1016/j.meegid.2019.03.015

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.

frontiers in Microbiology

MINI REVIEW
published: 06 May 2021
doi: 10.3389/fmicb.2021.655065

Contributions of Genetic Evolution to Zika Virus Emergence

*Su-Jhen Hung and Sheng-Wen Huang**

National Mosquito-Borne Diseases Control Research Center, National Health Research Institutes, Tainan, Taiwan

Mosquito-borne Zika virus (ZIKV) was considered an obscure virus causing only mild or self-limited symptoms until the explosive outbreaks in French Polynesia in 2013–2014 and in the Americas in 2015–2016, resulting in more than 700,000 cases of the disease, with occasional miscarriage and severe congenital birth defects, such as intrauterine growth restriction, fetal microcephaly, and other neurodevelopmental malformations. In this review, we summarized the evolution of ZIKV from a mundane virus to an epidemic virus. ZIKV has acquired a panel of amino acid substitutions during evolution when the virus spread from Africa, Asia, Pacific, through to the Americas. Robust occurrence of mutations in the evolution of ZIKV has increased its epidemic potential. Here we discussed the contributions of these evolutionary mutations to the enhancement of viral pathogenicity and host-mosquito transmission. We further explored the potential hypotheses for the increase in ZIKV activity in recent decades. Through this review, we also explored the hypotheses for the occurrence of the recent ZIKV epidemics and highlighted the potential roles of various factors including pathogen-, host-, vector-related, and environmental factors, which may have synergistically contributed to the ZIKV epidemics.

Keywords: Zika virus, evolution, virulence, mutation, emergence

OPEN ACCESS

Edited by:
Justin Jang Hann Chu, National University of Singapore, Singapore

Reviewed by:
Hongjie Xia, University of Texas Medical Branch at Galveston, United States
Sujan Shresta, La Jolla Institute for Immunology, United States

***Correspondence:**
Sheng-Wen Huang
joehuang@nhri.edu.tw

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *18 January 2021*
Accepted: *12 April 2021*
Published: *06 May 2021*

Citation:
Hung S-J and Huang S-W (2021) Contributions of Genetic Evolution to Zika Virus Emergence. Front. Microbiol. 12:655065. doi: 10.3389/fmicb.2021.655065

INTRODUCTION

Zika virus (ZIKV) belongs to the genus *Flavivirus* together with other important mosquito-borne human viruses, such as dengue virus (DENV), West Nile virus, Japanese encephalitis virus, and yellow fever virus. The flaviviruses have single positive-stranded 11-kb RNA genomes that the 5' and 3' untranslated regions flank a polyprotein coding region encoding three structural proteins [capsid, pre/membrane (prM), and envelope (E)] and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). The main mosquito vector during epidemics belongs to the genus *Aedes* (*Aedes aegypti* and *A. albopictus*) (Diallo et al., 2014; Guo et al., 2016; Epelboin et al., 2017; Rossi et al., 2018). ZIKV has continuously adapted to mosquitoes and non-human primates in a sylvatic cycle, which then results in a virus reservoir. When ZIKV opportunistically enters the human transmission cycle from the sylvatic cycle, the virus may initiate epidemics in humans. In addition to the *Aedes*-human transmission cycle, non-vectored transmission between humans has been hypothesized because the ZIKV genome can be found in saliva (Musso et al., 2015), urine (Gourinat et al., 2015), and even tears (Miner et al., 2016).

Zika virus was first isolated from non-human primates in the Zika forest of Uganda in 1947 (Dick et al., 1952), and in humans in 1954 (Macnamara, 1954). Since then, it has been continuously isolated in Uganda and Nigeria (Simpson, 1964; Moore et al., 1975), and only sporadic cases

with self-limited signs or mild symptoms were reported in Africa (Macnamara, 1954; Simpson, 1964; Moore et al., 1975; Fagbami, 1979) and South-east Asia (Olson and Ksiazek, 1981) until the first major outbreak, evidenced by serology and virology of patients with rash, fever, arthralgia, and conjunctivitis, occurred in Yap Island, Micronesia in 2007 (Duffy et al., 2009). Subsequently, ZIKV resumed low activity, and only a few countries including Cambodia (in 2008, 2009, and 2010), Nigeria (2011), Indonesia (2012), Russia (2012), and the Philippines (2012), reported ZIKV isolations from 2008 to 2012. Afterward, ZIKV caused another larger outbreak in French Polynesia in 2013–2014 (Cao-Lormeau et al., 2014), in which more than 30,000 individuals were observed for suspected infection. During this outbreak, ZIKV infections were first associated with patients with Guillain-Barré Syndrome (GBS), the neuropathic condition characterized by progressive weakness and diminished or absent myotatic reflexes (Walling and Dickson, 2013; Cao-Lormeau et al., 2016). Thereafter, ZIKV spread to several other Pacific islands, including Easter Island, Cook Island, Solomon Islands, New Caledonia, Vanuatu, Tonga, Fuji, Samoa, and American Samoa (Delatorre et al., 2018), and then to South American mainland. In May 2015, Brazil-originating ZIKV dissemination was reported (Zanluca et al., 2015), which then caused an explosive outbreak in September 2015. This outbreak casually linked ZIKV infection and birth defects or disabilities, such as severe microcephaly; the condition was termed congenital Zika syndrome (CZS). Simultaneously, ZIKV quickly spread from Brazil to the Caribbean and Central America and was eventually imported into the United States (Grubaugh et al., 2017). In 2016, at least 175,000 laboratory-confirmed ZIKV cases with more suspected cases from 48 countries and territories were reported in the Americas. Among these ZIKV-infected cases, World Health Organization reported 2,654 cases with CZS by February 2017, and 2,366 cases of these were in Brazil (World Health Organization, 2017). Although the prognosis of microcephaly needs to be more clearly examined, the patients with CZS usually have physical and learning disabilities. Therefore, CZS is a burden to families and societies. During the America epidemics, ZIKV strains continuously spread in Asia and caused an outbreak with 455 confirmed cases in Singapore in 2016 (Ho et al., 2017). Additionally, ZIKV locally disseminated and caused sporadic cases in the Philippines (Lim et al., 2017), Vietnam (Moi et al., 2017), and Thailand (Wongsurawat et al., 2018) in 2015–2016. Microcephaly was still noted among the clinical presentations of patients in Vietnam and Thailand, which indicated the general occurrence of CZS in America and Asia after 2015. Nonetheless, the reasons why ZIKV cases with CZS were not identified before 2015 is still unclear.

One of the plausible hypotheses is that ZIKV acquired some virulence mutations which contributed to the neurological infection in human fetal brains (**Table 1**). The nucleotide changes, along with amino acid substitutions, continuously accumulated over time and in different geographies, resulting in ZIKV evolution (Liu et al., 2019). The accumulated genetic diversity in the ZIKV population may increase the potential to cause severe diseases. To expand on this issue, we reviewed ZIKV evolution from the first isolation to recent emergent outbreaks and discussed the recent evidence on how the genetic changes affected ZIKV phenotypes, and on why ZIKV has become an emerging pathogen in recent outbreaks.

ZIKV GENETIC EVOLUTION: GENETIC CHANGES FROM OBSCURENESS TO A PROMINENT EMERGING PATHOGEN

The first ZIKV phylogenetic study was conducted after the 2007 Yap outbreak (Lanciotti et al., 2008). According to a phylogenetic study of the complete open reading frame sequence of the polyprotein, ZIKV has two lineages, the African (Nigeria, Senegal, and Uganda strains) and Asian (Malaysia 1966, Yap 2007, and Cambodia 2010) lineages (Haddow et al., 2012). According to the timescale of a phylogenetic tree estimation of ZIKV evolution (Delatorre et al., 2018), Asian ZIKV strains had two independent disseminations from Southeast Asia into the Pacific region–the first in Yap Island and the second in French Polynesia (**Figure 1**). With regards to the evidence that the Cambodia 2010 strain is the closest to the French Polynesia 2013 ZIKV outbreak strain, the former was sequentially introduced into the Pacific region and resulted in the 2013–2014 French Polynesia outbreak (**Figure 1**; Cao-Lormeau et al., 2014). Later, all of the Asian lineage strains isolated from American countries from 2014 including Brazil (Campos et al., 2015; Zanluca et al., 2015; Calvet et al., 2016), Colombia (Camacho et al., 2016), Puerto Rico (Lanciotti et al., 2016), and Guatemala (Lanciotti et al., 2016) showed phylogenetic and temporal closeness in the phylogenetic tree with >99% nucleotide identity with the French Polynesian strains (Lanciotti et al., 2016). Therefore, French Polynesian strains were suggested to simultaneously disseminate into the Easter Island, New Caledonia, the Cook Islands, and the Americas at approximately the same time, suggesting that these strains spread to the Americas and then caused a ZIKV epidemic globally in 2015–2016 (**Figure 1**). Simultaneously, the French Polynesian strain continued its dissemination in the Pacific countries including Samoa, Fiji, Tonga, and American Samoa in a stepping-stone process in 2015–2016 (Delatorre et al., 2018). Taken together, phylogenic evidence of ZIKV showed that the virus was first isolated in Africa and then was found in Asia. Afterward, ZIKV continuously evolved in Asia. ZIKV began to cause epidemics in the Pacific from 2000, including the 2007 Yap outbreak and the French Polynesia 2013–2014 outbreak. French Polynesia was suggested as the main origin for global ZIKV dissemination, which was sequentially introduced into the Americas where thousands of CZS cases were reported.

CONTRIBUTION OF ZIKV MUTATIONS TO FETAL MICROCEPHALY AND MOSQUITO TRANSMISSION

To date, severe CZS cases have been found to be primarily caused by Asian ZIKV strains (D'Ortenzio et al., 2016), raising an important scientific question. That is, whether the accumulated mutations in ZIKV, particularly those acquired when the African

TABLE 1 | Potential factors contributing to the recent emergence of Zika virus (ZIKV).

Potential factors		Description	References
Pathogen			
Mutation	prM-S139N	prM-S139N mutant causes a more severe microcephalic phenotype with a thinner cortex, more robust brain cell apoptosis, and more NPC differentiation disruption in mice	Yuan et al., 2017
Mutation	NS1-A982V	NS1-A982V mutation enhances ZIKV transmission in a mosquito-mouse-mosquito transmission cycle	Liu et al., 2017
		NS1-A982V mutation of ZIKV enhances the inhibition of interferon-beta production	Xia et al., 2018
Mutation	E-V763M	E-V763M mutation increases ZIKV replication, neurovirulence in neonatal mice, and maternal-to-fetal transmission	Shan et al., 2020
Mutation	C-T106A, prM-V123A, NS1-A982V, and NS5-M3392V	ZIKV with C-T106A, prM-V123A, NS1-A982V, and NS5-M3392V mutations has a fitness advantage	Liu et al., 2021
Host			
Genetics	Host genome background	The pathogenesis of discordant and dizygotic twins from ZIKV-infected mothers was compared, and host genetics was found to substantially affect the severity of a ZIKV infection, even when infected with the same strain	Caires-Junior et al., 2018
Immunity	Preexisting anti-flavivirus immunity	Previous DENV immunity had no or cross-protection impact against ZIKV infection	Gordon et al., 2019; Rodriguez-Barraquer et al., 2019; Carvalho et al., 2020
Environment			
Temperature	Climate change	Elevated temperatures can expand the geographic vector range, decrease the extrinsic incubation period of the pathogen, and increase the female mosquito biting rate	Morin et al., 2013; Paz and Semenza, 2016

prM-S139N, S139N mutation in the pre-membrane (prM) coding region; NS1-A982V, A982V mutation in the non-structural protein 1 (NS1) coding region; E-V763M, V763M mutation in the envelope (E) coding region; C-T106A, T106A mutation in the capsid (C) coding region; prM-V123A, V123A mutation in the pre-membrane (prM) coding region; NS5-M3392V, M3392V mutation in the non-structural protein 5 (NS5) coding region; DENV, Dengue virus; ZIKV, Zika virus.

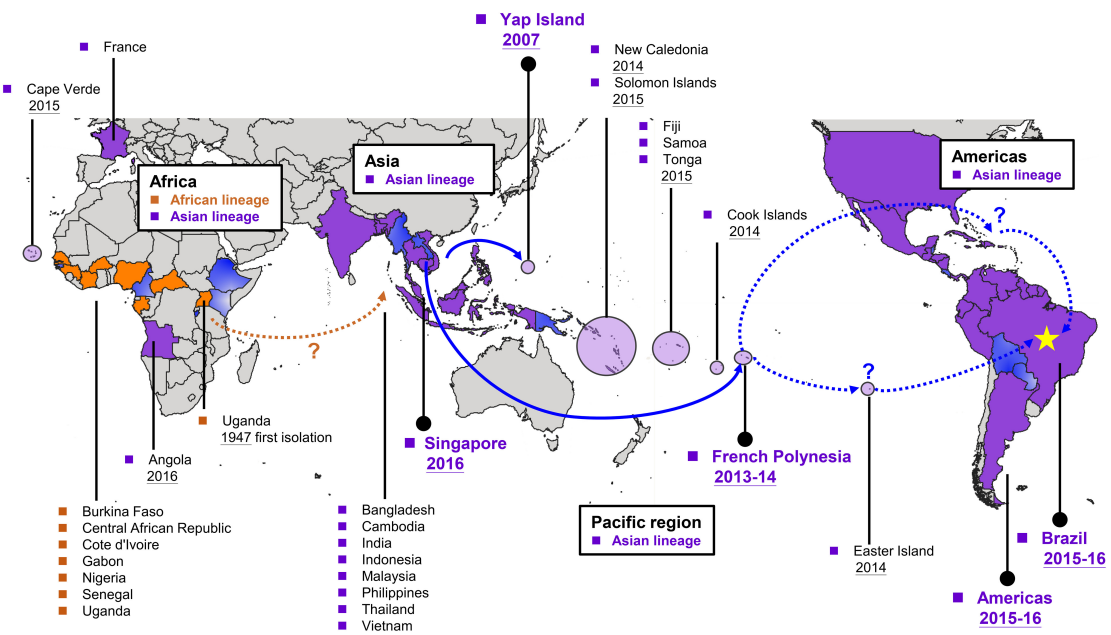


FIGURE 1 | Global dissemination of Zika virus (ZIKV). ZIKV originated from Africa and was then introduced to Asia. Thereafter, ZIKV continuously evolved in Asia and began to cause outbreaks in the Pacific region from 2007. The 2013–2014 French Polynesia strain became the common ancestor worldwide and resulted in recent ZIKV outbreaks in the Americas. According to phylogenetic analysis, the African lineage viruses were circulated mainly in African countries, whereas those circulated in Asia, the Pacific regions, and the Americas were Asian lineage viruses. Countries or areas that have locally acquired ZIKV cases are colored with orange (African lineage circulation), purple (Asian lineage circulation), or gradient blue (the circulating lineage has not been confirmed).

strain evolved to the Asian lineages, contributes to the increase in CZS severity or transmissibility in mosquito vectors. Thus, several studies have investigated the phenotypic differences between African and Asian strains; however, the results obtained did not agree with the intuitive hypotheses. *In vitro* tissue culture evidence revealed that the African strains exhibit higher growth rates and induce higher rates of cellular apoptosis than the Asian strains in primate cell lines, such as Vero, human embryonic kidney (HEK293) cells, human neuronal progenitor cells (NPCs) as well as neuroblastoma, glioblastoma, and monocyte-derived dendritic cells (Simonin et al., 2016; Anfasa et al., 2017; Bowen et al., 2017; Shao et al., 2017; Sheridan et al., 2017; Smith et al., 2018). Similarly, an *in vivo* model conferred that the African strains have higher virulence than the Asian strains in mice (Dowall et al., 2017; Shao et al., 2017). Moreover, mosquitoes are more susceptible to the African strains than the Asian strains (Weger-Lucarelli et al., 2016; Azar et al., 2017; Roundy et al., 2017). Cumulatively these studies demonstrate that the Asian strains, responsible for the ZIKV outbreaks in the Pacific and the Americas, were less virulent with lower vector competence than the African strains.

Due to this evidence that was counterintuitive to the hypothesis for ZIKV emergence and the epidemic trends of ZIKV after 2007, more recent studies have focused on phenotypic comparison of the ancestral (Cambodia 2010) strain and different contemporary Asian strains isolated after 2010 (Liu et al., 2017; Yuan et al., 2017). One study utilized a mouse embryonic microcephaly model to compare the virulence of the ancestral and contemporary strains (Yuan et al., 2017). *In vitro* NPC results indicated that contemporary Asian strains showed a substantial increase in viral growth and cellular apoptosis in contrast with the ancestral strain. Contemporary strains also exhibited higher virulence in mouse brain tissue, and sequentially induced severe microcephaly phenotypes in embryonic brains *in vivo* (Liu et al., 2017). The authors compared the transmissibility of the ancestral Cambodia strain and contemporary Asian strain between the interferon receptor-deficient adult mice and *A. aegypti* mosquitoes. Although the ancestral and contemporary strains exhibited similar growth rates in AG6 mice, the contemporary China 2016 strain not only enhanced the ZIKV NS1 viremia, but also exhibited higher infectivity rates in the mosquitoes feeding on ZIKV-infected AG6 mice. In addition, the contemporary strain exhibited a fitness advantage in contrast with the ancestral strain, in both mice and mosquitoes (Liu et al., 2021). Taken together, the contemporary strain exhibits higher neurovirulence, causing microcephaly, as seen in the mouse model, enhancing the viral infectivity in mosquitoes, and possessing superior fitness than the ancestral strain *in vivo*.

Considering the numerous evolutionary mutations that have been identified in the contemporary strains, via phylogenetic analysis, various studies have applied reverse genetics systems to assess the contributions made by these mutations of ZIKV phenotype (**Table 1**; Liu et al., 2017; Yuan et al., 2017; Shan et al., 2020; Liu et al., 2021). Results have shown that the prM-S139N mutation accelerates viral replication in human NPCs, as well as viral virulence in the neonatal mouse model (Yuan et al., 2017). Furthermore, prM-S139N mutants have a more severe microcephalic phenotype with a thinner cortex, more robust brain cell apoptosis, and more NPC differentiation disruption, in contrast with the wild-type virus. Meanwhile, the E-V763M mutation was found to increase ZIKV replication, neurovirulence in neonatal mice, and maternal-to-fetal transmission (Shan et al., 2020). Additionally, the NS1-A982V mutation enhanced ZIKV infectivity in *A. aegypti* mosquitoes and elevated NS1 antigenemia without altering infective virus propagation in AG6 mice (Liu et al., 2017). Nonetheless, the NS1-A982V mutation enhanced ZIKV transmission not only from AG6 mice to mosquitoes, but also in a mosquito-mouse-mosquito transmission cycle. Additionally, the NS1-A982V mutation caused the NS1 protein to bind to TBK1, and subsequently reduce its phosphorylation level, which may inhibit interferon-beta production (Xia et al., 2018). In addition to single mutation, contemporary Asian strains possess C-T106A, prM-V123A, NS1-A982V, and NS5-M3392V, which serve to synergistically increase viral fitness in mice and mosquitoes (Liu et al., 2021). Cumulatively, these findings indicate that not only several single mutations, but also multiple mutations, have altered ZIKV phenotypic characteristics, including disease severity, transmission, and viral fitness.

NON-MUTATION BASED POTENTIAL FACTORS RESPONSIBLE FOR THE RECENT ZIKV EMERGENCE

Although additional ZIKV mutations, responsible for enhancing ZIKV neurovirulence, mosquito infectivity, and fitness, are likely to be characterized in the future, the general consensus is that other factors, aside from genetic mutations, have also contributed to the large outbreaks that occurred in 2015–2016 in the Americas (**Table 1**). Due to the nature of error-prone RNA virus replication, mutations can readily occur during ZIKV evolution and may represent a founder's effect with a neutral fitness. Meanwhile, considering that the African strains are more virulent than the Asian strains, an alternative hypothesis is that all ZIKV strains have the rare capacity to cause severe neurological diseases through host-vector (urban), or host-host (sexual) transmission cycles (Rossi et al., 2018). Therefore, severe cases can only be detected by public health surveillance systems in a large outbreak; whereas in small outbreaks, such as the 2007 Yap State outbreak with approximately 5,000 human cases, the severe cases were not prevalent enough to be sufficiently recognized (Duffy et al., 2009). In contrast, with more than 30,000 cases of ZIKV diagnosed in the French Polynesia, and even greater numbers in the Americas, substantial rates of GBS or CZS cases were also detected during these devastating outbreaks. This alternative hypothesis may explain why infection by both African and Asian strains can result in fetus microcephaly and mosquito infectivity elevation. However, this hypothesis does not address one key question. That is, why was ZIKV activity substantially increased after the French Polynesia outbreak and robustly elevated during the American outbreak?

In addition to the pathogen, host factors including genetics and immunity represent major components in the epidemiologic

triad of infectious diseases (**Table 1**). In a study comparing the NPCs from three pairs of dizygotic twins with different CZS in ZIKV infection, the cells exhibited diverse gene profiles for major regulators of the neurodevelopmental program, including the mTOR and WNT pathways. Moreover, the NPCs from affected individuals were more susceptible to ZIKV infection with accelerated viral replication and reduced cell growth (Caires-Junior et al., 2018). Hence, host genetics substantially affect the severity of ZIKV infection, even for dizygotic twins infected with the same strain *in utero*. Nonetheless, no specific loci associated with the CZS has been identified, suggesting that CZS may be a multifactorial disease with oligogenic and/or epigenetic mechanisms. Despite host genetics, pre-existing flavivirus immunity was suggested to modulate subsequential ZIKV disease outcome. Early *in vitro* and mouse studies reported that preexisting anti-DENV immunity associated with ZIKV pathogenesis via antibody-dependent enhancement (ADE) (Stettler et al., 2016; Bardina et al., 2017; Rathore et al., 2019); however, clinical and epidemiological evidence for ADE has not been provided yet. In contrast, human studies demonstrated that previous DENV immunity had no or cross-protection impact against ZIKV infection (Halai et al., 2017; Moreira-Soto et al., 2017; Castanha et al., 2019; Pedroso et al., 2019; Katzelnick et al., 2020; Michlmayr et al., 2020; Tonnerre et al., 2020; Petzold et al., 2021). In the Nicaraguan and Brazil pediatric cohorts, prior DENV infection and pre-existing anti-NS1 DENV antibodies were associated with reduced risk of ZIKV infection and disease (Gordon et al., 2019; Rodriguez-Barraquer et al., 2019). One epidemiological study observed that the areas with large DENV epidemics within 6 years had lower rates of CZS, which suggesting the protection role of recent DENV exposure (Carvalho et al., 2020). In murine models, DENV-immune *Ifnar1*$^{-/-}$ mice displayed that DENV-reactive CD8+ T cells mediate cross-protection against subsequent ZIKV challenge (Wen et al., 2017a,b; Regla-Nava et al., 2018). Heterotypic cross-protection may drive ZIKV evolution under the selection of flavivirus-reactive CD8+ T cell immunity in Asia where ZIKV has been circulating for decades in the face of heterotypic flavivirus immunity. In addition, heterotypic cross-protection provides potential explanation why only few CZS patients have been reported in Africa, another potential ZIKV-circulating continent with a high prevalence of heterotypic flaviviruses (Amarasinghe et al., 2011; Lourenco et al., 2018; Hill et al., 2019). Nonetheless, one aspect that cannot be discounted is the possibility that the number of ZIKV cases are underestimated in certain African nations due to potential misdiagnosis of ZIKV infection in areas endemic for other similar diseases, such as dengue fever.

The environment accounts for the third major component in the epidemiologic triad (**Table 1**). The 2015 El Niño resulted in record-breaking global temperatures, accompanied by severe drought throughout northern and eastern South America, during the second half of the year. These elevated temperatures did not only decrease the extrinsic incubation period, but also expanded the host range and increased the female mosquito biting rate (Morin et al., 2013; Paz and Semenza, 2016). In addition, the expansion of the *Aedes spp.* mosquito range correlates with an increase in water storage during regional periods of drought (Pontes et al., 2000). According to the correlation that was made regarding climate change and *Aedes spp.* expansion, a model was recently developed and applied to predict the time required for each *A. aegypti* life cycle based on the biophysical response to environment conditions (Iwamura et al., 2020). Results demonstrated that the world became approximately 1.5% more suitable per decade for the development of *A. aegypti* from 1950 to 2000 and will accelerate to 3.2–4.4% by 2050, emphasizing the recent gradual expansion of *A. aegypti*. These studies simultaneously indicate that global climate change, including temperature elevation, potentially increase the *Aedes* spp. development and expansion. Hence, all three major components, namely, pathogen, host, and environment, likely contribute to the observed increase in ZIKV prevalence and, as such, must each be taken into account when considering the recent ZIKV emergence.

CONCLUSION

We believe that multiple components have likely synergistically contributed to the debilitating ZIKV outbreaks observed in recent years. Although the number of ZIKV cases reported globally has declined since 2017, surveillance data shows that ZIKV continues to be detected in several areas and in travelers from endemic regions. In future, global surveillance, differential diagnosis, vector control, and virus sequencing are required for prompt monitoring of the transmission dynamics and providing early warnings of possible ZIKV outbreaks in future. Moreover, ZIKV vaccine development is urgently needed to control the next wave of a potential outbreak; however, preexisting flavivirus antibodies in people with prior flavivirus infection or vaccination call for great caution in the design and implementation of vaccines.

AUTHOR CONTRIBUTIONS

S-WH and S-JH: conceptualization, data curation, writing – original draft, and writing – review and editing. S-WH: project administration, resources, supervision, and visualization. Both authors contributed to the article and approved the submitted version.

FUNDING

S-WH received funding from the National Health Research Institutes (Grant Number: MR-110-GP-04; http://www.nhri.org.tw/). The funders had no role in data collection, and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We would like to thank the Ministry of Health and Welfare in Taiwan and National Health Research Institutes for the research support.

REFERENCES

Amarasinghe, A., Kuritsky, J. N., Letson, G. W., and Margolis, H. S. (2011). Dengue virus infection in Africa. *Emerg. Infect. Dis.* 17:1349. doi: 10.3201/eid1708.101515

Anfasa, F., Siegers, J. Y., van der Kroeg, M., Mumtaz, N., Raj, V. S., de Vrij, F. M., et al. (2017). Phenotypic differences between Asian and African lineage Zika viruses in human neural progenitor cells. *mSphere* 2:e00292. doi: 10.1128/mSphere.00292-17

Azar, S. R., Roundy, C. M., Rossi, S. L., Huang, J. H., Leal, G., Yun, R., et al. (2017). Differential vector competency of Aedes albopictus populations from the Americas for Zika virus. *Am. J. Trop. Med. Hyg.* 97, 330–339. doi: 10.4269/ajtmh.16-0969

Bardina, S. V., Bunduc, P., Tripathi, S., Duehr, J., Frere, J. J., Brown, J. A., et al. (2017). Enhancement of Zika virus pathogenesis by preexisting antiflavivirus immunity. *Science* 356, 175–180. doi: 10.1126/science.aal4365

Bowen, J. R., Quicke, K. M., Maddur, M. S., O'Neal, J. T., McDonald, C. E., Fedorova, N. B., et al. (2017). Zika virus antagonizes type I interferon responses during infection of human dendritic cells. *PLoS Pathog.* 13:e1006164. doi: 10.1371/journal.ppat.1006164

Caires-Junior, L. C., Goulart, E., Melo, U. S., Araujo, B. H. S., Alvizi, L., Soares-Schanoski, A., et al. (2018). Discordant congenital Zika syndrome twins show differential in vitro viral susceptibility of neural progenitor cells. *Nat. Commun.* 9:475. doi: 10.1038/s41467-017-02790-9

Calvet, G., Aguiar, R. S., Melo, A. S., Sampaio, S. A., De Filippis, I., Fabri, A., et al. (2016). Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect. Dis.* 16, 653–660. doi: 10.1016/S1473-3099(16)00095-5

Camacho, E., Paternina-Gomez, M., Blanco, P. J., Osorio, J. E., and Aliota, M. T. (2016). Detection of autochthonous Zika virus transmission in Sincelejo, Colombia. *Emerg. Infect. Dis.* 22:927. doi: 10.3201/eid2205.160023

Campos, G. S., Bandeira, A. C., and Sardi, S. I. (2015). Zika virus outbreak, Bahia, Brazil. *Emerg. Infect. Dis.* 21:1885. doi: 10.3201/eid2110.150847

Cao-Lormeau, V. M., Blake, A., Mons, S., Lastère, S., Roche, C., Vanhomwegen, J., et al. (2016). Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. *Lancet* 387, 1531–1539. doi: 10.1016/S0140-6736(16)00562-6

Cao-Lormeau, V. M., Roche, C., Teissier, A., Robin, E., Berry, A. L., Mallet, H. P., et al. (2014). Zika virus, French polynesia, South pacific, 2013. *Emerg. Infect. Dis.* 20:1085. doi: 10.3201/eid2006.140138

Carvalho, M. S., Freitas, L. P., Cruz, O. G., Brasil, P., and Bastos, L. S. (2020). Association of past dengue fever epidemics with the risk of Zika microcephaly at the population level in Brazil. *Sci. Rep.* 10:1752. doi: 10.1038/s41598-020-58407-7

Castanha, P. M. S., Souza, W. V., Braga, C., Araujo, T. V. B., Ximenes, R. A. A., Albuquerque, M., et al. (2019). Perinatal analyses of Zika- and dengue virus-specific neutralizing antibodies: a microcephaly case-control study in an area of high dengue endemicity in Brazil. *PLoS Negl. Trop. Dis.* 13:e0007246. doi: 10.1371/journal.pntd.0007246

Delatorre, E., Fernández, J., and Bello, G. (2018). Investigating the role of easter island in migration of Zika virus from South Pacific to Americas. *Emerg. Infect. Dis.* 24:2119. doi: 10.3201/eid2411.180586

Diallo, D., Sall, A. A., Diagnes, C. T., Faye, O., Faye, O., Ba, Y., et al. (2014). Zika virus emergence in mosquitoes in southeastern Senegal, 2011. *PLoS One* 9:e109442. doi: 10.1371/journal.pone.0109442

Dick, G. W., Kitchen, S. F., and Haddow, A. J. (1952). Zika virus (I). Isolations and serological specificity. *Trans. R. Soc. Trop. Med. Hyg.* 46, 509–520. doi: 10.1016/0035-9203(52)90042-4

D'Ortenzio, E., Matheron, S., de Lamballerie, X., Hubert, B., Piorkowski, G., Maquart, M., et al. (2016). Evidence of sexual transmission of Zika virus. *N. Engl. J. Med.* 374, 2195–2198. doi: 10.1056/NEJMc1604449

Dowall, S. D., Graham, V. A., Rayner, E., Hunter, L., Atkinson, B., Pearson, G., et al. (2017). Lineage-dependent differences in the disease progression of Zika virus infection in type-I interferon receptor knockout (A129) mice. *PLoS Negl. Trop. Dis.* 11:e0005704. doi: 10.1371/journal.pntd.0005704

Duffy, M. R., Chen, T. H., Hancock, W. T., Powers, A. M., Kool, J. L., Lanciotti, R. S., et al. (2009). Zika virus outbreak on Yap Island, federated states of Micronesia. *N. Engl. J. Med.* 360, 2536–2543. doi: 10.1056/NEJMoa0805715

Epelboin, Y., Talaga, S., Epelboin, L., and Dusfour, I. (2017). Zika virus: an updated review of competent or naturally infected mosquitoes. *PLoS Negl. Trop. Dis.* 11:e0005933. doi: 10.1371/journal.pntd.0005933

Fagbami, A. H. (1979). Zika virus infections in Nigeria: virological and seroepidemiological investigations in Oyo State. *Epidemiol. Infect.* 83, 213–219. doi: 10.1017/s0022172400025997

Gordon, A., Gresh, L., Ojeda, S., Katzelnick, L. C., Sanchez, N., Mercado, J. C., et al. (2019). Prior dengue virus infection and risk of Zika: a pediatric cohort in Nicaragua. *PLoS Med* 16:e1002726. doi: 10.1371/journal.pmed.1002726

Gourinat, A. C., O'Connor, O., Calvez, E., Goarant, C., and Dupont-Rouzeyrol, M. (2015). Detection of Zika virus in urine. *Emerg. Infect. Dis.* 21, 84–86. doi: 10.3201/eid2101.140894

Grubaugh, N. D., Ladner, J. T., Kraemer, M. U., Dudas, G., Tan, A. L., Gangavarapu, K., et al. (2017). Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* 546, 401–405. doi: 10.1038/nature22400

Guo, X. X., Li, C. X., Deng, Y. Q., Xing, D., Liu, Q. M., Wu, Q., et al. (2016). Culex pipiens quinquefasciatus: a potential vector to transmit Zika virus. *Emerg. Microbes Infect.* 5:e102. doi: 10.1038/emi.2016.102

Haddow, A. D., Schuh, A. J., Yasuda, C. Y., Kasper, M. R., Heang, V., Huy, R., et al. (2012). Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Negl. Trop. Dis.* 6:e1477. doi: 10.1371/journal.pntd.0001477

Halai, U. A., Nielsen-Saines, K., Moreira, M. L., de Sequeira, P. C., Junior, J. P. P., de Araujo Zin, A., et al. (2017). Maternal zika virus disease severity, virus load, prior dengue antibodies, and their relationship to birth outcomes. *Clin. Infect. Dis.* 65, 877–883. doi: 10.1093/cid/cix472

Hill, S. C., Vasconcelos, J., Neto, Z., Jandondo, D., Zé-Zé, L., Aguiar, R. S., et al. (2019). Emergence of the Asian lineage of Zika virus in Angola: an outbreak investigation. *Lancet Infect. Dis.* 19, 1138–1147. doi: 10.1016/S1473-3099(19)30293-2

Ho, Z. J. M., Hapuarachchi, H. C., Barkham, T., Chow, A., Ng, L. C., Lee, J. M. V., et al. (2017). Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis. *Lancet Infect. Dis.* 17, 813–821. doi: 10.1016/S1473-3099(17)30249-9

Iwamura, T., Guzman-Holst, A., and Murray, K. A. (2020). Accelerating invasion potential of disease vector Aedes aegypti under climate change. *Nat. Commun.* 11:2130. doi: 10.1038/s41467-020-16010-4

Katzelnick, L. C., Bos, S., and Harris, E. (2020). Protective and enhancing interactions among dengue viruses 1-4 and Zika virus. *Curr. Opin. Virol.* 43, 59–70. doi: 10.1016/j.coviro.2020.08.006

Lanciotti, R. S., Kosoy, O. L., Laven, J. J., Velez, J. O., Lambert, A. J., Johnson, A. J., et al. (2008). Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg. Infect. Dis.* 14:1232. doi: 10.3201/eid1408.080287

Lanciotti, R. S., Lambert, A. J., Holodniy, M., Saavedra, S., and Signor, L. D. C. C. (2016). Phylogeny of Zika virus in western hemisphere, 2015. *Emerg. Infect. Dis.* 22:933. doi: 10.3201/eid2205.160065

Lim, S. K., Lim, J. K., and Yoon, I. K. (2017). An update on Zika virus in Asia. *Infect. Chemother.* 49, 91–100. doi: 10.3947/ic.2017.49.2.91

Liu, J., Liu, Y., Shan, C., Nunes, B. T. D., Yun, R., Haller, S. L., et al. (2021). Role of mutational reversions and fitness restoration in Zika virus spread to the Americas. *Nat. Commun.* 12:595. doi: 10.1038/s41467-020-20747-3

Liu, Y., Liu, J., Du, S., Shan, C., Nie, K., Zhang, R., et al. (2017). Evolutionary enhancement of Zika virus infectivity in Aedes aegypti mosquitoes. *Nature* 545, 482–486. doi: 10.1038/nature22365

Liu, Z. Y., Shi, W. F., and Qin, C. F. (2019). The evolution of Zika virus from Asia to the Americas. *Nat. Rev. Microbiol.* 17, 131–139. doi: 10.1038/s41579-018-0134-9

Lourenco, J., de Lourdes Monteiro, M., Valdez, T., Rodrigues, J. M., Pybus, O., and Faria, N. R. (2018). Epidemiology of the Zika virus outbreak in the Cabo Verde Islands, West Africa. *PLoS Curr.* 10:ecurrents.outbreaks.19433b1e4d007451c691f138e1e67e8c. doi: 10.1371/currents.outbreaks.19433b1e4d007451c691f138e1e67e8c

Macnamara, F. N. (1954). Zika virus: a report on three cases of human infection during an epidemic of jaundice in Nigeria. *Trans. R. Soc. Trop. Med. Hyg.* 48, 139–145. doi: 10.1016/0035-9203(54)90006-1

Michlmayr, D., Kim, E. Y., Rahman, A. H., Raghunathan, R., Kim-Schulze, S., Che, Y., et al. (2020). Comprehensive immunoprofiling of pediatric Zika reveals

key role for monocytes in the acute phase and no effect of prior dengue virus infection. *Cell Rep.* 31:107569. doi: 10.1016/j.celrep.2020.107569

Miner, J. J., Sene, A., Richner, J. M., Smit, A. M., Santeford, A., Ban, N., et al. (2016). Zika virus infection in mice causes panuveitis with shedding of virus in tears. *Cell Rep.* 16, 3208–3218. doi: 10.1016/j.celrep.2016.08.079

Moi, M. L., Nguyen, T. T. T., Nguyen, C. T., Vu, T. B. H., Tun, M. M. N., Pham, T. D., et al. (2017). Zika virus infection and microcephaly in Vietnam. *Lancet Infect. Dis.* 17, 805–806. doi: 10.1016/S1473-3099(17)30412-7

Moore, D. Á, Causey, O. R., Carey, D. E., Reddy, S., Cooke, A. R., Akinkugbe, F. M., et al. (1975). Arthropod-borne viral infections of man in Nigeria, 1964–1970. *Ann. Trop. Med. Parasitol.* 69, 49–64. doi: 10.1080/00034983.1975.11686983

Moreira-Soto, A., Sarno, M., Pedroso, C., Netto, E. M., Rockstroh, A., Luz, E., et al. (2017). Evidence for congenital zika virus infection from neutralizing antibody titers in maternal sera, northeastern Brazil. *J. Infect. Dis.* 216, 1501–1504. doi: 10.1093/infdis/jix539

Morin, C. W., Comrie, A. C., and Ernst, K. (2013). Climate and dengue transmission: evidence and implications. *Environ. Health Perspect.* 121, 1264–1272. doi: 10.1289/ehp.1306556

Musso, D., Roche, C., Nhan, T. X., Robin, E., Teissier, A., and Cao-Lormeau, V. M. (2015). Detection of Zika virus in saliva. *J. Clin. Virol.* 68, 53–55. doi: 10.1016/j.jcv.2015.04.021

Olson, J. G., and Ksiazek, T. G. (1981). Zika virus, a cause of fever in Central Java, Indonesia. *Trans. R. Soc. Trop. Med. Hyg.* 75, 389–393. doi: 10.1016/0035-9203(81)90100-0

Paz, S., and Semenza, J. C. (2016). El Niño and climate change–contributing factors in the dispersal of Zika virus in the Americas? *Lancet* 387:745. doi: 10.1016/S0140-6736(16)00256-7

Pedroso, C., Fischer, C., Feldmann, M., Sarno, M., Luz, E., Moreira-Soto, A., et al. (2019). Cross-protection of dengue virus infection against congenital zika syndrome, northeastern brazil. *Emerg. Infect. Dis.* 25, 1485–1493. doi: 10.3201/eid2508.190113

Petzold, S., Agbaria, N., Deckert, A., Dambach, P., Winkler, V., Drexler, J. F., et al. (2021). Congenital abnormalities associated with Zika virus infection-Dengue as potential co-factor? A systematic review. *PLoS. Negl. Trop. Dis.* 15:e0008984. doi: 10.1371/journal.pntd.0008984

Pontes, R. J., Freeman, J., Oliveira-Lima, L. W., Hodgson, J. C., and Spielman, A. (2000). Vector densities that potentiate dengue outbreaks in a Brazilian city. *Am. J. Trop. Med. Hyg.* 62, 378–383. doi: 10.4269/ajtmh.2000.62.378

Rathore, A. P. S., Saron, W. A. A., Lim, T., Jahan, N., and St. John, A. L. (2019). Maternal immunity and antibodies to dengue virus promote infection and Zika virus-induced microcephaly in fetuses. *Sci. Adv* 5:eaav3208. doi: 10.1126/sciadv.aav3208

Regla-Nava, J. A., Elong Ngono, A., Viramontes, K. M., Huynh, A. T., Wang, Y. T., Nguyen, A. T., et al. (2018). Cross-reactive Dengue virus-specific CD8+ T cells protect against Zika virus during pregnancy. *Nat. Commun.* 9:3042. doi: 10.1038/s41467-018-05458-0

Rodriguez-Barraquer, I., Costa, F., Nascimento, E. J. M., Castanha, P. M. S., Sacramento, G. A., Cruz, J., et al. (2019). Impact of preexisting dengue immunity on Zika virus emergence in a dengue endemic region. *Science* 363, 607–610. doi: 10.1126/science.aav6618

Rossi, S. L., Ebel, G. D., Shan, C., Shi, P. Y., and Vasilakis, N. (2018). Did Zika virus mutate to cause severe outbreaks? *Trends. Microbiol.* 26, 877–885. doi: 10.1016/j.tim.2018.05.007

Roundy, C. M., Azar, S. R., Brault, A. C., Ebel, G. D., Failloux, A. B., Fernandez-Salas, I., et al. (2017). Lack of evidence for Zika virus transmission by Culex mosquitoes. *Emerg. Microbes Infect.* 6:e90. doi: 10.1038/emi.2017.85

Shan, C., Xia, H., Haller, S. L., Azar, S. R., Liu, Y., Liu, J., et al. (2020). A Zika virus envelope mutation preceding the 2015 epidemic enhances virulence and fitness for transmission. *Proc. Natl. Acad. Sci. U.S.A.* 117, 20190–20197. doi: 10.1073/pnas.2005722117

Shao, Q., Herrlinger, S., Zhu, Y. N., Yang, M., Goodfellow, F., Stice, S. L., et al. (2017). The African Zika virus MR-766 is more virulent and causes more severe brain damage than current Asian lineage and dengue virus. *Development* 144, 4114–4124. doi: 10.1242/dev.156752

Sheridan, M. A., Yunusov, D., Balaraman, V., Alexenko, A. P., Yabe, S., Verjovski-Almeida, S., et al. (2017). Vulnerability of primitive human placental trophoblast to Zika virus. *Proc. Natl. Acad. Sci. U S A.* 114, E1587–E1596. doi: 10.1073/pnas.1616097114

Simonin, Y., Loustalot, F., Desmetz, C., Foulongne, V., Constant, O., Fournier-Wirth, C., et al. (2016). Zika virus strains potentially display different infectious profiles in human neural cells. *EBioMedicine* 12, 161–169. doi: 10.1016/j.ebiom.2016.09.020

Simpson, D. I. H. (1964). Zika virus infection in man. *Trans. R. Soc. Trop. Med. Hyg.* 58, 335–338. doi: 10.1016/0035-9203(64)90201-9

Smith, D. R., Sprague, T. R., Hollidge, B. S., Valdez, S. M., Padilla, S. L., Bellanca, S. A., et al. (2018). African and Asian Zika virus isolates display phenotypic differences both in vitro and in vivo. *Am. J. Trop. Med. Hyg.* 98, 432–444. doi: 10.4269/ajtmh.17-0685

Stettler, K., Beltramello, M., Espinosa, D. A., Graham, V., Cassotta, A., Bianchi, S., et al. (2016). Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. *Science* 353, 823–826. doi: 10.1126/science.aaf8505

Tonnerre, P., Melgaço, J. G., Torres-Cornejo, A., Pinto, M. A., Yue, C., Blümel, J., et al. (2020). Evolution of the innate and adaptive immune response in women with acute Zika virus infection. *Nat. Microbiol.* 5, 76–83. doi: 10.1038/s41564-019-0618-z

Walling, A., and Dickson, G. (2013). Guillain-Barré syndrome. *Am. Fam. Physician.* 87, 191–197.

Weger-Lucarelli, J., Ruckert, C., Chotiwan, N., Nguyen, C., Garcia Luna, S. M., Fauver, J. R., et al. (2016). Vector competence of American mosquitoes for three strains of Zika virus. *PLoS Negl. Trop. Dis.* 10:e0005101. doi: 10.1371/journal.pntd.0005101

Wen, J., Ngono, A. E., Regla-Nava, J. A., Kim, K., Gorman, M. J., Diamond, M. S., et al. (2017a). Dengue virus-reactive CD8+ T cells mediate cross-protection against subsequent Zika virus challenge. *Nat. Commun.* 8:1459. doi: 10.1038/s41467-017-01669-z

Wen, J., Tang, W. W., Sheets, N., Ellison, J., Sette, A., Kim, K., et al. (2017b). Identification of Zika virus epitopes reveals immunodominant and protective roles for dengue virus cross-reactive CD8+ T cells. *Nat. Microbiol.* 2:17036. doi: 10.1038/nmicrobiol.2017.36

Wongsurawat, T., Athipanyasilp, N., Jenjaroenpun, P., Jun, S. R., Kaewnapan, B., Wassenaar, T. M., et al. (2018). Case of microcephaly after congenital infection with Asian lineage Zika virus, Thailand. *Emerg. Infect. Dis.* 24:1758. doi: 10.3201/eid2409.180416

World Health Organization (2017). *Zika Situation Report*. Availble online at: https://www.who.int/emergencies/zika-virus/situation-report/en/ (accessed March 10, 2017).

Xia, H., Luo, H., Shan, C., Muruato, A. E., Nunes, B. T., Medeiros, D. B., et al. (2018). An evolutionary NS1 mutation enhances Zika virus evasion of host interferon induction. *Nat. Commun.* 9:414. doi: 10.1038/s41467-017-02816-2

Yuan, L., Huang, X. Y., Liu, Z. Y., Zhang, F., Zhu, X. L., Yu, J. Y., et al. (2017). A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. *Science* 358, 933–936. doi: 10.1126/science.aam7120

Zanluca, C., Melo, V. C. A. D., Mosimann, A. L. P., Santos, G. I. V. D., Santos, C. N. D. D., and Luz, K. (2015). First report of autochthonous transmission of Zika virus in Brazil. *Mem. Inst. Oswaldo Cruz.* 110, 569–572. doi: 10.1590/0074-02760150192

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 May 2021
doi: 10.3389/fmicb.2021.673855

Genomic Epidemiology of SARS-CoV-2 From Mainland China With Newly Obtained Genomes From Henan Province

Ning Song[1], *Guang-Lin Cui*[2] *and Qing-Lei Zeng*[3]*

[1] Center for Reproductive Medicine, Henan Key Laboratory of Reproduction and Genetics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2] Department of Clinical Laboratory, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [3] Department of Infectious Diseases, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Even though the COVID-19 epidemic in China has been successfully put under control within a few months, it is still very important to infer the origin time and genetic diversity from the perspective of the whole genome sequence of its agent, SARS-CoV-2. Yet, the sequence of the entire virus genome from China in the current public database is very unevenly distributed with reference to time and place of collection. In particular, only one sequence was obtained in Henan province, adjacent to China's worst-case province, Hubei Province. Herein, we used high-throughput sequencing techniques to get 19 whole-genome sequences of SARS-CoV-2 from 18 severe patients admitted to the First Affiliated Hospital of Zhengzhou University, a provincial designated hospital for the treatment of severe COVID-19 cases in Henan province. The demographic, baseline, and clinical characteristics of these patients were described. To investigate the molecular epidemiology of SARS-CoV-2 of the current COVID-19 outbreak in China, 729 genome sequences (including 19 sequences from this study) sampled from Mainland China were analyzed with state-of-the-art comprehensive methods, including likelihood-mapping, split network, ML phylogenetic, and Bayesian time-scaled phylogenetic analyses. We estimated that the evolutionary rate and the time to the most recent common ancestor (TMRCA) of SARS-CoV-2 from Mainland China were 9.25×10^{-4} substitutions per site per year (95% BCI: 6.75×10^{-4} to 1.28×10^{-3}) and October 1, 2019 (95% BCI: August 22, 2019 to November 6, 2019), respectively. Our results contribute to studying the molecular epidemiology and genetic diversity of SARS-CoV-2 over time in Mainland China.

Keywords: SARS-CoV-2, whole-genome sequence, tMRCA, evolutionary rate, Henan Province, Mainland China

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Arryn Craney,
Weill Cornell Medical Center, United States
Wei-Hua Chen,
Huazhong University of Science and Technology, China

***Correspondence:**
Qing-Lei Zeng
zengqinglei2009@163.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *28 February 2021*
Accepted: *23 April 2021*
Published: *20 May 2021*

Citation:
Song N, Cui G-L and Zeng Q-L (2021) Genomic Epidemiology of SARS-CoV-2 From Mainland China With Newly Obtained Genomes From Henan Province. Front. Microbiol. 12:673855. doi: 10.3389/fmicb.2021.673855

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes the severe respiratory disease coronavirus disease 2019 (COVID-19), which was first reported in Wuhan city, Hubei Province, China in December 2019 (Wu et al., 2020a; Zhou et al., 2020; Zhu et al., 2020), subsequently turning into a pandemic with devastating effects. As of February 20, 2021, there have

been 110 million confirmed infections and 2.4 million reported deaths worldwide (https://www.who.int/emergencies/diseases/novel-coronavirus-2019). So far, China has successfully managed to contain the epidemic. However, the number of newly confirmed cases has been rising rapidly outside of China, especially in the USA, India, Brazil, and Russia.

Henan Province, which borders with Hubei Province to the north, has the third-largest population in China (94 m people). It contains many large cities, including Zhengzhou (10 million people), Nanyang (10 million people), Zhoukou (8 million people), Shangqiu (7 million people), and Zhumadian (7 million people). Henan Province has the most transportation links with Hubei Province. The shortest travel time by high-speed railway from Zhengzhou (the capital city of Henan Province) to Wuhan (the capital city of Hubei Province) is <2 h. Moreover, most people from the southern regions of Henan Province choose to work in Wuhan. By December 30, 2020, Henan Province had reported 1,273 local confirmed cases of COVID-19, including 22 deaths. Almost all of the cases in Henan Province were reported in January and February 2020, as described by our previous studies (Zeng et al., 2020a,b).

Tracking the on-going evolution and transmission patterns of SARS-CoV-2 from the perspective of genomic epidemiology may further our understanding of the COVID-19 pandemic and could help improve public-health measures for containing the virus. The first genome sequence of SARS-CoV-2, which was isolated from a 41-year old man who worked at Wuhan Huanan Seafood Wholesale Market in the city of Wuhan, was posted on January 11, 2020 (Wu et al., 2020b). Previous studies on the genomic epidemiology of SARS-CoV-2 in China were based on individual province or limited amount of genome sequences (Lu et al., 2020; Nie et al., 2020; Geidelberg et al., 2021). However, with the increasing number of genome sequences available in the public database, it is imperative to perform such analysis with more sequences sampling from wider area in longer time span. Notably, there was only one whole-genome sequence reported from Henan deposited in the Global Initiative on Sharing All Influenza Data (GISAID) (http://gisaid.org/) (Elbe and Buckland-Merrett, 2017). In the present study, we generated 19 genome sequences of SARS-CoV-2 strains from 18 severe or critically ill patients in Henan Province using metagenomic sequencing. Together with other 710 genome sequences sampled from Mainland China with sampling dates between 24 December 2019 and 22 July 2020, we employed advanced comprehensive methods to investigate the genetic diversity, evolution, and transmission patterns of SARS-CoV-2 in China. The demographic, baseline, and clinical characteristics of these patients were also described. Our study may provide valuable information for investigating the impact of the early public-health intervention on SARS-CoV-2 transmission and evolution in China with high resolution than previous studies.

MATERIALS AND METHODS

Patients

According to the unified arrangement of the Henan provincial government, a total of 47 confirmed severe or critically ill patients were transported to and treated at the First Affiliated Hospital of Zhengzhou University between January 22, 2020 and March 7, 2020. All of these patients were diagnosed as severe or critically ill COVID-19 cases prior to their admission to the First Affiliated Hospital of Zhengzhou University. The classification criteria for disease severity were strictly in accordance with the national COVID-19 Control Plan.

Ethics Statement

This study was approved by the ethics committee of the First Affiliated Hospital of Zhengzhou University (ethical approval number: 2020-KY-116). Inform consent was waived as the samples used for the present study were collected after routine laboratory testing, and the study was considered less than minimal risk to subjects by the aforementioned committee.

Sample Collection and the Nucleic Acid Test of SARS-CoV-2

Throat swabs were collected from patients who were confirmed and hospitalized with COVID-19 at the First Affiliated Hospital of Zhengzhou University during February 2020 and screened for the presence of nucleic acid of SARS-CoV-2 in the Biosafety Level II Laboratory of the hospital. Patient demographic information was also collected on admission. The collection, transport, and nucleic acid test of specimens were carried out in strict accordance with relevant national regulations. Throat swab samples were inactivated at 56°C for 30 min. Total nucleic acid extraction was performed using the nucleic acid extraction kit (Shanghai BioGerm Medical Technology Co., Ltd.) following manufacturer's instructions. SARS-CoV-2 nucleic acid detection was conducted by a qRT-PCR method using the SARS-CoV-2 nucleic acid test kit (Shanghai BioGerm Medical Technology Co., Ltd.). The determination criteria for positive results are strictly in accordance with the manufacturer's instructions. The remaining positive nucleic acid samples were immediately stored at −80°C for subsequent sequencing analysis.

Next-Generation Sequencing of the Complete SARS-CoV-2 Genomes

Next-generation sequencing of the complete SARS-CoV-2 genomes was performed by the BGI Company (Shenzhen, China). Briefly, host DNA was removed from the positive samples using DNase I, and the concentration of RNA samples was measured by Qubit RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). DNA-depleted and purified RNA was used to construct the double-stranded (ds) circular DNA library with MGIEasy RNA Library preparation reagent set (MGI, Shenzhen, China), as follows: (1) RNA was fragmented by incubating with fragmentation buffer at 87°C for 6 min; (2) ds cDNA was synthesized using random hexamers with fragmented RNA; (3) ds cDNA was subjected to end repair, adaptor ligation, and 18-cycle PCR amplification; (4) PCR products were unique dual indexed (UDI), before going through circularization and rolling circle replication (RCR) to generate DNA nanoball (DNB)-based libraries. Negative controls prepared from nuclease-free water and total RNA isolated from human Michigan Cancer Foundation-7 (MCF-7) breast

cancer cells were included. DNB preps of clinical samples were sequenced on the ultra-high-throughput DNBSEQ-T1 platform (MGI, Shenzhen, China) with a paired-end 100 nt strategy, generating average of 100 Gb sequencing data for each sample.

Mutation Calling and Clade Assignment Analysis

To identify differences between 19 genome sequences from Henan Province and reference sequences, and to find out which clades they are from and where on the SARS-CoV-2 tree they fall, we used the Nextclade web tool (beta v0.10.0, https://clades.nextstrain.org/) to perform mutation calling and clade assignment. Here, clades are groups of related sequences that share a common ancestor. Of note, the Nextclade does not do a real phylogenetic analysis. Instead, clade assignment sequence-by-sequence was performed based on signature mutations. The clade-defining mutations (shown on https://clades.nextstrain.org/) are chosen such that assignment based on genotype works in most cases.

Collation of SARS-CoV-2 National Dataset

To investigate the full genetic characterization of the complete genomes of SARS-CoV-2 in Mainland China, 710 complete genomes of SARS-CoV-2 available from the GISAID (http://gisaid.org/) were downloaded (**Supplementary Table 1**) as of October 23, 2020, and aligned with 19 full-length genomes of SARS-CoV-2 determined in the present study from Henan Province. No statistical methods were used to predetermine the number of genomes in the present study as we downloaded all available genomes of human-obtained SARS-CoV-2 strains. Notably, genome sequences that were from environment or duplicate samples, and those without exact collection dates and sampling locations, and those contained >5% Ns after mapping to the Wuhan-Hu-1 reference (GenBank accession number: MN908947.3) were discarded. The dataset used in the present study was also not randomized. The final dataset ("dataset_729") included 729 genomes of SARS-CoV-2 strains from 17 provinces with sampling dates from December 24, 2019 to July 22, 2020. Of the 19 genomes of SARS-CoV-2 strains collected in the present study from Zhengzhou City for household registration in Henan Province, one was from Luohe City, one was from Luoyang City, one was from Nanyang City, two were from Shangqiu City, four were from Xinyang City, one was from Xuchang City, four were from Zhengzhou City, two were from Zhoukou City, and three were from Zhumadian City. We first aligned the collected dataset ("dataset_729") using MAFFT v7.222 (Katoh and Standley, 2013) under an automatic algorithm and then manually edited the alignment using BioEdit v7.2.5 (Hall, 1999).

Recombination Screening and Maximum-Likelihood Analysis

As recombination is a relatively frequent evolutionary mechanism in coronaviruses (Graham and Baric, 2010), we assessed the recombination of our dataset ("dataset_729") by the Phi-test approach using SplitsTree4 v4.16.2 (Huson and Bryant, 2006) and all available recombination detection methods using the Recombination Detection Program (RDP) v4.100 (Martin et al., 2015). The best-fit nucleotide substitution model for "dataset_729" was identified according to the Bayesian information criterion (BIC) method, Akaike Information Criterion (AIC) method, Corrected Akaike Information Criterion (AICc) method, and Decision Theory Selection (DT) method, with three (24 candidate models) substitution schemes in jModelTest v2.1.10 (Darriba et al., 2012). To evaluate the phylogenetic signals of "dataset_729," we performed likelihood-mapping analysis (Schmidt and Von Haeseler, 2007) using TREE-PUZZLE v5.3 (Schmidt et al., 2002), with 10,000 randomly chosen quartets for the dataset. Split network analysis was performed for "dataset_729" using a general time-reversible (GTR) characters transformation with a gamma distribution to describe among-site variation in the rate of nucleotide substitution (Γ), and a proportion of invariable (I) sites (GTR + Γ + I), with the NeighborNet method, which can be loosely thought of as a "hybrid" between the neighbor-joining (NJ) and split decomposition methods, implemented in SplitsTree4 v4.16.2 (Huson and Bryant, 2006). We estimated the maximum-likelihood (ML) phylogenetic tree for the dataset using RAxML v8.2.12 under the GTR + Γ + I nucleotide substitution model, which was identified as the best fitting model for ML inference by jModelTest v2.1.10 (Darriba et al., 2012). Branch support was inferred using 1 000 bootstrap replicates (Felsenstein, 1985), and trees were midpoint rooted. ML phylogenetic tree was visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), after which we regressed root-to-tip genetic divergence obtained from the ML phylogeny against sampling dates to investigate the temporal molecular evolutionary signals for the dataset using TempEst v1.5 (Rambaut et al., 2016).

Molecular Clock Phylogenetics

The Bayesian molecular clock phylogenies of SARS-CoV-2 for "dataset_729" were estimated under GTR + Γ + I nucleotide substitution model, a parametric exponential growth tree prior, and a strict molecular clock, which assumed constant evolutionary rates throughout the phylogeny, through a Markov chain Monte Carlo (MCMC) (Yang and Rannala, 1997) framework implemented in BEAST v1.10.4 (Drummond et al., 2012). A non-informative continuous-time Markov chain (CTMC) reference prior (Ferreira and Suchard, 2008) was used for the molecular clock rate. Bayesian analyses were run using BEAGLE v2.1.2 (Suchard and Rambaut, 2009) for computational enhancement. Bayesian analysis was run for 500 million MCMC steps with sampling parameters and trees every 50 000 generations. The convergence of the MCMC chains was inspected using Tracer v1.7.1 (Rambaut et al., 2018). Maximum clade credibility (MCC) summary tree was generated using TreeAnnotator v1.10.4 (Drummond et al., 2012) after discarding the first 10% as burn-in. MCC summary tree was visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/~figtree/).

Nucleotide Sequence Accession Numbers in GISAID

Nucleotide sequences of complete genome of SARS-Cov-2 identified in this study have been deposited in the GISAID under accession numbers EPI_ISL_1040031 to EPI_ISL_1040049.

TABLE 1 | Detail information on the SNP in the 19 sequences.

Sequence name	Clade	Missing sites	Gaps	Mutations	Variant type	Gene	Protein
FAHZU0002	19B	21	0	C8782T	Synonymous	ORF1a	Ser2839Ser
				G18598A	Missense	ORF1b	Ala1711Thr
				T28144C	Missense	ORF8	Leu84Ser
				C29095T	Synonymous	N	Phe274Phe
				C29586T	Missense	ORF10	Pro10Leu
FAHZU0007	19A	1065	0	C21219T	Synonymous	ORF1ab	Phe6985Phe
FAHZU0008	19A	68	0	T27432C	Synonymous	ORF7a	Ala13Ala
				C28253T	Synonymous	ORF8	Phe120Phe
FAHZU0010	19B	53	0	C8782T	Synonymous	ORF1a	Ser2839Ser
				T28144C	Missense	ORF8	Leu84Ser
FAHZU0011	19A	1195	0	G11083T	Missense	ORF1a	Leu3606Phe
FAHZU0012	19B	40	0	G29742A	Downstream	S	
				C8782T	Synonymous	ORF1a	Ser2839Ser
				C1077T	Missense	ORF1a	Pro271Leu
				G23438T	Missense	S	Ala626Ser
				T28144C	Missense	ORF8	Leu84Ser
				G28878A	Missense	N	Ser202Asn
FAHZU0014	19B	53	3*	C8782T	Synonymous	ORF1a	Ser2839Ser
				C19185T	Synonymous	ORF1b	Cys1906Cys
				C16375T	Missense	ORF1b	Pro970Ser
				T28144C	Missense	ORF8	Leu84Ser
FAHZU0017	19A	543	0	C29303T	Missense	N	Pro344Ser
				C15324T	Synonymous	ORF1b	Asn619Asn
FAHZU0018	19B	2406	0	C29095T	Synonymous	N	Phe274Phe
				T28144C	Missense	ORF8	Leu84Ser
				G18598A	Missense	ORF1b	Ala1711Thr
FAHZU0019	19B	24	0	C8782T	Synonymous	ORF1a	Ser2839Ser
				C29095T	Synonymous	N	Phe274Phe
				G18598A	Missense	ORF1b	Ala1711Thr
				T28144C	Missense	ORF8	Leu84Ser
				C29586T	Missense	ORF10	Pro10Leu
FAHZU0020	19A	117	0	C16596T	Synonymous	ORF1b	Tyr1043Tyr
				G26144T	Missense	ORF3a	Gly251Val
FAHZU0021	19A	51	0	C21219T	Synonymous	ORF1b	Phe2584Phe
				G11083T	Missense	ORF1a	Leu3606Phe
FAHZU0022	19A	266	0	C21707T	Missense	S	His49Tyr
				C9711T	Missense	ORF1a	Ser3149Phe
FAHZU0028	19B	1036	0	C1912T	Synonymous	ORF1a	Ser549Ser
				C8782T	Synonymous	ORF1a	Ser2839Ser
				T28144C	Missense	ORF8	Leu84Ser
				C16393T	Missense	ORF1b	Pro976Ser
				C18570T	Synonymous	ORF1b	Leu1701Leu
FAHZU0032	19A	70	0	A21141G	Synonymous	ORF1b	Leu2558Leu
				C28657T	Synonymous	N	Asp128Asp
				G3483T	Missense	ORF1a	Gly1073Val
				G26144T	Missense	ORF3a	Gly251Val
				C27928A	Missense	ORF8	Thr12Asn
FAHZU0033	19B	49	0	A1015G	Synonymous	ORF1a	Glu250Glu
				C8782T	Synonymous	ORF1a	Ser2839Ser
				G17594A	Missense	ORF1b	Ser1376Asn
				A26664G	Missense	M	Ile48Val
				T28144C	Missense	ORF8	Leu84Ser

(Continued)

TABLE 1 | Continued

Sequence name	Clade	Missing sites	Gaps	Mutations	Variant type	Gene	Protein
FAHZU0034	19A	47	0	C6982T	Synonymous	ORF1a	Cys2239Cys
				C19386T	Synonymous	ORF1b	Asp1973Asp
				G22081A	Synonymous	S	Gln173Gln
				T29483G	Missense	N	Ser404Ala
FAHZU0035	19B	33	0	C8782T	Synonymous	ORF1a	Ser2839Ser
				T18603C	Synonymous	ORF1b	His1712His
				C29095T	Synonymous	N	Phe274Phe
				G20683T	Missense	ORF1b	Val2406Phe
				C27925T	Missense	ORF8	Thr11Ile
				T28144C	Missense	ORF8	Leu84Ser
FAHZU0036	19A	1270	0	C486T	Missense	ORF1a	Ser74Leu
				C18512T	Missense	ORF1b	Pro1682Leu
				T18738C	Synonymous	ORF1b	Phe1757Phe

Mutations are called relative to the reference sequence Wuhan-Hu-1. Unsequenced regions at the 5′ and 3′ end are not shown. Clade names are assigned by the Nextclade.
[*]Three continuous gaps are inserted at position 21991.

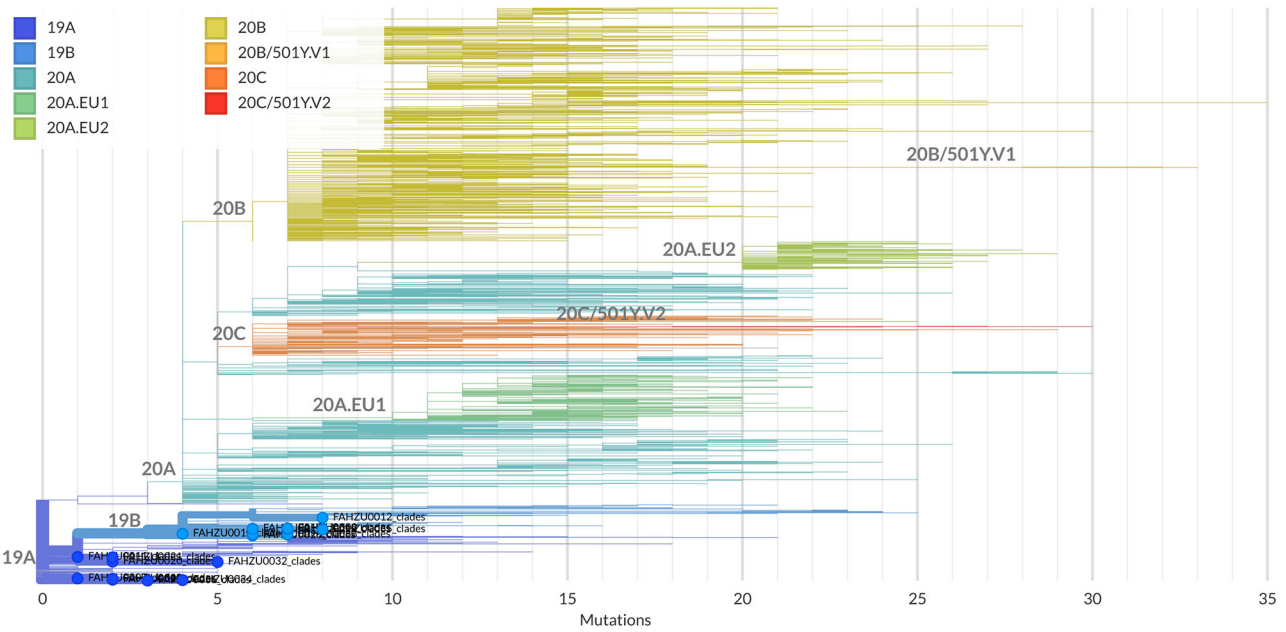


FIGURE 1 | Clade assignment of the 19 Henan sequences analyzed by the Nextclade. Currently, five major clades are defined: 19A and 19B emerged in Wuhan and have dominated the early outbreak; 20A emerged from 19A out of dominated the European outbreak in March and has since spread globally; 20B and 20C are large genetically distinct subclades 20A. The 19 Henan sequences are highlighted and marked with solid circles at the end of their branches.

RESULTS

Sequencing Result Summary

We collected 28 samples of nucleic acids extracted from 27 COVID-19 patients hospitalized in the first affiliated hospital of Zhengzhou University, which were then sent to BGI Company for next-generation sequencing. The total number of reading pairs among the 28 samples was between 209,838,632–1,540,364,746 (sequencing volume 41G∼308G), and the average sample had 743,557,625 read pairs (average sequencing volume is 148G), as shown in **Supplementary Table 2**. The number of SARS-CoV-2 sequences per million (NSPM) sequencing data was between 0.08 and 12,907, and the average NSPM value was 748. Eight samples had NSPM values <1.5, and only one sample had genome coverage >90%; the remaining 20 samples contained 18 genomes with coverage >90% (**Supplementary Table 2**). The sequencing depth was between 0.12X and 6148X, the average sequencing depth was 363X, and there were 19 samples with a depth of more than 4X that could assemble a genome with

TABLE 2 | Demographic, baseline, and clinical characteristics of the 18 patients with COVID-19 in Henan Province.

Variables	Patients (*N* = 18)
Age, years	
Mean ± SD	65.11 ± 16.78
Age group, n (%)	
<29 years	0 (0.00)
30-39 years	2 (11.11)
40-49 years	2 (11.11)
50-59 years	0 (0.00)
60-69 years	5 (27.78)
70-79 years	6 (33.33)
80-89 years	3 (16.67)
≥90 years	0 (0.00)
Sex, n (%)	
Female	4 (22.22)
Male	14 (77.78)
Clinical classification, n (%)	
Mild type	0 (0.00)
Moderate type	0 (0.00)
Severe type	6 (33.33)
Critically ill type	12 (66.67)
Exposure to confirmed cases, n (%)	
Yes	7 (38.89)
No	11 (61.11)
Travel to Hubei	
Yes	4 (22.22)
No	14 (77.73)
Comorbidities, n (%)	11 (61.11)
Cardiovascular and cerebrovascular diseases, cerebrovascular diseases	5 (27.78)
Endocrine system diseases	4 (22.22)
Respiratory system diseases	1 (5.55)
Malignant tumor	1 (5.55)
Nervous system diseases	0 (0.00)
Chronic kidney disease	1 (5.55)
Chronic liver disease	1 (5.55)
Chronic obstructive pulmonary disease	1 (5.55)
More than 1 comorbidity	4 (22.22)
Signs and symptoms at admission, n (%)	
Fever	18 (100)
Cough	13 (72.22)
Shortness of breath	11 (61.11)
Muscle ache	1 (5.55)
Headache and mental disorder symptoms	2 (11.11)
Sore throat	2 (11.11)
Rhinorrhea	1 (5.55)
Chest pain	0 (0.00)
Diarrhea	1 (5.55)
Nausea and vomiting	0 (0.00)
More than 1 sign or symptom	17 (94.44)
Chest radiograph and CT findings, n (%)	
Bilateral pneumonia	18 (100.00)
Unilateral pneumonia	0 (0.00)
No abnormal density shadow	0 (0.00)
Treatment, n (%)	
Antibiotic treatment	17 (94.44)
Antiviral treatment	18 (100.00)
Hormone therapy	13 (72.22)
Intravenous immunoglobulin therapy	14 (77.78)
Mechanical ventilation	13 (72.22)
ECMO	10 (55.56)
Convalescent plasma therapy	11 (61.11)
Traditional Chinese medicine	9 (50.00)
Clinical outcome, n (%)	
Negative and discharged	14 (77.78)
Died	4 (22.22)

CT, computed tomography; ECMO, extracorporeal membrane oxygenation; SD, standard deviation.

coverage >90% (**Supplementary Table 2**). In the present study, a total of 19 samples assembled a genome with coverage of more than 90%. Mutation sites with genome coverage over 90% and sequencing depth over 5X are shown in **Supplementary Figure 1**.

Detail information on the SNP in the 19 Henan sequences is shown in **Table 1** and **Supplementary Figure 2**. A total of 65 mutations were found, among which 45 sites were detected in the non-UTR region; 19 were synonymous mutations, and 23 were non-synonymous mutations. Of note, three sites were deleted for sample FAHZU0014. The number of mutations in each sample ranged from 1 to 6. Mutations were mostly distributed on ORF1ab. Most of the samples contained C8782T and T28144C mutations (FAHZU0018 was missing at site 8782 because of insufficient coverage at that site). It is worth noting that FAHZU0002 and FAHZU0019 are derived from the same patient with the same sequence similarity.

Clades defined by specific signature mutations are shown in **Figure 1**. The 19 Henan sequences were located in 19A and19B clades, which emerged in Wuhan and have dominated the early outbreak.

Demographic, Baseline, and Clinical Characteristics of the Patients

In this study, we only investigated the 18 severe or critically ill patients with COVID-19, for whom sequences of SARS-CoV-2 were successfully assembled with coverage of more than 90% by the aforementioned method. None of these patients were the medical staff. Demographic, baseline, and clinical characteristics of patients are shown in **Table 2**. Among them, 14 patients (77.78%) were male. Seven (38.89%) patients were exposed to confirmed cases. Four (22.22%) patients traveled to Hubei Province. Fourteen patients (77.78%) were aged 60–89 years, and 4 patients (22.22%) were aged 30–49 years. The minimum age was 30, and the maximum age

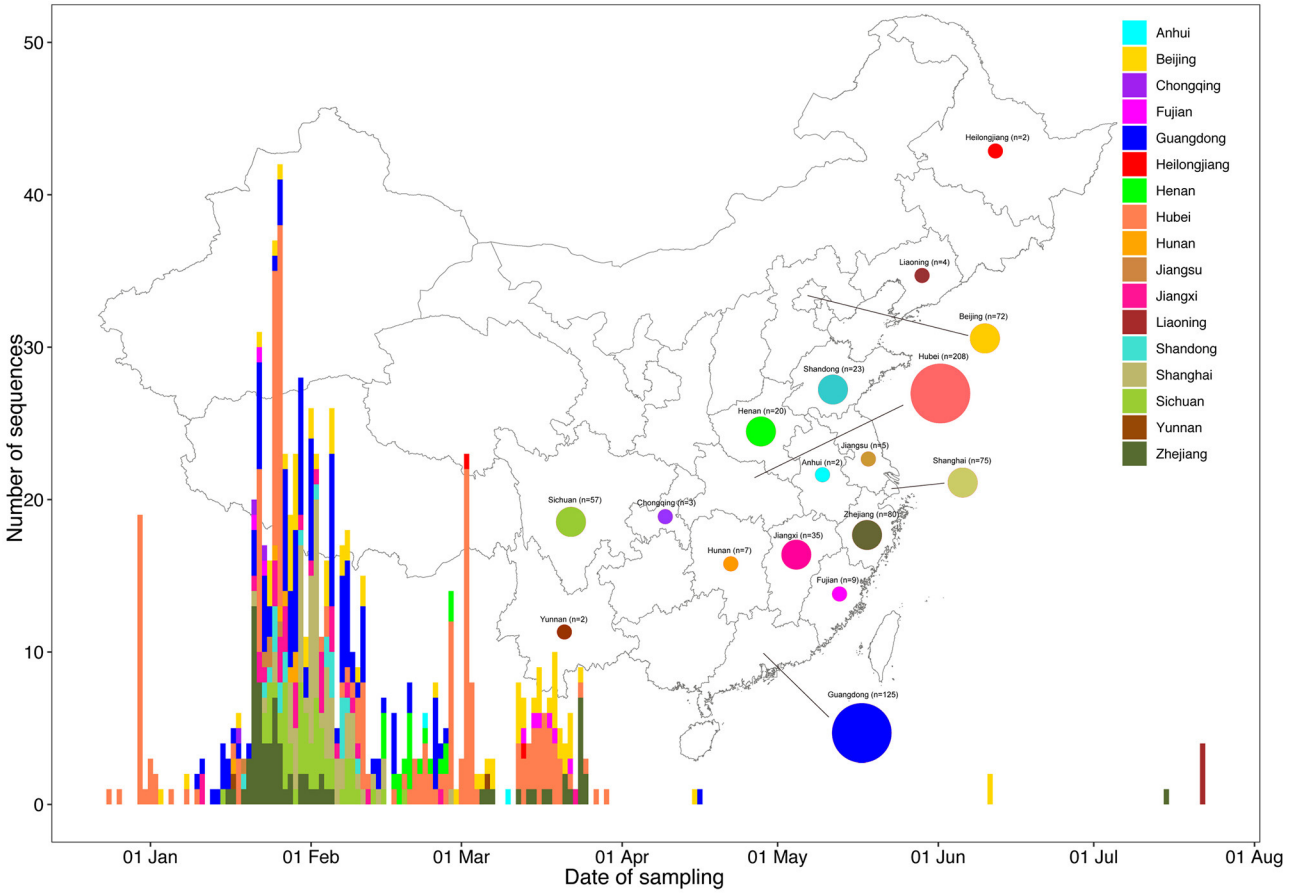


FIGURE 2 | Time series and geographic distribution of the 729 SARS-CoV-2 genomes from Mainland China by sampling date. The geographic distribution of the 729 SARS-CoV-2 genomes from Mainland China in the present study is shown at the provincial level. Colors indicate different sampling provinces from Mainland China.

was 83. Among all 18 patients, 11 (61.11%) had a history of chronic diseases, including cardiovascular and cerebrovascular diseases, cerebrovascular diseases, endocrine system diseases, respiratory system diseases, malignant tumor, chronic kidney disease, chronic liver disease, chronic obstructive pulmonary disease. The most common symptoms were fever and cough that were found in 18 (100.00%) and 13 patients (72.22%), respectively. Eleven patients (61.11%) had shortness of breath. In addition, 1 (5.55%) patient had muscle ache, 2 patients (11.11%) had a headache and mental disorder symptoms, 1 (5.55%) patient had rhinorrhea, and 1 (5.55%) patient had diarrhea. Seventeen patients (94.44%) had more than 1 sign or symptom. All the 18 patients had bilateral pneumonia, according to chest radiograph and CT findings. All patients received antiviral therapy. Seventeen patients (94.44%) were treated with antibiotics. Thirteen (72.22%) received hormone therapy. Fourteen patients (77.78%) received intravenous immunoglobulin therapy. Thirteen patients (72.22%) received mechanical ventilation. Ten patients (55.56%) received ECMO. Eleven (61.11%) received convalescent plasma therapy. Nine (50.00%) received traditional Chinese medicine. Fourteen patients (77.78%) were negative for nucleic acid tests and continued to receive treatment for other underlying diseases, and 4 patients (22.22%) died.

Sequences Distribution

As of October 23, 2020, 710 complete genomes of SARS-CoV-2 detected in mainland China were available from GISAID (http://gisaid.org/) (**Supplementary Table 1**). The sequences classified by sampling provinces through sampling dates are shown in **Figure 2**. Together with 19 sequences obtained from Henan Province in this study, our dataset finally ("Dataset_729") included 729 genome sequences of SARS-CoV-2 sampled from 17 provinces in mainland China (**Figure 2**. Anhui, $n = 2$; Beijing, $n = 72$; Chongqing, $n = 3$; Fujian, $n = 9$; Guangdong, $n = 125$; Heilongjiang, $n = 2$; Henan, $n = 20$; Hubei, $n = 208$; Hunan, $n = 7$; Jiangsu, $n = 5$; Jiangxi, $n = 35$; Liaoning, $n = 4$; Shandong, $n = 23$; Shanghai, $n = 75$; Sichuan, $n = 57$; Yunnan, $n = 2$; and Zhejiang, $n = 80$) with sampling dates between 24 December 2019 and 22 July 2020. The sequences were primarily from Hubei (208/729, 28.53%) and Guangdong (125/729, 17.15%). It is worth mentioning that the sequence distribution with reference to sampling time and province was very uneven; there were few

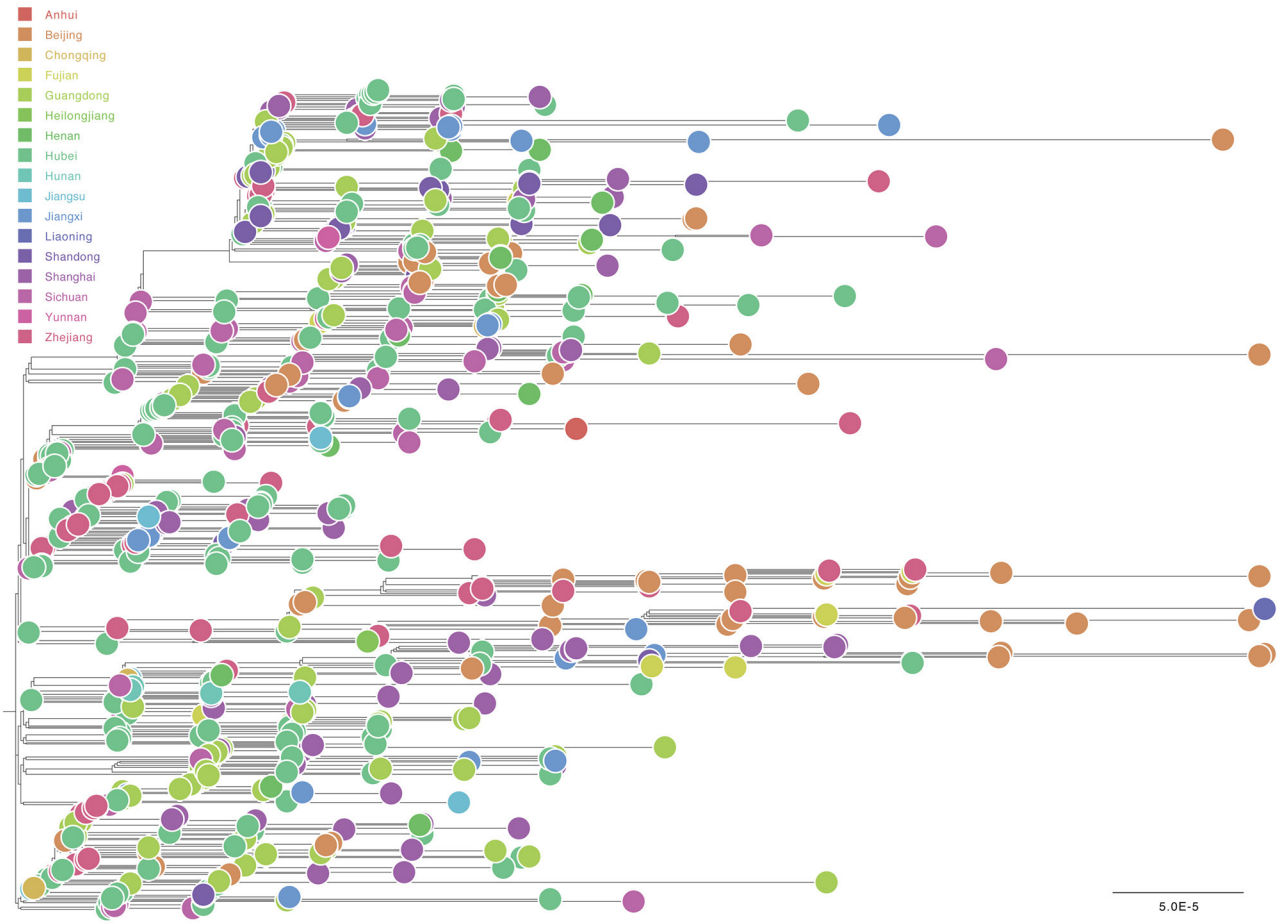


FIGURE 3 | Estimated maximum-likelihood phylogenetic tree of SARS-CoV-2 from Mainland China. Maximum-likelihood phylogenetic tree of SARS-CoV-2 for "dataset_729" from Mainland China is shown. Tree is midpoint rooted. Colors indicate different sampling provinces from Mainland China. The scale bar at the bottom indicates 0.00005 nucleotide substitutions per site.

sequences after April 1, 2020 due to the successful control of the epidemic in China.

Tree-Like Signals and Phylogenetic Analyses

For "dataset_729," a GTR+Γ+I nucleotide substitution model was the best-fit model based on the three substitution schemes (i.e., 24 candidate models) according to the AIC, AICc, BIC, and DT methods, and was thus used in subsequent likelihood-mapping and phylogenetic analyses. The PHI test of "dataset_729" showed no statistically significant evidence of recombination ($p = 1.0$). In addition, no evidence of recombination was found for "dataset_729" using RDP v4.100 (Martin et al., 2015). Our likelihood-mapping analysis revealed that the quartets from "dataset_729" were primarily distributed in the center (56.7%) rather than the corners (43%) or sides (0.3%) of the triangle, indicating a strong star-like topology signal and suggesting the possible rapid early spread of SARS-CoV-2 (**Supplementary Figure 3A**), which is in accordance with previous studies (Li et al., 2020b; Nie et al., 2020). The split network generated for "dataset_729" using the NeighborNet method revealed the existence of polytomies and thus was highly unresolved. This indicated that the phylogenetic relationship of our dataset was probably best represented by a star-like phylogenetic tree rather than a strictly bifurcating tree (**Supplementary Figure 3B**) that may be due to exponential epidemic spread, which was in accordance with the likelihood-mapping results. ML phylogenetic analysis of "dataset_729" also showed star-like topology (**Figure 3**), indicating the introduction of a new virus to an immunologically naive population, which was in accordance with the likelihood-mapping and split network results.

Notably, most genomic sequences of SARS-CoV-2 sampled from Henan Province were clustered with genomic sequences sampled from three other Chinese provinces: Guangdong, Sichuan, and Hubei (**Supplementary Figure 4**), thus indicating that SARS-CoV-2 had already spread from Hubei Province to other Chinese provinces, and also revealing that multiple independent introductions of SARS-CoV-2 from other Chinese

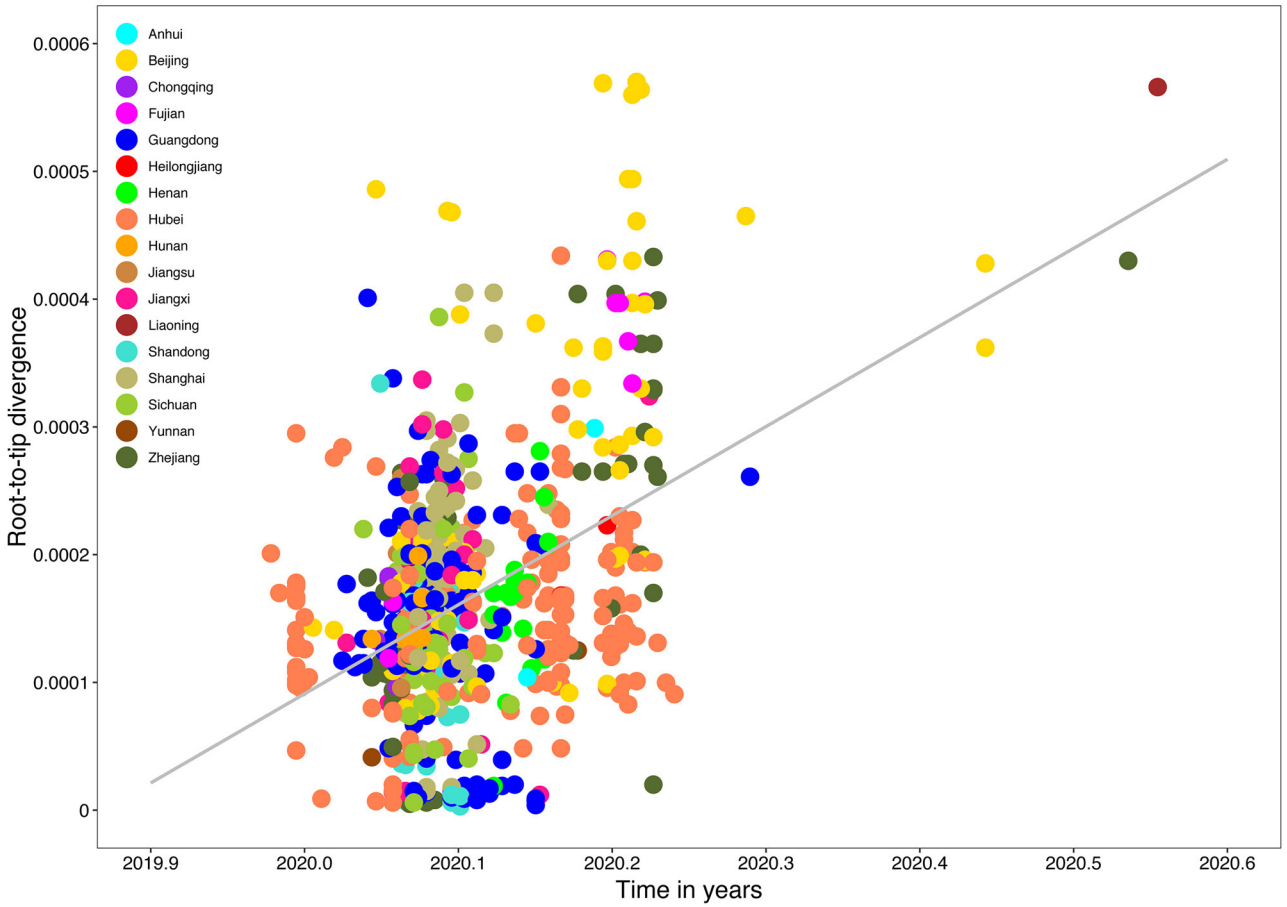


FIGURE 4 | Root-to-tip genetic divergence plot of SARS-CoV-2 from Mainland China. Root-to-tip genetic divergence for "dataset_729" from Mainland China in the Maximum likelihood tree (as shown in **Figure 3**) plotted against sampling date is shown. Colors indicate different sampling provinces from Mainland China. Gray color indicates linear regression line.

provinces into Henan Province had already occurred, although phylogenetic analyses were limited due to the low genetic variation of the virus.

Root-to-tip linear regression analyses between genetic divergence and sampling date using the best-fitting root, which minimizes the mean of the squares of the residuals, showed that "dataset_729" had a positive temporal signal (R^2 = 0.22; correlation coefficient = 0.47), thus suggesting a clocklike pattern of molecular evolution (**Figure 4**). The evolutionary rate and TMRCA date estimates of SARS-CoV-2 for "dataset_729" were 6.9735×10^{-4} substitutions per site per year and November 14, 2019, respectively. Based on Bayesian time-scaled phylogenetic analysis using the tip-dating method, the estimated TMRCA date and evolutionary rate estimates of SARS-CoV-2 for "dataset_729" under a strict molecular clock along with an exponential growth tree prior model with growth rate parameterization using the tip-dating method were October 1, 2019 [95% Bayesian credible interval (BCI): August 22, 2019 to November 6, 2019] and 9.25×10^{-4} substitutions per site per year (95% BCI: 6.75×10^{-4} to 1.28×10^{-3}), respectively. Still, these data were not consistent with the root-to-tip regression results using TempEst v1.5 (Rambaut et al., 2016). However, this posterior evolutionary rate is consistent with previous analyses (Nie et al., 2020; Geidelberg et al., 2021). The estimated TMRCA of the SARS-CoV-2 for "dataset_729" using Bayesian time-scaled phylogenetic analysis were around two months earlier than the result of the previous study (Lu et al., 2020), which may be due to the different datasets and Bayesian time-scaled phylogenetic models used. The estimated growth rate for "dataset_729" under the exponential growth tree prior model was 7.4 per year (95% BCI: 5.1–9.6). The estimates of the MCC phylogenetic relationships among the SARS-CoV-2 genomes for "dataset_729" from the Bayesian coalescent framework using the tip-dating method, as well as the exponential coalescent tree prior with growth rate parameterization and a strict molecular clock model, are displayed in **Figure 5**. As shown, "dataset_729" exhibited more genetic diversity than previous datasets (Li et al., 2020a,b,c; Nie et al., 2020).

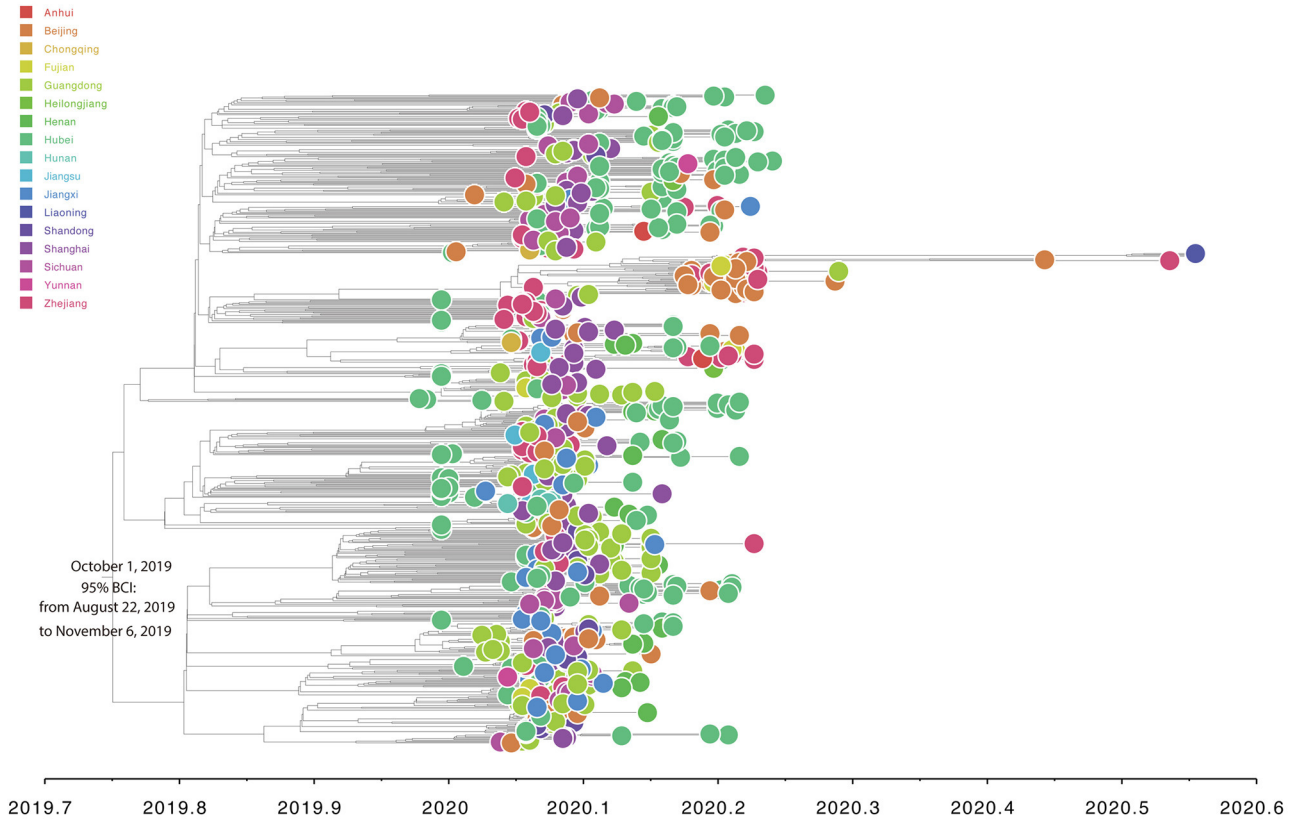


FIGURE 5 | Estimated Bayesian time-scaled maximum-clade-credibility phylogenetic tree of SARS-CoV-2 from Mainland China. Circle at the tip is colored according to sampling provinces from Mainland China. Note that the axis of abscissas is scaled by decimal date.

Discussion

In the present study, no mutations were found in the viral sequences from the same patient collected at two time points with a short interval. This might be due to the virus has a very low probability of mutating *in vivo* over a relatively limited time. However, it has been shown that the virus mutates in chronically infected patients (Kemp et al., 2021). For a better understanding of the impact of the minority viral population on SARS-CoV-2 evolution and transmission, more studies are needed to uncover the in-host variability of the virus during the course of the infection.

To investigate the epidemic spread of SARS-CoV-2 in Mainland China, we performed comprehensive evolutionary analyses of 729 genomes from "dataset_729." Our analyses of the genomic epidemiology of SARS-CoV-2 in Mainland China indicated that most infections from Henan Province resulted from virus importation from three other Chinese provinces, Guangdong, Zhejiang, and Hubei (**Figure 3** and **Supplementary Figure 4**). Also, multiple independent introductions of SARS-CoV-2 from other Chinese provinces into Henan Province were detected. Previous study have also reported multiple independent introductions of SARS-CoV-2 into Guangdong Province (Lu et al., 2020), and Weifang city of Shandong Province; yet, the epidemic spread of SARS-CoV-2 in Henan Province, Guangdong Province and Shandong Province was limited in size and duration. However, all these studies indicated that China has already contained the COVID-19 epidemic in a very short time period. The large-scale surveillance and intervention measures implemented in Henan Province effectively interrupted local transmission and ultimately contained the epidemic and limited eventual dissemination to other regions. Therefore, pathogen sequences have a large potential to inform epidemic surveillance and intervention efforts. It is worth noting that analyses of the phylogenetic structure should be carefully interpreted, as the number of mutations used for the phylogenetic structure was small. It is worth mentioning that the present study was based on a low and variable sampling of COVID-19 cases among different provinces in Mainland China and that COVID-19 cases from other provinces were still not sampled or sequenced or available at GISAID. Nationally, the obtained full-genome sequences of SARS-CoV-2 viruses represent only a tiny fraction of China's number of actual infections. Therefore, it is not suitable to draw conclusions on unknown early genetic diversity and geographic transmission routes in China based on such a small and under-sampled data set. Therefore, more genome sequences of SARS-CoV-2 are needed for a better understanding of the spread of the epidemic patterns of SARS-CoV-2 in China.

In conclusion, this study characterized the epidemic spread patterns of SARS-CoV-2 in Mainland China (including 17 provinces) based on genome data obtained from patients with COVID-19 between December 24, 2019 and July22, 2020. The genomes of SARS-CoV-2 obtained from Henan Province are important for regional and national efforts and other countries to understand how this virus is evolving and spreading around the world. Our results further the understanding of the molecular epidemiology and genetic diversity of SARS-CoV-2 over time in Mainland China. Our results also emphasize the importance of combining the SARS-CoV-2 genomes with all available epidemiological information for all phylogenetic analyses to gain insights into the roles of various surveillance and intervention measures for containing the spread of SARS-CoV-2 in China. Understanding antigenic evolution of SARS-CoV-2, especially in the viral spike, and more especially in the receptor-binding domain in real-time, is increasingly important for guiding prevention efforts and updating SARS-CoV-2 vaccines and treatments. More importantly, the public should also continue to take action in their area to reduce SARS-CoV-2 transmission.

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 studies involving human participants were reviewed and approved by the ethics committee of the First Affiliated Hospital of Zhengzhou University. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

Q-LZ and NS conceived and designed the study. NS and G-LC analyzed the data and wrote this manuscript. Q-LZ interpreted the data and provided critical comments. All authors reviewed and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 81970517), Talents Project of Health Science and Technology Innovation for Young and Middle-aged Investigators in Henan Province, China (No. 2020-OY-04), the Key Scientific Research Project of Henan Higher Education Institutions of China (No. 20B320028), and Basic Research Fund for Young Teachers of Zhengzhou University (Natural Science). The funders had no role in the design and conduct of the study, collection, management, analysis, interpretation of the data, preparation, review, approval of the manuscript, or the decision to submit the manuscript for publication.

ACKNOWLEDGMENTS

We gratefully acknowledge the Authors and Originating and Submitting Laboratories for their sequences and meta-data shared through GISAID, on which this research is based.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.673855/full#supplementary-material

Supplementary Figure 1 | The mutation site plot for the 19 samples with a coverage >90% obtained from the present study. Orange color presents mutation sites in each SARS-CoV-2 genome sequence.

Supplementary Figure 2 | Sequences view with the mapping information of their mutations, missing data, and gaps. Mutations are called relative to the reference sequence Wuhan-Hu-1. Ns (missing) displays a number of N characters (missing data) in the sequence. Each row displays a schema of the corresponding sequence by highlighting the differences to the reference Wuhan-Hu-1 along the genome. Positions are 1-based. Line markers on sequence views represent mutations colored by the resulting nucleotide, as shown at the top right of the figure. Unsequenced regions at the 5′ and 3′ end are indicated as light gray shading. The genome annotation view below the table displays the mapping between positions in the sequence, genes, and clade-defining mutations. It should be notedthat sometimes mutations are so close to each other that they overlap. Clade names are assigned by the Nextclade. Detailed information on the SNP is shown in **Supplementary Table 1**.

Supplementary Figure 3 | Likelihood-mapping and split network analyses of SARS-CoV-2 from Mainland China. Likelihood-mapping **(A)** and split network **(B)** analyses of SARS-CoV-2 for "dataset_729" from Mainland China with sampling dates between December 24, 2019 and July 22, 2020 are shown.

Supplementary Figure 4 | Detailed maximum-likelihood phylogenetic tree of SARS-CoV-2 from Mainland China. This tree shows the detailed information of **Figure 3**.

Supplementary Table 1 | Detail information on the 710 SARS-CoV-2 genomic sequences data set downloaded from GISAID.

Supplementary Table 2 | Assembly information of the 28 nucleic acid samples obtained from the present study.

REFERENCES

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

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

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Ferreira, M. R., and Suchard, M.A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* 36, 355–368. doi: 10.1002/cjs.5550360302

Geidelberg, L., Boyd, O., Jorgensen, D., Siveroni, I., Nascimento, F. F., Johnson, R., et al. (2021). Genomic epidemiology of a densely sampled COVID-19 outbreak in China. *Virus Evol.* 7:veaa102. doi: 10.1093/ve/veaa102

Graham, R. L., and Baric, R. S. (2010). Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J. Virol.* 84, 3134–3146. doi: 10.1128/JVI.01394-09

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

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

Kemp, S. A., Collier, D. A., Datir, R. P., Ferreira, I., Gayed, S., Jahun, A., et al. (2021). SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 592, 277–282. doi: 10.1038/s41586-021-03291-y

Li, X., Wang, W., Zhao, X., Zai, J., Zhao, Q., Li, Y., et al. (2020a). Transmission dynamics and evolutionary history of 2019-nCoV. *J. Med. Virol.* 92, 501–511. doi: 10.1002/jmv.25701

Li, X., Zai, J., Wang, X., and Li, Y. (2020b). Potential of large "first generation" human-to-human transmission of 2019-nCoV. *J. Med. Virol.* 92, 448–454. doi: 10.1002/jmv.25693

Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B. T., et al. (2020c). Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *J. Med. Virol.* 92, 602–611. doi: 10.1002/jmv.25731

Lu, J., Du Plessis, L., Liu, Z., Hill, V., Kang, M., Lin, H., et al. (2020). Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* 181, 997–1003.e1009. doi: 10.1016/j.cell.2020.04.023

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003

Nie, Q., Li, X., Chen, W., Liu, D., Chen, Y., Li, H., et al. (2020). Phylogenetic and phylodynamic analyses of SARS-CoV-2. *Virus Res.* 287:198098. doi: 10.1016/j.virusres.2020.198098

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

Schmidt, H. A., Strimmer, K., Vingron, M., and Von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504. doi: 10.1093/bioinformatics/18.3.502

Schmidt, H. A., and Von Haeseler, A. (2007). Maximum-likelihood analysis using TREE-PUZZLE. *Curr. Protoc. Bioinformatics* 17:6. doi: 10.1002/0471250953.bi0606s17

Suchard, M. A., and Rambaut, A. (2009). Many-core algorithms for statistical phylogenetics. *Bioinformatics* 25, 1370–1376. doi: 10.1093/bioinformatics/btp244

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et al. (2020a). A new coronavirus associated with human respiratory disease in China. *Nature* 580:E7. doi: 10.1038/s41586-020-2202-3

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et al. (2020b). Author Correction: a new coronavirus associated with human respiratory disease in China. *Nature* 580:E7. doi: 10.1038/s41586-020-2008-3

Yang, Z., and Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Mol. Biol. Evol.* 14, 717–724. doi: 10.1093/oxfordjournals.molbev.a025811

Zeng, Q. L., Li, G. M., Ji, F., Ma, S. H., Zhang, G. F., Xu, J. H., et al. (2020a). Clinical course and treatment efficacy of COVID-19 near Hubei Province, China: a multicentre, retrospective study. *Transbound Emerg. Dis.* 67, 2971–2982. doi: 10.1111/tbed.13674

Zeng, Q. L., Yu, Z. J., Gou, J. J., Li, G. M., Ma, S. H., Zhang, G. F., et al. (2020b). Effect of convalescent plasma therapy on viral shedding and survival in patients with coronavirus disease 2019. *J. Infect. Dis.* 222, 38–43. doi: 10.1093/infdis/jiaa228

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* 588:E6. doi: 10.1038/s41586-020-2951-z

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fmicb.2021.653986

Recurrent Dissemination of SARS-CoV-2 Through the Uruguayan–Brazilian Border

Daiana Mir[1†], Natalia Rego[2†], Paola Cristina Resende[3†], Fernando Tort[4†], Tamara Fernández-Calero[2,5†], Verónica Noya[2,6], Mariana Brandes[2], Tania Possi[6], Mailen Arleo[6], Natalia Reyes[6], Matías Victoria[4], Andres Lizasoain[4], Matías Castells[4], Leticia Maya[4], Matías Salvo[4], Tatiana Schäffer Gregianini[7], Marilda Tereza Mar da Rosa[7], Letícia Garay Martins[8], Cecilia Alonso[9], Yasser Vega[10], Cecilia Salazar[11], Ignacio Ferrés[11], Pablo Smircich[12], Jose Sotelo Silveira[13], Rafael Sebastián Fort[12], Cecilia Mathó[13], Ighor Arantes[14], Luciana Appolinario[3], Ana Carolina Mendonça[3], María José Benítez-Galeano[1], Camila Simoes[2], Martín Graña[2], Fernando Motta[3], Marilda Mendonça Siqueira[3], Gonzalo Bello[14], Rodney Colina[4]* and Lucía Spangenberg[2,15]**

[1] Unidad de Genómica y Bioinformática, Departamento de Ciencias Biológicas, Centro Universitario Regional Litoral Norte, Universidad de la República, Salto, Uruguay, [2] Unidad de Bioinformaítica, Institut Pasteur de Montevideo, Montevideo, Uruguay, [3] Laboratório de Vírus Respiratórios e Sarampo, Instituto Oswaldo Cruz – Fiocruz, Rio de Janeiro, Brazil, [4] Laboratorio de Virología Molecular, Departamento de Ciencias Biológicas, Centro Universitario Regional Litoral Norte, Universidad de la República, Salto, Uruguay, [5] Departamento de Ciencias Exactas y Naturales, Universidad Católica del Uruguay, Montevideo, Uruguay, [6] Laboratorio de Biología Molecular, Sanatorio Americano, Montevideo, Uruguay, [7] Laboratório Central de Saúde Pública, Centro Estadual de Vigilância em Saúde da Secretaria de Saúde do Estado do Rio Grande do Sul (LACEN/CEVS/SES-RS), Porto Alegre, Brazil, [8] Centro Estadual de Vigilância em Saúde da Secretaria de Saúde do Estado do Rio Grande do Sul, Porto Alegre, Brazil, [9] CENUR Este-Sede Rocha-Universidad de la República, Montevideo, Uruguay, [10] Laboratorio DILAVE/MGAP-INIA-Universidad de la República, Tacuarembó, Uruguay, [11] Laboratorio de Genómica Microbiana, Institut Pasteur de Montevideo, Montevideo, Uruguay, [12] Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, MEC, Laboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, [13] Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, MEC. Sección Biología Celular, Departamento de Biología Celular y Molecular, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, [14] Laboratorio de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz – Fiocruz, Rio de Janeiro, Brazil, [15] Departamento de Informática y Ciencias de la Computación, Universidad Católica del Uruguay, Montevideo, Uruguay

Uruguay is one of the few countries in the Americas that successfully contained the coronavirus disease 19 (COVID-19) epidemic during the first half of 2020. Nevertheless, the intensive human mobility across the dry border with Brazil is a major challenge for public health authorities. We aimed to investigate the origin of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strains detected in Uruguayan localities bordering Brazil as well as to measure the viral flux across this ∼1,100 km uninterrupted dry frontier. Using complete SARS-CoV-2 genomes from the Uruguayan–Brazilian bordering region and phylogeographic analyses, we inferred the virus dissemination frequency between Brazil and Uruguay and characterized local outbreak dynamics during the first months (May–July) of the pandemic. Phylogenetic analyses revealed multiple introductions of SARS-CoV-2 Brazilian lineages B.1.1.28 and B.1.1.33 into Uruguayan localities at the bordering region. The most probable sources of viral strains introduced to Uruguay were the Southeast Brazilian region and the state of Rio Grande do Sul. Some of the viral strains introduced in Uruguayan border localities between early May and mid-July were able to locally spread and

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch,
United States

Reviewed by:
Matthew M. Hernandez,
Icahn School of Medicine at Mount Sinai, United States
Piyush Baindara,
University of Missouri, United States
Xianding Deng,
University of California,
San Francisco, United States

***Correspondence:**
Luciía Spangenberg
lucia@pasteur.edu.uy
Rodney Colina
rodneycolina1@gmail.com
Gonzalo Bello
gbellobr@gmail.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 15 January 2021
Accepted: 08 April 2021
Published: 28 May 2021

Citation:
Mir D, Rego N, Resende PC, Tort F, Fernández-Calero T, Noya V, Brandes M, Possi T, Arleo M, Reyes N, Victoria M, Lizasoain A, Castells M, Maya L, Salvo M, Schäffer Gregianini T, Mar da Rosa MT, Garay Martins L, Alonso C, Vega Y, Salazar C, Ferrés I, Smircich P, Sotelo Silveira J, Fort RS, Mathó C, Arantes I, Appolinario L, Mendonça AC, Benitez-Galeano MJ, Simoes C, Graña M, Motta F, Siqueira MM, Bello G, Colina R and Spangenberg L (2021) Recurrent Dissemination of SARS-CoV-2 Through the Uruguayan–Brazilian Border. Front. Microbiol. 12:653986. doi: 10.3389/fmicb.2021.653986

originated the first outbreaks detected outside the metropolitan region. The viral lineages responsible for Uruguayan urban outbreaks were defined by a set of between four and 11 mutations (synonymous and non-synonymous) with respect to the ancestral B.1.1.28 and B.1.1.33 viruses that arose in Brazil, supporting the notion of a rapid genetic differentiation between SARS-CoV-2 subpopulations spreading in South America. Although Uruguayan borders have remained essentially closed to non-Uruguayan citizens, the inevitable flow of people across the dry border with Brazil allowed the repeated entry of the virus into Uruguay and the subsequent emergence of local outbreaks in Uruguayan border localities. Implementation of coordinated bi-national surveillance systems is crucial to achieve an efficient control of the SARS-CoV-2 spread across this kind of highly permeable borderland regions around the world.

Keywords: genomics, epidemiology, phylogeography, phylogenetics, SARS-CoV-2, Uruguay, Brazil

INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 19 (COVID-19), was first reported in South America on February 26, 2020 and rapidly spread through the region. South America is actually the second-worst affected region in the world, with more than 11 million SARS-CoV-2 cases and nearly 350,000 deaths confirmed as of December, 2020 (Worldometers, 2020). While the virus exponentially spread during the first half of 2020 in most South American countries, the rapid implementation of non-pharmaceutical interventions avoided an exponential growth of SARS-CoV-2 cases in Uruguay (Moreno et al., 2020; Valcarcel et al., 2020). Six months after the first four cases were reported on March 13, 2020, the country registered the lowest total and per capita numbers of SARS-CoV-2 cases (1,808 cases, 520 cases/million inhabitants) and deaths (45 deaths, 13 deaths/million inhabitants) in the region (Worldometers, 2020).

Despite Uruguay's success to control the early expansion of SARS-CoV-2, the intensive cross-border human mobility between Uruguay and neighboring countries heavily affected by the pandemic is a major concern for public health authorities, aiming to achieve long-term epidemic control. With an area of approximately 176,000 km^2 and 3.5 million inhabitants, Uruguay borders with Argentina to its west and southwest and with Brazil to its north and east. Of particular concern is the border with Brazil, a porous ∼1,100 km strip of land that separates the southernmost Brazilian state of Rio Grande do Sul (RS) and the Uruguayan departments of Artigas (AR), Rivera (RI), Cerro Largo (CL), Treinta y Tres (TT), and Rocha (RO). The Brazilian–Uruguayan border hosts about 170,000 people that live in twin cities located both sides of a dry border and that maintain an intense economic and social interdependence (Arnson et al., 2020).

With a total area of about 282,000 km^2 and 11.3 million inhabitants, RS is the fifth-most-populous Brazilian state and as of July 31, 2020, registered 66,692 SARS-CoV-2 cases and 1,876 deaths (saude, 2021). After an initial phase of relatively slow growth, the COVID-19 epidemic displayed a sharp increase in RS since early May that coincides with the detection of several outbreaks along Uruguayan border departments. The largest SARS-CoV-2 Uruguayan outbreaks (∼50–100 confirmed cases), outside the metropolitan region during the first months after the first detected cases in the country, were at RI and TT departments in May and June, respectively, while smaller outbreaks (∼10–20 confirmed cases) were detected in AR and CL by July (github, 2020). As of July 31, 2020, a total of 201 SARS-CoV-2 cases were reported in Uruguayan municipalities located along the border with Brazil, which represents 16% of the laboratory-confirmed cases in the country (github, 2020).

Previous phylogenetic analyses revealed the circulation of different predominant SARS-CoV-2 lineages in Brazil (B.1.1.28 and B.1.1.33) (Candido et al., 2020; Resende et al., 2020b) and Uruguay (A.5 and B.1) (Elizondo et al., 2020; Salazar et al., 2020), thus supporting independent viral seeding events and little viral exchanges between these neighboring countries during the very early phase of the epidemic. However, most Uruguayan SARS-CoV-2 samples previously analyzed were from Montevideo, the capital city of the country, and there is only scarce information concerning the virus strains circulating in Uruguayan localities bordering Brazil.

We generated 122 SARS-CoV-2 whole-genome sequences recovered from cases isolated at Uruguayan border departments (n = 54) as well as in the southernmost Brazilian state of RS (n = 68) to gain insight into the origin and dynamics of SARS-CoV-2 spread at the Brazilian–Uruguayan border. Our study provides important findings that demonstrate the relevance of building bi-national genomic surveillance workforces in countries with porous and dynamic bordering regions.

MATERIALS AND METHODS

SARS-CoV-2 Samples and Ethical Aspects

A total of 122 SARS-CoV-2 whole-genomes were recovered from nasopharyngeal–throat combined swab samples collected from deceased cases (n = 5), clinically ill or asymptomatic individuals that reside in five different Uruguayan departments (n = 54) at the Brazilian border and 41 different municipalities of the RS

Brazilian state (n = 68) (**Figure 1** and **Supplementary Table 1** in **Supplementary Material Section 2**). Uruguayan samples were collected between May 5 and July 26, 2020, and underwent testing at the Universidad de la República, CENUR Litoral Norte, Salto (Molecular Virology Lab); Sanatorio Americano Montevideo (SASA, Molecular Biology Lab); Universidad de la República CURE Este, Rocha (Molecular Ecology Lab); and Laboratory of DILAVE/MGAP-INIA-UdelaR (Tacuarembó). Sequencing of the Uruguayan samples was held at the Institut Pasteur de Montevideo (IPMON, Bioinformatics Unit) and Instituto de Investigaciones Biológicas Clemente Estable (IIBCE, Genomics Department). Brazilian samples were recovered from sentinel locations in RS state from March 9 to August 16, 2020, and sent to the central laboratory from RS state (LACEN-RS) for SARS-CoV-2 RT-PCR detection. Positive Brazilian samples were subsequently sent to the Laboratory of Respiratory Viruses and Measles, IOC, FIOCRUZ, WHO Regional Reference Laboratory for Coronavirus in the Americas. This study was approved by Ethics Committees in Uruguay (SASA Ethics Committee: CEI-SASA; **Supplementary Material Section 4**) and Brazil (FIOCRUZ-IOC Ethics Committee: 68118417.6.0000.5248 and Brazilian Ministry of Health SISGEN: A1767C3).

SARS-CoV-2 Identification of Positive Samples, Amplification, and Sequencing

Molecular detection of the virus was performed with different kits at each center (Bio-Manguinhos SARS-CoV-2 kit, OneStep RT-qPCR kit COVID-19 RT-PCR Real TM Fast, Coronavirus COVID-19 genesig Real-Time PCR assay), according to manufacturer's instructions (**Supplementary Material Section 1**). SARS-CoV-2 genomes were recovered using both long and short PCR amplicon protocols (Quick, 2020; Resende, 2020; Resende et al., 2020a) and three sequencing technologies: Oxford Nanopore Technologies (ONT), Illumina, and Ion Torrent (**Supplementary Table 1** in **Supplementary Material Section 2**). See **Supplementary Material Section 1** for details.

SARS-CoV-2 Whole-Genome Consensus Sequences

Whole-genome consensus sequences obtained from ONT were generated using an adaptation of the nCoV-2019 novel coronavirus ARTIC bioinformatics protocol[1] as in Resende et al. (2020c) with adjustments that are available at https://github.com/iferres/ncov2019-artic-nf. For Illumina and Ion Torrent, BCFtools was used for single-nucleotide polymorphism (SNP) calling (mpileup), SNP filtering, and consensus sequences reconstruction (consensus function) as described in detail in the **Supplementary Material Section 1**. Positions of interest were manually inspected to resolve undetermined bases. All genomes obtained in this study were uploaded at the EpiCoV database in the GISAID initiative under the accession numbers EPI_ISL_729794 to EPI_ISL_729861 (Brazilian genomes), and accession numbers of Uruguayan genomes are in **Supplementary Table 2** in **Supplementary Material Section 2**.

[1]https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html

SARS-CoV-2 Genotyping and Maximum Likelihood Phylogenetic Analyses

Uruguayan and Brazilian SARS-CoV-2 genome sequences were initially assigned to viral lineages according to Rambaut et al. (2020), using the pangolin web application[2] and later confirmed using maximum likelihood (ML) phylogenetic analyses. ML phylogenetic analyses were performed with the PhyML 3.0 program (Guindon et al., 2010), using an online web server (Guindon et al., 2005; **Supplementary Material Section 1**). Branch support was assessed by the approximate likelihood-ratio test based on a Shimodaira–Hasegawa-like procedure (SH-aLRT) with 1,000 replicates (Anisimova and Gascuel, 2006).

Phylogeographic Analyses

The previously generated ML trees were employed for the ancestral character state reconstruction (ACR) of epidemic locations with PastML (Ishikawa et al., 2019), using the marginal posterior probabilities approximation (MPPA) method with an F81-like model. We next constructed a time-scaled Bayesian phylogenetic tree for the Brazilian and Uruguayan sequences belonging to the B.1.1.28 and B.1.1.33 lineages using the Bayesian Markov chain Monte Carlo (MCMC) approach implemented in BEAST 1.10 (Suchard et al., 2018) with BEAGLE library v3 to improve computational time (**Supplementary Material Section 1**). Viral migrations were reconstructed using a reversible discrete phylogeographic model with a continuous-time Markov chain (CTMC) rate reference prior (Ferreira and Suchard, 2008).

Within-Host Diversity

Only those synapomorphic sites with at least 100 reads were kept for further analysis. As samples were obtained by three different sequencing technologies with different error profiles, a Shannon entropy value (H) was estimated per observation based on the four allele frequencies. We determined a linear model to estimate the contributions of sequencing technology, sample, and synapomorphic site to the observed A, C, G, and T frequencies (summarized as H): H exp 1/4 $\sim$ SEQ + SAMPLE + MUTATION. We used H exp 1/4 to better fit a normal distribution (**Supplementary Figure 1C** in **Supplementary Material Section 3**). Complementarily, allele frequencies, and H values were estimated and compared for six samples (from clade BR-UY-II33) where Illumina, Ion Torrent, and ONT sequencing data was available as samples were sequenced by the three technologies (**Supplementary Figures 1**, S2 in **Supplementary Material Section 3**). Details are found in **Supplementary Material Section 1**.

RESULTS

Prevalent SARS-CoV-2 Lineages at the Uruguayan–Brazilian Border

To understand the dynamics of SARS-CoV-2 spread at the Brazilian–Uruguayan border, we sequenced the viral genome

[2]https://pangolin.cog-uk.io

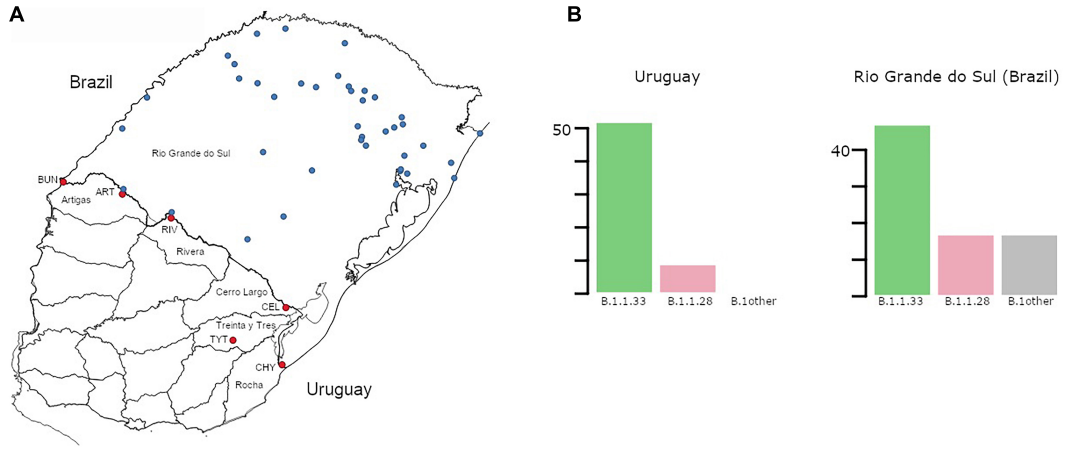


FIGURE 1 | (A) Map of Uruguay and Rio Grande do Sul (Brazil) showing the distribution of samples included in the study. Sampling locations in Uruguay and Rio Grande do Sul (RS) are marked in red and blue, respectively. Uruguayan departments bordering Brazil are explicitly named. Additionally, towns/cities within those departments are marked: BUN, Bella Unión; ART, Artigas; RIV, Rivera; CEL, Rio Branco; TYT, Treinta y Tres; CHY, Chuy. **(B)** Prevalence (number of cases) of the different SARS-CoV-2 lineages detected in Uruguayan departments at the border region and in RS. While no "B.1other" were found in Uruguay, 6 B.1 and 7 B.1.1 samples were found in RS. The total number of cases for Uruguay was 59 and for RS 81.

from 54 individuals diagnosed between May 5 and July 27 in the five Uruguayan border departments (AR, CL, RI, RO, and TT) and from 68 individuals diagnosed at 42 different municipalities from RS state collected between March 9 and August 16, 2020 (**Figure 1A** and **Supplementary Table 1** in **Supplementary Material Section 2**). These sequences were combined with a few SARS-CoV-2 whole-genome sequences from Uruguayan individuals diagnosed at RI (n = 5) and Montevideo (n = 1) and from Brazilian individuals sampled in RS (n = 13) retrieved from the EpiCoV database in the GISAID initiative. The resulting final dataset of 59 SARS-CoV-2 sequences from Uruguayan departments bordering Brazil represents 30% of all laboratory-confirmed cases (n = 201) in that region between March and July 2020. A low fraction of Uruguayan (∼4%) and no Brazilian individuals here sequenced reported international travel or contact with traveling people, indicating that most of them were locally infected. The SARS-CoV-2 genotyping of Uruguayans and Brazilians diagnosed at the border region revealed a quite homogenous pattern as all sequences belonged to the B.1 lineage characterized by the D614G mutation at the spike protein. Furthermore, all SARS-CoV-2 sequences from Uruguayan departments at the border region belonged to the dominant Brazilian lineages B.1.1.33 (85%) and B.1.1.28 (15%) (**Figure 1B**). Although different B.1 sub-clades were identified in RS, the Brazilian lineages B.1.1.33 (58%) and B.1.1.28 (21%) were also the most prevalent across all state regions (**Figure 1B**).

Identification of Major SARS-CoV-2 Uruguayan–Brazilian Clades

Phylogenetic trees of SARS-CoV-2 lineages B.1.1.28 and B.1.1.33 were inferred to explore the Uruguayan and Brazilian RS sequence clustering with sequences from other Brazilian states and worldwide. The ML phylogenetic analyses revealed that most SARS-CoV-2 sequences from Uruguay branched within four highly supported (aLRT = 1) Uruguayan clades (**Figure 2**). The largest B.1.1.33 Uruguayan clade TT-I_{33} (n = 28) comprises all sequences detected in TT between June 18 and July 2. The B.1.1.33 clades RI-I_{33} (n = 19) and RI-II_{33} (n = 3) comprise all sequences detected in RI between May 5 and June 5 and in late July, respectively (**Figure 2A**). The B.1.1.28 Uruguayan clade AR-I_{28} (n = 4) comprises all sequences detected in AR (Bella Unión City) in late July (**Figure 2B**). The B.1.1.33 sequence detected in March in Montevideo as well as the B.1.1.28 sequences detected in July in AR (Artigas City), CL, RO, and RI appeared as dyads or singletons.

The four major Uruguayan clades were nested among basal sequences from Brazil, forming three highly supported (aLRT = 1) Brazilian–Uruguayan clades (**Figure 2**). The largest clade (n = 39) designated as BR-UY-I_{33} comprises the Uruguayan clade TT-I_{33}, a group of basal Brazilian sequences (n = 10) isolated in São Paulo and Rio de Janeiro states between March 24 and June 6 and one sequence from Ireland. The second Brazilian–Uruguayan clade designated as BR-UY-II_{33} comprises the Uruguayan clades RI-I_{33} (n = 19) and RI-II_{33} (n = 3) and a group of Brazilian sequences (n = 8) isolated in different municipalities from RS between May 6 and July 28 (**Figure 2A**). The third Brazilian–Uruguayan clade designated as BR-UY-I_{28} comprises the Uruguayan clade AR-I_{28} and one Brazilian sequence isolated in the São Paulo state on March 20 (**Figure 2B**). Our analysis also identified one highly supported (aLRT = 1) monophyletic group designated as BR-RS-I_{33} (n = 12) that only comprises sequences from RS (**Figure 2A**).

Spatiotemporal Dissemination of SARS-CoV-2 Uruguayan–Brazilian Clades

To identify the number of independent introduction events of SARS-CoV-2 into Uruguay and their most probable

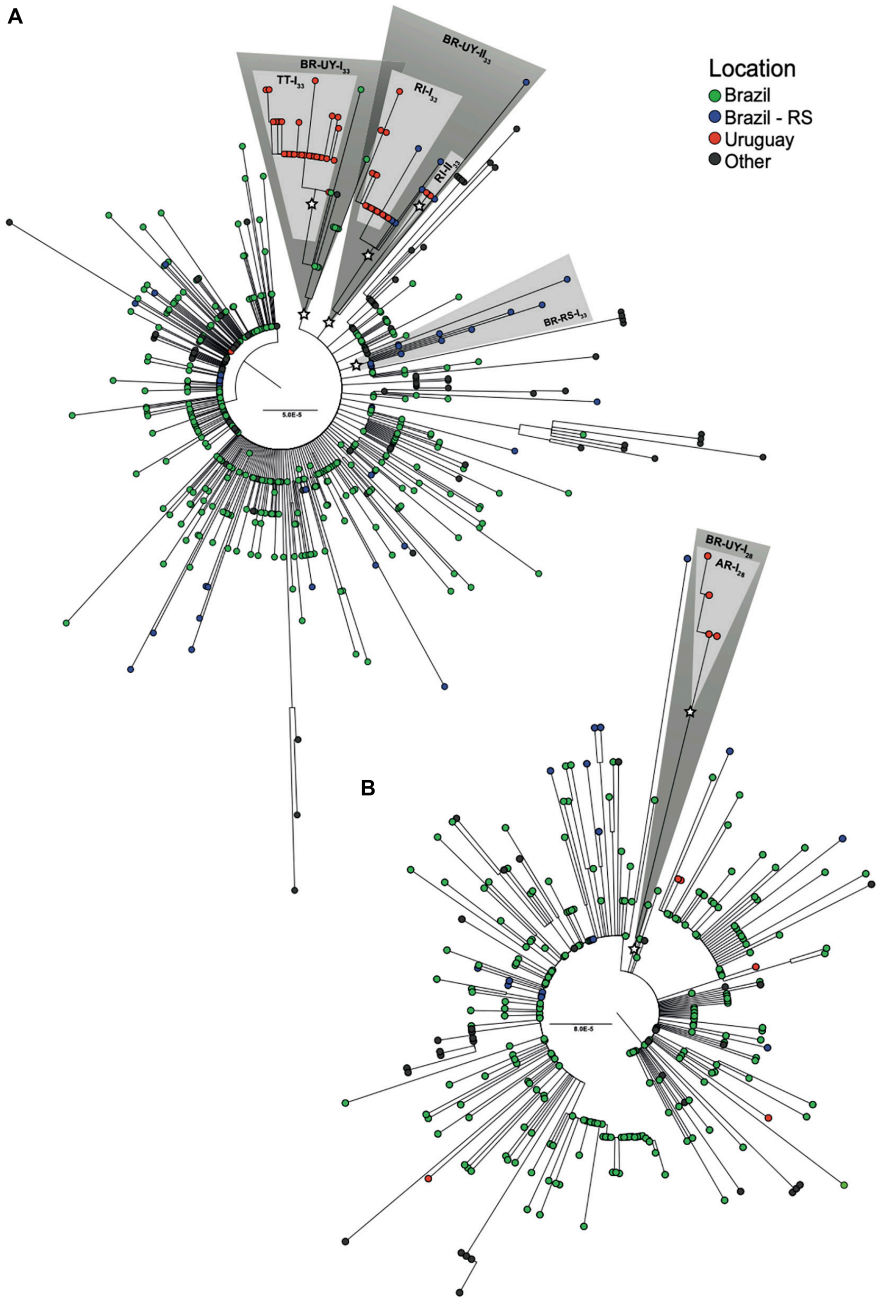


FIGURE 2 | Identification of major SARS-CoV-2 Uruguayan–Brazilian clades. ML phylogenetic trees of **(A)** 86 B.1.1.33 and **(B)** 23 B.1.1.28 genomes obtained in this study along with 492 and 275 worldwide reference sequences of the respective genotypes available in GISAID database. Tip circles are colored according to the sampling location. Node support (aLRT) values at key nodes are represented by asterisks (*). Shaded boxes highlight the position of clusters BR-UY-I_{33}, BR-UY-II_{33}, BR-RS-I_{33}, BR-UY-I_{28}, TT-I_{33}, RI-I_{33}, RI-II_{33}, and AR-I_{28}. The tree was rooted on midpoint and branch lengths are drawn to scale with the bars at the center indicating nucleotide substitutions per site. UY-AR, Artigas-Uruguay; UY-RI, Rivera-Uruguay; UY-TT, Treinta y Tres-Uruguay; BR-SE, Southeast Brazilian region; BR-RS, Rio Grande do Sul- Brazil.

source location, we used a ML-based probabilistic method of ancestral character state reconstruction implemented in the PastML program. Sequences were grouped according to country (Argentina, Brazil, Chile, and Uruguay) or region (North America, Europe, and Oceania) of origin. The ML phylogeographic analyses estimate nine separate viral introductions into Uruguay from Brazil: five of the lineage B.1.1.28 and four of the lineage B.1.1.33 (**Figure 3**). Three introductions of lineage B.1.1.33 and one introduction of lineage B.1.1.28 led to onward transmission to more than one individual and gave origin to the Uruguayan clades RI-I_{33}, RI-II_{33}, TT-I_{33}, and AR-I_{28}.

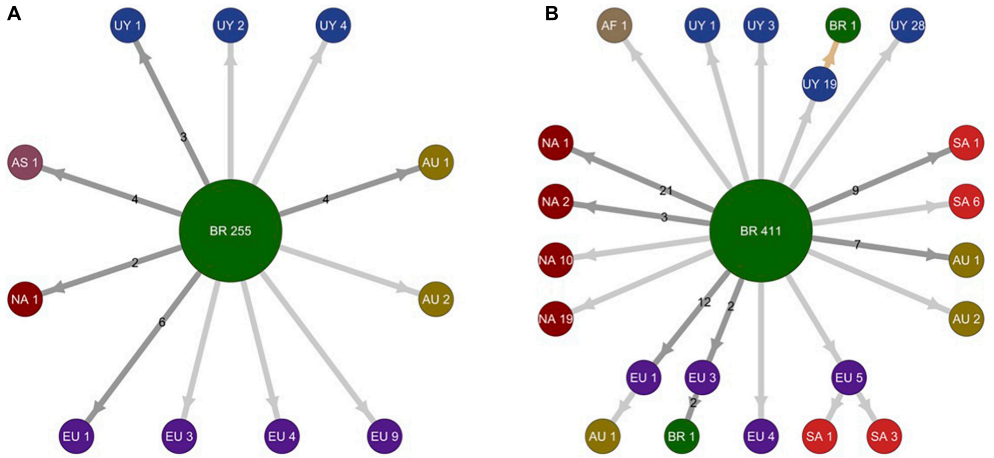


FIGURE 3 | Schematic representation of migration events during worldwide dissemination of SARS-CoV-2 lineages B.1.1.28 and B.1.1.33. The picture depicts the migration events in SARS-CoV-2 lineages B.1.1.28 **(A)** (*n* = 298 sequences) and B.1.1.33 **(B)** (*n* = 578 sequences) inferred by ancestral character reconstruction obtained through a maximum likelihood (ML) method implemented in PastML. Each node in the network is identified by location and number of sequences within different phylogenetic subclusters. Arrows indicate migration events deduced from location state changes across both trees (B.1.1.28 and B.1.1.33). Dark gray arrows identify multiple migration events and numbers in the arrows quantify the number of migration events connecting respective locations. Light gray arrows identify unique migration events. The light brown arrow points out to a single inferred migration event from Uruguay to Brazil. Nodes are colored according to their location. AF, Africa; AS, Asia; AU, Australia; BR, Brazil; EU, Europe; NA, North America; SA, South America; UY, Uruguay.

To more accurately infer the geographic source and timing of virus introductions in Uruguay, we conducted Bayesian-based phylogeographic and molecular clock analyses of Brazilian and Uruguayan sequences. Viral migrations were inferred using a discrete diffusion model between 13 locations: four Brazilian regions (Southeast, Northeast, North, and Central-West), three southern Brazilian states (RS, Paraná, and Santa Catarina), and six Uruguayan locations (AR, CL, MO, TT, RI, and RO). Bayesian analyses recovered the same Uruguayan and Brazilian clades previously detected by the ML analyses, with two exceptions: the clade RI-II$_{33}$ branched outside the clade BR-UY-II$_{33}$ and no Brazilian sequences branched with high support with clade AR-II$_{28}$. Bayesian analyses pointed that the Brazilian southeast region was the most probable source [posterior state probability (PSP) $\geq$ 0.96] of all B.1.1.28 and B.1.1.33 viruses introduced in Uruguay, with the exception of clade RI-I33, a B.1.1.28 sequence from RI that most probably originated in Brazil-RS (PSP $\geq$ 0.85), and the B.1.1.33 strain from Montevideo that probably originated in the North Brazilian region (PSP = 1) (**Figure 4** and **Supplementary Figures 3, 4**).

The temporal origin of the lineages B.1.1.28 and B.1.1.33 was traced back to February 2020 and the T_{MRCA} of major Brazilian–Uruguayan clades were inferred at the following: March 12 (95% HPD: March 09–21) for clade BR-UY-I$_{33}$, March 23 (95% HPD: March 14–31) for clade BR-RS-I$_{33}$, March 29 (95% HPD: March 12–April 13) for clade BR-UY-II$_{33}$, May 2 (95% HPD: April 27–May 04) for clade RI-I$_{33}$, May 26 (95% HPD: May 12–June 07) for clade TT-I$_{33}$, July 10 (95% HPD: June 27–July 18) for clade RI-II$_{33}$, and July 2 (95% HPD: June 17–July 16) for clade AR-I$_{28}$. The detection lag (time interval between cluster T_{MRCA} and the first detected case) of major B.1.1.33 Uruguayan outbreaks was estimated at 3 days (95% HPD: 1–8 days) for clade RI-I$_{33}$ and 23 days (95% HPD: 11–37 days) for clade TT-I$_{33}$. The control lag (time interval between the first detected case and the last transmission event) was estimated at 25 days (95% HPD: 20–28 days) for clade RI-I$_{33}$ and 10 days (95% HPD: 6–12 days) for clade TT-I$_{33}$. Though we obtained large confidence intervals, it is worth emphasizing that we were able to assess a remarkable 30% of each outbreak. A larger number of samples might not solve the issue (see discussion on cryptic circulation).

SARS-CoV-2 Lineage-Defining SNPs and Intra-Patient Viral Diversity

Severe acute respiratory syndrome coronavirus 2 Brazilian–Uruguayan clades here identified displayed several lineage-defining SNPs (single nucleotide polymorphisms/mutations) that were sequentially fixed during evolution and dissemination of each lineage (**Figure 4** and **Supplementary Table 3** in **Supplementary Material Section 2**). The Uruguayan/Brazilian lineages (BR-UY-I$_{33}$, BR-UY-II$_{33}$, and BR-UY-I$_{28}$) and the Brazilian lineage BR-RS-I$_{33}$ were all defined by one SNP. One SNP also characterized all sequences from clade TT-I$_{33}$ plus some basal Brazilian strains from São Paulo, two SNPs defined clade TT-I$_{33}$, and one additional SNP characterized all except one basal sequence of clade TT-I$_{33}$. Two SNPs characterized all sequences from clade RI-I$_{33}$ plus some basal Brazilian strains from RS, one SNP defined clade RI-I$_{33}$, and three SNPs defined clade RI-II$_{33}$. The clade AR-I$_{28}$ was defined by 10 SNPs (**Supplementary Table 3** in **Supplementary Material Section 2**).

To understand the dynamics of lineage-defining SNPs, we examined within-host diversity patterns by analyzing the mapped reads at every SNP genomic position in samples from the two major B.1.1.33 Brazilian–Uruguayan clades (**Supplementary Tables 4, 5** in

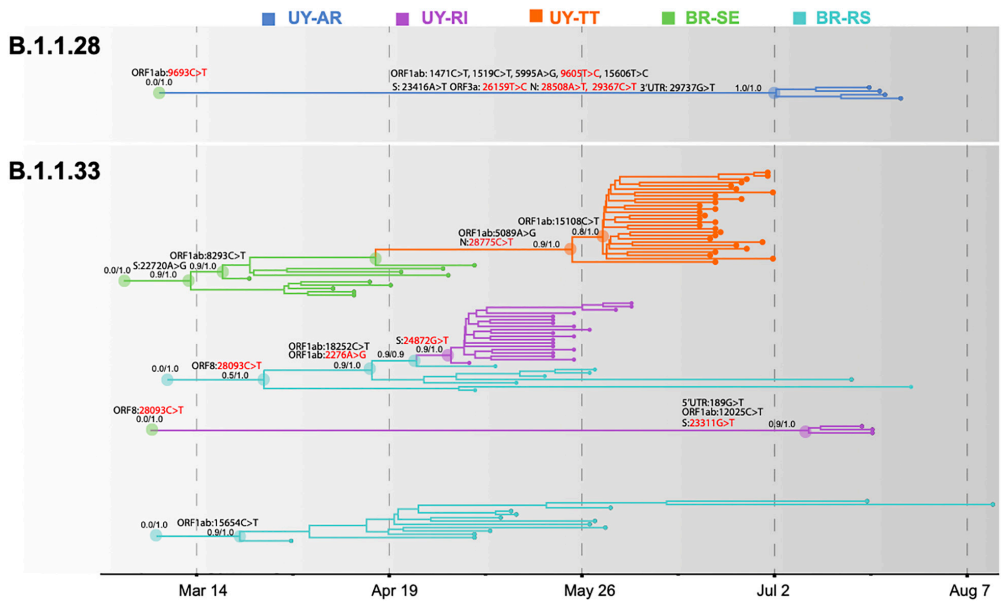


FIGURE 4 | Spatiotemporal dissemination of SARS-CoV-2 Uruguayan–Brazilian clades. Uruguayan–Brazilian clades inferred on the time-scaled Bayesian phylogeographic MCC tree (**Supplementary Figures 3, 4**) plotted independently. Branches are colored according to the most probable location state of their descendant nodes as indicated at the legend. Posterior probability/posterior state probability support values are indicated at key nodes. Synonymous (black) and non-synonymous (red) substitutions fixed at ancestral nodes are shown.

Supplementary Material Section 2). Our analysis of SNP positions reveals a clear-cut allocation of mutant alleles, which displayed either very low (<10%) or high (>80%) frequency within each clade (**Figures 5A,B**). Importantly, no sequences appeared in the 10–65% frequency range (**Figures 5A,B**). The most notable example was the mutation ORF1ab:15108C > T fixed during the dissemination of clade TT-I_{33} that appeared at a very low frequency (<5%) in the basal Uruguayan sample M70 (**Figure 5A**). Indeed, the estimated frequency of the mutant allele in basal sequences from a given clade was not higher than the corresponding frequency detected in sequences from a different clade (**Figure 5C**). While allele frequencies were calculated from three different technologies (Illumina, Ion Torrent, and ONT) and this could introduce some bias (**Supplementary Figures 1, 2** in **Supplementary Material Section 3**), for most samples only ONT data were available. It has been shown that in our conditions (we only kept for analyses those synapomorphic sites with at least 100× read coverage), ONT sequencing enables detection of within-specimen single-nucleotide variants at frequencies equal to and higher than ∼40% with adequate accuracy (Bull et al., 2020). Besides, considering intermediate allele frequencies as those in 40–60% range (Bull et al., 2020), small differences in accuracy and precision would not change our main finding regarding the absence of sequence with intermediate frequencies in this dataset.

DISCUSSION

Public health measures implemented by authorities were able to contain the early local expansion of SARS-CoV-2 in Uruguay, but the intensive human mobility across the contiguous dry borderland with Brazil can compromise long-term control of the epidemic. Indeed, our analyses showed that the COVID-19 epidemic in the Uruguayan–Brazilian border is mostly dominated by the Brazilian lineages B.1.1.33 and B.1.1.28 and that independent introductions of these lineages seeded the early outbreaks of SARS-CoV-2 detected in Uruguayan departments at the bordering region.

Our phylogeographic analyses support that both SARS-CoV-2 lineages most probably arose in the Southeast Brazilian region (PSP = 1) during February 2020, consistent with previous findings (Candido et al., 2020; Resende et al., 2020c), and were later disseminated at multiple times into the Southern region, making up 79% of all viral strains from the RS state. We distinguished two B.1.1.33 local sub-clades that together comprise 45% of all B.1.1.33 strains from RS: the clade BR-RS-I_{33} that arose by the end of March seems to have remained restricted to this Brazilian state, and the clade BR-UY-II_{33} that arose between late March and early April was later disseminated to Uruguay. These findings indicate that the COVID-19 epidemic in the Southern Brazilian state of RS was mostly seeded by epidemics from the Southeastern region.

The molecular epidemiologic profile of the COVID-19 epidemic at Uruguayan border localities (dominated by lineages B.1.1.33 and B.1.1.28) coincides with the observed pattern in RS state and differs from the profile of early outbreaks in the capital city of Montevideo (dominated by lineages A.5 and B.1) (Elizondo et al., 2020; Salazar et al., 2020). Our phylogeographic analyses support that the Southeast Brazilian region and RS were the most probable sources of six and two viral introductions, respectively, into Uruguay. This finding, however,

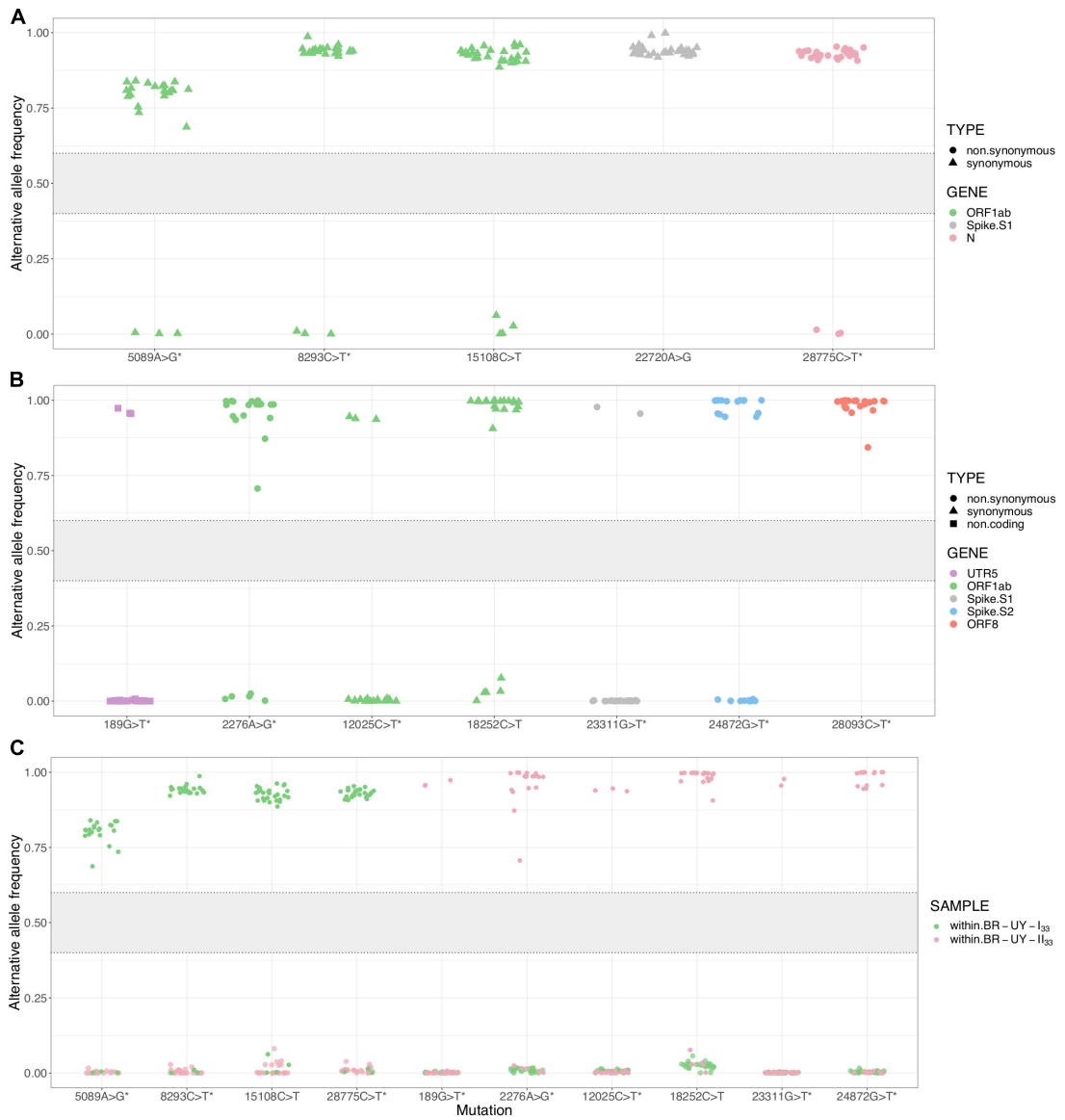


FIGURE 5 | Within-host diversity in synapomorphic sites for the SARS-CoV-2 clades BR-UY-I_{33} and BR-UY-II_{33}. **(A)** Alternative allele frequencies for the clade BR-UY-I_{33}, as observed from trimmed bam files. Positions and annotations follow Wuhan's reference sequence MN908947.3; corresponding gene annotation and mutation types are indicated by color and shape, respectively. The asterisks indicate the synapomorphic sites where the MUTATION explanatory variable shows a significant coefficient in the implemented linear model (p-value ≤ 0.05). In other words, the allele frequencies for these sites are robust to sequencing technology. **(B)** Same as **(A)**, but for the clade BR-UY-II_{33}. For both clades, the absence of alleles with intermediate frequencies ($0.4 < f < 0.6$) is notorious. **(C)** Alternative allele frequencies for all synapomorphies within both BR-UY-I_{33} and BR-UY-II_{33} clades. For each mutation, there is an overlap in the alternative allele frequencies of those samples that do not show the synapomorphy, regardless of the clade they belong to. The gray zone in the plot shows the intermediate allele frequency range (40–60%) as defined by Bull et al. (2020).

should be interpreted with caution because a potential sampling bias across Brazilian regions might influence phylogeographic reconstructions and the actual number of viral introductions from RS into Uruguay might be higher than here estimated (Lemey et al., 2020).

Our study supports that some of the viral strains introduced in the Uruguayan cities of Rivera, Treinta y Tres, and Bella Unión between early May and mid-July were able to spread locally. The most successfully disseminated local lineages were RI-I_{33} and TT-I_{33} that drove the major early outbreaks detected in Rivera and Treinta y Tres in May and June, respectively. Our analysis supports that SARS-CoV-2 Uruguayan clades displayed a variable period of local cryptic transmission, lasting between 1 and 3 weeks. Furthermore, the major Uruguayan clades took between 10 days and nearly a month to be controlled. Long periods of cryptic circulation and/or long times to control represent a potential risk of viral dissemination from Uruguayan border localities into the metropolitan region.

Our phylogenetic analyses identified several sub-clusters inside the B.1.1.28 and B.1.1.33 lineages defined by a set of synonymous and non-synonymous mutations with respect to the ancestral viruses that arose in Brazil in February 2020. The number of defining mutations ranges from one to 11 SNPs and correlates with the lineage T_{MRCA}, as lineages with the youngest T_{MRCA} displayed more SNPs than those with older ones. This is consistent with a molecular clock model (Duchene et al., 2020) and with the notion that most SNPs are probably evolutionarily neutral (van Dorp et al., 2020). Thus, while our study emphasizes the rapid genetic differentiation of regional SARS-CoV-2 lineages in South America and the relevance of molecular epidemiologic studies to track their emergence and spread, the observed SNPs should not be interpreted as an adaptive signature of higher viral transmissibility (MacLean et al., 2020).

Previous studies support that intra-host SARS-CoV-2 genetic diversity is transmissible, and information about within-host diversity could provide further resolution for identification of geographic transmission clusters (Lythgoe et al., 2020) and to trace how low-frequency mutations become fixed in local clusters (Popa et al., 2020). Our analyses of intra-host diversity, however, revealed that mutations present at low frequency (<10%) in basal strains of major Uruguayan–Brazilian clusters become highly prevalent (>80%) during subsequent transmissions, without evidence of basal samples with intermediate-frequency alternative alleles. These findings support that minor intra-host SARS-CoV-2 variants could become rapidly fixed due to stochastic events during inter-host transmissions and later perpetuated along transmission chains (Böhmer et al., 2020).

In summary, our study demonstrates for the first time a recurrent viral flux across the Uruguayan–Brazilian border, with multiple introductions of the SARS-CoV-2 Brazilian lineages B.1.1.28 and B.1.1.33 into Uruguayan localities at the border region and the potential of those viral introductions to ignite local outbreaks in Uruguay. These findings provide clear evidence that public health measures for viral control were not able to fully suppress the dissemination of SARS-CoV-2 along the Uruguayan-Brazilian border, and this possesses a major challenge for long-term epidemic control in Uruguay. The present work also highlights the relevance of bi-national consortiums for genomic surveillance of SARS-CoV-2 to obtain in-depth insights and better control of viral spread across highly permissive international borders.

DATA AVAILABILITY STATEMENT

Genome sequences were submitted to the EpiCoV/GISAID database, and accession numbers are specified in **Supplementary Tables 1, 2** and in **Supplementary Section 2**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the SASA Ethics Committee: CEI-SASA FIOCRUZ-IOC Ethics Committee: 68118417.6.0000.5248, and Brazilian Ministry of Health SISGEN: A1767C3. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LS made significant contributions in the post-processing of raw sequencing data, study design, coordination of Uruguayan group, funding acquisition, verifying underlying data, and writing the original draft. DM was in charge of phylogenetic and phylogeographic analyses, verified underlying data, and wrote original draft. NRg made significant contribution in the post-processing of raw sequencing data, verifying underlying data, variant analysis, and writing the original draft. TF-C was in charge of sequencing in Uruguay and read and corrected the manuscript. VN is the head of a diagnostic center in the Sanatorio Americano, obtained funding, and reviewed and edited the manuscript. MA, TP, and NRy worked on the diagnosis in Sanatorio Americano and read the manuscript. GB made significant contributions in the study design, coordination between Uruguayan and Brazilian institutions, coordination of phylogeographic and phylogenetic analyses, verification of underlying data, and writing the original draft. RC made significant contributions in the coordination of diagnostic centers in the Uruguayan non-metropolitan area, study design, and editing and review of the manuscript. FT worked on the diagnosis in CENUR-Norte, Uruguay, worked on nanopore sequencing in Uruguay centers, and edited and reviewed the manuscript. MB worked on nanopore sequencing in Uruguay centers and read the manuscript. MG made significant contributions in the technical support at the Institut Pasteur de Montevideo, helped in writing the grant for funding, and carefully edited and reviewed the manuscript. CSa made significant contributions in the bioinformatic analysis of sequencing data from Uruguayan centers and reviewed the manuscript. MV, AL, MC, LMay, and MSa worked on diagnosis in CENUR-Norte, Uruguay, and reviewed the manuscript. YV worked on diagnosis in Laboratorio DILAVE/MGAP-INIA-UdelaR-Tacuarembó, Uruguay, and reviewed the manuscript. CA worked on diagnosis in CURE-Rocha, Uruguay, and reviewed the manuscript. CSe worked on nanopore sequencing at Institut Pasteur de Montevideo, Uruguay, and reviewed the manuscript. IF made significant contributions in the software and reviewed the manuscript. PS and JS worked on Ion Torrent sequencing at Instituto de Investigaciones Biológicas Clemente Estable, Uruguay, and reviewed the manuscript. RF and CM made significant contributions in sequencing design, library preparation, and sequencing (Ion Torrent). MB-G worked on logistic support and reviewed the manuscript. IA worked on phylogenetic and phylogeographic analyses and reviewed the manuscript. PR coordinated the Brazilian SARS-CoV-2 sequencing team using Nanopore, Illumina, and Ion Torrent platforms at the National Reference Laboratory Fiocruz; analyzed the raw sequencing data; and revised and edited the manuscript.

TG and MM performed the extraction and the SARS-CoV-2 detection by real-time RT-PCR locally in the Rio Grande do Sul state and revised the manuscript. LMar is responsible for the epidemiological team in the Rio Grande do Sul state, collected all the epidemiological data from Brazilian samples included in this study, and revised the manuscript. FM is responsible for extraction and the SARS-CoV-2 confirmation by real-time RT-PCR at the National Reference Laboratory Fiocruz and revised the manuscript. LA and AM are responsible for whole-genome SARS-CoV-2 amplification, library construction, sequencing, and revision of the manuscript. MMS made significant contribution in financial support acquisition and revision and editing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Funding support in Brazil was from CGLab/MoH (General Laboratories Coordination of Brazilian Ministry of Health), CVSLR/FIOCRUZ (Coordination of Health Surveillance and Reference Laboratories of Oswaldo Cruz Foundation), CNPq COVID-19 MCTI 402457/2020-0, and INOVA VPPCB-005-FIO437 20-2. In Uruguay, this work was funded by the Manuel Perez Foundation that nucleated COVID-19 donation funds in the context of the project "Vigilancia epidemiológica del COVID-19 en las fronteras uruguayas y anaílisis de su transmisión en el interior del paiís." This work was also supported by FOCEM – Fondo para la Convergencia Estructural del Mercosur (COF 03/11). GB was funded by productivity research fellowships from the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Grant No. 302317/2017-1) and the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (Grant No. E-26/202.896/2018). Members of the Institut Pasteur de Montevideo were supported by FOCEM-Fondo para la Convergencia Estructural del Mercosur (COF03/11).

ACKNOWLEDGMENTS

We thank to the kind donors for their contribution to this project. We thank to Carolina Banchero and Rosina Segui for beautiful figure design. We thank to Hugo Naya for helping with statistical models and Luisa Berna for fruitful discussions. We are grateful to the Laboratory of Experimental Evolution of Viruses in charge Gonzalo Moratorio, the Laboratory of Microbial Genomics in charge Gregorio Iraola, the Unit of Molecular Biology in charge Carlos Robello, and all of the people from the Institut Pasteur de Montevideo, for making their facilities available at any time for our experiments and sharing their experience, as well as equipment, plastic supplies, and reagents. We thank Gabriela Ortiz and Paula Aguerrebere of the IAC (Instituto Asistencial Colectivo, Treinta y Tres) for their willingness and hard work in contacting patients. We also thank the workforce of the technical group of the Laboratory for Respiratory Viruses at LACEN/CEVS/SES-RS and the support of the Center for Scientific and Technological Development (CDCT/CEVS/SES-RS). Finally, we thank for efforts from different groups that contribute with SARS-CoV-2 genomes to the EpiCoV GISAID initiative (**Supplementary Tables 9, 10** in **Supplementary Material Section 2**).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.653986/full#supplementary-material

REFERENCES

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Arnson, C. J., Gedan, B., and Prusa, A. (2020). *Contagion: Brazil's COVID-19 Catastrophe Spills Over Borders*. Washington, DC: Wilson Center. Available online at: https://www.wilsoncenter.org/blog-post/contagion-brazils-covid-19-catastrophe-spills-over-borders (accessed december 2, 2020)

Böhmer, M. M., Buchholz, U., Corman, V. M., Hoch, M., Katz, K., Marosevic, D., et al. (2020). Articles Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *Lancet Infect. Dis.* 20, 920–928. doi: 10.1016/S1473-3099(20)30314-5

Bull, R. A., Adikari, T. N., Ferguson, J. M., Hammond, J. M., Stevanovski, I., Beukers, A. G., et al. (2020). Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat. Commun.* 11:6272. doi: 10.1038/s41467-020-20075-6

Candido, D. S., Claro, I. M., de Jesus, J. G., Souza, W. M., Moreira, F. R. R., Dellicour, S., et al. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 369, 1255–1260. doi: 10.1126/science.abd2161

Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., and Baele, G. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2. *bioRxiv* 077735. [pre-print]. doi: 10.1101/2020.05.04.077735

Elizondo, V., Harkins, G. W., Mabvakure, B., Smidt, S., Zappile, P., Marier, C., et al. (2020). SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay. *Emerg. Microb. Infect.* 2020:1863747. doi: 10.1080/22221751.2020.1863747

Ferreira, M., and Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *Canad. J. Statist.* 36:5550360302. doi: 10.1002/cjs.5550360302

github (2020). *Estadísticas epidemiológicas de Uruguay*. San Francisco: github. Available online at: https://guiad-covid.github.io/data/estadisticasuy/. (accessed december 2, 2020)

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33, W557–W559. doi: 10.1093/nar/gki352

Ishikawa, S., Zhukova, A., Iwasaki, W., and Gascuel, O. (2019). A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Mol. Biol. Evolut.* 36, 2069–2085. doi: 10.1093/molbev/msz131

Lemey, P., Hong, S. L., Hill, V., Baele, G., Poletto, C., Colizza, V., et al. (2020). Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat. Commun.* 11:5110. doi: 10.1038/s41467-020-18877-9

Lythgoe, K. A., Hall, M., Ferretti, L., de Cesare, M., MacIntyre-Cockett, G., Trebes, A., et al. (2020). Within-host genomics of SARS-CoV-2. *bioRxiv* 118992. [preprint]. doi: 10.1101/2020.05.28.118992

MacLean, O. A., Orton, R. J., Singer, J. B., and Robertson, D. L. (2020). No evidence for distinct types in the evolution of SARS-CoV-2. *Virus Evolut.* 6:veaa034. doi: 10.1093/ve/veaa034

Moreno, P., Moratorio, G., Iraola, G., Fajardo, A., Aldunate, F., Pereira, M., et al. (2020). An effective COVID-19 response in South America: the Uruguayan Conundrum. *medRxiv* [preprint]. doi: 10.1101/2020.07.24.20161802

Popa, A., Genger, J. W., Nicholson, M., Penz, T., Schmid, D., Aberle, S. W., et al. (2020). Mutational dynamics and transmission properties of SARS-CoV-2 superspreading events in Austria. *bioRxiv* 204339 [preprint]. doi: 10.1101/2020.07.15.204339

Quick, J. (2020). nCoV-2019 sequencing protocol V.1. *Protocols* 2020:bbmuik6w. doi: 10.17504/protocols.io.bbmuik6w

Rambaut, A., Holmes, E. C., O'Toole, Á, Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407. doi: 10.1038/s41564-020-0770-5

Resende, P. (2020). Long reads nanopore sequencing to recover SARS-CoV-2 whole genome V.3. *Protocols* 2020:bfy7jzn. doi: 10.17504/protocols.io.bfy7jpzn

Resende, P. C., Couto Motta, F., Roy, S., Appolinario, L., Fabri, A., Xavier, J., et al. (2020a). SARS-CoV-2 genomes recovered by long amplicon tiling multiplex approach using nanopore sequencing and applicable to other sequencing platforms. *bioRxiv* 069039. [preprint]. doi: 10.1101/2020.04.30.069039

Resende, P. C., Delatorre, E., Gräf, T., Mir, D., Couto Motta, F., Appolinario, L., et al. (2020c). Genomic surveillance of SARS-CoV-2 reveals community transmission of a major lineage during the early pandemic phase in Brazil. *bioRxiv* 2020:158006. doi: 10.1101/2020.06.17.158006

Resende, P. C., Delatorre, E., Gräf, T., Mir, D., Motta, F., and Appolinario, L. (2020b). Evolutionary dynamics and dissemination pattern of the SARS-CoV-2 lineage B.1.1.33 during the early pandemic phase in Brazil. *Front. Microbiol.* 11:615280. doi: 10.3389/fmicb.2020.615280

Salazar, C., Díaz-Viraqué, F., Pereira-Gómez, M., Ferrés, I., Moreno, P., Moratorio, G., et al. (2020). Multiple introductions, regional spread and local differentiation during the first week of COVID-19 epidemic in Montevideo, Uruguay. *bioRxiv* 086223 [preprint]. doi: 10.1101/2020.05.09.086223

saude (2021). *Painel Coronavírus RS*. Available online at: https://ti.saude.rs.gov.br/covid19/ (accessed April 30, 2021).

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 Evolut.* 4:1000520. doi: 10.1371/journal.pcbi.1000520

Valcarcel, B., Avilez, J. L., Torres-Roman, J. S., Poterico, J. A., Bazalar-Palacios, J., La Vecchia, C., et al. (2020). The effect of early-stage public health policies in the transmission of COVID-19 for South American countries. *Rev. Panam Salud Publica.* 44:e148. doi: 10.26633/RPSP.2020.148

van Dorp, L., Richard, D., Tan, C. C. S., Shaw, L. P., Acman, M., and Balloux, F. (2020). No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat. Commun.* 11:5986. doi: 10.1038/s41467-020-19818-2

Worldometers (2020). *COVID-19 CORONAVIRUS PANDEMIC*. Wuhan: Worldometers. Available online at: https://www.worldometers.info/coronavirus/ (accessed december 15, 2020)

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.

frontiers in Microbiology

REVIEW
published: 07 June 2021
doi: 10.3389/fmicb.2021.689979

Molecular Epidemiology, Evolution and Reemergence of Chikungunya Virus in South Asia

*Nadim Sharif[1], Mithun Kumar Sarkar[1], Rabeya Nahar Ferdous[2], Shamsun Nahar Ahmed[1], Md. Baki Billah[3], Ali Azam Talukder[1], Ming Zhang[4] and Shuvra Kanti Dey[1]**

[1] Department of Microbiology, Jahangirnagar University, Savar, Bangladesh, [2] Department of Microbiology, Bangladesh University of Health Sciences, Dhaka, Bangladesh, [3] Department of Zoology, Jahangirnagar University, Savar, Bangladesh, [4] Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA, United States

Chikungunya virus (CHIKV) is a vector (mosquito)-transmitted alphavirus (family *Togaviridae*). CHIKV can cause fever and febrile illness associated with severe arthralgia and rash. Genotypic and phylogenetic analysis are important to understand the spread of CHIKV during epidemics and the diversity of circulating strains for the prediction of effective control measures. Molecular epidemiologic analysis of CHIKV is necessary to understand the complex interaction of vectors, hosts and environment that influences the genotypic evolution of epidemic strains. In this study, different works published during 1950s to 2020 concerning CHIKV evolution, epidemiology, vectors, phylogeny, and clinical outcomes were analyzed. Outbreaks of CHIKV have been reported from Bangladesh, Bhutan, India, Pakistan, Sri Lanka, Nepal, and Maldives in South Asia during 2007–2020. Three lineages- Asian, East/Central/South African (ECSA), and Indian Ocean Lineage (IOL) are circulating in South Asia. Lineage, ECSA and IOL became predominant over Asian lineage in South Asian countries during 2011–2020 epidemics. Further, the mutant E1-A226V is circulating in abundance with *Aedes albopictus* in India, Bangladesh, Nepal, and Bhutan. CHIKV is underestimated as clinical symptoms of CHIKV infection merges with the symptoms of dengue fever in South Asia. Failure to inhibit vector mediated transmission and predict epidemics of CHIKV increase the risk of larger global epidemics in future. To understand geographical spread of CHIKV, most of the studies focused on CHIKV outbreak, biology, pathogenesis, infection, transmission, and treatment. This updated study will reveal the collective epidemiology, evolution and phylogenies of CHIKV, supporting the necessity to investigate the circulating strains and vectors in South Asia.

Keywords: chikungunya virus, evolution, epidemiology, phylogeny, South Asia

OPEN ACCESS

Edited by:
Kai Huang, University of Texas Medical Branch at Galveston, United States

Reviewed by:
Adriano de Bernardi Schneider, University of California, San Diego, United States
Eleonora Cella, University of Central Florida, United States
Jagadish Mahanta, Indian Council of Medical Research (ICMR), India

***Correspondence:**
Shuvra Kanti Dey
shuvradey@yahoo.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *01 April 2021*
Accepted: *10 May 2021*
Published: *07 June 2021*

Citation:
Sharif N, Sarkar MK, Ferdous RN, Ahmed SN, Billah BM, Talukder AA, Zhang M and Dey SK (2021) Molecular Epidemiology, Evolution and Reemergence of Chikungunya Virus in South Asia. Front. Microbiol. 12:689979. doi: 10.3389/fmicb.2021.689979

INTRODUCTION

Chikungunya virus (CHIKV) (family *Togaviridae*) is a mosquito-borne (arthropod-borne virus) *Alphavirus* that causes chikungunya fever in humans (International Committee on Taxonomy of Viruses (ICTV), 2021). Chikungunya virus was first characterized during 1952 to 1953 from a dengue-like outbreak in Newala district, Tanzania. CHIKV is considered as a member of

Semliki Forest virus (SFV) antigenic group and the word "chikungunya" is derived from Makonde word, Bantu language and translated as "bent over in pain," describing the posture of CHIKV infected patients (Robinson, 1955; Casals and Whitman, 1957; Spence and Thomas, 1959; Hammon et al., 1960; Jupp et al., 1988; Ekstrom et al., 1994; Gubler, 2001; Weaver et al., 2005; Weaver and Lecuit, 2015). About 3–5 million cases of CHIKV are reported every year globally (World Health and Organization (WHO), 2021). Although mortality due to CHIKV infection is rare, infection of CHIKV can cause severe and long term health conditions in patients (Deller and Russell, 1968). CHIKV can cause both symptomatic and asymptomatic infections in humans (Deller and Russell, 1968; Silva et al., 2018). Symptomatic CHIKV infection is categorized into acute, chronic and atypical depending on the manifestation of symptoms (Fourie and Morrison, 1979; Silva et al., 2018; Vairo et al., 2019). Several clinical symptoms of CHIKV infection overlap with the symptoms of dengue virus (DENV) and Zika virus (ZIKV) infections (Myers and Carey, 1967; Nimmannitya et al., 1969; Carey, 1971; Edwards et al., 2016; Furuya-Kanamori et al., 2016; Gaibani et al., 2016; Rodriguez-Morales et al., 2016; Jain et al., 2018; Ng et al., 2018; Silva et al., 2018; Zanotto and Leite, 2018). The co-circulation of CHIKV with other significant arboviruses, such as dengue virus, Zika virus, mayaro virus (MAYV), and yellow fever virus (YFV) in tropical regions of Asia with overlapping symptoms requires continuous epidemiological surveillance and effective differential diagnosis strategies (Halstead et al., 1969; Weaver and Lecuit, 2015; Vairo et al., 2019).

Chikungunya virus is a small (∼70 nm-diameter), enveloped virus with a linear, positive strand RNA genome of ∼11.8 kilo-bases (Khan et al., 2002; Kawashima et al., 2014). The RNA genome consists of one non-translated region (NTR) at 5′, two ORFs and another non-translated region (NTR) at 3′ end. Two polyproteins are encoded by two major open reading frames (ORFs) in CHIKV (Simizu et al., 1984; Schlesinger and Schlesinger, 1986). The positive-sense 5′ two-third RNA genome directly encodes a polyprotein containing four non-structural proteins (nsP1–4). The structural proteins are encoded by 3′ one-third of the genome (Simizu et al., 1984). The structural polyprotein converts into a capsid protein, two major envelope surface glycoproteins (E1 and E2) as well as two small peptides, E3 and 6K (Simizu et al., 1984).

Transmission of CHIKV involves two major cycles depending on the region of circulation. In African regions, CHIKV circulates mainly in a sylvatic/enzootic cycle involving forest dwelling mosquitoes and non-human primates (NHP) (Silva et al., 2018; Simo et al., 2019). The viruses rely on NHP as reservoir (e.g., monkeys and other vertebrates) hosts during inter-epidemic periods and transmitted by *Aedes* (e.g., *Aedes furcifer* and *Aedes africanus*) mosquitoes from reservoirs to human during epidemic (Jupp et al., 1988; Jupp and McIntosh, 1990; Pialoux et al., 2007; Powers and Logue, 2007; LaBeaud et al., 2011). On the contrary in urban cycles, a mosquito to (and from) human transmission is maintained. The urban cycle of CHIKV has been associated with several large epidemics of CHIKV across different continents including Asia, Europe, and North America (Weaver and Lecuit, 2015; Silva et al., 2018). Two significant species of mosquitoes namely, *Aedes aegypti* and *Aedes albopictus* (the "tiger" mosquito) are mainly involved in urban transmission of the disease (Soekiman, 1987; McGill, 1995; Reiter et al., 2006; Weaver and Lecuit, 2015; Wahid et al., 2017; Silva et al., 2018; Vairo et al., 2019). In temperate climates, *Ae. albopictus* mosquitoes thrive in high density (Soekiman, 1987; Reiter et al., 2006). *Ae. albopictus* are expanding into and adapting to new areas with the potential transmission capability of CHIKV and have been involved in recent epidemics in Asia (Soekiman et al., 1986a,b; Soekiman, 1987; Reiter et al., 2006; Weaver and Lecuit, 2015).

Persistent and larger epidemics of CHIKV infecting millions of people have been reported from Asia, but limited numbers of epidemiological research has been undertaken (Powers and Logue, 2007; Wahid et al., 2017; Vairo et al., 2019). South Asian regions are endemic for CHIKV epidemics. Continuous surveillance including phylogenetic, evolutionary, and epidemiologic analyses are required in these endemic regions to catch up the contemporary changes in the virus for developing effective diagnostics, treatments, and vaccines (Myers and Carey, 1967; Edwards et al., 2016; Rodriguez-Morales et al., 2016; Ng et al., 2018). In this review, updated epidemiology, evolution and phylogenomics of CHIKV in South Asia during 2004–2020 have been evaluated. Clinical features of chikungunya fever, transmission in temperate and tropical regions and the laboratory testing for the disease are also described in this study. Particularly, this study focuses on the recent trends of CHIKV epidemic in South Asia to create an integrated baseline for future studies.

MOLECULAR EPIDEMIOLOGY, PHYLOGENY AND EVOLUTION OF CHIKUNGUNYA VIRUS IN SOUTH ASIA AND REST OF THE WORLD

Starting from Africa, CHIKV has been transmitted globally. Recently, CHIKV infection has been detected from different countries on all continents, except Antarctica. Retrospective case studies have suggested that CHIKV epidemics have occurred during 1760s (Ross, 1956; Jupp et al., 1988). During early epidemics, they were inaccurately documented as dengue virus infection (Brighton et al., 1983; Silva and Dermody, 2017). CHIKV was first isolated and characterized from the serum of an infected patient with dengue like symptoms in Tanzania during 1952 to 1953 (Casals and Whitman, 1957; Johnson et al., 1977). CHIKV was detected in South Asia in a short time after identification in East Africa (1952) (Ross, 1956; Powers and Logue, 2007; Silva and Dermody, 2017; Wahid et al., 2017; Simo et al., 2019). After the first identification, local and occasional outbreaks of CHIKV were recorded for the following ∼50 years before 2004 in many countries in Asia (Mason and Haddow, 1957; Peyrefitte et al., 2007; Silva and Dermody, 2017; Simo et al., 2019).

Chikungunya virus has remained endemic in South Asian countries since 1960s (Powers and Logue, 2007; Wimalasiri-Yapa et al., 2019). The first significant CHIKV outbreak in Asia was

reported in the early 1960s in Bangkok, Thailand (Marchette et al., 1978; Powers and Logue, 2007). Outbreaks have been documented from seven countries out of eight in South Asia (Powers et al., 2000; Powers and Logue, 2007; Wimalasiri-Yapa et al., 2019). In SA and SEA, local and minor outbreaks of CHIKV were reported during the 1960s to 1980s in India, Indonesia, Malaysia, Cambodia, Vietnam, Myanmar, Pakistan, and Thailand (Pavri, 1964; Bedekar and Pavri, 1969; Burke et al., 1985; Kit, 2002; Porter et al., 2004; Zim et al., 2013; Weaver and Lecuit, 2015; Silva and Dermody, 2017; Wimalasiri-Yapa et al., 2019). Before 1985, most of the CHIKV outbreaks dated during the period 1961–1970 including several large cities in SEA namely, Kolkata and Bangkok as the main active sites of transmission of CHIKV (Brighton et al., 1983; Powers and Logue, 2007; Wahid et al., 2017). During 1985–2000, no significant outbreaks of CHIKV was documented in South Asian countries (Powers and Logue, 2007; Mascarenhas et al., 2018). Larger outbreaks involving more cases in India, Bangladesh, Pakistan, Sri Lanka, and Maldives have been documented recently (Powers and Logue, 2007; Wimalasiri-Yapa et al., 2019). About 85% of cases of CHIKV had been detected after 2000 in South Asian countries (Padbidri and Gnaneswar, 1979; Powers and Logue, 2007; Hapuarachchi et al., 2010; Manimunda et al., 2010; Haque et al., 2019; Vairo et al., 2019; Wimalasiri-Yapa et al., 2019). After 2007, CHIKV outbreaks in SEA and SA regions were larger and longer that infected millions of people in India, Bangladesh, Bhutan, Nepal, Thailand, and Philippines (**Supplementary Table 1**; Centers for Disease Control and Prevention (CDC), 2021). In SA, seven countries (Bangladesh, Bhutan, India, Pakistan, Sri Lanka, Nepal, and Maldives) out of eight have reported the local outbreaks and epidemics of CHIKV during 2010 to 2020 (**Figures 1A–C**, **2**; Mallhi et al., 2017; Wahid et al., 2017; Haque et al., 2019; Wimalasiri-Yapa et al., 2019; Centers for Disease Control and Prevention (CDC), 2021).

India has remained as the most affected country with CHIKV (Ravi, 2006; Powers and Logue, 2007; Wimalasiri-Yapa et al., 2019). Increased number of cases of CHIKV were reported after 2007 in India (Saxena et al., 2006; Mavalankar et al., 2007; Seyler et al., 2010; Murhekar et al., 2019). In 2007–2008, outbreak of CHIKV was reported in Bangladesh that continued in 2013, 2014, 2016, and 2017 (Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). In Pakistan, larger outbreaks were reported during 2010–2018 (Powers and Logue, 2007). Further, cases of CHIKV has increased in Maldives, Sri Lanka, Nepal, and Bhutan after 2010 (**Supplementary Table 1**; Hapuarachchi et al., 2010; Reller et al., 2013; Vairo et al., 2019; Wimalasiri-Yapa et al., 2019). The CHIKV epidemics in South Asia is affected significantly by the ongoing epidemics in other regions in the world. Both the emergence and reemergence of CHIKV in South Asia were linked with outbreaks in Africa. CHIKV was detected and reported repeatedly from various countries in Central and Southern Africa namely, the Central African Republic (CAR), Democratic Republic of Congo (DRC), Malawi, Sudan, Uganda, Zimbabwe, Kenya, and South Africa during the 1960s to the 1990s (Moore et al., 1974; Thonnon et al., 1999; Powers and Logue, 2007; Weaver and Lecuit, 2015; Silva and Dermody, 2017; Wahid et al., 2017; Silva et al., 2018; Simo et al., 2019). After 2005, the outbreaks in Africa significantly contributed in reemergence of CHIKV in South Asia. Reemergence of CHIKV occurred in India in 2007, 2009, 2011, 2015, 2017, and 2019, in Bangladesh during 2009–2011, 2013–2015, 2016, 2017, and 2019, and in Pakistan during 2009, 2011, 2013–2016 (Mallhi et al., 2017; Wahid et al., 2017; Haque et al., 2019; Wimalasiri-Yapa et al., 2019; Centers for Disease Control and Prevention (CDC), 2021).

As of 2020, local and autochthonous outbreaks have been reported from about 114 countries and territories (Centers for Disease Control and Prevention (CDC), 2021). The first larger outbreak occurred in coastal Kenya in 2004 (Powers and Logue, 2007; Sergon et al., 2007; Gudo et al., 2016; Silva and Dermody, 2017; Simo et al., 2019). In 2004, two large outbreaks started in Kenya (Powers and Logue, 2007; Weaver and Forrester, 2015; Vairo et al., 2019). This outbreak in Kenya transmitted to the Union of the Comoros by January 2005 (Pastorino et al., 2004). From there CHIKV was transmitted to the surrounding locations in the Indian Ocean. During 2005–2007, about 500,000 cases (1/3 of the population) were documented on La Reunion Island (Josseran et al., 2006; Powers and Logue, 2007; Renault et al., 2007; Mascarenhas et al., 2018). From Indian Ocean, the CHIKV epidemic spread to India, infecting ∼1.5 million individual. Further, the virus was transmitted to Indonesia, Maldives, Sri Lanka, Myanmar, Thailand, and other countries in Asia (Laras et al., 2005; Parola et al., 2006; Silva and Dermody, 2017). First autochthonous transmission in the Americas occurred in 2013 (Yactayo et al., 2016; Humphrey et al., 2017; Mascarenhas et al., 2018). As of 2020, millions of cases of CHIKV have been reported from Latin American countries including Brazil, Bolivia, Colombia, Argentina, Cuba, Costa Rica, Ecuador, and Peru (Powers and Logue, 2007; Rodrigues Faria et al., 2016; Yactayo et al., 2016; Cunha et al., 2020; Centers for Disease Control and Prevention (CDC), 2021; Pan American and Health Organization (PAHO), 2021; World Health and Organization (WHO), 2021). Further, CHIKV infection has transmitted to the Oceania/Pacific islands including the Marshall Islands, American Samoa, Cook Islands, Samoa, French Polynesia, and Kiribati in 2014 (**Figures 1A–C**; Powers and Logue, 2007; Staples et al., 2009; Vairo et al., 2019; Centers for Disease Control and Prevention (CDC), 2021).

According to previous lineage systems, four lineages namely, East-, Central- and South African lineage (ECSA), Asian Urban lineage (AUL), West African lineage (WA), and Indian Ocean lineage (IOL) were defined (Powers and Logue, 2007; Silva and Dermody, 2017). Recently, the most updated classification system based on 1,066 genomes sampled between February 1953 and December 2019 included nine lineages namely, Asian urban (AUL), AUL-America (AUL-Am), South America (SAL), Middle Africa (MAL), Indian Ocean (IOL), East Africa (EAL), Africa and Asia (AAL), Sister Taxa to ECSA (sECSA) and West Africa (WA) and named according to the regions of origin (de Bernardi Schneider et al., 2019; Nextstrain, 2021; Spicher et al., 2021). Further, genotypes of CHIKV was specified based on partial or complete E1 gene sequencing (Presti et al., 2016; Phadungsombat et al., 2020). Molecular epidemiologic and evolutionary analysis

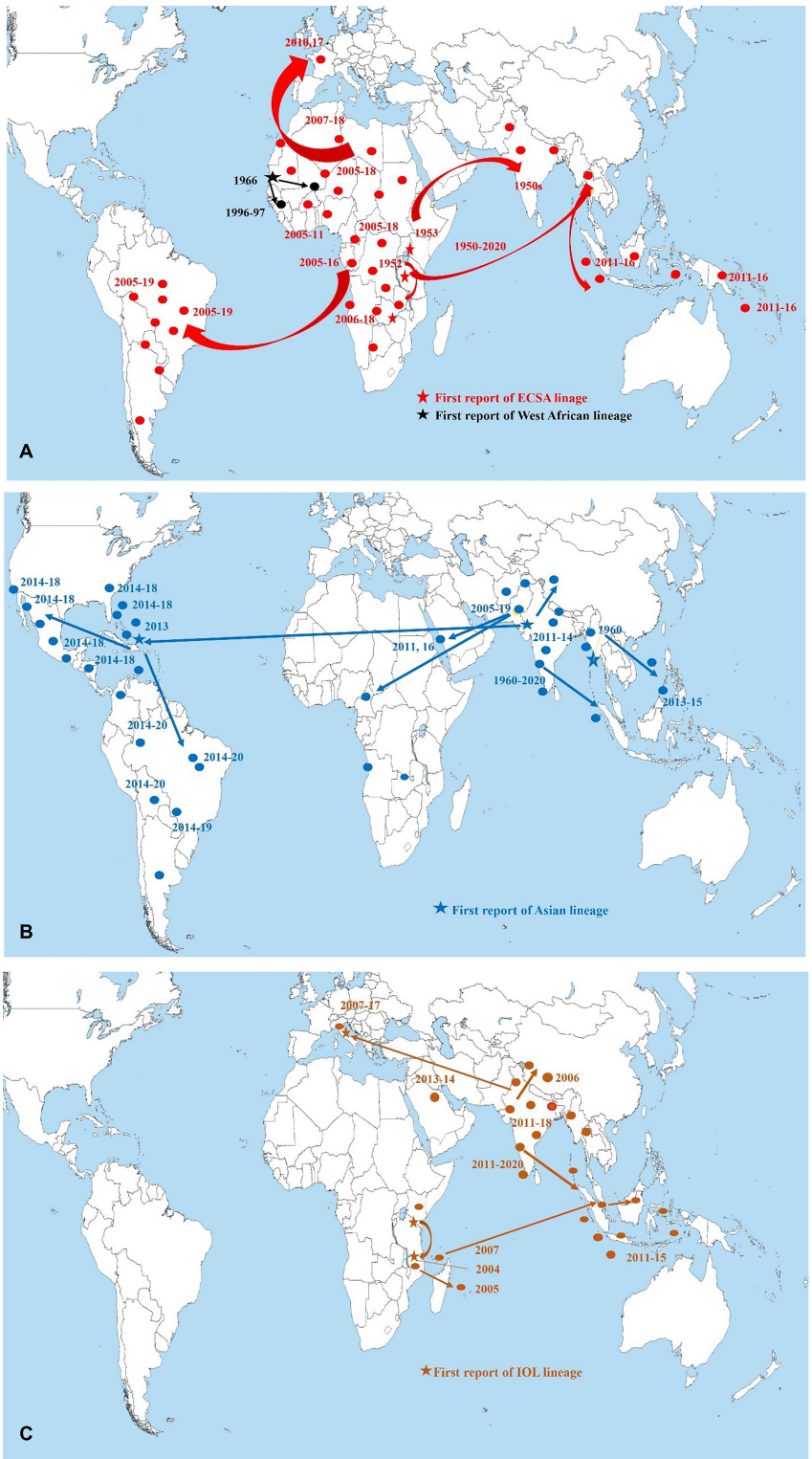


FIGURE 1 | Map depicting worldwide distribution of chikungunya virus lineages during 1950s to 2020. Star signs indicate the first report/start of large outbreak of chikungunya virus. Circles indicate local outbreaks. **(A)** Red circles, arrows, and stars indicate ECSA lineage, black circles, arrows, and stars indicate West African lineage. **(B)** Navy blue circles, arrows, and stars indicate Asian lineage. **(C)** Brown circles, arrows, and stars indicate IOL lineage (Centers for Disease Control and Prevention (CDC), 2021; European Centre for Disease Prevention and Control (ECDC), 2021; Pan American and Health Organization (PAHO), 2021; World Health and Organization (WHO), 2021). The arrows indicated the probable spread of CHIKV based on the information extracted from Centers for Disease Control and Prevention (CDC), 2021; European Centre for Disease Prevention and Control (ECDC), 2021; Pan American and Health Organization (PAHO), 2021; World Health and Organization (WHO), 2021. Data were extracted directly from the databases and websites and cross-checked. The map was built using ArcGIS Pro software.

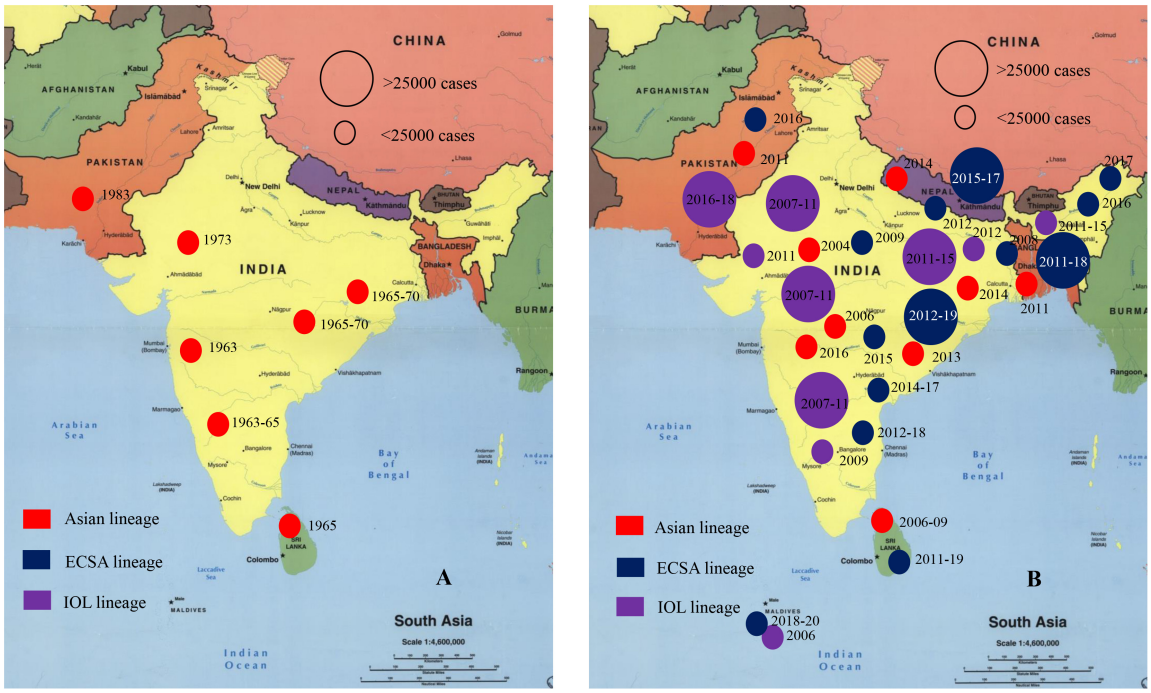


FIGURE 2 | Onset of chikungunya virus outbreak in South Asia regions. **(A)** Indicates the case number of CHIKV patients during 1950s to 2003 in South Asia. **(B)** Indicates the case number of CHIKV patients during 2004 to 2020 in South Asian countries (Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). Star represents capital and large cities.

confirmed circulation of several lineages of CHIKV in South Asia (Powers and Logue, 2007; Presti et al., 2016).

East-, Central- and South African lineage is considered as the ancestor and has circulated in South Asia from Africa at early 1960s. The second confirmed lineage, AUL, was first detected in outbreaks in Asian countries (Thailand, India, Cambodia, Vietnam, Malaysia, Taiwan, Myanmar, and Indonesia) during 1958 to 1973 and named as Asian lineage (Powers and Logue, 2007; Presti et al., 2016). Another distinct lineage called Indian Ocean Lineage (IOL) evolved from ECSA lineage (Schuffenecker et al., 2006; Presti et al., 2016; Silva and Dermody, 2017; Pyke et al., 2018) was detected after 2004. Phylogenetic and mutational analysis revealed the presence of Asian lineage, IOL and Africa and Asia lineages in South Asian countries in recent time (Wahid et al., 2017; Deeba et al., 2020). Asian lineage was the most predominant during 1960s to 2000 in South Asian and South East Asian countries (Yadav et al., 2003; Powers and Logue, 2007; Wimalasiri-Yapa et al., 2019). After 2004, the IOL transmitted rapidly in the South Asian countries. After 2005 outbreaks, IOL lineage has been reported from most of the outbreaks (80%) in South Asian countries (Powers and Logue, 2007; Wimalasiri-Yapa et al., 2019). Since 2005, outbreak associated with IOL lineage have emerged and reemerged every year in South Asian countries (Wimalasiri-Yapa et al., 2019; Phadungsombat et al., 2020). Besides, ESCA, AUL, and AAL lineages are circulating in South Asia in a low frequency after 2005 outbreaks. Evolutionary analysis has revealed that during 2010 to 2020, outbreaks in Bangladesh, Bhutan, India, Pakistan, and Sri Lanka have been associated with ECSA-IOL lineage (Powers and Logue, 2007; Wahid et al., 2017; Melan et al., 2018; Deeba et al., 2020). During 2017 outbreaks in Bangladesh, only the ECSA-IOL lineage was reported (**Supplementary Table 1**). Other countries of South East Asia, namely, Bhutan, Myanmar, and Vietnam had reported the presence of IOL lineage during recent outbreaks (Powers and Logue, 2007; Staples et al., 2009; Pyke et al., 2018; Wimalasiri-Yapa et al., 2019).

Molecular evolutionary analysis confirmed the divergence of lineages from each other in previous studies and Nextstrain project and recently published phylogenies and evolutionary analysis (de Bernardi Schneider et al., 2019; Deeba et al., 2020; Nextstrain, 2021; Spicher et al., 2021). The single nucleotide variants of CHIKV can change the stability and fold of locally stable RNA structures. Besides, the 3′ untranslated regions of CHIKV was found to contain non-structural RNA elements and evolutionary conserved regions (de Bernardi Schneider et al., 2019; Deeba et al., 2020; Spicher et al., 2021). Difference among lineages and origin of one lineage from other can be traced by analyzing duplication events and changes of architecture in 3′UTR (de Bernardi Schneider et al., 2019). An estimation of average evolutionary divergence over sequence pairs within CHIKV lineages was calculated in previous studies by following the maximum likelihood model (de Bernardi Schneider et al., 2019). The number of base substitutions per site was expressed from averaging over all sequence pairs within each group and found that AUL was most divergent (substitution per site was 0.0128) followed by MAL (0.0107)

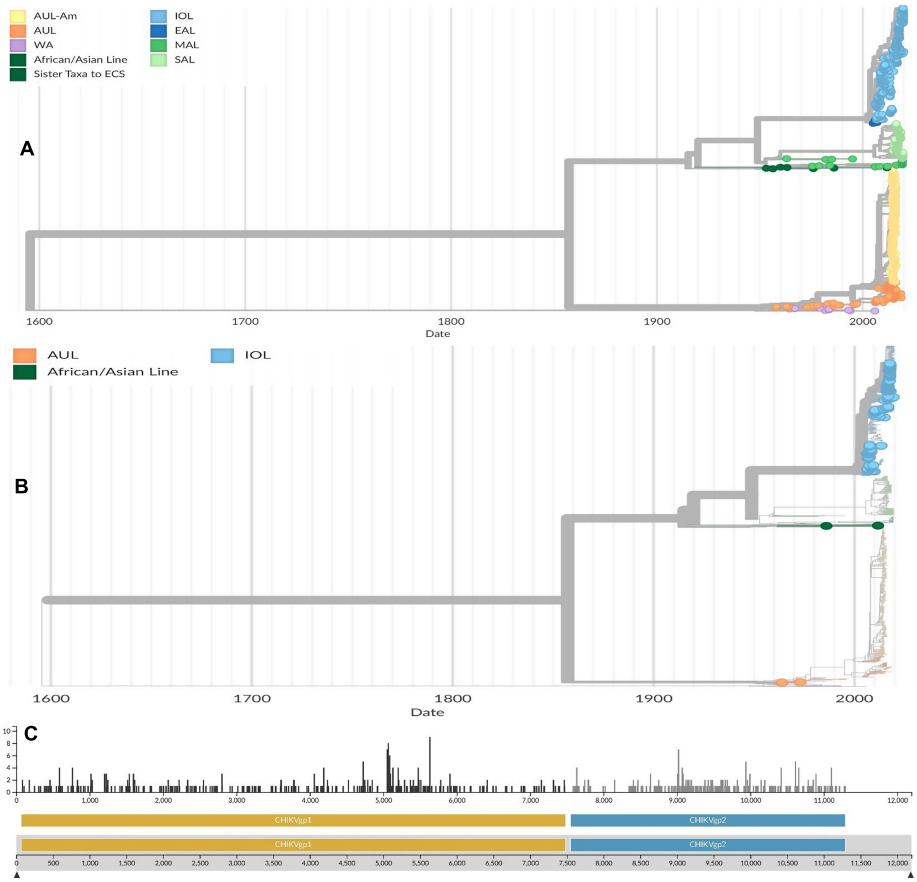


FIGURE 3 | (A) Phylogenetic tree of globally distributed 1,066 genomes of chikungunya viruses sampled between February 1953 and December 2019. The phylogenetic tree was retrieved and from Nextstrain and modified based on the most updated information on CHIKV (de Bernardi Schneider et al., 2019; Spicher et al., 2021). **(B)** Phylogenetic tree of 157 whole genomes of CHIKV circulating in South Asia (India-88, Bangladesh-40, Sri Lanka-22, and Pakistan-7) sampled between February 1953 and December 2019. The trees were conducted using Maximum likelihood model. The trees were computed using bootstrap value of 1,000. Reference sequences were selected using temporal and spatial emergence of isolates during outbreaks in South Asia (de Bernardi Schneider et al., 2019; Spicher et al., 2021). **(C)** Mutational analysis of available reference genomes of CHIKV circulating in South Asian countries. The horizontal scale indicated the nucleotide base position in 5′ to 3′ direction, while the vertical scale indicated the number of substitutions per position in the genome (Nextstrain, 2021). Lineages are indicated as Asian urban (AUL), AUL-America (AUL-Am), South America (SAL), Middle Africa (MAL), Indian Ocean (IOL), East Africa (EAL), Africa and Asia (AAL), Sister Taxa to ECSA (sECSA), and West Africa (WA) (Nextstrain, 2021). Source: https://nextstrain.org/community/ViennaRNA/CHIKV.

and WA (0.0102), while SAL was least divergent (0.003). In four countries, the available 157 whole genome of CHIKV in Nextstrain (India-88, Bangladesh-40, Sri Lanka-22, and Pakistan-7) had an estimated $2.63e^{-4}$ substitution per site per year (**Figures 3A–C**). The number of mutations including substitutions are higher in CHIKVgp1 within 5,000 bases to 6,000 bases position, while in CHIKVgp2 the frequency of mutation is about 1.9 per position within 8,500 bases to 11,000 bases.

TRANSMISSION OF CHIKV

Chikungunya virus is transmitted in humans by infected mosquitoes (Diallo et al., 1999). CHIKV is an enzootic virus in tropical regions of Africa and Asia (Paul and Singh, 1968; Mourya, 1987; Silva et al., 2018). Emergence and reemergence of CHIKV is significantly regulated by the transmission of the virus through vectors. To understand the reemergence potential in South Asia, this review covered the transmission of CHIKV. Generally, an uninfected mosquito takes in CHIKV from infected viremic person during ingesting the blood (Mourya and Banerjee, 1987; Monteiro et al., 2019). The virus is replicated inside the mosquito midgut. When CHIKV carrying mosquitoes bite a healthy individual, the virus is transmitted inside his/her body (Silva et al., 2018). The virus also replicates inside newly infected person body (Silva and Dermody, 2017). If another uninfected mosquito bite the newly infected person after he has become viremic, the mosquito will take in CHIKV and start another cycle (Silva et al., 2018; Onyango et al., 2020). The complete transmission cycle from human to mosquito and back to humans can be completed within a week (Diallo et al., 1999). Mosquitoes can act as vectors

of CHIKV. Both vertical and horizontal transmissions of the virus can occurred in mosquitoes (Mavale et al., 2010; Jain et al., 2016). For successful transmission from arthropod vectors to a human, CHIKV must replicate inside the vectors and reach the salivary glands within 1 week (Lim et al., 2018).

Transmission of CHIKV is maintained by sylvatic cycle in the African and urban cycle in the Asian regions (Jupp and McIntosh, 1990; Staples et al., 2009; Monteiro et al., 2019). In South Asia, the urban mosquito *Ae. aegypti* and *Ae. albopictus* have been reported to be the most significant vector (Soekiman et al., 1986a,b; Banerjee et al., 1988; Mourya et al., 1994; Diallo et al., 1999; Scolari et al., 2019). Regional large outbreaks in South Asia are caused by these urban and peridomestic mosquitoes (Soekiman et al., 1986a,b; Scolari et al., 2019; Wimalasiri-Yapa et al., 2019). *Ae. albopictus* have a great adaption capacity in new ecological niches, as a result it can expand its enzootic range globally (Silva et al., 2018). In urban cycles in South Asia, the onset of epidemics are dependent on environmental factors, viral genetics, mosquito ecology, human behavior, and presence of competent vectors (Soekiman et al., 1986a,b). During the 2005–2006 Indian Ocean Islands epidemic, a substitution point mutation originated at position 226 in the E1 glycoprotein (outer membrane protein) of CHIKV, replacing an Alanine to Valine (Weaver and Lecuit, 2015; Silva et al., 2018). This mutation in ECSA genotype of CHIKV enhanced the vector specificity and epidemic potential of CHIKV (Kumar et al., 2008). The new mutants of CHIKV namely, IOL of ECSA genotype became capable of surviving in and transmitting by *Ae. albopictus* (Tsetsarkin et al., 2016). The E1-A226V substitution increases viral infectivity in *Ae. albopictus* midgut cells without compromising viral replication. This mutant strain initiated autochthonous cases of CHIKV more rapidly through *Ae. albopictus* in South Asian countries (Soekiman et al., 1986a,b; Vazeille et al., 2007; Scolari et al., 2019). Further, *Ae. furcifer–taylori* is the main group of vectors detected during epidemics associated with sylvatic cycle (Jupp and McIntosh, 1990; Scolari et al., 2019). *Ae. furcifer, Aedes taylori, Aedes luteocephalus, Ae. africanus*, and *Aedes neoafricanus* are the major species of vectors involved in sylvatic cycles for many years (Mourya and Banerjee, 1987; Jupp and McIntosh, 1990; Monteiro et al., 2019; Scolari et al., 2019). Numerous field and laboratory works have been undertaken on roles of mosquito vectors in the transmission of CHIKV, but less is known about the importance of vertebrate hosts in viral maintenance (Powers and Logue, 2007; Silva et al., 2018). Laboratory animal studies and serosurveys confirmed the presence of CHIKV specific antibodies in potential vertebrate reservoirs. Significant levels of antibody against CHIKV have been detected in wild non-human primates (Silva et al., 2018).

Vectors and vertebrates have significant roles in inter-epidemic periods both in the sylvatic and urban transmission. In sylvatic cycles, non-human primate (NHP) species including Guinea baboons, Chacma baboons, African green monkeys, patas monkeys, red-tail monkeys, guenons, bushbabies, and mandrills may have significant roles as amplifiers hosts or reservoirs of CHIKV (Silva et al., 2018). On the contrary, in urban cycle, the mosquitoes play main roles probably by *trans*-ovarian (vertical transmission) cycles (Silva et al., 2018).

CLINICAL FEATURES OF PATIENTS INFECTED WITH CHIKV

To understand the complete epidemiological prospects of CHIKV burden in South Asian countries, studies on the clinical manifestations in patients infected with CHIKV are required. Generally, the incubation period of CHIKV in human ranges from 3 days to 7 days (Munasinghe et al., 1966; Powers and Logue, 2007; Staples et al., 2009; Onyango et al., 2020). Most of the studies on CHIKV infection clinical presentation reported that about 70–93% of the patients develop symptoms, 3–25% seropositive patients may be asymptomatic, and 2–7% patients may develop atypical symptoms (Powers and Logue, 2007; Staples et al., 2009; Silva et al., 2018; Suhrbier, 2019; Wimalasiri-Yapa et al., 2019). The most reported triad of clinical signs and symptoms for CHIKV infection from documented epidemics and outbreaks includes fever, arthralgia (joint pain), and a rash (itchy rash) (Munasinghe et al., 1966; Riswari et al., 2016). Most of the time the triad is accompanied by other symptoms of the CHIKV infection. Generally, epidemics of CHIKV infection result in two clinical outcomes of illness including the acute phase and chronic phase (Silva et al., 2018; Suhrbier, 2019). In most of the cases, fever accompany with the joint pain and rash. Rash is reported from 50 to 60% cases (Silva and Dermody, 2017). The non-itchy rash becomes visible during 2–5 days of post-infection (Powers and Logue, 2007; Silva et al., 2018). After fever, the most significant clinical presentation of CHIKV infection is the severe joint pain (arthralgia) (Powers and Logue, 2007; Suhrbier, 2019). Arthralgia is reported from about 90 to 98% of CHIKV cases (Brighton and Simson, 1984; Silva and Dermody, 2017). Besides triad, weakness, malaise, headache, chills, retro-orbital pain, photophobia, lumbar back pain, conjunctivitis, pharyngitis, lymphadenopathy and myalgia are other common symptoms reported with CHIKV infection (**Table 1**; Silva and Dermody, 2017; Silva et al., 2018; Suhrbier, 2019; Wimalasiri-Yapa et al., 2019). Most infections completely resolve within weeks or months but there have been documented cases of CHIKV-induced arthralgia persisting for several years with up to 12% of patients with CHIKV disease developing chronic joint problems (Powers and Logue, 2007; Silva et al., 2018). A comparison of clinical manifestations associated with CHIKV in South Asia and rest of the world is presented in **Figure 4**.

Long term sequelae associated with CHIKV have been reported in previous studies worldwide (Andrei and De Clercq, 1993; Suhrbier, 2019; Wimalasiri-Yapa et al., 2019). This condition is called CHIKV induced chronic arthralgia (Fourie and Morrison, 1979; Tesh, 1982). Aged persons with prior rheumatological disease are the main victims of chromic symptoms (Tesh, 1982; Adebajo, 1996; Arpino et al., 2009; Barr et al., 2018; Suhrbier, 2019). CHIKV can cause numerous unusual clinical complications in patients (Arpino et al., 2009; Economopoulou et al., 2009; Acevedo et al., 2017;

TABLE 1 | Distribution of typical and atypical acute clinical manifestations of patients infected with chikungunya virus during 2004 to 2020 in Africa and Asia (Centers for Disease Control and Prevention (CDC), 2021; European Centre for Disease Prevention and Control (ECDC), 2021; Pan American and Health Organization (PAHO), 2021; World Health and Organization (WHO), 2021).

Organ/System	Percentage of patients (*N* = 1022108)	Typical acute clinical manifestations	Atypical acute clinical manifestations (%)
Systemic	85–100%	Fever	Lymphadenopathy (5)
		Asthenia	
Musculoskeletal	70–90%	Arthralgia	Articular destruction (7)
		Arthritis	Bleeding gums
		Myalgia	
		Joint edema	
		Tenosynovitis	
		Backache	
		Relapsing-remitting polyarthralgias	
Skin	65–75%	Rash	Bullous dermatosis
		Erythema	Hyperpigmentation
		Bodyache	Stomatitis
			Melena
			Xerosis (10)
Neurological	45–55%	Headache	Meningoencephalitis
		Fatigue	Encephalopathy Seizures
			Sensorineural Abnormalities
			Guillain-Barré syndrome
			Paresis
			Palsies
			Neuropathy (30–40)
Cardiovascular			Myocarditis
			Pericarditis
			Arrhythmias
			Hypotension
			Cardiomyopathy
			Heart failure (20)
Gastrointestinal			Nausea Vomiting
			Abdominal pain
			Anorexia
			Diarrhea (<5)
Hematological	5–10%	Lymphopenia	Hemorrhage (<5)
		Thrombocytopenia	
Respiratory			Dyspnea
			Pneumonia
			Respiratory
			Failure (14–25)
Ocular	5–10%	Retro-orbital pain	Conjunctivitis Photophobia
		Photosensitivity	Retinitis
			Optic neuritis
			Uveitis (<5)
Hepatic			Hepatitis
			Hepatomegaly
			Altered function
			Liver failure (<5)
Renal			Nephritis
			Albuminuria
			Hematuria
			Acute renal failure (20–25)

Godaert et al., 2017; Barr et al., 2018). Occasionally, atypical clinical manifestations have been documented from patients infected with CHIKV in South Asia (Suhrbier, 2019; Wimalasiri-Yapa et al., 2019). Atypical symptoms including myocarditis with

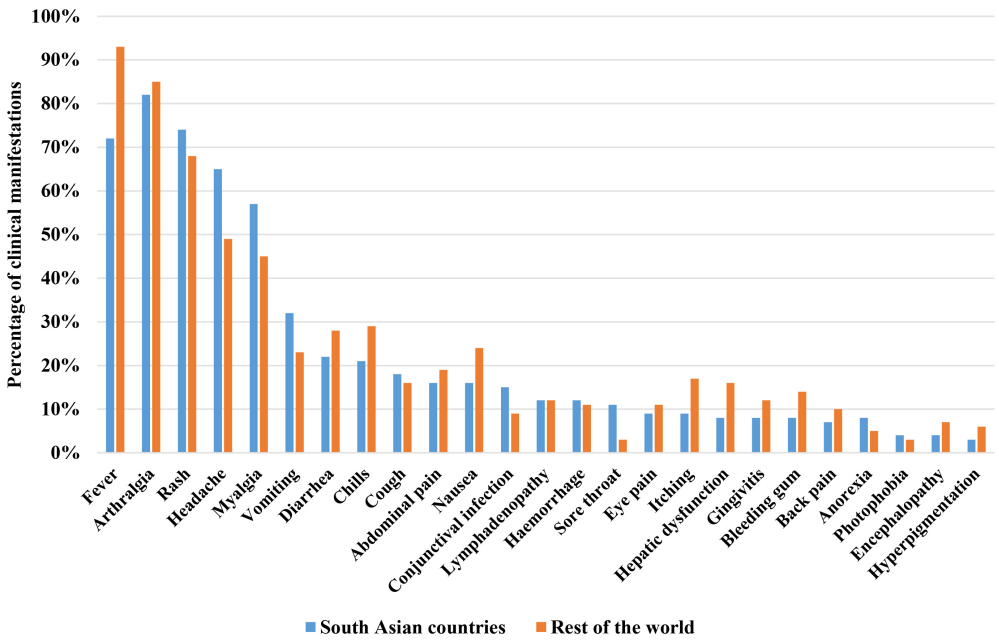


FIGURE 4 | Comparison of clinical manifestations of patients infected with CHIKV in South Asia and rest of the world (Centers for Disease Control and Prevention (CDC), 2021; European Centre for Disease Prevention and Control (ECDC), 2021; Pan American and Health Organization (PAHO), 2021; World Health and Organization (WHO), 2021).

sinus tachycardia, cardiomegaly, ventricular ectopics, abnormal electrocardiograms and finally congestive heart failure have been reported from Asia (Deller and Russell, 1968; Obeyesekere and Hermon, 1972, 1973; Rajapakse et al., 2010; Silva et al., 2018; Deeba et al., 2019; Wimalasiri-Yapa et al., 2019). In 2015, aggressive clinical course developing shock, severe purpuric lesions, a distinct area large of necrosis in the nasal region, bullous dermatosis, acronecrosis of upper limb, rapid onset of septic shock, and multi-organ failure have been detected in CHIKV infected patients (Powers and Logue, 2007; Rajapakse et al., 2010; Bonifay et al., 2018; Suhrbier, 2019). Vertically infected neonates develop various clinical symptoms (Robillard et al., 2006; Beserra et al., 2019). At the time of birth, about half of the viremic mothers transmit the virus to neonates that results in serious consequences to the neonates (Contopoulos-Ioannidis et al., 2018; van Enter et al., 2018; Shen et al., 2020). About 50% infected neonates develop symptoms within 3–7 days. Numerous clinical symptoms including fever, rashes, sepsis- like illness, poor feeding, diffuse limb edema, irritability (hyperalgesia), respiratory distress, meningoencephalitis, nervous system abnormalities, and hemorrhagic and cardiac manifestations have been detected in neonates (Powers and Logue, 2007; van Enter et al., 2018; Kumar et al., 2019).

DIAGNOSIS OF CHIKV INFECTION

Diagnosis of CHIKV infection is conducted on the basis of clinical, epidemiological, and laboratory criteria (Silva et al., 2018; Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). Clinical manifestations of CHIKV infection including abrupt onset of high fever and severe joint pain with rash are the main factors of clinical diagnosis. CHIKV infection is difficult to distinguish and diagnose based on only clinical findings in regions where CHIKV co-circulates with DENV and ZIKV (Calisher, 1999; Cabral-Castro et al., 2016). Diagnostic criteria based on epidemiological findings can include the recent (within last 14 days) travel history of the suspects in areas with CHIKV outbreak or endemics (Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). Laboratory diagnosis of CHIKV infection may include virus isolation, virus characterization, viral RNA detection, and serology (Silva et al., 2018; Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). Immunochromatographic assay targeting the E1 antigen of virus from sera of patients can detect different CHIKV genotypes (Silva et al., 2018; Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). Molecular methods including RT-PCR, RT-LAMP, qRT-PCR are most reliable to diagnose CHIKV because of high sensitivity and specificity (Calvo et al., 2016). In endemic regions of DENV, ZIKV, and *Leptospira* infections differential diagnosis by novel multiplex molecular methods have been introduced to detect CHIKV. Among them, RT-LAMP assay is the most promising in differentiating between ZIKV, CHIKV, and DENV infections (Silva et al., 2018; Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). Further, RT-qPCR assays are also used in differential diagnosis among

TABLE 2 | Commercially available Chikungunya Virus (CHIKV) diagnostic assays in South Asia with principle of functions (Gaibani et al., 2016; Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021).

Manufacturer	Country of origin	Principle	IgM	IgG
Abcam	Germany	IgM human ELISA kit	+	/
CTK Biotech	United States	CHIK IgM combo rapid test	+	/
CTK Biotech	United States	RecombiLISA CHIK IgM	+	
DRG*	Germany	CHIK IgM micro-capture ELISA	+	/
DRG	Germany	ELISA	/	+
Euroimmun	Germany	Anti-CHIKV IIFT	+	+
Euroimmun	Germany	Anti-CHIKV ELISA	+	+
GenWay	Germany	IgM-capture ELISA	+	/
GenWay	Germany	ELISA	/	+
IBL international	Germany	IgM micro-capture ELISA	+	/
IBL international	Germany	IgG-capture ELISA	/	+
InBios	United States	CHIKjj-MAC-ELISA	+	+
Novatec	Germany	IgM-capture ELISA	+	/
Novatec	Germany	ELISA	/	+
SD Diagnostics	South Korea	CHIKa IgM ELISA	+	/
SD Diagnostics	South Korea	SD BIOLINE Chikungunya IgM	+	/

**Only for research purposes.*

ZIKV, CHIKV, and DENV (Carey, 1971; Thaung et al., 1975; Furuya-Kanamori et al., 2016; Villamil-Gómez et al., 2016). Serological methods, such as ELISA and plaque reduction neutralization testing (PRNT) are performed generally (Ayu et al., 2010). Various commercial serological test kits are available worldwide (**Table 2**; Thein et al., 1992; Gaibani et al., 2016; Jain et al., 2018). Diagnosis of CHIKV is prioritized during epidemics, but sporadic cases are often neglected. Further, testing of CHIKV is affected by number and severity of sick people or people suspected of CHIKV infection. Serological test kits and molecular diagnosis can be integrated to detect the sporadic and travelers CHIKV infection rapidly and accurately (Silva et al., 2018; Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). In South Asia, the co-circulation of DENV-CHIKV is one of the main concerns in diagnosis approaches. After the rainy season in 2019, a larger outbreak of DENV was reported to infect about 0.2 million people in Bangladesh (Centers for Disease Control and Prevention (CDC), 2021; World Health and Organization (WHO), 2021). During the outbreak of DENV in Bangladesh CHIKV remained under-diagnosed or undiagnosed that misrepresented the actual burden of CHIKV. Appropriate etiological diagnosis should be achieved through combined clinical, epidemiological and laboratory approaches conducted by expert health professionals.

THE RISK OF CHIKV EPIDEMICS IN SOUTH ASIA IN FUTURE

The previous epidemiological studies support for a larger outbreak of CHIKV in South Asian countries in future. Due to the cyclic nature of CHIKV infection, the epidemics reappear in every 3–4 years in the endemic regions. Further, high density of vectors and co-circulation of CHIKV-DENV during the same seasons are the major concerns in South Asian countries (Thaung et al., 1975; Furuya-Kanamori et al., 2016; Villamil-Gómez et al., 2016). There are various monitoring systems for CHIKV in South Asian countries, but they are only applied during epidemics. As a result, the real disease burden of CHIKV in South Asia still remains underrated. Further, the number of existing genotypic characterization of CHIKV is not enough to point out the diversity in South Asia. The evolutionary and epidemiologic analysis in this article supports for a severe and prolonged epidemics of CHIKV in south Asian countries in near future. To manage such an epidemic in future, this study suggests to conduct routine genotypic surveillance and genomic characterizations of CHIKV to assess the actual diversity of the virus in South Asia. Further, to mitigate the risk of larger epidemic, integrated approaches including epidemiologic characterizations, vector surveillance, evolutionary analysis and effective routine diagnosis are required to reduce the risk of future outbreaks and associated health burden.

CONCLUSION

In conclusion, this study finds that CHIKV has become a consistent health burden in South Asia. Tropical regions of South Asia and South America have been the main focal point of CHIKV transmission after 2007. Recently, larger epidemics of CHIKV involving millions of people have been reported from India, Bangladesh, Nepal, Bhutan, and Pakistan. Three lineages of CHIKV namely, Asian, ECSA and IOL are circulating in South Asia during 2011–2020. After 2011, ECSA lineage and IOL lineage of genotype ECSA has become predominant in South Asian countries. Prevalence of E1-A226V mutants and density of vectors namely, *Aedes aegypti* and *Aedes albopictus* remain high in South Asian countries. This study provides a comprehensive analysis on the updated phylogenomic, evolution and molecular epidemiology of CHIKV in South Asian countries, which will not only provide exact scenario of CHIKV but also help in developing better treatment, diagnosis and preventive measures. Further, this study adds integrated knowledge on recent diagnosis, clinical characteristics and transmission of CHIKV in South Asia. The rapid spread of CHIKV in recent years urges the utmost need to take control measures, as well as to search for options to develop vaccines. In future, more studies focusing the molecular characterizations and evolution of CHIKV, as well as vector-pathogen interaction should be conducted to understand the CHIKV infection in depth. This study will work as an updated database for future studies focusing molecular epidemiology, evolution, phylogeny, diagnosis, vaccine development and prevention of CHIKV in South Asia.

AUTHOR CONTRIBUTIONS

NS performed the systematic and data collection, provided with the illustrations, and was a major contributor in writing the manuscript. MS performed the data analysis and was a major

contributor in revising the manuscript. RF performed the data analysis and was a minor contributor in revising the manuscript. SA performed the data minor analysis. MB performed the minor revision. AT performed the critical evaluation and verification of the manuscript. MZ was a major contributor in revising the manuscript. SD conceptualized the review article and provided oversight, critical evaluation, and verification of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors wish to thank the faculty members of Department of Microbiology, Jahangirnagar University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.689979/full#supplementary-material

Supplementary Table 1 | Documented local outbreaks of chikungunya virus with the distribution of vector species during 1950s to 2020 from more than 110 countries and territories worldwide. Data were retrieved from Centers for Disease Control and Prevention (https://www.cdc.gov/chikungunya/index.html), World Health Organization (https://www.who.int/news-room/fact-sheets/detail/chikungunya), European Centre for Disease Prevention and Control (https://www.ecdc.europa.eu/en/chikungunya-monthly), and Pan American Health Organization (https://www.paho.org/hq/index.php?option=com_topics&view=rdmore&cid=5855&Itemid=40931&lang=en).

REFERENCES

Acevedo, N., Waggoner, J., Rodriguez, M., Rivera, L., Landivar, J., Pinsky, B., et al. (2017). Zika virus, chikungunya virus, and dengue virus in cerebrospinal fluid from adults with neurological manifestations. *Front. Microbiol.* 8:42. doi: 10.3389/fmicb.2017.00042

Adebajo, A. O. (1996). Rheumatic manifestations of tropical diseases. *Curr. Opin. Rheumatol.* 8, 85–89. doi: 10.1097/00002281-199601000-00015

Andrei, G., and De Clercq, E. (1993). Molecular approaches for the treatment of hemorrhagic fever virus infections. *Antiviral. Res.* 22, 45–75. doi: 10.1016/0166-3542(93)90085-W

Arpino, C., Curatolo, P., and Rezza, G. (2009). Chikungunya and the nervous system: what we do and do not know. *Rev. Med. Virol.* 19, 121–129. doi: 10.1002/rmv.606

Ayu, S. M., Lai, L. R., Chan, Y. F., Hatim, A., Hairi, N. N., Ayob, A., et al. (2010). Seroprevalence survey of chikungunya virus in Bagan Panchor. *Am. J. Trop. Med. Hyg.* 83, 1245–1248. doi: 10.4269/ajtmh.2010.10-0279

Banerjee, K., Mourja, D. T., and Malunjkar, A. S. (1988). Susceptibility & transmissibility of different geographical strains of Aedes aegypti mosquitoes to Chikungunya virus. *Indian J. Med. Res.* 87, 134–138.

Barr, K. L., Khan, E., Farooqi, J. Q., Imtiaz, K., Prakoso, D., Malik, F., et al. (2018). Evidence of chikungunya virus disease in Pakistan since 2015 with patients demonstrating involvement of the central nervous system. *Front. Public Health.* 6:186. doi: 10.3389/fpubh.2018.00186

Bedekar, S. D., and Pavri, K. M. (1969). Studies with Chikungunya virus. *Indian J. Med. Res.* 57, 1193–1197.

Beserra, F. L. C. N., Oliveira, G. M., Marques, T. M. A., Farias, L. A. B. G., Santos, J. R. D., Daher, E. D. F., et al. (2019). Clinical and laboratory profiles of children with severe chikungunya infection. *Rev. Soc. Bras. Med. Trop.* 52:e20180232. doi: 10.1590/0037-8682-0232-2018

Bonifay, T., Prince, C., Neyra, C., Demar, M., Rousset, D., Kallel, H., et al. (2018). Atypical and severe manifestations of chikungunya virus infection in French Guiana: A hospital-based study. *PLoS One.* 13:e0207406. doi: 10.1371/journal.pone.0207406

Brighton, S. W., Prozesky, O. W., and de la Harpe, A. L. (1983). Chikungunya virus infection. A retrospective study of 107 cases. *S. Afr. Med. J* 63, 313–315.

Brighton, S. W., and Simson, I. W. (1984). A destructive arthropathy following Chikungunya virus arthritis – a possible association. *Clin. Rheumatol.* 3, 253–258. doi: 10.1007/BF02030766

Burke, D. S., Nisalak, A., and Nimmannitya, S. (1985). Disappearance of Chikungunya virus from Bangkok. *Trans. R. Soc. Trop. Med. Hyg.* 79, 419–420. doi: 10.1016/0035-9203(85)90398-0

Cabral-Castro, M. J., Cavalcanti, M. G., Peralta, R. H. S., and Peralta, J. M. (2016). Molecular and serological techniques to detect co-circulation of DENV, ZIKV and CHIKV in suspected dengue-like syndrome patients. *J. Clin. Virol.* 82, 108–111. doi: 10.1016/j.jcv.2016.07.017

Calisher, C. H. (1999). "Chikungunya, O'nyong nyong and Mayaro viruses (Togaviridae)," in *Encyclopedia of Virology*, 2nd Edn, eds A. Granoff and R. G. Webster (London: Elsevier Science & Technology Books), 236–241.

Calvo, E. P., Sánchez-Quete, F., Durán, S., Sandoval, I., and Castellanos, J. E. (2016). Easy and inexpensive molecular detection of dengue, chikungunya and zika viruses in febrile patients. *Acta. Trop.* 163, 32–37. doi: 10.1016/j.actatropica.2016.07.021

Carey, D. E. (1971). Chikungunya and dengue: a case of mistaken identity? *J. Hist. Med. Allied Sci.* 26, 243–262. doi: 10.1093/jhmas/XXVI.3.243

Casals, J., and Whitman, L. (1957). Mayaro virus: a new human disease agent. I. Relationship to other arbor viruses. *Am. J. Trop. Med. Hyg.* 6, 1004–1011. doi: 10.4269/ajtmh.1957.6.1004

Centers for Disease Control and Prevention (CDC) (2021). *Chikungunya Virus*. Georgia: CDC.

Contopoulos-Ioannidis, D., Newman-Lindsay, S., Chow, C., and LaBeaud, A. D. (2018). Mother-to-child transmission of Chikungunya virus: A systematic review and meta-analysis. *PLoS Negl. Trop. Dis.* 12:e0006510. doi: 10.1371/journal.pntd.0006510

Cunha, M. S., Costa, P. A., Correa, I. A., de Souza, M. R., Calil, P. T., da Silva, G. P. D., et al. (2020). Chikungunya Virus: An Emergent Arbovirus to the South American Continent and a Continuous Threat to the World. *Front. Microbiol.* 11:1297. doi: 10.3389/fmicb.2020.01297

de Bernardi Schneider, A., Ochsenreiter, R., Hostager, R., Hofacker, I. L., Janies, D., and Wolfinger, M. T. (2019). Updated phylogeny of Chikungunya virus suggests lineage-specific RNA architecture. *Viruses* 11:798. doi: 10.3390/v11090798

Deeba, F., Haider, M. S. H., Ahmed, A., Tazeen, A., Faizan, M. I., Salam, N., et al. (2020). Global transmission and evolutionary dynamics of the Chikungunya virus. *Epidemiol. Infect.* 148:e63. doi: 10.1017/S0950268820000497

Deeba, I. M., Hasan, M. M., Al Mosabbir, A., Siam, M. H. B., Islam, M. S., Raheem, E., et al. (2019). Manifestations of Atypical Symptoms of Chikungunya during the Dhaka Outbreak (2017) in Bangladesh. *Am. J. Trop. Med. Hyg.* 100, 1545–1548. doi: 10.4269/ajtmh.19-0122

Deller, J. J. Jr., and Russell, P. K. (1968). Chikungunya disease. *Am. J. Trop. Med. Hyg.* 17, 107–111. doi: 10.4269/ajtmh.1968.17.107

Diallo, M., Thonnon, J., Traore-Lamizana, M., and Fontenille, D. (1999). Vectors of chikungunya virus in Senegal: current data and transmission cycles. *Am. J. Trop. Med. Hyg.* 60, 281–286. doi: 10.4269/ajtmh.1999.60.281

Economopoulou, A., Dominguez, M., Helynck, B., Sissoko, D., Wichmann, O., Quenel, P., et al. (2009). Atypical Chikungunya virus infections: clinical manifestations, mortality and risk factors for severe disease during the 2005–2006 outbreak on Reunion. *Epidemiol. Infect.* 137, 534–541. doi: 10.1017/S0950268808001167

Edwards, T., Signor, L. D. C. C., Williams, C., Donis, E., Cuevas, L. E., and Adams, E. R. (2016). Co-infections with chikungunya and dengue viruses, Guatemala, 2015. *Emerg. Infect. Dis.* 22:2003. doi: 10.3201/eid2211.161017

Ekstrom, M., Liljestrom, P., and Garoff, H. (1994). Membrane protein lateral interactions control Semliki Forest virus budding. *EMBO J.* 13, 1058–1064. doi: 10.1002/j.1460-2075.1994.tb06354.x

European Centre for Disease Prevention and Control (ECDC). (2021). *Chikungunya worldwide overview*. Available online at: https://www.ecdc.europa.eu/en/chikungunya-monthly (accessed Jan 03, 2021).

Fourie, E. D., and Morrison, J. G. (1979). Rheumatoid arthritic syndrome after chikungunya fever. *S. Afr. Med. J.* 56, 130–132.

Furuya-Kanamori, L., Liang, S., Milinovich, G., Magalhaes, R. J. S., Clements, A. C., Hu, W., et al. (2016). Co-distribution and co-infection of chikungunya and dengue viruses. *BMC Infect. Dis.* 16:84. doi: 10.1186/s12879-016-1417-2

Gaibani, P., Landini, M. P., and Sambri, V. (eds) (2016). "Diagnostic methods for CHIKV based on serological tools," in *Chikungunya Virus*, (New York: Humana Press), 63–73. doi: 10.1007/978-1-4939-3618-2_6

Godaert, L., Najioullah, F., Bartholet, S., Colas, S., Yactayo, S., Cabié, A., et al. (2017). Atypical clinical presentations of acute phase chikungunya virus infection in older adults. *J. Am. Geriatr. Soc.* 65, 2510–2515. doi: 10.1111/jgs.15004

Gubler, D. J. (2001). Human arbovirus infections worldwide. *Ann. N. Y. Acad. Sci.* 951, 13–24. doi: 10.1111/j.1749-6632.2001.tb02681.x

Gudo, E. S., Black, J. F., and Cliff, J. L. (2016). Chikungunya in Mozambique: a forgotten history. *PLoS Negl. Trop. Dis.* 10:e0005001. doi: 10.1371/journal.pntd.0005001

Halstead, S. B., Nimmannitya, S., and Margiotta, M. R. (1969). Dengue and chikungunya virus infection in man in Thailand, 1962–1964. II. Observations on disease in outpatients. *Am. J. Trop. Med. Hyg.* 18, 972–983. doi: 10.4269/ajtmh.1969.18.972

Hammon, W. M., Rundnick, A., and Sather, G. E. (1960). Viruses associated with epidemic hemorrhagic fevers of the Philippines and Thailand. *Science* 131, 1102–1103. doi: 10.1126/science.131.3407.1102

Hapuarachchi, H. C., Bandara, K. B. A. T., Sumanadasa, S. D. M., Hapugoda, M. D., Lai, Y. L., Lee, K. S., et al. (2010). Re-emergence of Chikungunya virus in South-east Asia: virological evidence from Sri Lanka and Singapore. *J. Gen. Virol.* 91, 1067–1076. doi: 10.1099/vir.0.015743-0

Haque, F., Rahman, M., Banu, N. N., Sharif, A. R., Jubayer, S., Shamsuzzaman, A. K. M., et al. (2019). An epidemic of chikungunya in northwestern Bangladesh in 2011. *PLoS One.* 14:e0212218. doi: 10.1371/journal.pone.0212218

Humphrey, J. M., Cleton, N. B., Reusken, C. B., Glesby, M. J., Koopmans, M. P., and Abu-Raddad, L. J. (2017). Urban chikungunya in the Middle East and North Africa: a systematic review. *PLoS Negl. Trop. Dis.* 11:e0005707. doi: 10.1371/journal.pntd.0005707

International Committee on Taxonomy of Viruses (ICTV) (2021). *Togaviridae*. Available online at: https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/togaviridae/872/genus-alphavirus (accessed Jan 03, 2021).

Jain, J., Kushwah, R. B. S., Singh, S. S., Sharma, A., Adak, T., Singh, O. P., et al. (2016). Evidence for natural vertical transmission of chikungunya viruses in field populations of Aedes Aegypti in Delhi and Haryana states in India-a preliminary report. *Acta Trop.* 162, 46–55. doi: 10.1016/j.actatropica.2016.06.004

Jain, J., Okabayashi, T., Kaur, N., Nakayama, E., Shioda, T., Gaind, R., et al. (2018). Evaluation of an immunochromatography rapid diagnosis kit for detection of chikungunya virus antigen in India, a dengue-endemic country. *Virol. J.* 15:84. doi: 10.1186/s12985-018-1000-0

Johnson, B. K., Chanas, A. C., Shockley, P., Squires, E. J., Gardner, P., Wallace, C., et al. (1977). Arbovirus isolations from, and serological studies on, wild and domestic vertebrates from Kano Plain, Kenya. Trans. *R. Soc. Trop. Med. Hyg.* 71, 512–517. doi: 10.1016/0035-9203(77)90146-8

Josseran, L., Paquet, C., Zehgnoun, A., Caillere, N., Le Tertre, A., Solet, J. L., et al. (2006). Chikungunya disease outbreak, Reunion Island. *Emerg. Infect. Dis.* 12, 1994–1995. doi: 10.3201/eid1212.060710

Jupp, P. G., and McIntosh, B. M. (1990). Aedes furcifer and other mosquitoes as vectors of chikungunya virus at Mica, northeastern Transvaal, South Africa. *J. Am. Mosq. Control Assoc.* 6, 415–420.

Jupp, P. G., McIntosh, B. M., and Monath, T. P. (1988). "Chikungunya virus disease," in *The Arboviruses: Epidemiology and Ecology. Volume II*, ed. T. P. Monath, (Boca Raton, FL: CRC Press), 137–157. doi: 10.1201/9780429280245-7

Kawashima, K. D., Suarez, L. A. C., Labayo, H. K. M., Liles, V. R., Salvoza, N. C., Klinzing, D. C., et al. (2014). Complete genome sequence of chikungunya virus isolated in the Philippines. *Genome Announc.* 2, e336–e314. doi: 10.1128/genomeA.00336-14

Khan, A. H., Morita, K., del Carmen, Parquet, M., Hasebe, F., Mathenge, E. G., et al. (2002). Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation siteThe GenBank accession number of the sequence reported in this paper is AF369024. *J. Gen. Virol.* 83, 3075–3084. doi: 10.1099/0022-1317-83-12-3075

Kit, L. S. (2002). Emerging and re-emerging diseases in Malaysia. *Asia. Pac. J. Public Health* 14, 6–8.

Kumar, N. P., Joseph, R., Kamaraj, T., and Jambulingam, P. (2008). A226V mutation in virus during the 2007 chikungunya outbreak in Kerala, India. *J. Gen. Virol.* 89, 1945–1948. doi: 10.1099/vir.0.83628-0

Kumar, S., Agrawal, G., Wazir, S., Kumar, A., Dubey, S., Balde, M., et al. (2019). Experience of perinatal and Neonatal Chikungunya virus (CHIKV) infection in a tertiary care neonatal centre during outbreak in North India in 2016: a case series. *J. Trop. Pediatr.* 65, 169–175. doi: 10.1093/tropej/fmy032

LaBeaud, A., Bashir, F., and King, C. H. (2011). Measuring the burden of arboviral diseases: the spectrum of morbidity and mortality from four prevalent infections. *Popul. Health Metr.* 9:1. doi: 10.1186/1478-7954-9-1

Laras, K., Sukri, N. C., Larasati, R. P., Bangs, M. J., Kosim, R., Djauzi, et al. (2005). Tracking the re-emergence of epidemic chikungunya virus in Indonesia. *Trans. R. Soc. Trop. Med. Hyg.* 99, 128–141. doi: 10.1016/j.trstmh.2004.03.013

Lim, E. X. Y., Lee, W. S., Madzokere, E. T., and Herrero, L. J. (2018). Mosquitoes as suitable vectors for alphaviruses. *Viruses* 10:84. doi: 10.3390/v10020084

Mallhi, T. H., Khan, Y. H., Khan, A. H., Tanveer, N., Khan, O. H., and Aftab, R. A. (2017). Commentary: outbreak of Chikungunya in Pakistan. *Front. Public Health.* 5:261. doi: 10.3389/fpubh.2017.00261

Manimunda, S. P., Sugunan, A. P., Rai, S. K., Vijayachari, P., Shriram, A. N., Sharma, S., et al. (2010). Outbreak of chikungunya fever, Dakshina Kannada District, South India, 2008. *Am. J. Trop. Med. Hyg.* 83, 751–754. doi: 10.4269/ajtmh.2010.09-0433

Marchette, N. J., Rudnick, A., Garcia, R., and MacVean, D. W. (1978). Alphaviruses in Peninsular Malaysia. I. Virus isolations and animal serology. *Southeast Asian J. Trop. Med. Public Health.* 9, 317–329.

Mascarenhas, M., Garasia, S., Berthiaume, P., Corrin, T., Greig, J., Ng, V., et al. (2018). A scoping review of published literature on chikungunya virus. *PLoS One.* 13:e0207554. doi: 10.1371/journal.pone.0207554

Mason, P. J., and Haddow, A. J. (1957). An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952–53; an additional note on Chikungunya virus isolations and serum antibodies. *Trans. R. Soc. Trop. Med. Hyg.* 51, 238–240. doi: 10.1016/0035-9203(57)90022-6

Mavalankar, D., Shastri, P., and Raman, P. (2007). Chikungunya epidemic in India: a major public-health disaster. *Lancet Infect. Dis.* 7, 306–307. doi: 10.1016/S1473-3099(07)70091-9

Mavale, M., Parashar, D., Sudeep, A., Gokhale, M., Ghodke, Y., Geevarghese, G., et al. (2010). Venereal transmission of chikungunya virus by Aedes aegypti mosquitoes (Diptera: culicidae). *Am. J. Trop. Med. Hyg.* 83, 1242–1244. doi: 10.4269/ajtmh.2010.09-0577

McGill, P. E. (1995). Viral infections: alpha-viral arthropathy. *Baillieres Clin. Rheumatol.* 9, 145–150. doi: 10.1016/S0950-3579(05)80151-7

Melan, A., Aung, M. S., Khanam, F., Paul, S. K., Riaz, B. K., Tahmina, S., et al. (2018). Molecular characterization of chikungunya virus causing the 2017 outbreak in Dhaka, Bangladesh. *N. Microbes New Infect.* 24, 14–16. doi: 10.1016/j.nmni.2018.03.007

Monteiro, V. V. S., Navegantes-Lima, K. C., de Lemos, A. B., Da Silva, G. L., de Souza Gomes, R., Reis, J. F., et al. (2019). Aedes–Chikungunya Virus Interaction: Key Role of Vector Midguts Microbiota and Its Saliva in the Host Infection. *Front. Microbial.* 10:492. doi: 10.3389/fmicb.2019.00492

Moore, D. L., Reddy, S., Akinkugbe, F. M., Lee, V. H., David-West, T. S., Causey, et al. (1974). An epidemic of chikungunya fever at Ibadan, Nigeria, 1969. *Ann. Trop. Med. Parasitol.* 68, 59–68. doi: 10.1080/00034983.1974.11686925

Mourya, D. T. (1987). Absence of transovarial transmission of Chikungunya virus in Aedes aegypti & Ae. albopictus mosquitoes. *Indian J. Med. Res.* 85, 593–595.

Mourya, D. T., and Banerjee, K. (1987). Experimental transmission of chikungunya virus by Aedes vittatus mosquitoes. *Indian J. Med. Res.* 86, 269–271.

Mourya, D. T., Gokhale, M. D., Malunjkar, A. S., Bhat, H. R., and Banerjee, K. (1994). Inheritance of oral susceptibility of Aedes aegypti to chikungunya virus. *Am. J. Trop. Med. Hyg.* 51, 295–300. doi: 10.4269/ajtmh.1994.51.295

Munasinghe, D. R., Amarasekera, P. J., and Fernando, C. F. (1966). An epidemic of dengue-like fever in Ceylon (chikungunya) – a clinical and haematological study. *Ceylon Med. J.* 11, 129–142.

Murhekar, M., Kanagasabai, K., Shete, V., Joshua, V., Ravi, M., Kirubakaran, B. K., et al. (2019). Epidemiology of chikungunya based on laboratory surveillance

data—India, 2016–2018. *Trans. R. Soc. Trop. Med. Hyg.* 113, 259–262. doi: 10.1093/trstmh/try141

Myers, R. M., and Carey, D. E. (1967). Concurrent isolation from patient of two arboviruses, chikungunya and dengue type 2. *Science* 157, 1307–1308. doi: 10.1126/science.157.3794.1307

Nextstrain (2021). *Molecular epidemiology of Chikungunya virus*. Available online at: https://nextstrain.org/community/ViennaRNA/CHIKV?f_country=Bangladesh,Pakistan,Sri_Lanka,India (accessed May 01, 2021).

Ng, D. H., Ho, H. J., Chow, A., Wong, J., Kyaw, W. M., Tan, A., et al. (2018). Correlation of clinical illness with viremia in Zika virus disease during an outbreak in Singapore. *BMC Infect. Dis.* 18:301. doi: 10.1186/s12879-018-3211-9

Nimmannitya, S., Halstead, S. B., Cohen, S. N., and Margiotta, M. R. (1969). Dengue and chikungunya virus infection in man in Thailand, 1962–1964. I. Observations on hospitalized patients with hemorrhagic fever. *Am. J. Trop. Med. Hyg.* 18, 954–971. doi: 10.4269/ajtmh.1969.18.954

Obeyesekere, I., and Hermon, Y. (1972). Myocarditis and cardiomyopathy after arbovirus infections (dengue and chikungunya fever). *Br. Heart. J.* 34, 821–827. doi: 10.1136/hrt.34.8.821

Obeyesekere, I., and Hermon, Y. (1973). Arbovirus heart disease: myocarditis and cardiomyopathy following dengue and chikungunya fever – a follow-up study. *Am. Heart. J.* 85, 186–194. doi: 10.1016/0002-8703(73)90459-6

Onyango, M. G., Ciota, A. T., and Kramer, L. D. (2020). The Vector-Host-Pathogen Interface: The Next Frontier in the Battle Against Mosquito-Borne Viral Diseases? *Front. Cell Infect. Microbiol.* 10:547. doi: 10.3389/fcimb.2020.564518

Padbidri, V. S., and Gnaneswar, T. T. (1979). Epidemiological investigations of chikungunya epidemic at Barsi, Maharashtra state, India. *J. Hyg. Epidemiol. Microbiol. Immunol.* 23, 445–451.

Pan American and Health Organization (PAHO) (2021). *Chikungunya: Epidemiological alerts and updates*. Available online at: https://www.paho.org/hq/index.php?option=com_topics&view=rdmore&cid=5855&Itemid=40931&lang=en (accessed Jan 03, 2021).

Parola, P., de Lamballerie, X., Jourdan, J., Rovery, C., Vaillant, V., Minodier, P., et al. (2006). Novel chikungunya virus variant in travelers returning from Indian Ocean islands. *Emerg. Infect. Dis.* 12, 1493–1499. doi: 10.3201/eid1210.060610

Pastorino, B., Muyembe-Tamfum, J. J., Bessaud, M., Tock, F., Tolou, H., Durand, J. P., et al. (2004). Epidemic resurgence of Chikungunya virus in democratic Republic of the Congo: identification of a new central African strain. *J. Med. Virol.* 74, 277–282. doi: 10.1002/jmv.20168

Paul, S. D., and Singh, K. R. (1968). Experimental infection of *Macaca radiata* with Chikungunya virus and transmission of virus by mosquitoes. *Indian J. Med. Res.* 56, 802–811.

Pavri, K. M. (1964). Presence of chikungunya antibodies in human sera collected from Calcutta and Jamshedpur before 1963. *Indian J. Med. Res.* 52, 698–702.

Peyrefitte, C. N., Rousset, D., Pastorino, B. A., Pouillot, R., Bessaud, M., Tock, F., et al. (2007). Chikungunya virus, Cameroon, 2006. *Emerg. Infect. Dis.* 13, 768–771. doi: 10.3201/eid1305.061500

Phadungsombat, J., Imad, H., Rahman, M., Nakayama, E. E., Kludkleeb, S., Ponam, T., et al. (2020). A Novel Sub-Lineage of Chikungunya Virus East/Central/South African Genotype Indian Ocean Lineage Caused Sequential Outbreaks in Bangladesh and Thailand. *Viruses* 12:1319. doi: 10.3390/v12111319

Pialoux, G., Gaüzère, B. A., Jauréguiberry, S., and Strobel, M. (2007). Chikungunya, an epidemic arbovirosis. *Lancet Infect. Dis.* 7, 319–327. doi: 10.1016/s1473-3099(07)70107-x

Porter, K. R., Tan, R., Istary, Y., Suharyono, W., Sutaryo, Widjaja, S., et al. (2004). A serological study of Chikungunya virus transmission in Yogyakarta, Indonesia: evidence for the first outbreak since 1982. *Southeast Asian J. Trop. Med. Public Health.* 35, 408–415.

Powers, A. M., Brault, A. C., Tesh, R. B., and Weaver, S. C. (2000). Reemergence of Chikungunya and O'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. *J. Gen. Virol.* 81, 471–479. doi: 10.1099/0022-1317-81-2-471

Powers, A. M., and Logue, C. H. (2007). Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *J. Gen. Virol.* 88, 2363–2377. doi: 10.1099/vir.0.82858-0

Presti, A. L., Cella, E., Angeletti, S., and Ciccozzi, M. (2016). Molecular epidemiology, evolution and phylogeny of Chikungunya virus: an updating review. *Infect. Genet. Evol.* 41, 270–278. doi: 10.1016/j.meegid.2016.04.006

Pyke, A. T., Moore, P. R., and McMahon, J. (2018). New insights into chikungunya virus emergence and spread from Southeast Asia. *Emerg. Microbes. infect* 7:26. doi: 10.1038/s41426-018-0024-2

Rajapakse, S., Rodrigo, C., and Rajapakse, A. (2010). Atypical manifestations of chikungunya infection. *Trans. R. Soc. Trop. Med. Hyg.* 104, 89–96. doi: 10.1016/j.trstmh.2009.07.031

Ravi, V. (2006). Re-emergence of chikungunya virus in India. *Indian J. Med. Microbiol.* 24, 83–84. doi: 10.1016/s0255-0857(21)02403-8

Reiter, P., Fontenille, D., and Paupy, C. (2006). Aedes albopictus as an epidemic vector of chikungunya virus: another emerging problem? *Lancet Infect. Dis.* 6, 463–464. doi: 10.1016/s1473-3099(06)70531-x

Reller, M. E., Akoroda, U., Nagahawatte, A., Devasiri, V., Kodikaarachchi, W., Strouse, J. J., et al. (2013). Chikungunya as a cause of acute febrile illness in southern Sri Lanka. *PLoS One.* 8:e82259. doi: 10.1371/journal.pone.0082259

Renault, P., Solet, J. L., Sissoko, D., Balleydier, E., Larrieu, S., Filleul, L., et al. (2007). A major epidemic of chikungunya virus infection on Reunion Island, France, 2005–2006. *Am. J. Trop. Med. Hyg.* 77, 727–731. doi: 10.4269/ajtmh.2007.77.727

Riswari, S. F., Ma'roef, C. N., Djauhari, H., Kosasih, H., Perkasa, A., Yudhaputri, F. A., et al. (2016). Study of viremic profile in febrile specimens of chikungunya in Bandung, Indonesia. *J. Clin. Virol.* 74, 61–65. doi: 10.1016/j.jcv.2015.11.017

Robillard, P. Y., Boumahni, B., Gerardin, P., Michault, A., Fourmaintraux, A., Schuffenecker, I., et al. (2006). Vertical maternal fetal transmission of the chikungunya virus. Ten cases among 84 pregnant women. *Presse Med.* 35, 785–788. doi: 10.1016/s0755-4982(06)74690-5

Robinson, M. C. (1955). An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952–53. I. Clinical features. *Trans. R. Soc. Trop. Med. Hyg.* 49, 28–32. doi: 10.1016/0035-9203(55)90080-8

Rodrigues Faria, N., Lourenço, J., Marques, de Cerqueira, E., Maia, de Lima, M., et al. (2016). Epidemiology of Chikungunya Virus in Bahia, Brazil, 2014-2015. *PLoS Curr.* 8:ecurrents.outbreaks.c97507e3e48efb946401755d468c28b2. doi: 10.1371/currents.outbreaks.c97507e3e48efb946401755d468c28b2

Rodriguez-Morales, A. J., Villamil-Gómez, W. E., and Franco-Paredes, C. (2016). The arboviral burden of disease caused by co-circulation and co-infection of dengue, chikungunya and Zika in the Americas. *Travel. Med. Infect. Dis.* 4, 177–179. doi: 10.1016/j.tmaid.2016.05.004

Ross, R. W. (1956). The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. *J. Hyg.* 54, 177–191. doi: 10.1017/s0022172400044442

Saxena, S. K., Singh, M., Mishra, N., and Lakshmi, V. (2006). Resurgence of chikungunya virus in India: an emerging threat. *Euro. Surveill.* 11, E060810–E060812. doi: 10.2807/esw.11.32.03019-en

Schlesinger, M., and Schlesinger, S. (1986). "Formation and assembly of alphavirus glycoproteins," in *The Togaviridae and Flaviviridae*, eds S. Schlesinger, and M. J. Schlesinger, (New York: Plenum Publishing Corp), 121–148. doi: 10.1007/978-1-4757-0785-4_5

Schuffenecker, I., Iteman, I., Michault, A., Murri, S., Frangeul, L., Vaney, M. C., et al. (2006). Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med.* 3:e263. doi: 10.1371/journal.pmed.0030263

Scolari, F., Casiraghi, M., and Bonizzoni, M. (2019). Aedes spp. and their microbiota: a review. *Front. Microbial.* 10:2036. doi: 10.3389/fmicb.2019.02036

Sergon, K., Yahaya, A. A., Brown, J., Bedja, S. A., Agata, N., Allaranger, Y., et al. (2007). Seroprevalence of chikungunya virus infection on Grande Comore Island, Union of the Comoros, March 2005. *Am. J. Trop. Med. Hyg.* 76, 1189–1193. doi: 10.4269/ajtmh.2007.76.1189

Seyler, T., Hutin, Y., Ramanchandran, V., Ramakrishnan, R., Manickam, P., and Murhekar, M. (2010). Estimating the burden of disease and the economic cost attributable to chikungunya, Andhra Pradesh, India, 2005–2006. *Trans. R. Soc. Trop. Med. Hyg.* 104, 133–138. doi: 10.1016/j.trstmh.2009.07.014

Shen, J. Y., Li, M., Xie, L., Mao, J. R., Zhou, H. N., Wang, P. G., et al. (2020). Perinatal Vertical Transmission of Chikungunya Virus in Ruili, a Town on the Border between China and Burma. *Virol. Sin* 36, 145–148. doi: 10.1007/s12250-020-00245-y

Silva, L. A., and Dermody, T. S. (2017). Chikungunya virus: epidemiology, replication, disease mechanisms, and prospective intervention strategies. *J. Clin. Investig.* 127, 737–749. doi: 10.1172/JCI84417

Silva, J. V. Jr., Ludwig-Begall, L. F., de Oliveira-Filho, E. F., Oliveira, R. A., Durães-Carvalho, R., Lopes, T. R., et al. (2018). A scoping review of Chikungunya virus infection: epidemiology, clinical characteristics, viral co-circulation complications, and control. *Acta. Trop.* 188, 213–224. doi: 10.1016/j.actatropica.2018.09.003

Simizu, B., Yamamoto, K., Hashimoto, K., and Ogata, T. (1984). Structural proteins of Chikungunya virus. *J. Virol.* 51, 254–258. doi: 10.1128/jvi.51.1.254-258.1984

Simo, F. B. N., Bigna, J. J., Well, E. A., Kenmoe, S., Sado, F. B. Y., Weaver, S. C., et al. (2019). Chikungunya virus infection prevalence in Africa: a contemporaneous systematic review and meta-analysis. *Public Health.* 166, 79–88. doi: 10.1016/j.puhe.2018.09.027

Soekiman, S. (1987). A study on susceptibility of Indonesia colonies of Aedes aegypti and Aedes albopictus mosquitoes to experimental infection with dengue type 3 and chikungunya viruses. *Kobe J. Med. Sci.* 33, 19–34.

Soekiman, S., Matsumura, T., and Yamanishi, H. (1986b). Multiplication of chikungunya virus in salivary glands of Aedes albopictus (Oahu strain) mosquitoes: an electron microscopic study. *JPN. J. Med. Sci. Biol.* 39, 207–211. doi: 10.7883/yoken1952.39.207

Soekiman, S., Konishi, E. J., and Matsumura, T. (1986a). Susceptibility of Indonesia colonies of Aedes aegypti and Aedes albopictus mosquitoes to experimental infection with chikungunya virus. *Kobe J. Med. Sci.* 32, 127–132.

Spence, L. P., and Thomas, L. (1959). Application of haemagglutination and complement fixation techniques to the identification and serological classification of arthropod-borne viruses; studies on Chikungunya and Makonde viruses. *Trans. R. Soc. Trop. Med. Hyg.* 53, 248–255. doi: 10.1016/0035-9203(59)90004-5

Spicher, T., Delitz, M., Schneider, A. D. B., and Wolfinger, M. T. (2021). Dynamic molecular epidemiology reveals lineage-associated single-nucleotide variants that alter RNA structure in Chikungunya virus. *Genes* 12:239. doi: 10.3390/genes12020239

Staples, J. E., Breiman, R. F., and Powers, A. M. (2009). Chikungunya fever: an epidemiological review of a re-emerging infectious disease. *Clin. Infect. Dis.* 49, 942–948. doi: 10.1086/605496

Suhrbier, A. (2019). Rheumatic manifestations of chikungunya: emerging concepts and interventions. *Nat. Rev. Rheumatol.* 15, 597–611. doi: 10.1038/s41584-019-0276-9

Tesh, R. B. (1982). Arthritides caused by mosquito-borne viruses. *Annu. Rev. Med.* 33, 31–40. doi: 10.1146/annurev.me.33.020182.000335

Thaung, U., Ming, C. K., Swe, T., and Thein, S. (1975). Epidemiological features of dengue and chikungunya infections in Burma. *Southeast Asian J. Trop. Med. Public Health.* 6, 276–283.

Thein, S., La Linn, M., Aaskov, J., Aung, M. M., Aye, M., Zaw, A., et al. (1992). Development of a simple indirect enzyme-linked immunosorbent assay for the detection of immunoglobulin M antibody in serum from patients following an outbreak of chikungunya virus infection in Yangon, Myanmar. *Trans. R. Soc. Trop. Med. Hyg.* 86, 438–442. doi: 10.1016/0035-9203(92)90260-J

Thonnon, J., Spiegel, A., Diallo, M., Diallo, A., and Fontenille, D. (1999). Chikungunya virus outbreak in Senegal in 1996 and 1997. *Bull. Soc. Pathol. Exot.* 92, 79–82.

Tsetsarkin, K. A., Chen, R., and Weaver, S. C. (2016). Interspecies transmission and chikungunya virus emergence. *Curr. Opin. Virol.* 16, 143–150. doi: 10.1016/j.coviro.2016.02.007

Vairo, F., Haider, N., Kock, R., Ntoumi, F., Ippolito, G., and Zumla, A. (2019). Chikungunya: epidemiology, pathogenesis, clinical features, management, and prevention. *Infect. Dis. Clin. North. Am.* 33, 1003–1025. doi: 10.1016/j.idc.2019.08.006

van Enter, B. J., Huibers, M. H., van Rooij, L., Steingrover, R., van Hensbroek, M. B., Voigt, R. R., et al. (2018). Perinatal outcomes in vertically infected neonates during a chikungunya outbreak on the island of Curacao. *Am. J. Trop. Med. Hyg.* 99, 1415–1418. doi: 10.4269/ajtmh.17-0957

Vazeille, M., Moutailler, S., Coudrier, D., Rousseaux, C., Khun, H., Huerre, M., et al. (2007). Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito. *Aedes albopictus. PLoS One.* 2:e1168. doi: 10.1371/journal.pone.0001168

Villamil-Gómez, W. E., Rodríguez-Morales, A. J., Uribe-García, A. M., González-Arismendy, E., Castellanos, J. E., Calvo, E. P., et al. (2016). Zika, dengue, and chikungunya co-infection in a pregnant woman from Colombia. *Int. J. Infect. Dis.* 51, 135–138. doi: 10.1016/j.ijid.2016.07.017

Wahid, B., Ali, A., Rafique, S., and Idrees, M. (2017). Global expansion of chikungunya virus: mapping the 64-year history. *Int. J. Infect. Dis.* 58, 69–76. doi: 10.1016/j.ijid.2017.03.006

Weaver, S. C., and Forrester, N. L. (2015). Chikungunya: Evolutionary history and recent epidemic spread. *Antivir. Res.* 120, 32–39. doi: 10.1016/j.antiviral.2015.04.016

Weaver, S. C., Frey, T. K., Huang, H. V., Kinney, R. M., Rice, C. M., Roehrig, J. T., et al. (2005). "Togaviridae," in *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, C. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger, and L.A. Ball, (Amsterda: Elsevier), 999–1008.

Weaver, S. C., and Lecuit, M. (2015). Chikungunya virus and the global spread of a mosquito-borne disease. *N. Engl. J. Med.* 372, 1231–1239. doi: 10.1056/NEJMra1406035

Wimalasiri-Yapa, B. R., Stassen, L., Huang, X., Hafner, L. M., Hu, W., Devine, G. J., et al. (2019). Chikungunya virus in Asia–Pacific: a systematic review. *Emerg. Microbes. infect.* 8, 70–79. doi: 10.1080/22221751.2018.1559708

World Health and Organization (WHO) (2021). *Chikungunya*. Available online at: https://www.who.int/news-room/fact-sheets/detail/chikungunya (accessed Jan 03, 2021).

Yactayo, S., Staples, J. E., Millot, V., Cibrelus, L., and Ramon-Pardo, P. (2016). Epidemiology of Chikungunya in the Americas. *J. Infect. Dis.* 214, S441–S445. doi: 10.1093/infdis/jiw390

Yadav, P., Shouche, Y. S., Munot, H. P., Mishra, A. C., and Mourya, D. T. (2003). Genotyping of Chikungunya virus isolates from India during 1963–2000 by reverse transcription-polymerase chain reaction. *Acta. Virol.* 47, 125–127.

Zanotto, P. M. D. A., and Leite, L. C. D. C. (2018). The challenges imposed by Dengue, Zika, and Chikungunya to Brazil. *Front. Immunol.* 9:1964. doi: 10.3389/fimmu.2018.01964

Zim, M. M., Sam, I. C., Omar, S. S., Chan, Y. F., AbuBakar, S., and Kamarulzaman, A. (2013). Chikungunya infection in Malaysia: comparison with dengue infection in adults and predictors of persistent arthralgia. *J. Clin. Virol.* 56, 141–145. doi: 10.1016/j.jcv.2012.10.019

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fmicb.2021.655228

Host Adaptive Evolution of Avian-Origin H3N2 Canine Influenza Virus

*Fucheng Guo[1,2], Ayan Roy[3], Ruichen Wang[2], Jinjin Yang[2], Zhipeng Zhang[2], Wen Luo[2], Xuejuan Shen[2,4], Rui-Ai Chen[2,4], David M. Irwin[5,6] and Yongyi Shen[1,2,4,7]**

[1] Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China, [2] Center for Emerging and Zoonotic Diseases, College of Veterinary Medicine, South China Agricultural University, Guangzhou, China, [3] Department of Biotechnology, Lovely Professional University, Phagwara, India, [4] Zhaoqing Branch Center of Guangdong Laboratory for Lingnan Modern Agricultural Science and Technology, Zhaoqing, China, [5] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada, [6] Banting and Best Diabetes Centre, University of Toronto, Toronto, ON, Canada, [7] Key Laboratory of Zoonosis Prevention and Control of Guangdong Province, Guangzhou, China

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Haixia Xiao,
Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, China
Yuhai Bi,
Institute of Microbiology, Chinese Academy of Sciences, China

***Correspondence:**
Yongyi Shen
shenyy@scau.edu.cn

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *18 January 2021*
Accepted: *11 May 2021*
Published: *14 June 2021*

Citation:
Guo F, Roy A, Wang R, Yang J, Zhang Z, Luo W, Shen X, Chen R-A, Irwin DM and Shen Y (2021) Host Adaptive Evolution of Avian-Origin H3N2 Canine Influenza Virus. Front. Microbiol. 12:655228. doi: 10.3389/fmicb.2021.655228

Since its first isolation in around 2007, the avian-origin H3N2 canine influenza virus (CIV) has become established and continues to circulate in dog populations. This virus serves as a useful model for deciphering the complex evolutionary process of interspecies transmission of influenza A virus (IAV) from one species to its subsequent circulation in another mammalian host. The present investigation is a comprehensive effort to identify and characterize genetic changes that accumulated in the avian-origin H3N2 CIV during its circulation in the dog. We revealed that H3N2 CIV experiences greater selection pressure with extremely high global non-synonymous to synonymous substitution ratios per codon (dN/dS ratio) for each gene compared to the avian reservoir viruses. A total of 54 amino acid substitutions were observed to have accumulated and become fixed in the H3N2 CIV population based on our comprehensive codon-based frequency diagram analysis. Of these substitutions, 11 sites also display high prevalence in H3N8 CIV, indicating that convergent evolution has occurred on different lineages of CIV. Notably, six substitutions, including HA-G146S, M1-V15I, NS1-E227K, PA-C241Y, PB2-K251R, and PB2-G590S, have been reported to play imperative roles in facilitating the transmission and spillover of IAVs across species barriers. Most of these substitutions were found to have become fixed in around 2015, which might have been a favorable factor that facilitating the spread of these CIV lineages from South Asia to North America and subsequent further circulation in these areas. We also detected 12 sites in six viral genes with evidence for positive selection by comparing the rates of non-synonymous and synonymous substitutions at each site. Besides, our study reports trends of enhanced ongoing adaptation of H3N2 CIV to their respective host cellular systems, based on the codon adaptation index analysis, which points toward increasing fitness for efficient viral replication. In addition, a reduction in the abundance of the CpG motif, as evident from an analysis of relative dinucleotide abundance, may contribute

to the successful evasion of host immune recognition. The present study provides key insights into the adaptive changes that have accumulated in the avian-origin H3N2 viral genomes during its establishment and circulation into dog populations.

Keywords: influenza A viruses, H3N2 canine influenza virus (CIV), interspecies transmission, genetic change, positive selection, codon adaptation index

INTRODUCTION

Influenza A virus (IAV) belongs to the family Orthomyxoviridae. According to the antigenicity of surface glycoproteins, namely, hemagglutinin (HA) and neuraminidase (NA), IAVs can be subtyped into multiple HxNy subtypes, including 18 HA (H1–H18) and 11 NA (N1–N11) (Mostafa et al., 2018). IAVs have been isolated from a wide range of animal hosts, including birds, humans, horses, whales, minks, pigs, and dogs, with specific subtypes predominating in each species, and these viruses constantly pose threats to both human and animal health (Yoon et al., 2014). Aquatic birds are thought to be the reservoir hosts for IAVs, except H17N10 and H18N11, which are bat-origin (Yoon et al., 2014).

The past two decades have witnessed the interspecific transmission and circulation of two IAV subtypes in dog populations [termed canine influenza virus (CIV)], namely, the equine-origin H3N8 influenza virus (EIV) and the avian-origin H3N2 influenza virus (AIV) (Borland et al., 2020). Emerging around 1999, the H3N8 strain variant of EIV was first isolated in the southeastern United States in 2004 and was the first IAV to cause epidemic disease in dogs (Crawford et al., 2005; Anderson et al., 2012). Since its emergence, H3N8 CIV circulated continuously among dogs in the United States until 2016 (Borland et al., 2020) and caused sporadic outbreaks in the dogs in the United Kingdom and Australia (Daly et al., 2008; Kirkland et al., 2010). Another CIV in canine, H3N2 CIV, was isolated from dogs around 2007 and rapidly spread into several areas of Southeast Asia (Li et al., 2010; Bunpapong et al., 2014). This CIV lineage then spread to the United States and Canada, through viral importation events from Asia in 2015 and 2017, respectively, subsequently causing more than 1,000 infections in dogs in these areas (Pulit-Penaloza et al., 2017; Voorhees et al., 2017, 2018; Weese et al., 2019). H3N2 CIV has now successfully colonized to become an enzootic virus throughout South-East Asia and North America and occasionally causes epizootics in pet and sheltered dogs (He et al., 2019; **Supplementary Figure 1**).

Evolution drives the cross-species transmission of IAVs (Guo et al., 2019). H3N2 CIV arose from a single cross-species transfer event, with all of its gene segments having an avian origin (Li et al., 2010; Zhu et al., 2015), and has successfully established and maintained its circulation in its new host environment. Thus, H3N2 CIV serves as an informative and relatively reliable model for understanding how IAVs emerge and adapt in new hosts. Many previous studies have focused on the phylogenetic history of H3N2 CIV (Zhu et al., 2015; Voorhees et al., 2017, 2018; He et al., 2019). In addition, several putative or confirmed adaptive mutations that separate H3N2 CIV from avian reservoir viruses or lead to antigenic changes among circulation clades had been reported previously (Yang et al., 2013; Lin et al., 2016; He et al., 2019; Lyu et al., 2019; Wu et al., 2021). Recently, the evolutionary dynamics on population size, selection pressure, and nucleotide substitution rates on virus genes of H3N2 CIV has been estimated (Shen et al., 2021). However, the genetic factors associated with its adaptation remain unclear. In the present study, we explored and thoroughly characterized the process by which H3N2 CIV adapted to its new host environment in real time (i.e., epidemiological time), after the spillover event from its avian source specie from multiple perspectives including analyses that assess selection pressure, amino acid mutation, codon usage, and dinucleotide distributions. Information provided by the present analysis, in combination with existing knowledge, promises to provide detailed insights into the potential genetic factors that may be responsible for host adaptation and facilitate an understanding of the complex patterns of adaptive evolution in establishing stable lineage descendants.

MATERIALS AND METHODS

Dataset Collection and Processing

All influenza sequences used in this study were downloaded from the Influenza Virus Resource at the National Center for Biotechnology Information (NCBI)[1] and the Global Initiative on Sharing Avian Influenza Data[2]. Redundant sequences, laboratory strains, and short sequences (<85% of the corresponding gene) were removed. MAFFT v7.221 was used to generate individual gene codon-based alignments followed by manual alignment to codon position. Specially, aligned nucleotide sequences of MP and NS segments were edited such that all of the codons from the first open reading frame (ORF) (M1 or NS1) were followed by the codons from the second ORF (M2 or NEP/NS2) to avoid repetition of nucleotides between these two ORFs. For the PA and PB1 segments, only the longest ORF was used.

Avian IAV Dataset

More than 18,000 avian-isolated complete genomes of IAVs were downloaded. For the HA and NA gene segments, only subtypes H3Nx and HxN2 were considered for further curation, respectively. The large datasets were reduced using methods applied in a previous study (Zhu et al., 2015) to make this analysis tractable. Specifically, gene-wise datasets were clustered based on an identity of greater than 98% using the CD-HIT software package (Li and Godzik, 2006), with one sequence from each cluster selected for the maximum likelihood (ML) phylogenetic

[1] www.ncbi.nlm.nih.gov/genomes/FLU

[2] www.gisaid.org

analysis of all H3N2 CIV strains for each gene segment using IQ-TREE with default settings (Nguyen et al., 2015). Subsequently, the subset of avian sequences closely related to H3N2 CIV was selected and expanded to the original number of taxa (before using CD-HIT) (Zhu et al., 2015) for subsequent analyses. In total, 2867 taxa for PB2, 3410 taxa for PB1, 4407 taxa for PA, 3229 taxa for NS, 3311 taxa for NP, 2282 taxa for NA, 2564 taxa for MP, and 603 taxa for HA were used in the downstream analyses (**Supplementary Table 1** and **Supplementary Sheet 1**).

H3N2 and H3N8 CIV Datasets

All available H3N2 CIV and H3N8 CIV strains were downloaded and pretreated as described above. Recombinant strains and some other lineages were excluded as previously described (Chen et al., 2018). The final dataset comprised of 264, 253, 261, 258, 252, 250, 247, and 247 unique HA, MP, NA, NP, NS, PA, PB1, and PB2 coding sequences, respectively, for H3N2 CIV (**Supplementary Table 1** and **Supplementary Sheet 2**). For H3N8 CIV, the final dataset encompassed 108, 204, 93, 73, 203, 54, 55, and 56 unique HA, MP, NA, NP, NS, PA, PB1, and PB2 coding sequences, respectively (**Supplementary Table 1** and **Supplementary Sheet 3**).

Sequence Analysis

Selection Analysis

Rates of non-synonymous and synonymous substitutions (dN/dS) for each segment from H3N2 AIV, H3N2 CIV, and H3N8 CIV were estimated using the SLAC (Single Likelihood Ancestry Counting) method (Kosakovsky Pond and Frost, 2005) with ML trees inferred by IQ-TREE as the input reference trees. To test the potential impact of sampling bias on the results, a reshuffling test was performed. Specially, for each alignment, 50% of the sequences were randomly selected, and the SLAC analysis was performed based on the sub-datasets. The reshuffling test was performed three times for each alignment, and a similar trend among the reshuffled sub-datasets would suggest that sampling bias did not affect the results. SLAC, FEL (Fixed Impact Probability) (Kosakovsky Pond and Frost, 2005), MEME (Evolutionary Mixed Effects Model) (Murrell et al., 2012), and FUBAR (Fast, Unconstrained Bayesian Approximation) (Murrell et al., 2013) methods implemented at the DATAMONKEY[3] server were used to identify codons under positive selection (Weaver et al., 2018). Sites that were reported by at least two methods were considered as positively selected. The significant result is chosen by p-value < 0.1 for SLAC, FEL, and MEME, and posterior probability > 0.9 for FUBAR (He et al., 2019).

Identification of Effective Substitutions Over the Epidemiological Time Period

The year-wise frequencies of amino acids at the codon level were computed for each gene segment of H3N2 CIV. For each gene segment, if the number of sequences for a year was less than 3, then they were manually combined with the sequences belonging to the next year in the dataset to reduce calculation bias. Subsequently, a "proportion switch" was defined as the replacement of one abundant amino acid by another in successive years, $n(\omega, j_{th}, a_k)$ was defined as the number of sequences with amino acid a_k at the j_{th} position, and the amino acid proportion $f\ (\omega, j_{th}, a_k)$ at the j_{th} position at year ω was given by $f\ (\omega, j_{th}, a_k) = n(\omega, j_{th}, a_k)/N(\omega)$ (Shih et al., 2007; Pu et al., 2018). A proportion switch between two different amino acids a_m and a_n at a given site j_{th} between years ω and $\omega + 1$ was reported when both of the following conditions were met: (i) $f\ (\omega, j_{th}, a_m) + f\ (\omega, j_{th}, a_n) > 0.8$ and $(\omega + 1, j_{th}, a_m) + f\ (\omega + 1, j_{th}, a_n) > 0.75$ and (ii) $[n(\omega, j_{th}, a_m) - n(\omega, j_{th}, a_n)]$ and $[n(\omega + 1, j_{th}, a_m) - n(\omega + 1_{th}, j, a_n)]$ had opposite signs or zero in absolute value (Shih et al., 2007; Pu et al., 2018). Finally, we defined an "effective substitution" as a novel amino acid accumulating to a high proportion (>75%) in the following epidemiological time after the proportion switch.

Measurement of Codon Usage and Dinucleotide Distributions

ORFs were concatenated (HA+MP+NA+NP+NS +PA+PB1+PB2) to assess the adaptation of the H3N2 CIV to the host microenvironment (Cristina et al., 2015). Base compositional features of the virus genome, including proportion of G and C at the first (GC1), second (GC2), and third (GC3) positions of codons and overall dinucleotide distribution, were calculated using the CodonW program[4] (**Supplementary Table 2**). Dinucleotides with relative abundance > 1.25 were considered to be overrepresented, whereas dinucleotides with relative abundance < 0.78 were inferred as underrepresented (Kunec and Osterrieder, 2016). The GC3 values (x-axis) of the viral genes were then plotted against the respective GC12 values (mean of GC1 and GC2) (y-axis) to generate neutrality plots to explore the magnitude of the genomic compositional constraint and natural selection operating on virus genomes (Sueoka, 1988). To evaluate the fitness of the codon usage of H3N2 CIV to the host expression system, we employed the codon adaptation index (CAI) using the CAIcal web server[5] (Sharp and Li, 1987; Puigbo et al., 2008), with host genomes (duck and canine) downloaded from the Ensembl database[6], as references.

Statistical Analysis

Statistical significance was evaluated using the Student's t-test, with a significance level of 5% (0.05). To investigate the evolution of viral genome across the evolutionary timescale, linear regressions were performed between CAI values, dinucleotide abundance, and their respective collection dates. The presence of a significant regression coefficient was considered as supportive of adaptation over time (Franzo et al., 2017). All statistical analyses were performed using the SPSS software package (IBM Corp; version 23.0).

RESULTS AND DISCUSSION

The subsequent spread of a newly emergent IAV within a new host population requires a period of adaptation (Guo et al., 2018).

[3]http://www.datamonkey.org/analyses

[4]http://sourceforge.net/projects/codonw/

[5]http://genomes.urv.es/CAIcal/

[6]http://www.ensembl.org

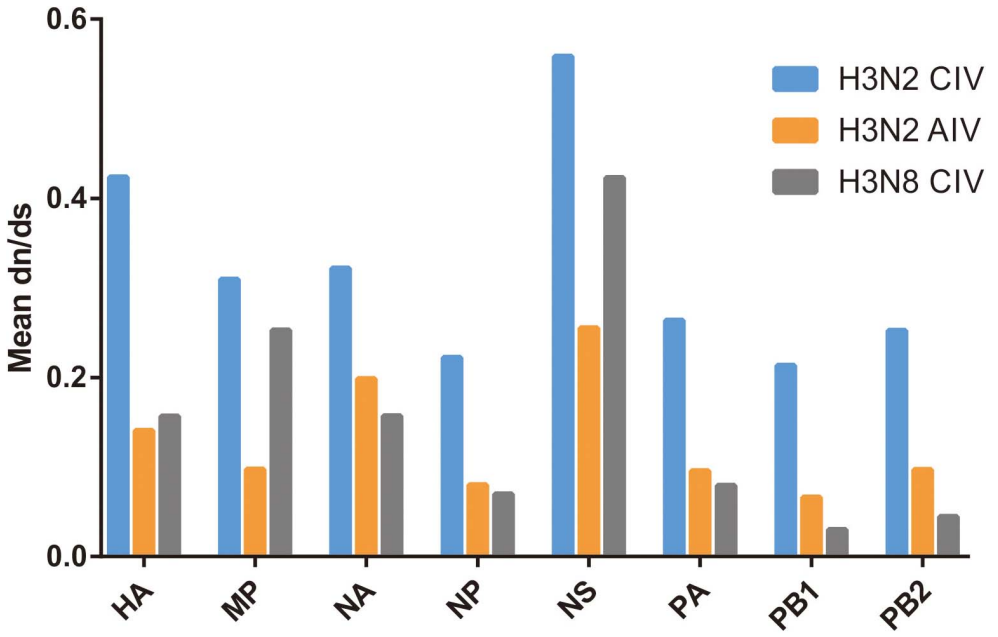


FIGURE 1 | Comparative analysis of the global dN/dS ratio for each gene segment of H3N2 CIV (blue), H3N2 AIV (orange), and H3N8 CIV (gray) based on the completed alignments.

This study explores and characterizes the host evolution process of H3N2 CIV over its epidemiological time, after its interspecific transmission event from an avian IAV source and aims to unravel the riddles of the emergence and subsequent adaptation of an IAV in a new host environment. Here, supported by previous reports (Voorhees et al., 2018; He et al., 2019), our analysis revealed that with its establishment in canine populations, H3N2 CIV experiences greater levels of selection pressure than H3N2 AIV or H3N8 CIV, as evident from the significantly higher dN/dS substitution ratios seen for each gene (**Figure 1**). The reliability of our results was further demonstrated by the reshuffling test which revealed similar trends among the reshuffled sub-datasets (**Supplementary Figure 2**), indicating that sampling bias did not influence the results. Higher dN/dS ratios reflect both adaptive pressure and relaxed selective constraint that might facilitate the accumulation of favorable genetic changes associated with adaptive finesse to a new host population (Joseph et al., 2018). Similarly, changes in adaptive pressure leading to higher dN/dS rations have also been reported for both H2N2 pandemic strains and European avian-like H1N1 swine influenza virus (SIV) strains during their evolution after separating from their reservoir avian IAVs strains (Joseph et al., 2015, 2018).

To investigate how the adaptive pressure acts on the evolution of H3N2 CIV, we undertook a comprehensive frequency diagram analysis by real-time scanning of amino acid frequency changes at the codon level for each gene. Such investigations tend to be useful in visualizing the temporal dynamics of substitutions reflecting evolution at any amino acid site (Shih et al., 2007). A total of 54 effective substitutions were observed during the circulation of H3N2 CIV (**Table 1**, **Figure 2**, and **Supplementary Figure 3**), signifying the potential functional importance of these sites and the impact of positive selection on them. Notably, most of these effective substitutions displayed extremely low frequency in H3N2 AIV (**Table 1**), which implies an increasing genetic distance between H3N2 CIV and its reservoir viruses. Some of the effective substitutions, such as HA-V418I, NA-L390S, M2-R18K, NS1-E227K, and PB2-K251R, were also seen in the majority of the AIV sequences with a proportion of greater than 75% (**Table 1**). We speculate that mutations had occurred during the host-shift event of H3N2 CIV, but later reversed back, possibly due to selection pressures operational at different stages of the circulation phase of this lineage of newly acquired CIV (Long et al., 2019). However, it is interesting to note that 11 out of the 54 (20.37%) effective substitutions in H3N2 CIV, including HA-G146S, HA-V242I, HA-V418I, M1-V15I, M2-R18K, PA-N347D, PB1-V200I, PB2-K251R, PB2-I292T, PB2-V511I, and PB2-G590S, are fixed in H3N8 CIV (**Table 1**), indicating that convergent evolution has occurred on different lineages of CIVs at these sites. Such convergent sites may act as host markers associated with adaptation of IAVs in canine population and demand sincere attention.

Intriguingly, 41 of the 54 effective substitutions (75.92%) took less than 5 years since their first occurrence to become fixed (**Supplementary Figure 4**), which is considerably less than the conditional fixation time for a neutral mutation (which is expected to 219 years) (Li, 1997). Such observations emphasize the important role of positive selection operating on these residues (Li, 1997). In addition, it is noteworthy that most of these substitutions were fixed around 2015 (**Figure 2** and **Supplementary Figure 3**), a time when H3N2 CIV was spreading from South-East Asia to North America (Pulit-Penaloza et al., 2017; Voorhees et al., 2017, 2018; Weese et al., 2019). Whether

TABLE 1 | Effective substitutions that occurred during the circulation of H3N2 CIV.

Gene	Codon position[a]	H3N2 AIV proportion	H3N8 CIV proportion	Experimental validation[b]	Location/phenotype[c]
HA	P4L	1.82% (11/603)	0.00% (0/108)		
	I25M	1.16% (7/603)	0.00% (0/108)		
	G146S	0.00% (0/603)	100.00% (108/108)	S146G	Antigenic epitope A Potential to alter the virulence of H1N1pdm09 in swine (Henningson et al., 2015)
	N188D	11.44% (69/603)	0% (0/108)	S188N	Antigenic epitope B, Receptor-binding site (190 helix) Increased virulence in mammals (Matos-Patron et al., 2015)
	V242I	5.64% (34/603)	98.15% (106/108)		Antigenic epitope D
	R261H	0.83% (5/603)	0.00% (0/108)		
	K326R	0.83% (5/603)	0.00% (0/108)		
	V418I	84.25% (508/603)	100.00% (108/108)		
NA	T16A	1.67% (38/2282)	N/A		Transmembrane helix
	V50I	1.10% (25/2282)			
	Y67H	29.10% (664/2282)			
	I153T	0.00% (0/2282)			Located near catalytic sites 151 and 152 (McAuley et al., 2019).
	H155Y	20.60% (470/2282)			
	V263T	0.04% (1/2282)			
	R283Q	7.80% (178/2282)			
	S311N	7.93% (181/2282)			
	D313N	0.35% (8/2282)			
	R338K	35.89% (819/2282)			
	E357D	0.00% (0/2282)			
	L390S	96.23% (2196/2282)			
M1	**V15I**	1.56% (40/2564)	90.20% (184/204)	V15I/T	Increased virulence in mammals (Chen et al., 2007)
	R95K	0.59% (15/2564)	4.90% (10/204)		
	S207N	0.90% (23/2564)	4.90% (10/204)		
M2	G14E	2.11% (54/2564)	0.00% (0/204)		
	R18K	94.27% (2417/2564)	100.00% (204/204)		
	V27I	5.81% (149/2564)	4.90% (10/204)		
NS1	T197I	1.30% (42/3229)	0.49% (1/203)		
	P212S	1.95% (63/3229)	2.96% (6/203)		
	E227K	85.60% (2764/3229)	0.00% (0/203)	E227R	Human host marker (Finkelstein et al., 2007)
NS2	D27G	5.33% (172/3229)	0.00% (0/203)		
	L40F	1.18% (38/3229)	0.49% (1/203)		
PA	Y65H	0.00% (0/4407)	0.00% (0/54)		
	C241Y	0.32% (14/4407)	0.00% (0/54)	C241Y	Enhance the replicative ability of an H5N1 virus in A549 cells and enhance its pathogenicity in mice (Yamaji et al., 2015)
	E243D	0.02% (1/4407)	0.00% (0/54)		
	N347D	0.25% (11/4407)	96.30% (52/54)		
	S388G	58.13% (2562/4407)	0.00% (0/54)		
	R401K	0.59% (26/4407)	0.00% (0/54)		
	G684E	0.43% (19/4407)	0.00% (0/54)		
PB1	E97K	0.62% (21/3410)	0.00% (0/55)		
	R187K	0.09% (3/3410)	0.00% (0/55)	R187K	May contribute higher pathogenicity in mice for H9N2 (Liu et al., 2016)
	V200I	7.39% (252/3410)	94.55% (52/55)		
	S216N	4.13% (141/3410)	0.00% (0/55)		
	V218I	0.06% (2/3410)	0.00% (0/55)		
	T434S	0.00% (0/3410)	0.00% (0/55)		
	A661T	0.18% (6/3410)	0.00% (0/55)		

(Continued)

TABLE 1 | Continued

Gene	Codon position[a]	H3N2 AIV proportion	H3N8 CIV proportion	Experimental validation[b]	Location/phenotype[c]
PB2	M76I	0.14% (4/2867)	0.00% (0/56)		
	K251R	97.84% (2805/2867)	92.86% (52/56)	K251R	Increased virulence in mice (Prokopyeva et al., 2016)
	I292T	11.58% (332/2867)	94.64% (53/56)		
	S334N	0.17% (5/2867)	0.00% (0/56)		
	V338I	7.12% (204/2867)	0.00% (0/56)		
	V511I	9.66% (277/2867)	100.00% (56/56)		
	G590S	2.55% (73/2867)	96.43% (54/56)	GQ590/591SR/K	Increased polymerase activity (Mehle and Doudna, 2009)
	T598A	0.70% (20/2867)	0.00% (0/56)		
	S714I	0.00% (1/2867)	0.00% (0/56)	S714R	Increased polymerase activity, increased virulence in mammals, mammalian host marker (Gabriel et al., 2005; Gabriel et al., 2007)

[a]Mutations observed to be fixed during the circulation of H3N2 CIV. Mutations showing higher proportion in H3N8 CIV (>75%) are marked with bold. [b]Experimental validation of mutations provided in previous studies. [c]Associated location or phenotype of the mutations validated by previous studies.

the accumulation of these substitutions has contributed to the further expansion of this lineage of CIV needs to be further investigated. Besides, considering the high genetic similarity between the H3N2 CIV sequences recently isolated in South-East Asia and North America (Lyu et al., 2019), it is possible that a novel favorable genotype consisting of these adaptive mutations might have become dominant in both South-East Asia and North America. Our conjecture is also supported by the observation that most of these substitutions have become fixed with proportions extremely close to 1 (**Figure 2** and **Supplementary Figure 3**) in recent years, indicating that there is little regional difference in these substitutions.

We noted that 20 of the 54 (37.03%) effective substitutions occurred in the HA and NA genes (**Table 1**, **Figure 2**, and **Supplementary Figure 3**), which suggests that a greater level, relative to other viral genes, of adaptive pressure has acted on these two genes. This suggestion is also supported by our dN/dS analysis (**Figure 1**). HA is responsible for the attachment of the virus to the sialic acid receptors on the cell surface, whereas NA catalyzes the separation of the new virion from infected cells and facilitates viral movement through mucus (Mostafa et al., 2018). It has been reported that a functional optimal balance between HA and NA is required for efficient replication and transmission of IAVs (Mitnaul et al., 2000; Yen et al., 2011). The N188D substitution, located at the receptor-binding region (190 helix) of HA protein (**Figure 3**), is a notable replacement (Skehel and Wiley, 2000). A recent study revealed that alterations of amino acid at this site have significant effects on the binding capacity of A(H1N1)pdm09 (Matos-Patron et al., 2015). In addition, N188D, G146S, and V242I mutations are located at the antigenic epitope B, antigenic epitope A, and antigenic epitope D of the HA protein, respectively (**Figure 3**), where substitutions at these sites may lead to antigenic drift in IAVs and the subsequent evasion of host immune responses (Shih et al., 2007).

For the NA protein, a T16A substitution occurs at a site located in the transmembrane region of this protein (**Figure 4**; McAuley et al., 2019). The transmembrane region anchors the NA protein to membrane and is considered to be part of the signal peptide that permits the transport of this protein across the endoplasmic reticulum membrane (McAuley et al., 2019). Therefore, adaptive mutations in this region might alter the interactions of the NA protein to the membrane for better stability in new hosts. Interestingly, this site had previously been reported to be under positive selection in avian-origin European H1N1 SIV strains (Joseph et al., 2018), thus signifying a potential functional role of the amino acid residues at this site in the adaptation of IAVs between avian and mammalian hosts. In addition, the I153T and H155Y are located close to the enzymatic site of the NA protein (**Figure 4**; McAuley et al., 2019). Variation in this enzymatic region, and relative framework residues, tends to change the aspects of viral replication, transmissibility, and its susceptibility to antiviral inhibitors (McAuley et al., 2019). Taken together, mutations at these functionally important sites in HA and NA might play crucial roles in regulating receptor-binding specificity and affinity, as well as facilitating evasion from host antibodies, thus conferring enhanced fitness to H3N2 CIV in new host environments. However, functional experiments are needed to reach definite conclusions.

Effective evasion from host immune responses along with rapid viral replication and assembly are other factors critical to the virulence and host range of IAVs (Long et al., 2019). Replication and transcription of influenza viruses are catalyzed by the viral polymerase complex composed of the PB2, PB1, and PA proteins, while the M1 protein plays an essential role in viral assembly, budding, and morphogenesis, and the NS1 protein is the major viral interferon antagonist for IAVs (Mostafa et al., 2018). In the present study, we observed that multiple adaptive mutations have occurred and have been fixed in these genes that might play crucial roles in facilitating the transmission and spillover of IAVs across species barriers (**Table 1**). The NS1-227K/R had been reported to be a mammalian marker and exists in all human pandemic influenza viruses while NS1-227E has been noted to be highly conserved in avian strains (Finkelstein et al., 2007). Given the role of the NS1 protein in suppressing host cellular immune responses, the NS1-E227K mutation might be a molecular

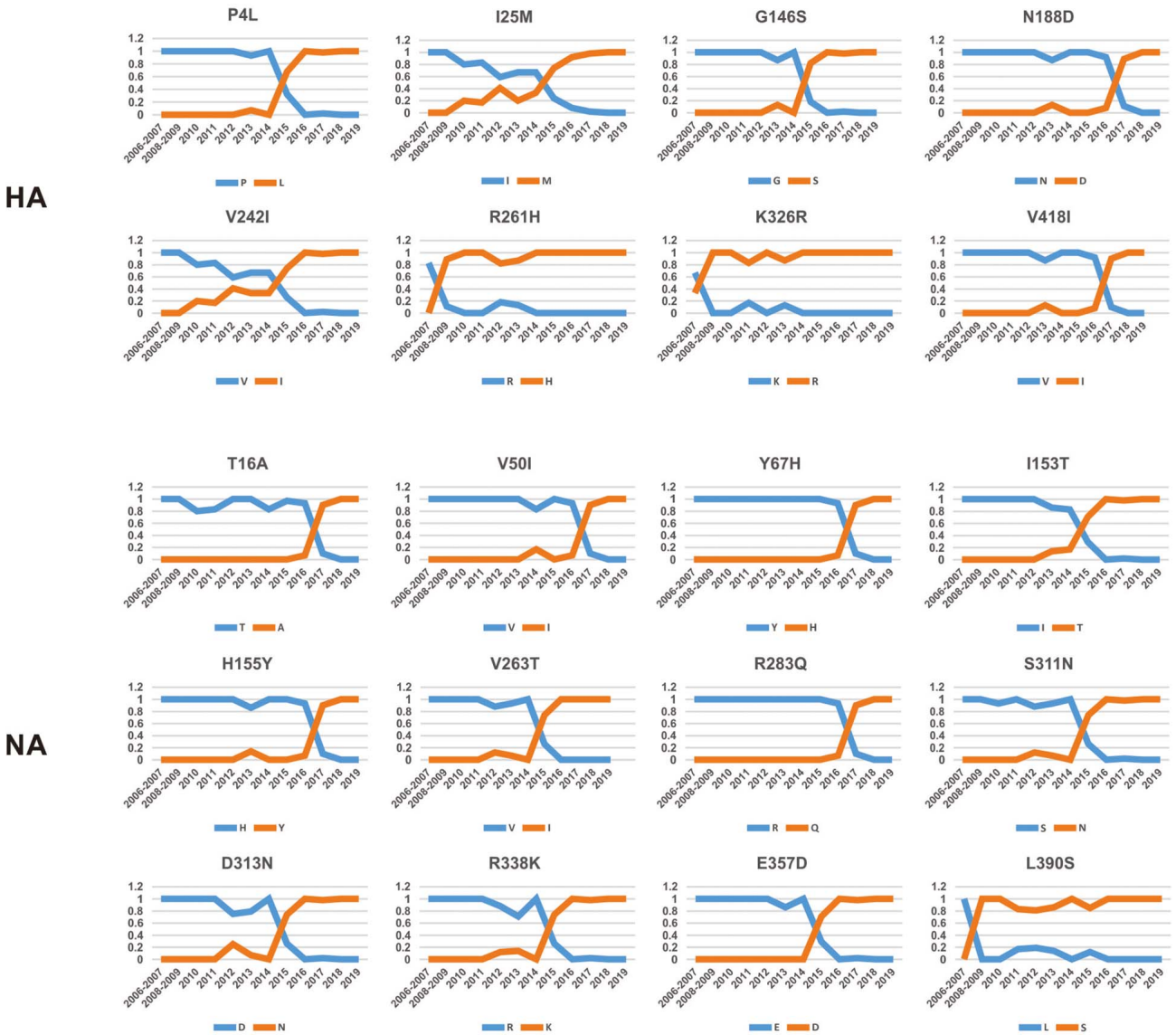


FIGURE 2 | Dynamic changes in the amino acid frequencies for sites in each effective substitution site in the HA and NA gene segments of H3N2 CIV during the circulation phase.

determinant to hone the function of the immune suppression by H3N2 CIV. In addition, M1-V15I (Chen et al., 2007), PA-C241Y (Yamaji et al., 2015), PB1-R187K (Liu et al., 2016), PB2-K251R (Prokopyeva et al., 2016), and PB2-G590S (Mehle and Doudna, 2009) substitutions have all been reported to enhance pathogenicity and virulence of AIVs in mammalian hosts due to their contributions to improving replicative efficacy and assembly fitness. Notably, no substitution was found to be fixed in the NP gene, which may indicate that the adaptation of H3N2 CIV to canines is less reflected by changes in the NP segment (**Table 1**). Such an observation is consistent with higher selective constraints (as evident from the lower dN/dS ratio) in the NP gene compared with other gene segments (**Figure 1**).

Analyses to identify sites being suffered by positive selection in viruses have been wildly used to characterize signatures of adaptive evolution (Joseph et al., 2015, 2018; He et al., 2019). He et al. (2019) identified four positively selected sites (4, 218, 436, and 453) in HA and one positively selected site (222) in NA for H3N2 CIV using the dN/dS method. The present analysis involves a larger sequence dataset associated with H3N2 CIV whole genomes and identified 12 residues in six viral gene segments that are suggested to have experienced positive selection (**Table 2**). Of these 12 sites, three (HA-436, HA-453, and NA-222) were previously reported by He et al. (2019). Interestingly, real-time scanning of amino acid frequency at these sites, both from the current study and from the report by He et al. (2019), clearly demonstrates that no

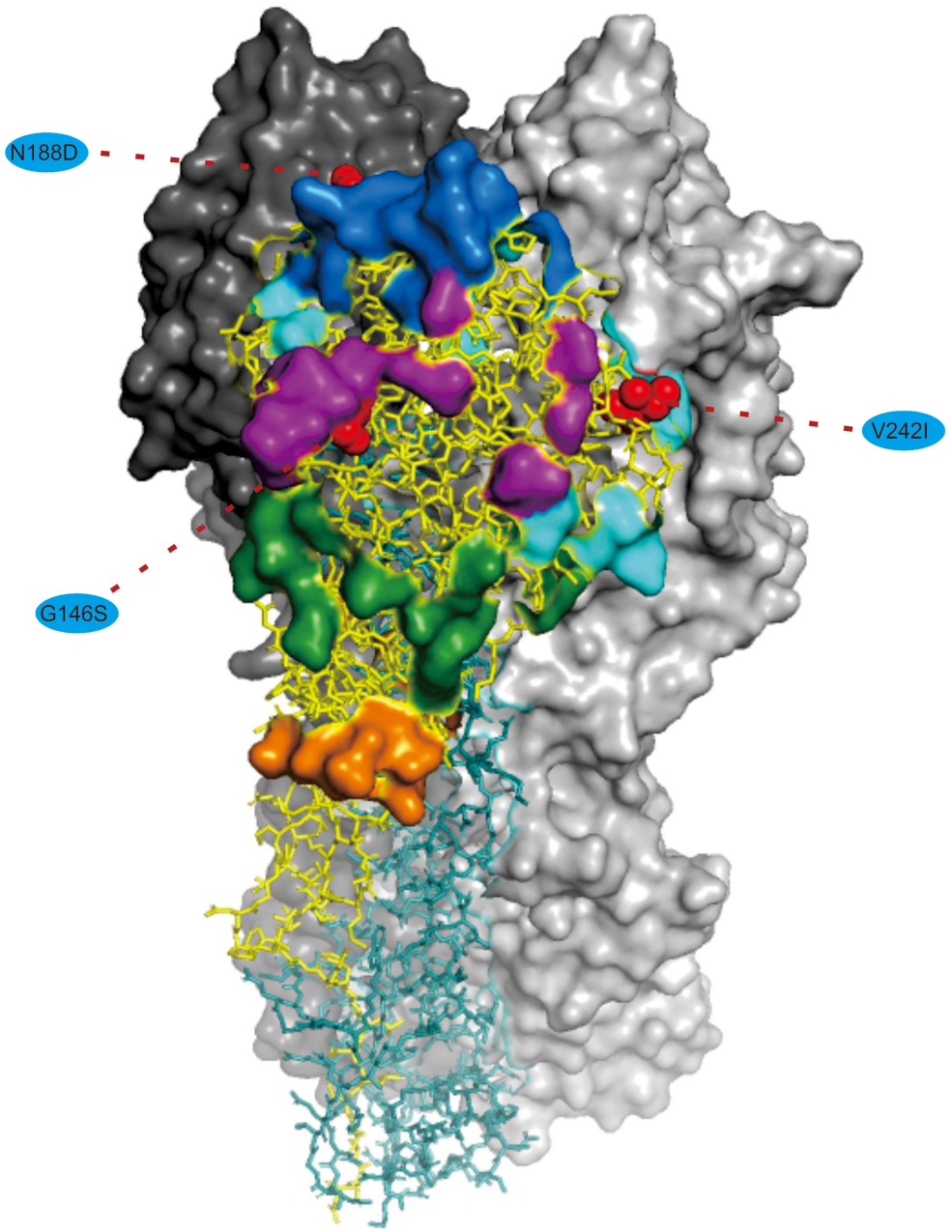


FIGURE 3 | Mapping of selected effective substitutions onto the known 3D structure of HA (PDB ID: 6N4F). The monomers representing the HA1 and HA2 subunits are represented in yellow and blue, respectively. A trimer complex is shown in surface representation with the antigenic sites highlighted. Antigenic epitopes A, B, C, D, and E are marked in purple, blue, orange, sky blue, and green, respectively. Numbers in colored ellipse represent codon alignment number (H3 numbering). The G146S, N188D and V242I substitutions are shown as red spheres that are located at the antigenic epitopes A, B, and D, respectively.

significant change in frequency has occurred at these sites, with the exception of HA-4 (**Supplementary Figure 5**). Although sites M2-22, M2-29, NA-36, and NA-222 had undergone one or more substitutions, none have been fixed in the examined circulation time, in conflict with the expectations of positive selection acting on these residues. On the other hand, most of the effective substitutions observed in the current study were not detected as having signals of positive selection with the dN/dS method. These observations highlight the sensitivity and robustness of the frequency diagram analysis method in revealing footprints of natural selection in comparison to inferences derived from dN/dS analysis.

Host selection pressure on viral replicative efficacy and adaptation leaves a footprint in the base composition of a viral genome (Greenbaum et al., 2008; Guo et al., 2020). The genetic code is degenerate, and the non-random usage

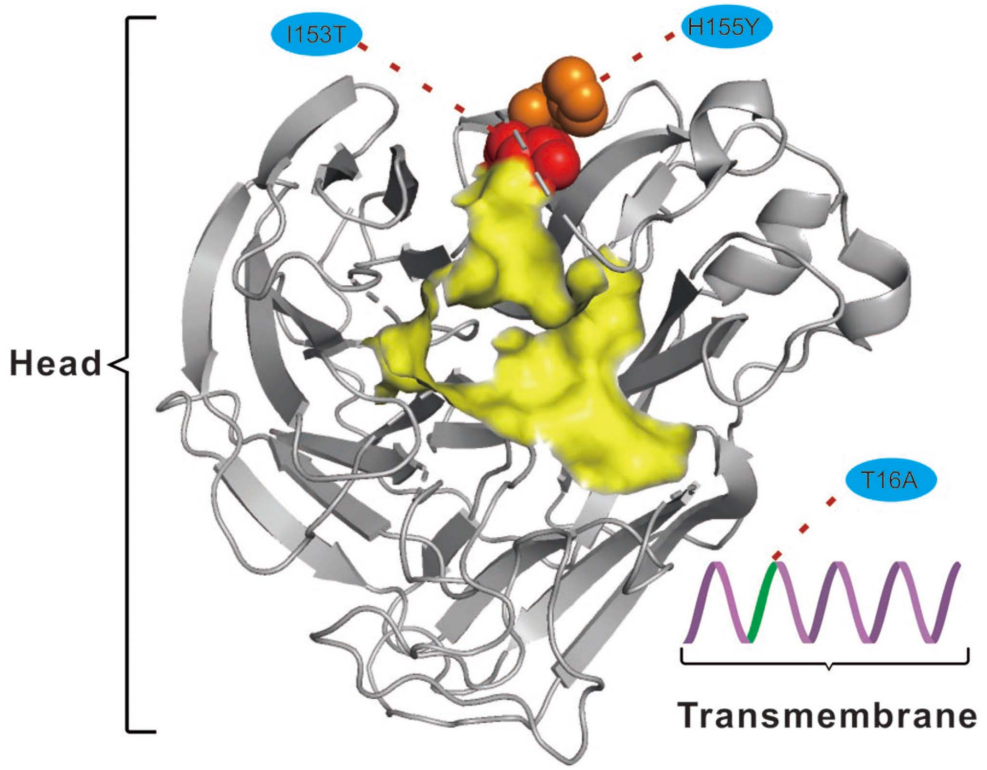


FIGURE 4 | Mapping of selected effective substitutions onto the known 3D structure of NA (PDB ID: 6N4D). The I153T (represented as a red sphere) and H155Y (represented as an orange sphere) substitutions are located close to the active site (highlighted in yellow) on the head region (shown as gray cartoon) of the NA protein. The T16A (shown as green cartoon) substitution is located in the transmembrane region (shown as a purple cartoon) of the NA protein.

TABLE 2 | Sites under positive selection in the H3N2 CIV gene segments.

		Test methods							
		MEME		FAUBAR		FEL		SLAC	
Segment	Codon	ω^+	*p*-value	dN-dS	Post.Pro	dN-dS	*p*-value	dN-dS	*p*-value
HA	141		NS	3.602	0.906	3.004	0.083		ns
	436	3.66	0.08	5.632	0.943	3.353	0.067		ns
	453	5.38	0.03	8.087	0.987	5.244	0.022	10	0.0609
M1	15	3.36	0.09	8.87	0.984	3.36	0.067		ns
M2	22	6.58	0.02	11.762	0.987	6.582	0.01	12.6	0.0642
	29	3.51	0.08	5.444	0.959		ns		ns
NA	36	3.21	0.1	4.943	0.96	3.208	0.073		ns
	222		NS	3.665	0.901	2.842	0.092		ns
NS1	71	3.25	0.09	6.967	0.963	3.251	0.071		ns
PA	99		NS	4.911	0.953	3.076	0.079		ns
	237		NS	4.787	0.935	2.762	0.097		ns
PB1	723	4.06	0.06	4.67	0.953	4.059	0.044		ns

Position in HA is based on the H3 numbering. MEME $p < 0.1$, FEL $p < 0.1$, SLAC $p < 0.1$, FAUBAR Post.Pro > 0.9. NS, non-significant.

of synonymous codons leads to codon usage bias (Shah and Gilchrist, 2011). Codon usage is believed to be mainly shaped by a "mutation-selection-drift balance" (Shah and Gilchrist, 2011). Viruses, owing to their small genomes, rely on host cellular machinery for processes such as replication and protein synthesis and show a strong relationship with host codon usage. It is believed that a higher similarity of their codon usage with the host codon usage profile provides them with

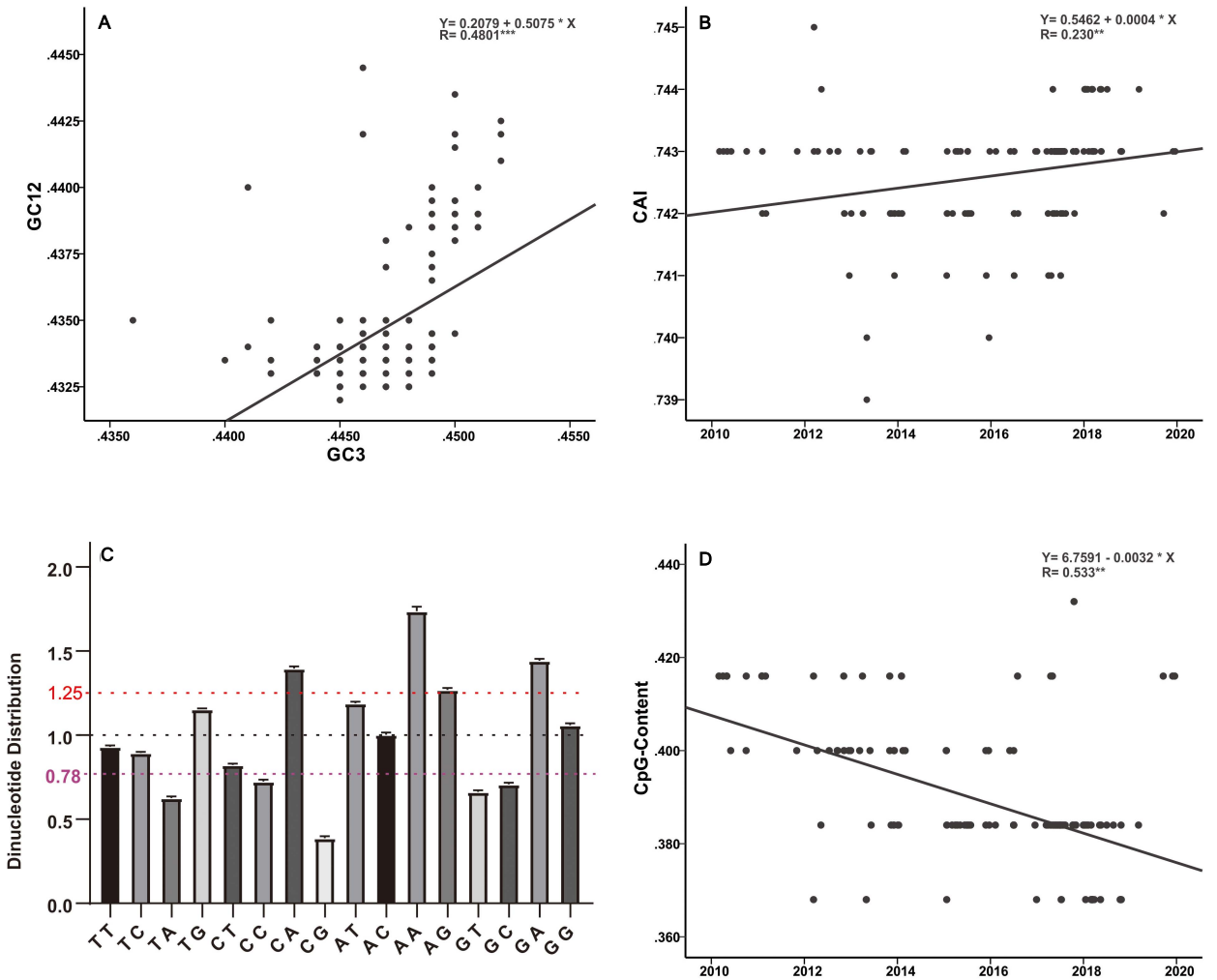


FIGURE 5 | Base composition analysis of the H3N2 CIV genomes. **(A)** GC3 values are plotted against the GC12 values to generate a neutrality plot. **(B)** CAI values calculated with respect to the canine host are plotted according to collection date. **(C)** Relative dinucleotide abundance analysis for the H3N2 CIV. Dotted lines in red, black, and purple signify the optimal values for overrepresentation, average representation, and underrepresentation of the dinucleotides, respectively. **(D)** CpG contents plotted according to sampling date of the H3N2 CIV.

an advantage as they would be adapted to host tRNA pool and thus could efficiently generate new progeny (Tian et al., 2018). Such conjectures have been verified by previous functional experiments (Carlini, 2004; Mueller et al., 2006; Coleman et al., 2008). In addition, it has been reported that current H3N2 viruses enhance replication in interferon (IFN)-treated human cells, by gradually skewing codon usage toward the interferon-altered tRNA pool (Smith et al., 2018). In the present analysis, the slope of the regression line pertaining to the neutrality plot between GC3 values (*x*-axis) and GC12 (*x*-axis) of the H3N2 CIV genes is 0.5075 (**Figure 5A**). It has been suggested that the slope of these plots reflects the degree of compositional constraint operating on the genes of interest (Sueoka, 1988). Such an observation indicates that both of the compositional constraint (50.75%) and natural selection (49.25%) have a similar magnitude in influencing the codon usage of H3N2 CIV genomes. CAI analysis reveals the viral adaptability in codon usage to that of corresponding host(s), and a higher CAI indicates strong adaptation to host cellular machinery (Sharp and Li, 1987). Consistent with a previous study (Li et al., 2018), we found that H3N2 CIV exhibited better adaptation to the duck's codon usage pattern than that of the canine, which is evident from the observation that H3N2 CIV showed a significantly higher mean CAI value to the duck host than that to the canine host (0.818464 ± 0.000953 for duck host versus 0.742636 ± 0.000894 for canine host, $p < 0.001$) (**Supplementary Table 2**), possibly due to a long-term circulation in the avian environment before transmission to canines. However, the present findings reveal a reduction in the distance between H3N2 CIV and the canine codon bias over time (**Figure 5B**), thus indicating increasing fitness to the canine usage pattern for H3N2 CIV. Similar observations have been reported for canine parvovirus type 2 and H1N1/pdm2009 viruses, where the host shift event

from their reservoir host to the new host environment leads to subsequent adaptation in codon usage to better fit the cellular machinery of the new hosts (Franzo et al., 2017; Guo et al., 2020).

Analysis of the relative abundance of the 16 possible dinucleotides in the H3N2 CIV genomes revealed a distinct usage preference among the dinucleotides (**Figure 5C** and **Supplementary Table 2**). It was noted that the CpA, ApA, and GpA dinucleotides were significantly overrepresented (relative abundance > 1.25). In contrast, TpA, CpC, GpT, GpC, and CpG dinucleotides were significantly underrepresented (relative abundance < 0.78). Notably, only the CpG dinucleotide was observed to be extremely underrepresented (relative abundance < 0.50) (**Figure 5C** and **Supplementary Table 2**). Extensive analyses of the relative abundance of the CpG dinucleotide over epidemiological time showed that H3N2 CIV has experienced a continuous decrease in CpG dinucleotide content (**Figure 5D**). It is believed that CpG motifs in viral genomes are treated as viral molecular patterns and are recognized by the host's intracellular pattern recognition receptor Toll-like receptor 9 (TLR9), leading to the stimulation of immune responses against the viral pathogens. Thus, suppression of CpG dinucleotides points toward a strategy by the viruses to avoid host immune signals (Sivori et al., 2004; Jimenez-Baranda et al., 2011). In addition, a tendency of sustained reduction of CpG motifs in the H3N2 CIV genome is indicative of these viral pathogens continuing to improve their response to evade host immune signals (**Figure 5D** and **Supplementary Table 2**). Such observations have previously been reported for the 1918 human H1N1, 1968 human H3N2, and 1963 horse H3N8 viruses during their circulation phases (Greenbaum et al., 2008; Jimenez-Baranda et al., 2011).

CONCLUSION

Influenza A virus endemics in animals such as canines and cats should be carefully monitored due to their large global population, with many of them having close contact with humans, thus they pose potential risks to public health. Recently, in 2016, a veterinarian in a New York animal shelter was infected through a transmission event for a feline H7N2 influenza virus (Belser et al., 2017). Such instance forewarns potential viral transmission through close contact between humans and their companion animals (Belser et al., 2017). In addition, the fact that the respiratory tract of dogs harbors receptors for both avian-adapted and mammal-adapted influenza viruses (Tate et al., 2014) and the increasing evidence of canine infection with swine, human, and avian IAVs of different subtypes (Chen et al., 2018; Zhao et al., 2020) point toward a capacity for canines to serve as "mixing vessels" to generate novel gene constellations, which may lead to future pandemics in the human population.

The present study deals with the identification and profiling of genetic changes during the process of adaptation of H3N2 CIV over epidemiological time. We observed that H3N2 experiences a higher level of adaptive selection pressure that yielded 54 effective substitutions that became fixed in the viral population. Some of these mutations have been experimentally verified to play important roles in allowing AIVs to cross species barriers, while others should be treated as candidate sites for further functional experiments. Most of the substitutions became fixed around 2015, which might have facilitated the invasion and further circulation of this lineage of CIV from South Asia to North America. In addition, the present findings revealed evidence of convergent evolution in the different CIV lineages. We also detected multiple sites under positive selection using the dN/dS analysis method and showed that the frequency diagram analysis method might exhibit higher sensitivity and robustness for revealing the impact of natural selection. Codon usage analysis, as executed in the present study, indicated that H3N2 CIV is better adapted to the duck's cellular machinery, in comparison to canine host, thus supporting existing knowledge regarding the transmission route of the virus from an avian to canine host. In the present study, an increasing fitness of the virus to the host microenvironment, following its emergence in the canine population, was also detected by employing codon usage and dinucleotide distribution analyses. Such observations reveal that the adaptive evolution of H3N2 CIV, after its original incursion, further increased adaptation to the canine host environment.

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 author/s.

AUTHOR CONTRIBUTIONS

YS designed and supervised the study. FG, AR, RW, JY, ZZ, WL, and XS generated the data. FG analyzed the data. FG, YS, AR, and DI wrote and prepared the manuscript. R-AC commented on and revised the drafts of the manuscript. All the authors have read and agreed to submission of the manuscript.

FUNDING

This work was supported by Guangdong Major Project of Basic and Applied Basic Research (No. 2020B0301030007), the National Natural Science Foundation of China (31822056), the Guangdong Science and Technology Innovation Leading Talent Program (2019TX05N098), the 111 Project (D20008), the Department of Education of Guangdong Province (2019KZDXM004 and 2019KCXTD001), the Department of Science and Technology of Guangdong Province (2020B1111320002), Department of Agriculture of Guangdong Province, and the Xijiang innovation team of Zhaoqing city.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.655228/full#supplementary-material

Supplementary Figure 1 | Regional distribution of the 301 H3N2 CIV strains isolated from 2005 to 2019 analyzed in the present study.

Supplementary Figure 2 | Comparative analysis of the global dN/dS ratio for each gene segment of H3N2 CIV (blue), H3N2 AIV (orange), and H3N8 CIV (gray) based on the reshuffled sub-datasets represented by circles.

Supplementary Figure 3 | Dynamic changes in the amino acid frequency for each effective substitution of the MP, NS, PA, PB1, and PB2 gene segments of H3N2 CIV during the circulation phase.

Supplementary Figure 4 | Time required (years) for the fixation of 54 effective substitutions in H3N2 CIV since their first occurrence (only continuous accumulations have been considered).

Supplementary Figure 5 | Dynamic changes in amino acid frequency for sites under positive selection (dN/dS > 1) during the circulation phase of H3N2 CIV. [a]Positive selection sites detected in current study. [b]Positive selection sites detected by He et al. (2019).

Supplementary Table 1 | Strain information (strain name, accession number, host, subtype, location) for the selection and amino acid frequency analyses associated with AIV (Sheet 1), H3N2 CIV (Sheet 2), and H3N8 CIV (Sheet 3). Actual sequence number for each dataset that were used for identifying effective substitutions over epidemiological time for H3N2 CIV has also been provided in Sheet 4.

Supplementary Table 2 | Strain name, collection date, digital date, location, base composition and CAI values for 220 completely sequenced H3N2 CIV genomes.

REFERENCES

Anderson, T. C., Bromfield, C. R., Crawford, P. C., Dodds, W. J., Gibbs, E. P. J., and Hernandez, J. A. (2012). Serological evidence of H3N8 canine influenza-like virus circulation in USA dogs prior to 2004. *Vet. J.* 191, 312–316. doi: 10.1016/j.tvjl.2011.11.010

Belser, J. A., Pulit-Penaloza, J. A., Sun, X., Brock, N., Pappas, C., Creager, H. M., et al. (2017). A novel A(H7N2) influenza virus isolated from a veterinarian caring for cats in a New York City animal shelter causes mild disease and transmits poorly in the ferret model. *J. Virol.* 91:e00672-17. doi: 10.1128/JVI.00672-17

Borland, S., Gracieux, P., Jones, M., Mallet, F., and Yugueros-Marcos, J. (2020). Influenza A virus infection in cats and dogs: a literature review in the light of the "One Health" Concept. *Front. Public Health* 8:83. doi: 10.3389/fpubh.2020.00083

Bunpapong, N., Nonthabenjawan, N., Chaiwong, S., Tangwangvivat, R., Boonyapisitsopa, S., Jairak, W., et al. (2014). Genetic characterization of canine influenza A virus (H3N2) in Thailand. *Virus Genes* 48, 56–63. doi: 10.1007/s11262-013-0978-z

Carlini, D. B. (2004). Experimental reduction of codon bias in the Drosophila alcohol dehydrogenase gene results in decreased ethanol tolerance of adult flies. *J. Evol. Biol.* 17, 779–785. doi: 10.1111/j.1420-9101.2004.00725.x

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, Y., Trovao, N. S., Wang, G., Zhao, W., He, P., Zhou, H., et al. (2018). Emergence and Evolution of Novel Reassortant Influenza A Viruses in Canines in Southern China. *mBio* 9:e00909-18. doi: 10.1128/mBio.00909-18

Coleman, J. R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., and Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science* 320, 1784–1787. doi: 10.1126/science.1155761

Crawford, P. C., Dubovi, E. J., Castleman, W. L., Stephenson, I., Gibbs, E. P., Chen, L., et al. (2005). Transmission of equine influenza virus to dogs. *Science* 310, 482–485. doi: 10.1126/science.1117950

Cristina, J., Moreno, P., Moratorio, G., and Musto, H. (2015). Genome-wide analysis of codon usage bias in ebolavirus. *Virus Res.* 196, 87–93. doi: 10.1016/j.virusres.2014.11.005

Daly, J. M., Blunden, A. S., Macrae, S., Miller, J., Bowman, S. J., Kolodziejek, J., et al. (2008). Transmission of equine influenza virus to English foxhounds. *Emerg. Infect. Dis.* 14, 461–464. doi: 10.3201/eid1403.070643

Finkelstein, D. B., Mukatira, S., Mehta, P. K., Obenauer, J. C., Su, X., Webster, R. G., et al. (2007). Persistent host markers in pandemic and H5N1 influenza viruses. *J. Virol.* 81, 10292–10299. doi: 10.1128/JVI.00921-07

Franzo, G., Tucciarone, C. M., Cecchinato, M., and Drigo, M. (2017). Canine parvovirus type 2 (CPV-2) and Feline panleukopenia virus (FPV) codon bias analysis reveals a progressive adaptation to the new niche after the host jump. *Mol. Phylogenet. Evol.* 114, 82–92. doi: 10.1016/j.ympev.2017.05.019

Gabriel, G., Abram, M., Keiner, B., Wagner, R., Klenk, H. D., and Stech, J. (2007). Differential polymerase activity in avian and mammalian cells determines host range of influenza virus. *J. Virol.* 81, 9601–9604. doi: 10.1128/JVI.00666-07

Gabriel, G., Dauber, B., Wolff, T., Planz, O., Klenk, H. D., and Stech, J. (2005). The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18590–18595. doi: 10.1073/pnas.0507415102

Greenbaum, B. D., Levine, A. J., Bhanot, G., and Rabadan, R. (2008). Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathog.* 4:e1000079. doi: 10.1371/journal.ppat.1000079

Guo, F., Luo, T., Pu, Z., Xiang, D., Shen, X., Irwin, D. M., et al. (2018). Increasing the potential ability of human infections in H5N6 avian influenza A viruses. *J. Infect.* 77, 349–356. doi: 10.1016/j.jinf.2018.07.015

Guo, F., Yang, J., Pan, J., Liang, X., Shen, X., Irwin, D. M., et al. (2020). Origin and evolution of H1N1/pdm2009: a codon usage perspective. *Front. Microbiol.* 11:1615. doi: 10.3389/fmicb.2020.01615

Guo, F. C., Li, Y. L., Yu, S., Liu, L., Luo, T. T., Pu, Z. Q., et al. (2019). Adaptive evolution of human-isolated H5Nx Avian Influenza A Viruses. *Front. Microbiol.* 10:1328. doi: 10.3389/fmicb.2019.01328

He, W., Li, G., Zhu, H., Shi, W., Wang, R., Zhang, C., et al. (2019). Emergence and adaptation of H3N2 canine influenza virus from avian influenza virus: an overlooked role of dogs in interspecies transmission. *Transbound. Emerg. Dis.* 66, 842–851. doi: 10.1111/tbed.13093

Henningson, J. N., Rajao, D. S., Kitikoon, P., Lorusso, A., Culhane, M. R., Lewis, N. S., et al. (2015). Comparative virulence of wild-type H1N1pdm09 influenza A isolates in swine. *Vet. Microbiol.* 176, 40–49. doi: 10.1016/j.vetmic.2014.12.021

Jimenez-Baranda, S., Greenbaum, B., Manches, O., Handler, J., Rabadan, R., Levine, A., et al. (2011). Oligonucleotide motifs that disappear during the evolution of influenza virus in humans increase alpha interferon secretion by plasmacytoid dendritic cells. *J. Virol.* 85, 3893–3904. doi: 10.1128/JVI.01908-10

Joseph, U., Linster, M., Suzuki, Y., Krauss, S., Halpin, R. A., Vijaykrishna, D., et al. (2015). Adaptation of Pandemic H2N2 Influenza A Viruses in Humans. *J. Virol.* 89, 2442–2447. doi: 10.1128/Jvi.02590-14

Joseph, U., Vijaykrishna, D., Smith, G. J. D., and Su, Y. C. F. (2018). Adaptive evolution during the establishment of European avian-like H1N1 influenza A virus in swine. *Evol. Appl.* 11, 534–546. doi: 10.1111/eva.12536

Kirkland, P. D., Finlaison, D. S., Crispe, E., and Hurt, A. C. (2010). Influenza virus transmission from horses to dogs, Australia. *Emerg. Infect. Dis.* 16, 699–702. doi: 10.3201/eid1604.091489

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

Kunec, D., and Osterrieder, N. (2016). Codon pair bias is a direct consequence of Dinucleotide Bias. *Cell Rep.* 14, 55–67. doi: 10.1016/j.celrep.2015.12.011

Li, G., Wang, R., Zhang, C., Wang, S., He, W., Zhang, J., et al. (2018). Genetic and evolutionary analysis of emerging H3N2 canine influenza virus. *Emerg. Microbes Infect.* 7:73. doi: 10.1038/s41426-018-0079-0

Li, S., Shi, Z., Jiao, P., Zhang, G., Zhong, Z., Tian, W., et al. (2010). Avian-origin H3N2 canine influenza A viruses in Southern China. *Infect. Genet. Evol.* 10, 1286–1288. doi: 10.1016/j.meegid.2010.08.010

Li, W. (1997). *Molecular Evolution*. (Sunderland, MA: Sinauer Associates), 487.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Lin, Y., Xie, X., Zhao, Y., Kalhoro, D. H., Lu, C., and Liu, Y. (2016). Enhanced replication of avian-origin H3N2 canine influenza virus in eggs, cell cultures and mice by a two-amino acid insertion in neuraminidase stalk. *Vet. Res.* 47:53. doi: 10.1186/s13567-016-0337-x

Liu, Q. T., Liu, Y. Z., Yang, J., Huang, X. M., Han, K. K., Zhao, D. M., et al. (2016). Two genetically similar H9N2 Influenza A Viruses show different pathogenicity in mice. *Front. Microbiol.* 7:1737. doi: 10.3389/fmicb.2016.01737

Long, J. S., Mistry, B., Haslam, S. M., and Barclay, W. S. (2019). Host and viral determinants of influenza A virus species specificity. *Nat. Rev. Microbiol.* 17, 67–81. doi: 10.1038/s41579-018-0115-z

Lyu, Y., Song, S., Zhou, L., Bing, G., Wang, Q., Sun, H., et al. (2019). Canine Influenza Virus A(H3N2) Clade with Antigenic Variation, China, 2016-2017. *Emerg. Infect. Dis.* 25, 161–165. doi: 10.3201/eid2501.171878

Matos-Patron, A., Byrd-Leotis, L., Steinhauer, D. A., Barclay, W. S., and Ayora-Talavera, G. (2015). Amino acid substitution D222N from fatal influenza infection affects receptor-binding properties of the influenza A(H1N1)pdm09 virus. *Virology* 484, 15–21. doi: 10.1016/j.virol.2015.05.012

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

Mehle, A., and Doudna, J. A. (2009). Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21312–21316. doi: 10.1073/pnas.0911915106

Mitnaul, L. J., Matrosovich, M. N., Castrucci, M. R., Tuzikov, A. B., Bovin, N. V., Kobasa, D., et al. (2000). Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *J. Virol.* 74, 6015–6020. doi: 10.1128/jvi.74.13.6015-6020.2000

Mostafa, A., Abdelwhab, E. M., Mettenleiter, T. C., and Pleschka, S. (2018). Zoonotic potential of Influenza A Viruses: a comprehensive overview. *Viruses* 10:497. doi: 10.3390/v10090497

Mueller, S., Papamichail, D., Coleman, J. R., Skiena, S., and Wimmer, E. (2006). Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J. Virol.* 80, 9687–9696. doi: 10.1128/jvi.00738-06

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., et al. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* 30, 1196–1205. doi: 10.1093/molbev/mst030

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

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Prokopyeva, E. A., Sobolev, I. A., Prokopyev, M. V., and Shestopalov, A. M. (2016). Adaptation of influenza A(H1N1)pdm09 virus in experimental mouse models. *Infect. Genet. Evol.* 39, 265–271. doi: 10.1016/j.meegid.2016.01.022

Pu, Z., Xiang, D., Li, X., Luo, T., Shen, X., Murphy, R. W., et al. (2018). Potential pandemic of H7N9 Avian Influenza A Virus in human. *Front. Cell. Infect. Microbiol.* 8:414. doi: 10.3389/fcimb.2018.00414

Puigbo, P., Bravo, I. G., and Garcia-Vallve, S. (2008). CAIcal: a combined set of tools to assess codon usage adaptation. *Biol. Direct* 3:38. doi: 10.1186/1745-6150-3-38

Pulit-Penaloza, J. A., Simpson, N., Yang, H., Creager, H. M., Jones, J., Carney, P., et al. (2017). Assessment of molecular, antigenic, and pathological features of canine Influenza A(H3N2) viruses that emerged in the United States. *J. Infect. Dis.* 216(Suppl. 4), S499–S507. doi: 10.1093/infdis/jiw620

Shah, P., and Gilchrist, M. A. (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10231–10236. doi: 10.1073/pnas.1016719108

Sharp, P. M., and Li, W. H. (1987). The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295. doi: 10.1093/nar/15.3.1281

Shen, H. X., Li, X., Yang, D. Q., Ju, H. B., Ge, F. F., Wang, J., et al. (2021). Phylogenetic analysis and evolutionary dynamics of H3N2 canine and feline influenza virus strains from 2006 to 2019. *J. Med. Virol.* 93, 3496–3507. doi: 10.1002/jmv.26767

Shih, A. C. C., Hsiao, T. C., Ho, M. S., and Li, W. H. (2007). Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6283–6288. doi: 10.1073/pnas.0701396104

Sivori, S., Falco, M., Della Chiesa, M., Carlomagno, S., Vitale, M., Moretta, L., et al. (2004). CpG and double-stranded RNA trigger human NK cells by Toll-like receptors: induction of cytokine release and cytotoxicity against tumors and dendritic cells. *Proc. Natl. Acad. Sci. U.S.A.* 101, 10116–10121. doi: 10.1073/pnas.0403744101

Skehel, J. J., and Wiley, D. C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* 69, 531–569. doi: 10.1146/annurev.biochem.69.1.531

Smith, B. L., Chen, G., Wilke, C. O., and Krug, R. M. (2018). Avian Influenza Virus PB1 Gene in H3N2 Viruses Evolved in Humans To Reduce Interferon Inhibition by Skewing Codon Usage toward Interferon-Altered tRNA Pools. *mBio* 9:e01222-18. doi: 10.1128/mBio.01222-18

Sueoka, N. (1988). Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 85, 2653–2657. doi: 10.1073/pnas.85.8.2653

Tate, M. D., Job, E. R., Deng, Y. M., Gunalan, V., Maurer-Stroh, S., and Reading, P. C. (2014). Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses* 6, 1294–1316. doi: 10.3390/v6031294

Tian, L., Shen, X., Murphy, R. W., and Shen, Y. (2018). The adaptation of codon usage of +ssRNA viruses to their hosts. *Infect. Genet. Evol.* 63, 175–179. doi: 10.1016/j.meegid.2018.05.034

Voorhees, I. E. H., Dalziel, B. D., Glaser, A., Dubovi, E. J., Murcia, P. R., Newbury, S., et al. (2018). Multiple Incursions and Recurrent Epidemic Fade-Out of H3N2 Canine Influenza A Virus in the United States. *J. Virol.* 92:e00323-18. doi: 10.1128/JVI.00323-18

Voorhees, I. E. H., Glaser, A. L., Toohey-Kurth, K., Newbury, S., Dalziel, B. D., Dubovi, E. J., et al. (2017). Spread of Canine Influenza A(H3N2) Virus, United States. *Emerg. Infect. Dis.* 23, 1950–1957. doi: 10.3201/eid2312.170246

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., and Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* 35, 773–777. doi: 10.1093/molbev/msx335

Weese, J. S., Anderson, M. E. C., Berhane, Y., Doyle, K. F., Leutenegger, C., Chan, R., et al. (2019). Emergence and Containment of Canine Influenza Virus A(H3N2), Ontario, Canada, 2017-2018. *Emerg. Infect. Dis.* 25, 1810–1816. doi: 10.3201/eid2510.190196

Wu, M. H., Su, R. S., Gu, Y. X., Yu, Y. N., Li, S., Sun, H. P., et al. (2021). Molecular Characteristics, Antigenicity, Pathogenicity, and Zoonotic Potential of a H3N2 Canine Influenza Virus Currently Circulating in South China. *Front. Microbiol.* 12:628979. doi: 10.3389/fmicb.2021.628979

Yamaji, R., Yamada, S., Le, M. Q., Ito, M., Sakai-Tagawa, Y., and Kawaoka, Y. (2015). Mammalian adaptive mutations of the PA protein of highly pathogenic avian H5N1 influenza virus. *J. Virol.* 89, 4117–4125. doi: 10.1128/JVI.03532-14

Yang, G. H., Li, S. J., Blackmon, S., Ye, J. Q., Bradley, K. C., Cooley, J., et al. (2013). Mutation tryptophan to leucine at position 222 of haemagglutinin could facilitate H3N2 influenza A virus infection in dogs. *J. Gen. Virol.* 94, 2599–2608. doi: 10.1099/vir.0.054692-0

Yen, H. L., Liang, C. H., Wu, C. Y., Forrest, H. L., Ferguson, A., Choy, K. T., et al. (2011). Hemagglutinin-neuraminidase balance confers respiratory-droplet transmissibility of the pandemic H1N1 influenza virus in ferrets. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14264–14269. doi: 10.1073/pnas.1111000108

Yoon, S. W., Webby, R. J., and Webster, R. G. (2014). Evolution and ecology of influenza A viruses. *Curr. Top. Microbiol. Immunol.* 385, 359–375. doi: 10.1007/82_2014_396

Zhao, S., Schuurman, N., Tieke, M., Quist, B., Zwinkels, S., van Kuppeveld, F. J. M., et al. (2020). Serological Screening of Influenza A Virus Antibodies in Cats and Dogs Indicates Frequent Infection with Different Subtypes. *J. Clin. Microbiol.* 58:e01689-20. doi: 10.1128/JCM.01689-20

Zhu, H., Hughes, J., and Murcia, P. R. (2015). Origins and evolutionary dynamics of H3N2 canine Influenza Virus. *J. Virol.* 89, 5406–5418. doi: 10.1128/JVI.03395-14

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 24 June 2021
doi: 10.3389/fmicb.2021.689658

Coxsackievirus A16 in Southern Vietnam

Le Nguyen Truc Nhu[1], Le Nguyen Thanh Nhan[2], Nguyen To Anh[1], Nguyen Thi Thu Hong[1], Hoang Minh Tu Van[1,3], Tran Tan Thanh[1], Vu Thi Ty Hang[1], Do Duong Kim Han[1], Nguyen Thi Han Ny[1], Lam Anh Nguyet[1], Du Tuan Quy[2], Phan Tu Qui[4], Truong Huu Khanh[2], Nguyen Thanh Hung[2], Ha Manh Tuan[3], Nguyen Van Vinh Chau[4], Guy Thwaites[1,5], H. Rogier van Doorn[1,5] and Le Van Tan[1]**

[1] Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam, [2] Children's Hospital 1, Ho Chi Minh City, Vietnam, [3] Children's Hospital 2, Ho Chi Minh City, Vietnam, [4] Hospital for Tropical Disease, Ho Chi Minh City, Vietnam, [5] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:
Justin Jang Hann Chu, National University of Singapore, Singapore

Reviewed by:
Sheng-Wen Huang, National Health Research Institutes, Taiwan
Zhenglun Liang, National Institutes for Food and Drug Control, China

***Correspondence:**
Le Nguyen Truc Nhu
nhulnt@oucru.org
Le Van Tan
tanlv@oucru.org

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 01 April 2021
Accepted: 11 May 2021
Published: 24 June 2021

Citation:
Nhu LNT, Nhan LNT, Anh NT, Hong NTT, Van HMT, Thanh TT, Hang VTT, Han DDK, Ny NTH, Nguyet LA, Quy DT, Qui PT, Khanh TH, Hung NT, Tuan HM, Chau NVV, Thwaites G, Doorn HRv and Tan LV (2021) Coxsackievirus A16 in Southern Vietnam. Front. Microbiol. 12:689658. doi: 10.3389/fmicb.2021.689658

Background: Hand, Foot and Mouth Disease (HFMD) is a major public health concern in the Asia-Pacific region. Most recent HFMD outbreaks have been caused by enterovirus A71 (EV-A71), coxsackievirus A16 (CVA16), CVA10, and CVA6. There has been no report regarding the epidemiology and genetic diversity of CVA16 in Vietnam. Such knowledge is critical to inform the development of intervention strategies.

Materials and Methods: From 2011 to 2017, clinical samples were collected from in- and outpatients enrolled in a HFMD research program conducted at three referral hospitals in Ho Chi Minh City (HCMC), Vietnam. Throat or rectal swabs positive for CVA16 with sufficient viral load were selected for whole genome sequencing and evolutionary analysis.

Results: Throughout the study period, 320 CVA16 positive samples were collected from 2808 HFMD patients (11.4%). 59.4% of patients were male. The median age was 20.8 months (IQR, 14.96–31.41). Patients resided in HCMC (55.3%), Mekong Delta (22.2%), and South East Vietnam (22.5%). 10% of CVA16 infected patients had moderately severe or severe HFMD. CVA16 positive samples from 153 patients were selected for whole genome sequencing, and 66 complete genomes were obtained. Phylogenetic analysis demonstrated that Vietnamese CVA16 strains belong to a single genogroup B1a that clusters together with isolates from China, Japan, Thailand, Malaysia, France and Australia. The CVA16 strains of the present study were circulating in Vietnam some 4 years prior to its detection in HFMD cases.

Conclusion: We report for the first time on the molecular epidemiology of CVA16 in Vietnam. Unlike EV-A71, which showed frequent replacement between subgenogroups B5 and C4 every 2–3 years in Vietnam, CVA16 displays a less pronounced genetic alternation with only subgenogroup B1a circulating in Vietnam since 2011. Our collective findings emphasize the importance of active surveillance for viral circulation in HFMD endemic countries, critical to informing outbreak response and vaccine development.

Keywords: coxsackievirus A16, hand foot mouth disease, picornavirus, evolution, Vietnam

INTRODUCTION

First described in 1957 (Esposito and Principi, 2018), Hand, Foot and Mouth Disease (HFMD) is a common childhood illness caused by a wide range of enteroviruses (EVs) of the genus *Enterovirus* of the family *Picornaviridae* (Cai et al., 2019). The disease is characterized with lesions on the skin and oral mucosa, and is generally self-limited within a few days. However, clinical complications may occur, and include encephalitis, myocarditis, aseptic meningitis, pulmonary edema, acute flaccid paralysis and even death (Chen Q. et al., 2019). HFMD is now endemic in the Asia-Pacific region. Currently, there is no effective antiviral to treat the severe patients, while only EV-A71vaccine has been successfully developed and used in Mainland China.

Historically, most of the HFMD outbreaks have been caused by EV-A71 and Coxsackievirus A16 (CVA16) (Cai et al., 2019). EV-A71 can be associated with severe clinical outcomes while CVA16 is responsible for milder illness. In recent years, CVA6 and CVA10 have also been recognized as emerging pathogens of HFMD. EV-A71 and CVA16 are phylogenetically closely related (Brown et al., 2020). Genetically, CVA16 is divided into 2 genogroups A and B, with genogroup B being further divided into B1a, B1b, B1c, and B2 (Noisumdaeng et al., 2019). Recently new genogroups (C and D) have been reported in Peru, France and China (Chen L. et al., 2019).

In Vietnam, the first HFMD outbreak was reported in 2005, predominantly caused by EV-A71 and CVA16 (Phan et al., 2007). In 2008, HFMD became a notifiable disease in Vietnam, and a major outbreak was recorded in 2011–2012 with over 200,000 children admitted to hospitals and more than 200 deaths (Van Hoang et al., 2019b). Although HFMD cases have been reported nationwide, the number of reported, severe and fatal cases is greater in southern regions of Vietnam.

There are limited data regarding the epidemiology and genetic diversity of CVA16 from Vietnam. Likewise, globally, only 200 complete genomes of CVA16 have been deposited in GenBank or Virus Pathogen resource database[1]. Hence, to inform the development and implementation of intervention strategies, especially HFMD vaccines, we studied the molecular epidemiology of CVA16 in Vietnam.

MATERIALS AND METHODS

Study Program

Clinical samples used in this study were obtained from patients enrolled in a HFMD research program conducted at 3 referral hospitals in Ho Chi Minh City (HCMC), Viet Nam between 2011 and 2017 (Le et al., 2020). For the period between August 2011 and June 2013, the study was conducted in the pediatric intensive care unit (PICU) of the Hospital for Tropical Diseases (Geoghegan et al., 2015). Accordingly, only severe patients (i.e., those with grade 2b1 or above) were enrolled (grading system according to the guidelines issued by the Vietnamese Ministry of Health, **Supplementary Table 1**). In the second stage, July 2013–December 2017, the study sites were expanded to Children's Hospital 1 and Children's Hospital 2 and the recruitment covered patients seen in outpatient clinics, or admitted to the infectious disease wards or PICUs of these three study sites (Van Hoang et al., 2019a).

We collected information regarding demographic and clinical grades (**Supplementary Table 1**) on enrolment to the study and (if inpatients) daily until discharge or day 7 of hospitalization (whichever came first). Additionally, we sampled acute throat- and rectal swabs at enrolment from each participant. For the analysis of the present study, any throat or rectal swabs positive for CVA16 with a cycle threshold of ≤ 30 were selected for whole genome sequencing.

Random PCR and Library Preparation

Whole genome sequencing of selected samples was performed using a previously described MiSeq-based approach (Nguyen et al., 2016). Briefly, 110μl of the selected swab in viral transport medium was centrifuged at 13,000 rpm for 10 min to remove host cells or large cell debris. The collected supernatants were treated with DNase at 37°C for 30 min. Viral nucleic acid was then extracted using QIAamp viral RNA kit (Qiagen, Hilden, Germany) and recovered in 50 μL elution buffer provided with the kit. cDNA was synthesized from 10 ul of viral nucleic acid using Super Script III enzyme (Invitrogen, Carlsbad, CA, United States) and FR26RV-Endoh primer. This was followed by the conversion of cDNA into double-stranded (ds) DNA using exo-Klenow (Invitrogen), and random amplification of the resulting dsDNA using Platinum PCR Supermix (Invitrogen) and FR20RV primer. PCR products were then purified using AMPure (Beckman Coulter, Indianapolis, IN, United States) and processed to the library preparation step using Nextera XT DNA kit (Illumina, San Diego, CA, United States), following the manufacturer's instructions. Finally, the prepared library was sequenced using MiSeq reagent kit v3 (600 cycles, Illumina), which produced read lengths up to 2 × 300 bp, in a MiSeq platform (Illumina) (Nguyen et al., 2016).

Sequence Assembly/Editing and Recombinant Detection

Genomic reads coming out of the MiSeq platform were trimmed to remove adaptors. Consensus sequences of CVA16 genomes were then generated using reference-mapping method. This step involved mapping of all reads from each sample to a reference genome (accession number: KX595295) and generating consensus sequences using default settings, and manual editing of the obtained consensuses using Geneious software package version 8.1.5 (Biomatters, Auckland, New Zealand). The obtained CVA16 sequences were submitted to the National Center for Biotechnology Information (GenBank accession numbers: MZ043537-MZ043565 and MW999251-MW999294).

For phylogenetic analysis, we retrieved all CVA16 genomes available from Virus pathogen database and analysis resource. This dataset also includes all CVA16 sequences available in GenBank, and 6 previously reported VP1 sequences from

[1] https://www.viprbrc.org

Vietnam (Phan et al., 2007) in 2005. Identical sequences were removed. Sequences alignment was carried out using MUSCLE method available in Geneious. We then applied RDP4 [Recombination Detection Program version 4 (Martin et al., 2015)] to detect any recombinants in the dataset. Potential recombinants were confirmed by at least 3 of 4 selected methods, including Chimera, GENECONV, Maxchi, Bootscan, and Siscan (using default settings) with p-value < 0.05. Recombinant strains were then removed from further analysis.

Phylogenetic Analysis

The relatedness between the Vietnamese and global CVA16 strains was investigated by reconstructing maximum-likelihood phylogenetic trees for viral capsid protein 1 (VP1) and complete coding sequences (CDS) utilizing IQ-TREE software version 1.4.3 (L. T. Nguyen et al., 2015). For these analyses, we used general time reversible and Tamura-Nei 93 nucleotide substitution models, for CDS and VP1 dataset, respectively, with gamma distribution among sites (four rate categories), as suggested by the IQ-TREE software.

For phylogeographic and timescale phylogenetic investigations of the Vietnamese CVA16 strains, any sequences with insufficient temporal signal, indicated by TempEst v1.5 (Rambaut et al., 2016), were excluded from analysis. Time-scale phylogenetic analysis was performed for both the 68 CDS and 90 VP1 sequence datasets obtained from the present study using BEAST package v1.8.3 (Drummond and Rambaut, 2007). For phylogeographic analysis, since the VP1 dataset more widely represented for the geographic locations of the study participants, we first focused our analysis on the VP1 sequences. Accordingly, all Vietnamese sequences were divided into 4 regions, including HCMC, Mekong Delta (Long An, Ben Tre, Dong Thap, Kien Giang, An Giang, Soc Trang, and Ca Mau), South East area (Dong Nai, Tay Ninh, Binh Duong, Lam Dong, and Binh Thuan), and Central Vietnam (Binh Dinh). Running the analysis with individual provinces would lead to inconclusive results because of the small number of samples. Similar analysis was also carried out for the CDS dataset.

We used Tamura-Nei 93 (Tamura and Nei, 1993) nucleotide substitution model for VP1 dataset (general time reversible for CDS dataset (Lanave et al., 1984)), with gamma distribution and 4 rate categories, strict molecular clock model and Bayesian skyline plot (10 groups). We employed a 1,300 million step Bayesian Markov chain Monte Carlo implemented in BEAST package, sampling every 1,30,000 steps for CDS dataset (and 100 million sampling every 10,000 steps for VP1 dataset). Convergence was evaluated by Tracer v1.5.1 (Rambaut et al., 2018) with 10% burn-in and effective sample size above 150. Maximum-clade credibility (MCC) trees were summarized with TreeAnnotator (BEAST package) and visualized in Figtree version 1.4.2 (Drummond and Rambaut, 2007).

Standard Biosecurity and Institutional Safety Procedures

All laboratory procedures described in the present study were conducted in a laboratory compliant with Good Clinical Laboratory Practice and biosafety level-2 laboratory certified by Vietnamese Ministry of Health. Accordingly, clinical samples are stored in biosecurity complied BSL2 repository. And testing of clinical samples was conducted in ISO 15189 certified laboratory. A protocol biosafety risk assessment was conducted and a risk management strategy for each laboratory procedure was developed and implemented. Staff working in the laboratory attended annual refreshment training on institutional biosafety and biosecurity.

RESULTS

Demographics

During 2011–2017, there were 2,813 HFMD patients enrolled in the study. Of these, 2,397 were positive for EV by real-time RT-PCR. CVA16 was detected in 322 patients, accounting for 13.43% of EV positive cases. Other common EVs detected included EV-A71, CVA6, and CVA10; accounting for 26.86, 24.53, and 10.18% of the PCR confirmed cases, respectively. During the study period, CVA16 was not detected in 2011, likely attributable the sampling bias of the study during this year. Then the proportion of CVA16 increased between 2014 and 2015 followed by a drop in 2016, and then increased again in 2017 (**Figure 1**).

320/322 CVA16 patients with available demographic and clinical information were included for analysis in the present study. Most (59.4%) of these patients were male, in agreement with previous reports (Le et al., 2020; Min et al., 2021). The median age was 20.82 months. 55.3% of the 320 patients resided in HCMC, the rest in Mekong Delta (22.2%) and South East Vietnam (22.5%). 32/320 (10%) of the CVA16 infected

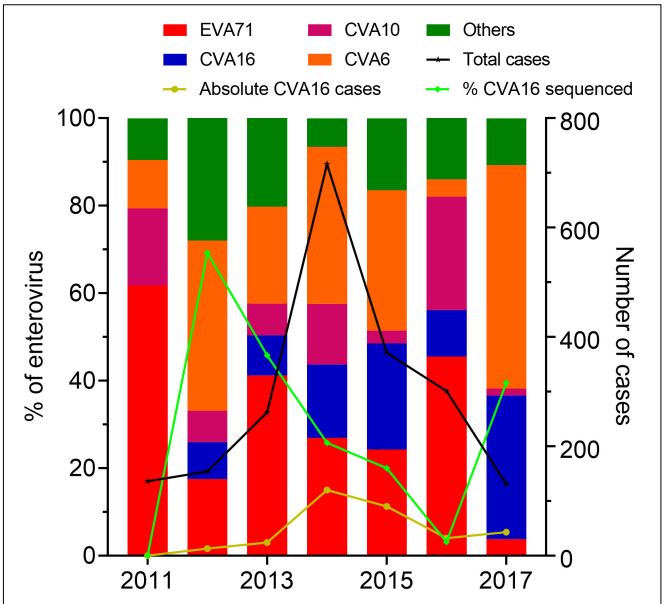


FIGURE 1 | Enterovirus prevalence among HFMD cases in Vietnam during 2011–2017. The left *Y* axis shows the proportions of enterovirus serotypes detected among cases enrolled in the present study and the proportions of CVA16 cases included for whole genome sequencing. The right *Y* axis shows number of HFMD cases enrolled in the study between 2011 and 2017.

patients had moderately severe or severe (grade 2b1 and above) HFMD (**Table 1**).

Phylogenetics

Results of Whole Genome Sequencing and Recombination Assessment

Of the 320 CVA16 patients included for analysis, 153 had a throat or rectal swab with sufficient viral load (Cp $\leq$ 30) for whole genome sequencing. Subsequently, 96 VP1 sequences were successfully recovered, including 81 from patients with mild (grade 1 or 2a) and 15 from those with severe diseases (grade 2b1, 2b2, or 3) (**Table 1**). Of these, 70 also had CDS successfully recovered. One CDS were indicated as recombinants (**Supplementary Figure 1**), and seven (6 VP1 and 1 CDS) sequences were identical sequences. We removed the identical sequences and recombinant sequences from subsequent analysis, and thus included 90 non-identical VP1 sequences and 68 CDS for detailed phylogenetic investigations.

Phylogenetic Analysis

Phylogenetic analyses of 1,066 global CVA16 VP1 sequences, including 90 Vietnamese sequences of the present study and 6 previously reported sequences from Vietnam in 2005, showed that CVA16 was divided into 2 genogroups, A and B, in accordance to previous studies (Perera et al., 2007; Noisumdaeng et al., 2018). Genogroup B subsequently split into B1 and B2 (subgroups a, b and c). All Vietnamese CVA16 sequences from this study belonged to group B1a, clustering together with strains collected in Vietnam in 2005, and from China, Thailand, Malaysia, Japan, France, and Germany (**Figure 2**).

Phylogeographics and Phylodynamics of CVA16 in Southern Vietnam

To investigate the dispersal of CVA16 in southern Vietnam, we performed detailed phylogeographic analysis for HCMC and other regions, including Mekong Delta, South East and Central Vietnam. The results revealed that CVA16 is dispersed widely across the southern regions of Vietnam during the study (**Figure 3**). Additionally, VP1-based time-scale phylogenetic analysis revealed that the time to most recent common ancestor of CVA16 sequences generated as part of the present study was estimated to be around the third quarter of 2008 (95% CI, April 2007–October 2009), 4 years before their first detection in HFMD cases enrolled in the clinical study from 2012 onward.

Bayesian skyline plot assessment revealed that the genetic diversity of CVA16 strains circulating in southern Vietnam remained stable during the study period (2011–2017), with some slight fluctuations corresponding to the increase in CVA16 cases reported in 2014–2015 and 2017 (**Figure 4**). The estimated evolution rate of CVA16 circulating in southern Vietnam was about 5.54×10^{-3} substitutions per site per year (95% CI, 4.57×10^{-3}–6.59×10^{-3}). Similar results of the time to most recent common ancestor, evolution rate and genetic diversity were obtained when the time-scale phylogenetic analysis was carried out for CDS dataset (**Supplementary Figures 2, 3**).

DISCUSSION

We report for the first time the detailed molecular epidemiology of CVA16 infection in Vietnam between 2011 and 2017. We showed that together with EV-A17, CVA10, and CVA6, CVA16 is one of the most predominant HFMD pathogens associated with severe clinical phenotypes in Vietnam during the study period (2011–2017). Of these, EV-A71 was the leading cause (Van Hoang et al., 2019a; Le et al., 2020). Additionally, despite its relatively stable genetic diversity, CVA16 dispersed widely between HCMC and other neighboring provinces. This reflected the endemicity of the pathogen and the burden posed by HFMD in southern Vietnam as a whole (Le et al., 2020, 2018).

A total of 32 CVA16 patients had severe HFMD (grade 2b1 or above), accounting for 10% of 320 CVA16 infected patients of the present study. While the data emphasize that CVA16 infection causes severe clinical consequences, this high proportion of

TABLE 1 | Demographic information and clinical grades of CVA16 cases in Vietnam 2011–2017.

		Total (*n* = 320)	Sequenced group (*n* = 96)	Not sequenced group (*n* = 224)	*p*-value*
Gender	Male	190 (59.4%)	54 (56.3%)	136 (60.7%)	0.46
	Female	130 (40.6%)	42 (43.8%)	88 (39.3%)	
Age (months)	Median	20.82	20.93	20.43	0.86
	IQR	14.96–31.41	14.23–33.71	15.23–29.4	
Provinces	HCMC	177 (55.3%)	47 (49.0%)	130 (58.0%)	0.42
	Mekong delta	71 (22.2%)	24 (25.0%)	47 (21.0%)	
	South East	71 (22.2%)	25 (26.0%)	46 (20.5%)	
	Central	1 (0.3%)	0 (0.0%)	1 (0.3%)	
Highest grade	1	152 (47.5%)	45 (46.9%)	107 (47.8%)	0.08
	2a	136 (42.5%)	36 (37.5%)	100 (44.6%)	
	2b1	20 (6.3%)	8 (8.3%)	12 (5.4%)	
	2b2	2 (0.6%)	2 (2.1%)	0 (0.0%)	
	3	10 (3.1%)	5 (5.2%)	5 (2.2%)	

Data on outcome could not be retrieved for severe patients of stage 1, 2011–2012 periods, so outcome comparison was not performed since it was likely to be biased.
**Comparisons between sequenced and not sequenced groups. Fisher or Wilcoxon rank-sum test was used to compared between groups when appropriate.*

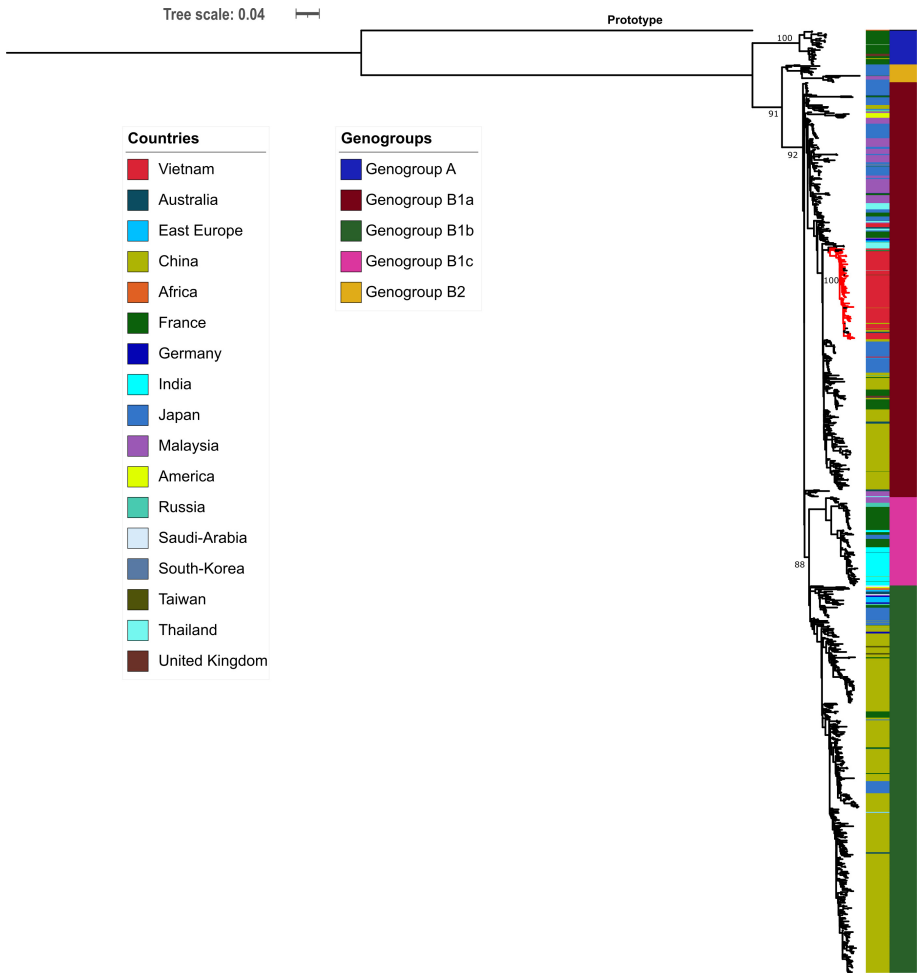


FIGURE 2 | Maximum likelihood tree of CVA16 of Vietnam and global strains based on VP1 sequence. Vietnamese isolates (red-colored branch) belong to genogroup B1a containing strains from China, Thailand, Malaysia, Japan, France, and Germany.

clinical complications associated with CVA16 infection should be interpreted with caution. During the first stage of our study (2011–2012), we only focused our sampling on severe patients admitted to the PICU of the Hospital for Tropical Diseases. This group accounted for 12/32 (38%) of the severe cases of the present study. As such, our finding may have been biased by the over representation of the PICU admitted patients during the first phase of the study. Additionally, we were not able to include individuals with mild or asymptomatic infection who did not present to our study sites. As a consequence, our sampling biases hampered informative analysis aiming at identifying the association between viral strains and/or genetic variations and clinical severities. However, sub-analysis did not reveal such association in our dataset, supporting previous report (Singh et al., 2002; Shih et al., 2000).

Between 2011 and 2017, EV-A71 circulation was observed every 2 or 3 years (Van Hoang et al., 2019a; Le et al., 2020). Likewise, CVA6 and CVA16 replaced each other to become the dominant serotypes between the interval peaks of EV-A71. HFMD is a disease of young children, especially those without pre-existing immunity. Thus, the observed cyclic patterns of the common causes is likely attributable to the requirement of a sufficient number of susceptible newborn babies to sustain the ongoing transmission in the community.

EV-A71 subgenogroup replacements have frequently occurred in endemic countries (Le et al., 2018; Van Hoang et al., 2019a). In contrast, the CVA16 subgenogroup B1a was the only subgenogroup circulating in southern Vietnam during 2011–2017, and was genetically related to the B1a strains detected in HCMC in 2005, in Asia (China, Thailand, Malaysia and Japan) and in Europe (France and Germany). Together, the data pointed to the wide dispersal of B1a within Vietnam and globally. The genetic characteristics of CVA16 subgenogroup circulating in HCMC and Vietnam between 2006 and 2010 remain unknown.

Our estimated evolution rate for the VP1 dataset of the Vietnamese CVA16 sequences was 5.43×10^{-3} substitutions per site per year, supporting a recent report from China (Han et al., 2020), and our reported data (5.12×10^{-3} substitutions per site per year) for EV-A71 in Vietnam (Van Hoang et al., 2019a). In contrast, CVA6 showed slightly higher evolution rate

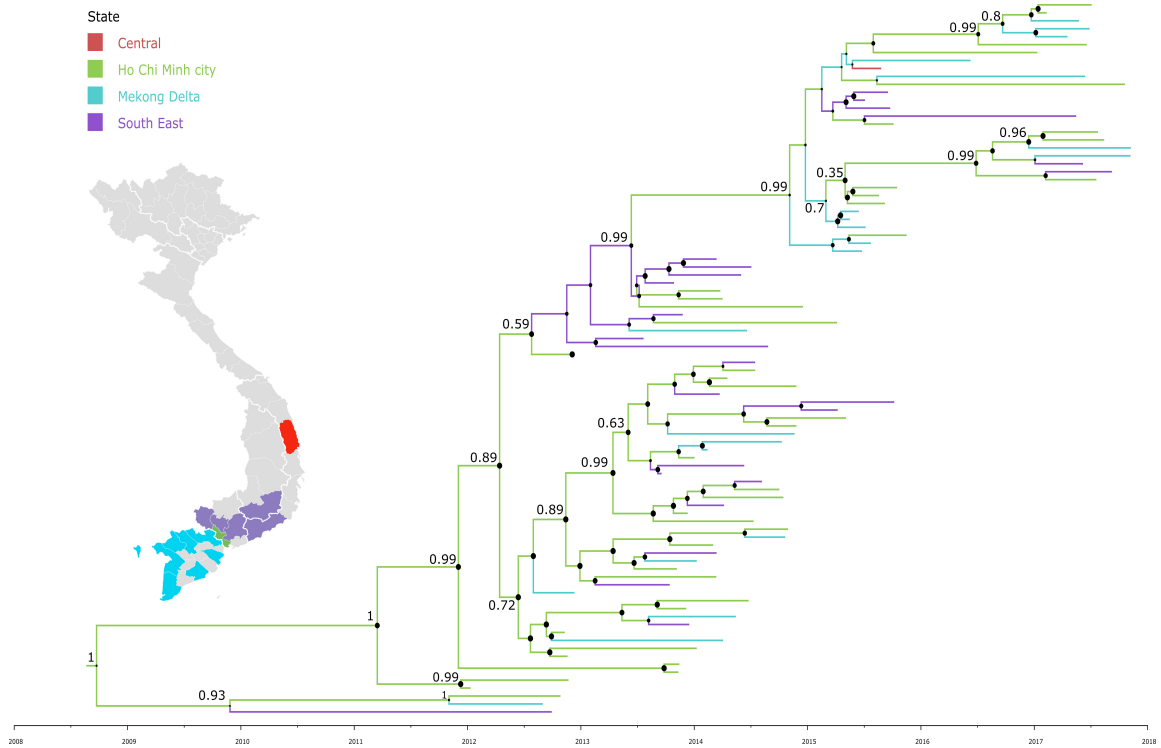


FIGURE 3 | Maximum clade credibility trees demonstrating the phylogeography of CVA16 isolates in Vietnam. The tree was constructed using VP1 sequences of Vietnamese sequences and branches are colored by regions.

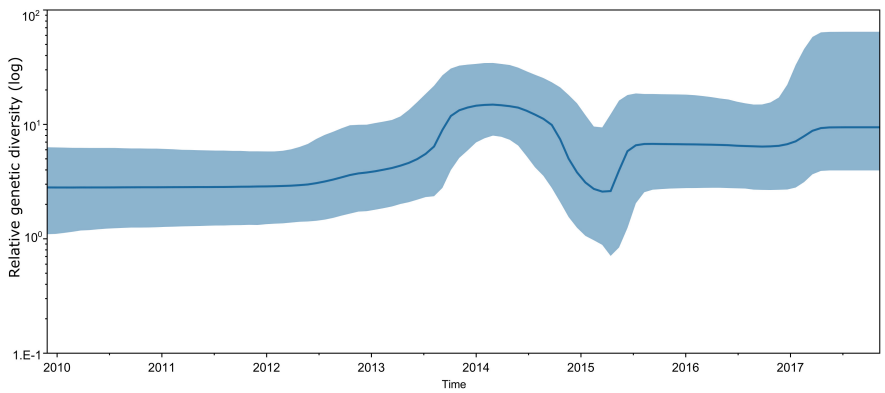


FIGURE 4 | Estimated relative genetic diversity of CVA16 in Vietnam by Bayesian skyline plot using VP1 sequence dataset.

(T. A. Nguyen et al., 2018), 7.42×10^{-3} substitutions per site per year. This might be partly explained by the more recent emergence of CVA6, compared to EV-A71 and CVA16, which warrants further investigations.

Our Bayesian-based analysis showed the most recent common ancestor (tMRCA) of CVA16 in Vietnam was dated to around August–September 2008. This suggested that CVA16 strains sequenced from patients enrolled in the present study had been circulating in Vietnam for 4 years before it was detected in HFMD cases enrolled in the present study from 2012 onward, supporting previous report regarding cryptic circulation of HFMD pathogens prior to causing community transmission (Tee et al., 2010). Recombination is a common phenomenon occurring as part of the evolutionary process of EVs (Simmonds and Welch, 2006). Likewise, we detected one recombinant CVA16 in our dataset. These collective findings in turn emphasize the importance of active surveillance for circulating HFMD strains in endemic countries, critical to informing public health authorities.

Our study had some limitations. As outlined above, from 2011 to 2012 our focus was patients admitted PICU of the Hospital for Tropical diseases. And we were not able to include individuals with mild or asymptomatic infection who did not present to our

study sites. Therefore, we may have underestimated the diversity of CVA16 and EV serotypes detected in HFMD patients during this period, and have overestimated the frequency of severe manifestations from CVA16 infections. Additionally, the three hospitals in HCMC where our study was based are responsible for receiving HFMD from southern Vietnam. Therefore, while our findings may be generalized for southern Vietnam, the epidemiology of CVA16 in central and the north of Vietnam warrants further research.

In summary, our study demonstrates that CVA16 is an endemic HFMD pathogen with a single subgenogroup B1a circulating in southern Vietnam during 2011–2017. Despite its moderate evolution rate and stable genetic diversity, CVA16 was dispersed widely in southern Vietnam during the study period. Viral circulation in Vietnam happened some time in 2008, 4 years prior to its detection in HFMD cases enrolled in the present study from 2012 onward. These collective findings emphasize the importance of active surveillance for viral circulation in HFMD endemic countries, critical to informing outbreak response.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in GenBank repository, accession numbers MZ043537 – MZ043565 and MW999227 – MW999294.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board of Children's Hospital 1, Children's Hospital 2 and Hospital for Tropical Diseases, and the Oxford Tropical Research Ethics Committee approved the study.

AUTHOR CONTRIBUTIONS

LNha, NH, HT, NC, GT, HD, and LT designed the study. HV, DH, DQ, PQ, TK, and LNha recruited the patients. LNhu, NA, NH, TT, VH, NN, and LN performed laboratory experiments. LNhu, NA, NH, and TT analyzed the data. LNhu drafted the manuscript. LT reviewed and edited the manuscript. All authors read the final manuscript and agreed with its contents.

FUNDING

This study was funded by the Wellcome Trust of Great Britain (106680/B/14/Z and 204904/Z/16/Z).

ACKNOWLEDGMENTS

We are indebted to Le Kim Thanh for her logistic support. We thank the patients for their participations in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.689658/full#supplementary-material

REFERENCES

Brown, D. M., Zhang, Y., and Scheuermann, R. H. (2020). Epidemiology and sequence-based evolutionary analysis of circulating non-polio enteroviruses. *Microorganisms* 8, 1–23. doi: 10.3390/microorganisms8121856

Cai, K., Wang, Y., Guo, Z., Yu, H., Li, H., Zhang, L., et al. (2019). Clinical characteristics and managements of severe hand, foot and mouth disease caused by enterovirus A71 and coxsackievirus A16 in Shanghai. *China. BMC Infect. Dis.* 19:285. doi: 10.1186/s12879-019-3878-6

Chen, L., Yao, X. J., Xu, S. J., Yang, H., Wu, C. L., Lu, J., et al. (2019). Molecular surveillance of coxsackievirus A16 reveals the emergence of a new clade in mainland China. *Arch. Virol.* 164, 867–874. doi: 10.1007/s00705-018-4112-3

Chen, Q., Zhang, Q., and Hu, Z. (2019). Profiles of human enteroviruses associated with hand, foot, and mouth disease in nanjing, China. *Disaster Med. Public Health Prep.* 13, 740–744. doi: 10.1017/dmp.2018.155

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Esposito, S., and Principi, N. (2018). Hand, foot and mouth disease: Current knowledge on clinical manifestations, epidemiology, aetiology and prevention. *Eur. J. Clin. Microbiol. Infect. Dis.* 37, 391–398. doi: 10.1007/s10096-018-3206-x

Geoghegan, J. L., Tan, L., Van, Kühnert, D., Halpin, R. A., Lin, X., Simenauer, A., et al. (2015). Phylodynamics of enterovirus A71-associated hand, foot, and mouth disease in viet nam. *J. Virol.* 89, 8871–8879. doi: 10.1128/jvi.00706-15

Han, Z., Song, Y., Xiao, J., Jiang, L., Huang, W., Wei, H., et al. (2020). Genomic epidemiology of coxsackievirus A16 in mainland of China, 2000–18. *Virus Evol.* 6, 1–16. doi: 10.1093/ve/veaa084

Lanave, C., Preparata, G., Saccone, C., and Serio, G. (1984). A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20, 86–93. doi: 10.1007/bf02101990

Le, N. T. N., Nguyen, T. T. H., Le, N. T. N., Lam, A. N., Nguyen, T. H. N., Tran, T. T., et al. (2018). Severe enterovirus A71 associated hand, foot and mouth disease. Vietnam, 2018: preliminary report of an impending outbreak. *Eurosurveillance* 23:1800590. doi: 10.2807/1560-7917.ES.2018.23.46.1800590

Le, N. T. N., Truong, H. K., Nguyen, T. T. H., Van Hoang, M. T., Le, N. T. N., Nguyen, T. H. N., et al. (2020). Clinical, etiological and epidemiological investigations of hand, foot and mouth disease in Southern Vietnam during 2015 – 2018. *PLoS Negl. Trop. Dis.* 14:e0008544. doi: 10.1371/journal.pntd.0008544

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003

Min, N., Ong, Y. H. B., Han, A. X., Ho, S. X., Yen, E. W. P., Ban, K. H. K., et al. (2021). An epidemiological surveillance of hand foot and mouth disease in paediatric patients and in community: a singapore retrospective cohort study, 2013–2018. *PLoS Negl. Trop. Dis.* 15:e0008885. doi: 10.1371/journal.pntd.0008885

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nguyen, T. A., Le, N. T. N., Hoang, M. T., Van, Nguyen, T. T. H., Tran, T. T., et al. (2018). Emerging coxsackievirus A6 causing hand. foot and mouth disease, vietnam. *Emerg. Infect. Dis.* 24, 17–19.

Nguyen, T. A., Tran, T. T., Hoang, M. T., Van, Nghiem, M. N., Le, N. T. N., et al. (2016). Development and evaluation of a non-ribosomal random PCR and next-generation sequencing based assay for detection and sequencing of hand, foot and mouth disease pathogens. *Virol. J.* 13:125. doi: 10.1186/s12985-016-0580-9

Noisumdaeng, P., Korkusol, A., Prasertsopon, J., Sangsiriwut, K., Chokephaibulkit, K., Mungaomklang, A., et al. (2019). Longitudinal study on enterovirus A71 and coxsackievirus A16 genotype/subgenotype replacements in hand, foot and mouth disease patients in Thailand, 2000–2017. *Int. J. Infect. Dis.* 80, 84–91. doi: 10.1016/j.ijid.2018.12.020

Noisumdaeng, P., Sangsiriwut, K., Prasertsopon, J., Klinmalai, C., Payungporn, S., Mungaomklang, A., et al. (2018). Complete genome analysis demonstrates multiple introductions of enterovirus 71 and coxsackievirus A16 recombinant strains into Thailand during the past decade. *Emerg. Microbes Infect.* 7:1. doi: 10.1038/s41426-018-0215-x

Perera, D., Yusof, M. A., Podin, Y., Ooi, M. H., Thao, N. T. T., Wong, K. K., et al. (2007). Molecular phylogeny of modern coxsackievirus A16. *Arch. Virol.* 152, 1201–1208. doi: 10.1007/s00705-006-0934-5

Phan, V. T., Nguyen, T. T. T., Perera, D., Truong, Th. K, Nguyen, T. K. T., Tang, C. T., et al. (2007). Epidemiologic and virologic investigation of hand, foot, and mouth disease, southern Vietnam, 2005. *Emerg. Infect. Dis.* 13, 1733–1741.

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.* 2:1. doi: 10.1093/ve/vew007

Shih, S. R., Ho, M. S., Lin, K. H., Wu, S. L., Chen, Y. T., Wu, C. N., et al. (2000). Genetic analysis of enterovirus 71 isolated from fatal and non-fatal cases of hand, foot and mouth disease during an epidemic in Taiwan, 1998. *Virus Res.* 68, 127–136. doi: 10.1016/S0168-1702(00)00162-3

Simmonds, P., and Welch, J. (2006). Frequency and dynamics of recombination within different species of human enteroviruses. *J. Virol.* 80, 483–493. doi: 10.1128/jvi.80.1.483-493.2006

Singh, S., Poh, C. L., and Chow, V. T. K. (2002). Complete sequence analyses of enterovirus 71 strains from fatal and non-fatal cases of the hand, foot and mouth disease outbreak in Singapore (2000). *Microbiol. Immunol.* 46, 801–808. doi: 10.1111/j.1348-0421.2002.tb02767.x

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526. doi: 10.1093/oxfordjournals.molbev.a040023

Tee, K. K., Lam, T. T.-Y., Chan, Y. F., Bible, J. M., Kamarulzaman, A., Tong, C. Y. W., et al. (2010). Evolutionary genetics of human enterovirus 71: origin, population dynamics, natural selection, and seasonal periodicity of the VP1 gene. *J. Virol.* 84, 3339–3350. doi: 10.1128/JVI.01019-09

Van Hoang, M. T., Nguyen, T. A., Nguyen, T. T. H., Le, N. T. N., Lam, A. N., Tran, T. T., et al. (2019a). Enterovirus A71 phenotypes causing hand, foot and mouth disease. Vietnam. *Emerg. Infect. Dis.* 25, 788–791. doi: 10.3201/eid2504.181367

Van Hoang, M. T., Nguyen, T. A., Tran, T. T., Vu, T. T. H., Le, N. T. N., Nguyen, T. H. N., et al. (2019b). Clinical and aetiological study of hand, foot and mouth disease in southern Vietnam, 2013–2015: inpatients and outpatients. *Int. J. Infect. Dis.* 80, 1–9. doi: 10.1016/j.ijid.2018.12.004

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 29 June 2021
doi: 10.3389/fmicb.2021.676314

Meta-Analysis and Structural Dynamics of the Emergence of Genetic Variants of SARS-CoV-2

*Nicolas Castonguay[1], Wandong Zhang[2,3] and Marc-André Langlois[1,4]**

[1] Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada, [2] Department of Cellular and Molecular Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada, [3] Human Health Therapeutics Research Centre, National Research Council Canada, Ottawa, ON, Canada, [4] uOttawa Center for Infection, Immunity and Inflammation (CI3), Ottawa, ON, Canada

The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in late December 2019 in Wuhan, China, and is the causative agent for the worldwide COVID-19 pandemic. SARS-CoV-2 is a positive-sense single-stranded RNA virus belonging to the betacoronavirus genus. Due to the error-prone nature of the viral RNA-dependent polymerase complex, coronaviruses are known to acquire new mutations at each cycle of genome replication. This constitutes one of the main factors driving the evolution of its relatively large genome and the emergence of new genetic variants. In the past few months, the identification of new B.1.1.7 (United Kingdom), B.1.351 (South Africa), and P.1 (Brazil) variants of concern (VOC) has highlighted the importance of tracking the emergence of mutations in the SARS-CoV-2 genome that impact transmissibility, virulence, and immune and neutralizing antibody escape. Here we analyzed the appearance and prevalence trajectory over time of mutations that appeared in all SARS-CoV-2 genes from December 2019 to April 2021. The goal of the study was to identify which genetic modifications are the most frequent and study the dynamics of their propagation, their incorporation into the consensus sequence, and their impact on virus biology. We also analyzed the structural properties of the spike glycoprotein of the B.1.1.7, B.1.351, and P.1 variants for its binding to the host receptor ACE2. This study offers an integrative view of the emergence, disappearance, and consensus sequence integration of successful mutations that constitute new SARS-CoV-2 variants and their impact on neutralizing antibody therapeutics and vaccines.

Keywords: SARS-CoV-2, COVID-19, variants, evolution, Coronavirus, B1.351 variant, B.1.1.7 variant, P.1 variant

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Siddappa N. Byrareddy,
University of Nebraska Omaha, United States
Svetlana Khaiboullina,
University of Nevada, Reno, United States

***Correspondence:**
Marc-André Langlois
langlois@uottawa.ca

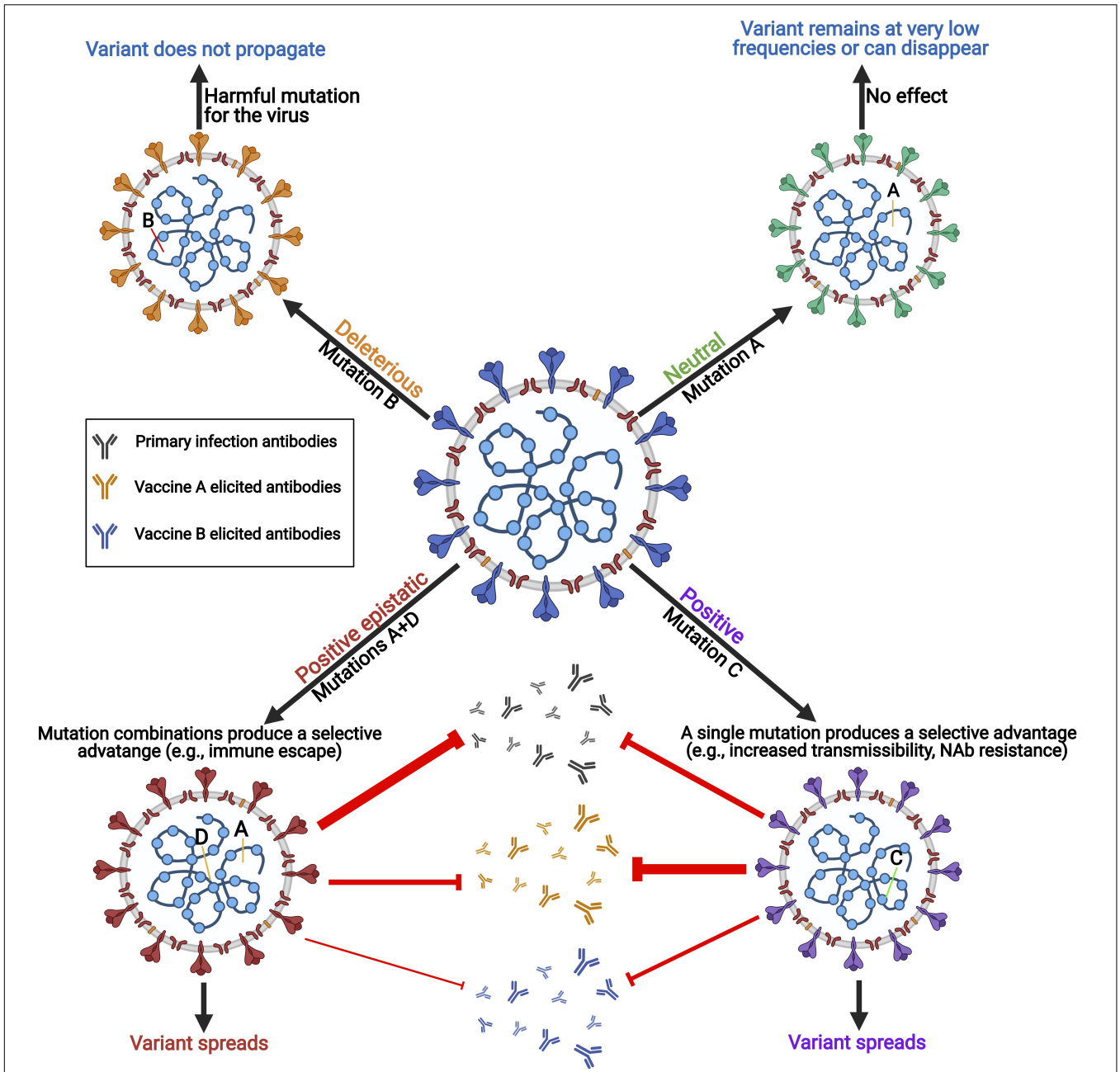
Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *05 March 2021*
Accepted: *27 May 2021*
Published: *29 June 2021*

Citation:
Castonguay N, Zhang W and Langlois M-A (2021) Meta-Analysis and Structural Dynamics of the Emergence of Genetic Variants of SARS-CoV-2. Front. Microbiol. 12:676314. doi: 10.3389/fmicb.2021.676314

HIGHLIGHTS

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological agent of COVID-19, which has caused > 3.4 million deaths worldwide as of April 2021. Mutations occur in the genome of SARS-CoV-2 during viral replication and affect viral infectivity, transmissibility, and virulence. In early March 2020, the D614G mutation in the spike protein emerged, which increased viral transmissibility and is now found in over 90% of all SARS-CoV-2 genomic sequences in GISAID database. Between October and December 2020, B.1.1.7 (United Kingdom), B.1.351 (South Africa), and P.1 (Brazil) variants of concern (VOCs) emerged, which have increased



GRAPHICAL ABSTRACT | Errors are regularly made when SARS-CoV-2 replicates its RNA genome. The viral polymerase complex is error-prone with imperfect proofreading abilities. These errors or mutations often lead to deleterious or neutral effects on the virus. However, sometimes these mutations have a positive effect and create genetic variants of the virus with different features including increased transmissibility, pathogenicity, and immune escape capabilities. When mutations work collaboratively to create a new virus feature, this is called epistasis.

neutralizing antibody escape capabilities because of mutations in the receptor-binding domain of the spike protein. Characterizing mutations in these variants is crucial because of their effect on adaptive immune responses, neutralizing antibody therapy, and their impact on vaccine efficacy. Here we tracked and analyzed mutations in SARS-CoV-2 genes since the beginning of the pandemic and investigated their functional impact on the spike of these three VOCs.

INTRODUCTION

In late December 2019, a new betacoronavirus known as severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) emerged in the city of Wuhan in the province of Hubei, China (Zhu et al., 2020). SARS-CoV-2 is the etiological viral agent for the worldwide COVID-19 pandemic resulting in more than 162 million infected and 3.4 million deaths worldwide as

of April 2021 (Dong et al., 2020; Wu F. et al., 2020). SARS-CoV-2 is an enveloped, positive-sense single-stranded RNA (+ ssRNA) virus with a genome length of 29,811 nucleotides (Kim et al., 2020). The mutation rates of RNA viruses are generally higher than that of DNA viruses because of the low fidelity of their viral RNA polymerases (Drake, 1993; Duffy, 2018). Mutations occur when viral replication enzymes introduce errors in the viral genome resulting in the creation of premature termination codons, deletions, and insertions of nucleotides that can alter open reading frames and result in amino acid substitutions in viral proteins. These mutations, combined with the selective pressure of the human immune system, lead to the selection and evolution of viral genomes (Drake, 1993; Peck and Lauring, 2018). However, coronaviruses are one of the few members of the RNA virus family that possess limited, but measurable, proofreading ability via the 3′- to 5′-exoribonuclease activity of the non-structural viral protein 14 (nsp14) (Minskaia et al., 2006; Eckerle et al., 2010). Coronaviruses are therefore expected to evolve through genetic drift much slower than other RNA viruses that do not have this ability, such as influenza viruses (Eckerle et al., 2010; Ma et al., 2015). Additionally, SARS-CoV-2 and other coronaviruses have low known occurrences of recombination between family members (i.e., genetic shift), and therefore are mostly susceptible to genetic drift (Rausch et al., 2020).

SARS-CoV-2 has reached pandemic status due to its presence on every continent and has since maintained a high level of transmissibility across hosts of various ethnical and genetic backgrounds (Chu et al., 2020; Dong et al., 2020). Moreover, SARS-CoV-2 infections have been reported to naturally infect minks, ferrets, cats, tiger, and dogs, which allows the virus to replicate in completely new hosts and mutate to produce new variants and possibly new strains (Li X. et al., 2020; Shi et al., 2020). In March 2020, the now dominant D614G mutation first emerged in the spike protein (S) of SARS-CoV-2. The S protein is present as a trimer at the surface of the viral envelope and is responsible for attachment of the virus to the human angiotensin-converting enzyme 2 (hACE2), the entry receptor for SARS-CoV-2 into human cells (Walls et al., 2020). Published evidence has now shown that D614G increases viral fitness, transmissibility, and viral load but does not directly affect COVID-19 pathogenicity (Korber et al., 2020; Li Q. et al., 2020; Plante et al., 2020; Zhang et al., 2020). Additionally, emerging evidence indicates that D614G may have epistatic interactions that exacerbate the impact of several other independent mutations (Li Q. et al., 2020). Mutations in the S protein, and particularly in the receptor-binding domain (RBD), are of very high concern given that they can directly influence viral infectivity, transmissibility, and resistance to neutralizing antibodies and T cell responses.

New mutations are frequently and regularly detected in the genome of SARS-CoV-2 through whole genome sequencing; however, very few of these mutations make it into the transmitted viral consensus sequence. The reference strain is generally regarded as the dominant transmitted strain at a given time. Its sequence is determined by aligning large numbers of recently sequenced genomes and establishing a consensus sequence composed of the highest-frequency nucleotide for each position in the viral genome. A genetic variant is a version of the reference strain that has acquired one or several new mutations and acts as the founder for further genetic diversification and evolution. Mutations arise regularly in the reference strain, but few are longitudinally conserved. Genetic variants are therefore the rare successful offshoots of the reference strain.

Some variants rise rapidly in frequency and then collapse and disappear, while others will rise and overtake the frequency of the reference strain and become the new reference. There are three main genetic variants that have emerged in the past few months with sustained upward frequency trajectories. The first is the United Kingdom variant, also known as B.1.1.7/501Y.V1 (B1.1.7). It was first detected in September 2020 and is now present worldwide and poised to become the new reference strain (Rambaut et al., 2020; O'Toole et al., 2021). The South African variant, also known as B.1.351/501Y.V2 (B.1.351), was first reported in October 2020 and is now increasing in prevalence in South Africa, Europe, and North America (O'Toole et al., 2021; Tegally et al., 2021). The Brazilian variant, P.1/501Y.V3 (P.1), was first detected in travelers from Brazil that landed in Japan in January 2021 (Faria et al., 2021). It has since been identified in 42% of specimens in the Amazonian city of Manaus and in the United States at the end of January 2021 (Faria et al., 2021). These three variants are associated with increased resistance to neutralizing antibodies (Nabs), and all possess the N501Y mutation, which is a mutation in the RBD that is critical for the spike protein to interact with hACE2 (Yi et al., 2020; Greaney et al., 2021; Wang et al., 2021). This mutation is reported to cause increased resistance to Nabs, increased transmissibility, and increased virulence in animal models (Gu et al., 2020). In addition to the N501Y mutation, both the South African and Brazil variants possess RBD mutations K417N(T) and E484K, which are also associated with further increased Nabs escape capabilities (Greaney et al., 2021; Wang et al., 2021).

Here we present a retrospective metadata analysis of mutations throughout the SARS-CoV-2 genome that reached at least a 1% worldwide frequency between December 2019 and January 2021. We specifically investigated their frequency trajectory over time and their fixation into the reference sequencing using the Global Initiative on Sharing Avian Influenza Data (GISAID) (Shu and McCauley, 2017). Additionally, we analyzed mutations in the S protein of the B.1.1.7, B.1.351, and P.1 variants and illustrated their impact on molecular interactions between the S protein and hACE2 and their potential impact on Nabs.

MATERIALS AND METHODS

Data Collection and Mutational Analysis

Genomes uploaded to the GISAID EpiCoV^{TM} server database were analyzed from December 1, 2019, to December 31, 2020, and viral sequences with submission dates from December 1, 2019 to January 6, 2021 were selected. We first selected recurring mutations that were present in more than 500 reported

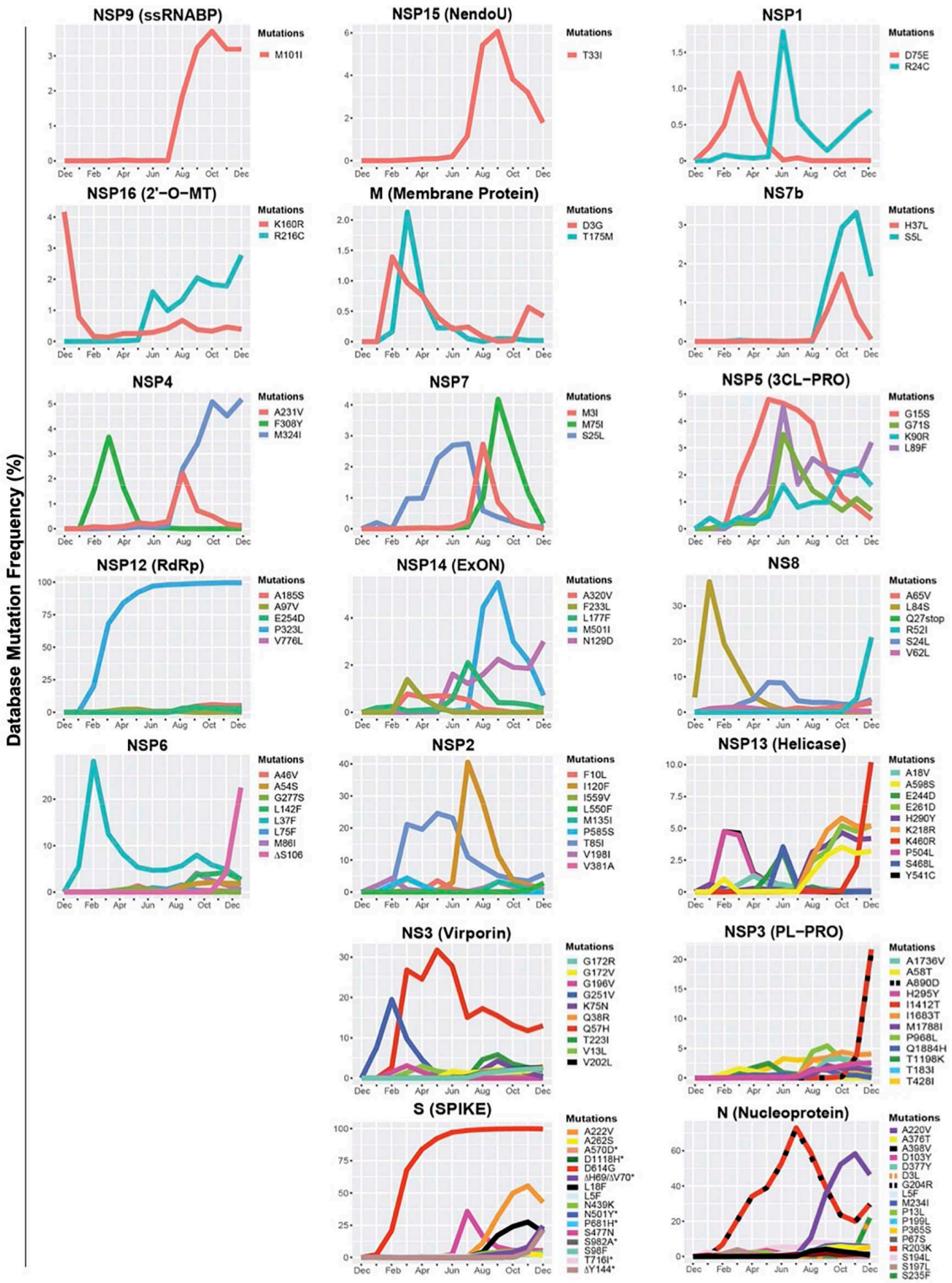


FIGURE 1 | Variations in mutations and mutation frequencies in severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) genes. The occurrence and frequency of mutations in various SARS-CoV-2 genes are presented for the period between December 2019 and January 1, 2021. Plotted are mutations that reached at least a 1% worldwide frequency. SARS-CoV-2 genes are represented with the function of the genes in parentheses. Graphs were generated using RStudio. *Overlapping curves.

TABLE 1 | Non-synonymous mutations in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) genes with a worldwide frequency ≥ 0.01%.

Protein (gene)	Genome nucleotide mutation	Codon	AA mutation	Frequency (%)*	Effect
S	A23403G	G**A**T to G**G**T	D614G	99.70	Moderate increase in Transmissibility Covid-19 Genomic UK Consortium, 2021
	C22227T	G**C**T to G**T**T	A222V	42.79	No effect Covid-19 Genomic UK Consortium, 2021
	C21615T	**C**TT to **T**TT	L18F	19.86	N/A
	G22992A	A**G**C to A**A**C	S477N	5.36	Spike-hACE2 complex stability and Nab interference Zahradník et al., 2021
	C22879A	AA**C** to AA**A**	N439K	4.09	Antibody escape ability Li Q. et al., 2020; Covid-19 Genomic UK Consortium, 2021
	C21575T	**C**TT to **T**TT	L5F	1.44	N/A
	C21855T	T**C**T to T**T**T	S98F	2.74	N/A
	G22346T	**G**CT to **T**CT	A262S	1.64	N/A
	A23063T	**A**AT to **T**AT	N501Y	80.16	Increase affinity for hACE2 Covid-19 Genomic UK Consortium, 2021
	C23709T	A**C**A to A**T**A	T716I	75.37	N/A
	C23271A	G**C**T to G**A**T	A570D	76.30	N/A
	G24914C	**G**CA to **C**AC	D1118H	74.82	N/A
	C23709A	C**C**T to C**A**T	P681H	79.18	N/A
	T24506G	**T**CA to **G**CA	S982A	74.61	N/A
	21765-21770del	**N/A**	ΔH69/ΔV70	75.13	Antibody escape Covid-19 Genomic UK Consortium, 2021
	21991-21993del	**N/A**	ΔY144	76.91	Decrease infectivity, antibody escape Li Q. et al., 2020
NSP1(ORF1ab)	C335T	**C**GC to **T**GC	R24C	0.38	N/A
NSP2	G1210T	AT**G** to AT**T**	M135I	0.01	N/A
(ORF1ab)	C1059T	A**C**C to A**T**C	T85I	13.95	N/A
	T1947C	G**T**T to G**C**T	V381A	1.08	N/A
	C2453T	**C**TC to **T**TC	L550F	5.93	N/A
NSP3	C7926T	G**C**A to G**T**A	A1736V	0.12	N/A
(ORF1ab)	T7767C	A**T**C to A**C**C	I1683T	0.19	N/A
	C5622T	C**C**T to C**T**T	P968L	0.09	N/A
	G8371T	CA**G** to CA**T**	Q1884H	0.05	N/A
	C4002T	A**C**T to A**T**T	T428I	0.44	N/A
	C5388A	G**C**T to G**A**T	A890D	76.25	N/A
	G8083A	AT**G** to AT**A**	M1788I	0.68	N/A
	C3602T	**C**AC to **T**AC	H295Y	0.30	N/A
NSP4	C9246T	G**C**T to G**T**T	A231V	0.12	N/A
(ORF1ab)	G9526T	AT**G** to AT**T**	M324I	0.14	N/A
NSP5	A10323G	A**A**G to A**G**G	K90R	3.07	N/A
(ORF1ab)	G10097A	**G**GT to **A**GT	G15S	0.35	N/A
	C10319T	**C**TT to **T**TT	L89F	1.44	N/A
NSP6	C11109T	G**C**T to G**T**T	A46V	0.15	N/A
(ORF1ab)	C11396T	**C**TT to **T**TT	L142F	0.02	N/A
	G11083T	TTG to TT**T**	L37F	1.26	N/A
	C11195T	**C**TT to **T**TT	L75F	0.13	N/A
	G11230T	AT**G** to AT**T**	M86I	0.03	N/A
	G11801A	**G**GT to **A**GT	G277S	0.07	N/A
	G11132T	**G**CT to **T**CT	A54S	0.02	N/A
	11288-11296 del	N/A	ΔS106/ΔG3676/ ΔF3677	87.61	N/A
NSP7	G12067T	AT**G** to AT**T**	M75I	0.15	N/A
(ORF1ab)	C11916T	T**C**A to T**T**A	S25L	0.10	N/A
NSP9(ORF1ab)	G12988T	AT**G** to AT**T**	M101I	0.23	N/A
NSP12(ORF1ab)	C14408T	C**C**T to C**T**T	P323L	99.49	Improves processivity by interaction with NSP8 Kannan et al., 2020
	G15766T	**G**TG to **T**TG	V776L	0.14	N/A
	G13993T	**G**CT to **T**CT	A185S	0.09	N/A
	G15598A	**G**TC to **A**TC	V720I	0.17	N/A
	G14202T	GA**G** to GA**T**	E254D	0.01	N/A
	C13730T	G**C**T to G**T**T	A97V	0.35	N/A

(Continued)

TABLE 1 | Continued

Protein (gene)	Genome nucleotide mutation	Codon	AA mutation	Frequency (%)*	Effect
NSP13	C16289T	G**C**T to G**T**T	A18V	0.03	N/A
(ORF1ab)	G17019T	GA**G** to GA**T**	E261D	0.22	N/A
	C17104T	**C**AT to **T**AT	H290Y	0.23	N/A
	C17747T	C**C**T to C**T**T	P504L	0.07	Increases hydrophobicity of 2A domain Ugurel et al., 2020
	C17639T	T**C**A to T**T**A	S468L	0.02	N/A
	A17615G	A**A**G to A**G**G	K460R	15.82	N/A
	A16889G	A**A**A to A**G**A	K218R	0.08	N/A
	G18028T	**G**CA to **T**CA	A598S	0.26	N/A
NSP14	C18998T	G**C**A to G**T**A	A320V	0.01	N/A
(ORF1ab)	C18568T	**C**TC to **T**TC	L177F	0.22	N/A
	G19542T	AT**G** to AT**T**	M501I	0.21	N/A
	A18424G	**A**AT to **G**AT	N129D	1.37	N/A
NSP15(ORF1ab)	C19718T	A**C**A to A**T**A	T33I	0.06	N/A
NSP16	A21137G	A**A**G to A**G**G	K160R	4.69	N/A
(ORF1ab)	A21390G	TT**A** to TT**G**	R216C	1.30	N/A
N	C28932T	G**C**T to G**T**T	A220V	0.38	N/A
	G28580T	**G**AT to **T**AT	D103Y	0.02	N/A
	G28883C	**G**GA to **C**GA	G204R	75.15	Destabilizes N protein Wu S. et al., 2020
	C28311T	C**C**C to C**T**C	P13L	2.00	N/A
	C28869T	C**C**A to C**T**A	P199L	5.55	N/A
	G28882A	A**G**C to A**A**A	R203K	81.96	Destabilizes N protein Wu S. et al., 2020
	C28883A	AG**C** to AA**A**	R203K	81.96	Destabilizes N protein Wu S. et al., 2020
	C28854T	T**C**A to T**T**A	S194L	0.33	Destabilizes N protein Wu S. et al., 2020
	C28977T	T**C**T to T**T**T	S235F	76.34	N/A
	G28280C	**G**AT to **C**TA	D3L	75.53	N/A
	A28281T	G**A**T to C**T**A	D3L	75.53	N/A
	T28282A	GA**T** to CT**A**	D3L	75.53	N/A
	G28975C	AT**G** to AT**C**	M234I	7.58	N/A
	C28472T	**C**CT to **T**CT	P67S	1.38	N/A
	G29399A	**G**CT to **A**CT	A376T	0.09	N/A
	C29366T	**C**CA to **T**CA	P365S	0.15	N/A
	G29402T	**G**AT to **T**AT	D377Y	1.61	N/A
	C28887T	A**C**T to A**T**T	T205I	7.00	N/A
	C29466T	G**C**A to G**T**A	A398V	0.51	N/A
M	A26530G	G**A**T to G**G**T	D3G	0.01	N/A
	C27046T	A**C**G to A**T**G	T175M	0.02	N/A
NS3	G25907T	G**G**T to G**T**T	G172V	1.41	N/A
(ORF3a)	G25617T	AA**G** to AA**T**	K75N	0.02	N/A
	G25563T	CA**G** to CA**T**	Q57H	14.36	N/A
	C26060T	A**C**T to A**T**T	T223I	0.38	N/A
	G25429T	**G**TA to **T**TA	V13L	0.05	N/A
	G25996T	**G**TA to **T**TA	V202L	0.17	N/A
	A25505G	C**A**A to C**G**A	Q38R	0.04	N/A
	G25906C	**G**GT to **C**GT	G172R	5.31	N/A
NS7b	A27866T	C**A**T to C**T**A	H37L	0.01	N/A
(ORF7b)	AT27867A	CA**T** to CT**A**	H37L	0.01	N/A
	C27769T	T**C**A to T**T**A	S5L	0.07	N/A
NS8	C28087T	G**C**T to G**T**T	A65V	0.77	N/A
(ORF8)	T28144C	T**T**A to T**C**A	L84S	0.14	N/A
	C27964T	T**C**A to T**T**A	S24L	1.38	N/A
	G28077T	**G**TG to **T**TG	V62L	0.24	N/A
	C27972T	**C**AA to **T**AA	Q27stop	76.05	Inactivates NS8 Pereira, 2020
	G28048T	A**G**A to A**T**A	R52I	76.00	N/A

**Mutation frequency as of April 30, 2021.*

genomes by August 2020, and another selection was made in January 2021 to capture recurring mutations present in more than 4,000 reported genomes in GISAID. This strategy allowed us to study mutations reaching at least 1% in worldwide frequency. We filtered through 309,962 genomes for the analysis of selected mutations.

The worldwide frequency of the hCoV-19/Wuhan strain, and hCoV-19/D614G, B.1.1.7, B.1.351, and P.1 variants were analyzed from December 1, 2019 to April 30, 2021. For the analysis of the mutations in B.1.1.7, B.1.351, and P.1 variants, we used the GISAID EpiCoV^{TM} server database. Viral sequences on GISAID with submission dates between December 1, 2019 and April 30, 2021 were selected for the analysis, and we filtered through 1,090,689 genomes for the analysis of the variants on GISAID EpiCoV^{TM} server database. Only complete SARS-CoV-2 genomes (28 to 30 kbps) isolated from human hosts were analyzed. MUSCLE alignment tool on UGene and SnapGene was used to determine the nucleotide mutations and codon changes of the non-synonymous and synonymous mutations by sequence alignments to the NCBI

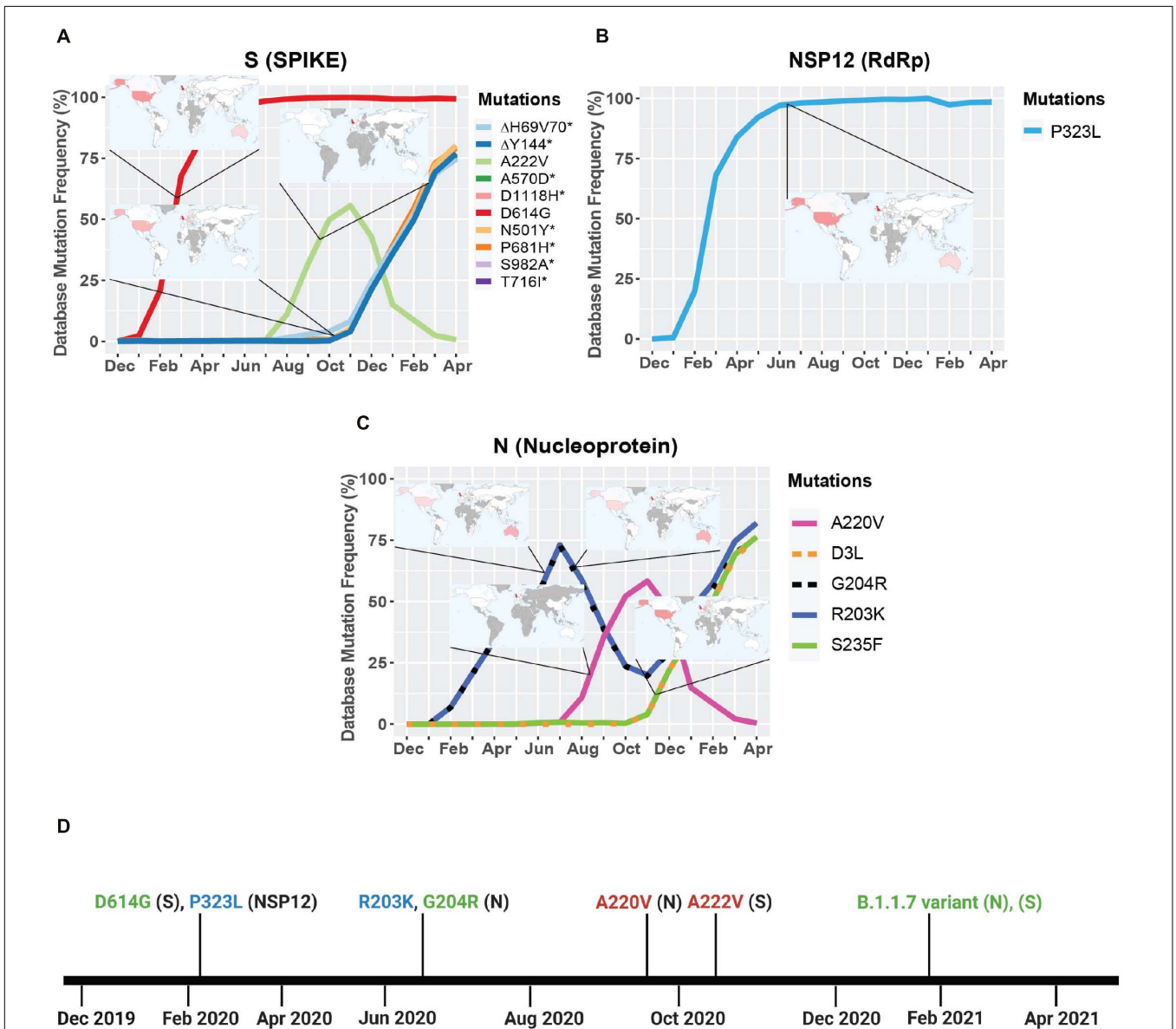


FIGURE 2 | Geographic location and timeline of dominant mutations in NSP12, S, and N genes. **(A)** Frequency of S protein mutations with corresponding geographic maps. **(B)** Frequency of RdRp mutations with corresponding geographic maps. **(C)** Frequency of nucleoprotein mutations with corresponding geographic maps. **(D)** Timeline of the appearance of mutations reaching a frequency higher than 50% worldwide between December 2019 and April 2021. For the geomaps: a low frequency of reported cases of the mutations is represented in white, while higher frequencies are represented from pink to red, and gray represents no data. All maps were taken from Global Initiative on Sharing Avian Influenza Data (GISAID). Graphs were generated using RStudio and Biorender. *Overlapping curves.

SARS-CoV-2 reference genome (NC_045512). All genomes uploaded to the GISAID database that were used in this study are presented in **Supplementary Table 1**. Graphs of mutations and variants were plotted with RStudio using the ggplot2 command. The timeline and genome illustrations were produced in Biorender.

Structural Modeling

Mutations in the spike protein in complex with hACE2 were analyzed using a mutagenesis tool for PyMOL (PDB: 7A94). Visualizations of mutations in the B.1.1.7, B.1.351, and P.1 variants were produced using the spike protein closed conformation (PDB: 6ZGE), interaction with hACE2 (PDB: 7A94), interaction with C102 Nab (PDB: 7K8M), and interaction with C121 Nab (PDB: 7K8X). Figures and rendering were prepared with PyMOL.

RESULTS

Identification of Emerging Mutations in Various Severe Acute Respiratory Syndrome Coronavirus 2 Genes

Emerging mutations in the SARS-CoV-2 genome were investigated to illustrate the fluctuations of these mutations during a period of 12 months. We tracked genome mutations with a worldwide frequency greater than 1% from December 2019 to December 31, 2020 in the GISAID database. Genes NSP8, NSP10, NS6, NS7a, and E are not illustrated in **Figure 1** given that they did not display mutations with frequencies sufficiently high to meet our inclusion criteria during the study period. This is indicative that these are some of the most conserved sequences of the SARS-CoV-2 genome. Viral proteins that have low mutation frequencies could be due to

TABLE 2 | Synonymous, non-synonymous, and deletion mutations in the B.1.1.7 (United Kingdom) variant.

Variant	Protein (gene)	Genome nucleotide mutation	S or NS	AA mutation Rambaut et al., 2020	Domain	Frequency* (%)	Effect
B.1.1.7	*NSP2 (ORF1ab)*	C913T	S			N/A	N/A
	NSP3 (ORF1ab)	C3267T	NS	T183I		76.14	N/A
		C5388A	NS	A890D		76.24	N/A
		C5986T	S			N/A	N/A
		T6954C	NS	I1412T		75.21	N/A
	NSP6 (ORF1ab)	11288-11296del	NS	ΔS106/ ΔG107/ΔF108		87.61	N/A
	NSP12 (ORF1ab)	C14676T	S			N/A	N/A
		C15279T	S			N/A	N/A
	NSP13 (ORF1ab)	C16176T	S			N/A	N/A
	S	21765-21770del	NS	ΔH69/ΔV70	NTD	75.13	Antibody escape Covid-19 Genomic UK Consortium, 2021
		21991-21993del	NS	ΔY144	NTD	76.91	Decreases infectivity, Antibody escape Li Q. et al., 2020
		A23063T	NS	N501Y	RBD	80.16	Increases affinity for hACE2 Wu et al., 2017
		C23271A	NS	A570D	SD1	76.30	
		A23403G	NS	D614G	SD2	99.29	Moderate increase in transmissibility Li Q. et al., 2020; Wu et al., 2017
		C23709A	NS	P681H	SD2	79.18	N/A
		C23709T	NS	T716I		75.37	N/A
		T24506G	NS	S982A	HR1	74.61	N/A
		G24914C	NS	D1118H		74.82	N/A
	NS8 (ORF8)	C27972T	NS	Q27stop		76.05	Inactivation of NS8 Pereira, 2020
		G28048T	NS	R52I		76.00	N/A
		A28111G	NS	Y73C		76.07	N/A
	M	T26801C	S			N/A	N/A
	N	G28280C	NS	D3L		75.53	N/A
		A28281T	NS	D3L		75.53	N/A
		T28282A	NS	D3L		75.53	N/A
		C28977T	NS	S235F		76.34	N/A

**Mutation frequency as of April 30, 2021.*

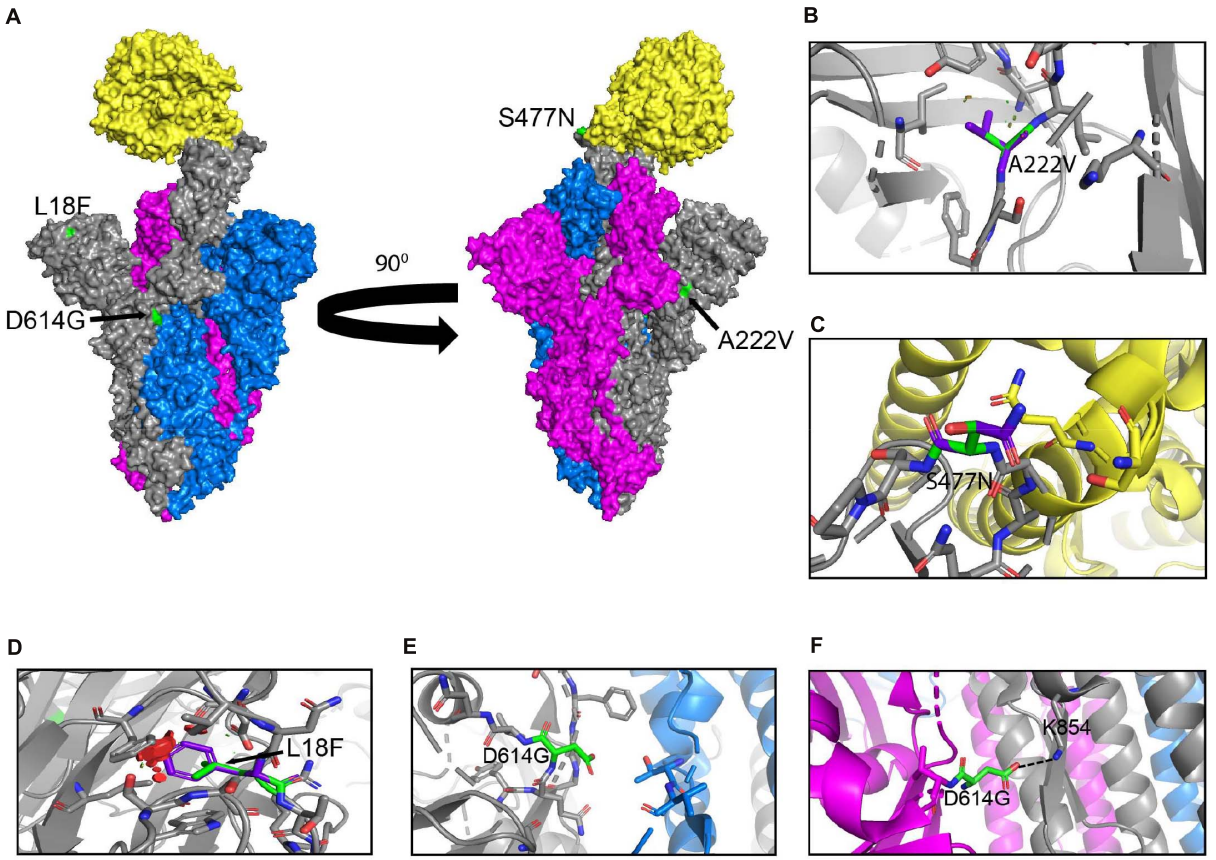


FIGURE 3 | Structural rendering of the most frequent mutations in the S protein. **(A)** Surface representation of human angiotensin-converting enzyme 2 (hACE2) (yellow) in complex with S protein trimers illustrated in gray, blue, and magenta. Interactions of high frequency mutations are presented as follows: **(B)** A222V, **(C)** S477N, **(D)** L18F, **(E)** D614G in the open conformation, and **(F)** D614G in the closed conformation. Reference sequence residues are illustrated in green, and the mutated amino acid is represented in purple. The red markers illustrate the steric clash when the mutations are inserted into the structure. Graphs were generated using PyMOL.

evolutionary constraints. Some of these viral proteins may be evolutionarily conserved and play important roles in virus assembly and stability, genome replication, and viral release from infected cells. Our analysis highlights the fixation of the D614G mutation in the S protein and that of the P323L mutation in the RNA-dependent RNA polymerase (RdRp) (**Figure 1**). These are the only mutations to have successfully become part of the reference sequence as of December 2020. They appeared to have emerged simultaneously in January 2020 and became present in more than 90% of all sequenced genomes by June 2020. Additionally, some other mutations also emerged rapidly but then stabilized in frequency or faded out. For example, Q57H (NS3), R203K (N), and G204R (N) are mutations that emerged rapidly and appeared to have stabilized at a frequency of 15% to 40%. Others like I120F (NSP2), L37F (NSP6), S477N (S), and L84S (NS8) illustrate mutations that emerged rapidly and then faded-out just as quickly. We also demonstrate that most genes in SARS-CoV-2 have mutations with overall frequencies lower than 10% (**Figure 1**). These mutations are also summarized in **Table 1**, which illustrates nucleotide substitution producing the amino acid change, the frequency, and their respective effects. Globally, there is an uneven effort to test for SARS-CoV-2 infections in the population and sequence the viral genomes in those infected. As a result, emerging mutations go unreported until they reach

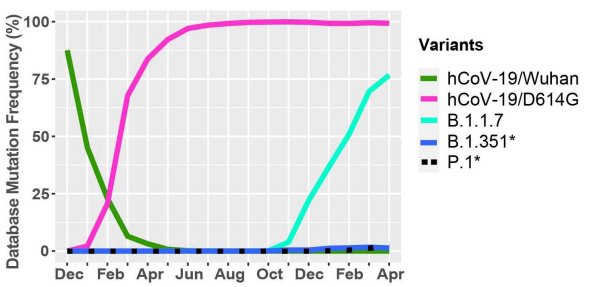


FIGURE 4 | Worldwide frequency of B.1.1.7, B.1.351, P.1 and D614G variants. Database variants frequency were analyzed from December 2019 to April 30, 2021. hCoV-19/Wuhan is the original reference strain and hCoV-19/D614G is the current reference strain containing the D614G mutation in the S protein. B.1.1.7, B.1.351, and P.1 represent the United Kingdom, South African, and Brazilian variants. *Overlapping curves.

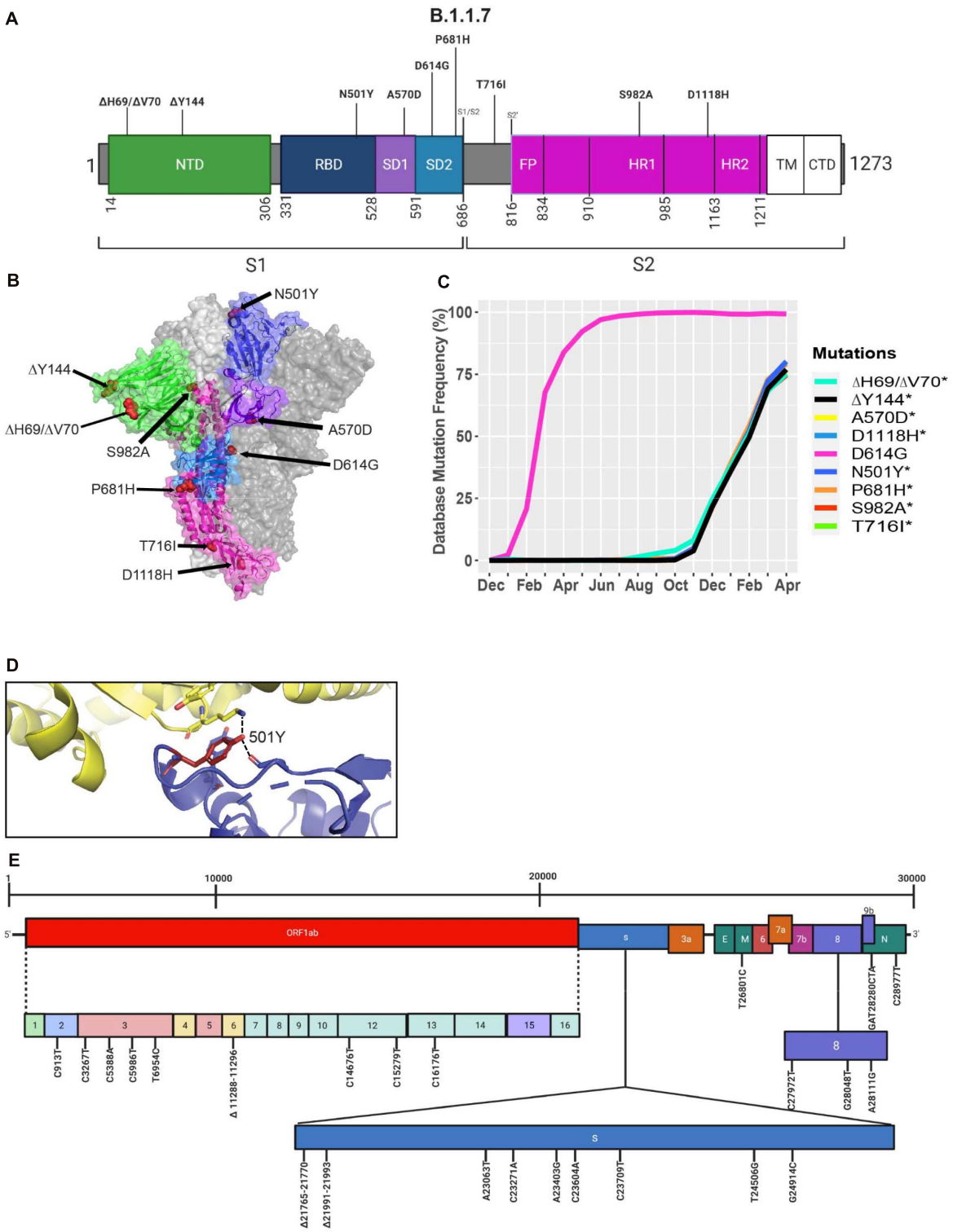


FIGURE 5 | Analysis of mutations in the B.1.1.7 variant. **(A)** Mutation map of the spike protein of B.1.1.7. **(B)** Structural representation of spike with ACE2. B.1.1.7 S protein mutations are presented in red. N-terminal domain (NTD) (green), receptor-binding domain (RBD) (blue), SD1 (purple), subdomain 2 (SD2) (light blue), and S2 (magenta) are illustrated. The other S protein monomers are displayed in gray and white. **(C)** Frequency of the mutations in the S protein, B.1.1.7 variant from December 2019 to April 30, 2021. **(D)** Interaction of the N501Y (red) mutation in the RBD (blue) of S protein with hACE2 (yellow). The dash lines indicate interactions with adjacent residues. **(E)** Genome of the SARS-CoV-2 B.1.1.7 variant with identified nucleotide substitutions and deletions. Graphs were generated using Biorender, PyMOL, and RStudio. *Overlapping curves.

countries with more intensive sequencing capabilities. Therefore, mutations presented here vastly underrepresent the global landscape of mutation frequency dynamics.

Geographic Localization and Timeline of the Viral Genes With Mutations Higher Than 50% Frequency

We next turned our attention to viral genes where at least one mutation reached a frequency higher than 50%. Only three genes met this criterion: NSP12, S, and N. We then took their mutation graphs and added worldwide geographic data provided from GISAID. Geomaps are useful to determine if a mutation is a localized and regional event or is found worldwide. In the S protein, D614G is found worldwide with initially higher reported cases in the United States, United Kingdom, and Australia, probably due to more intense large-scale testing and sequencing, while A222V is mostly reported in the United Kingdom, but has not yet been reported in South America, Central and East African regions (**Figure 2A**). Like D614G, P323L in the RdRp is also found worldwide, with higher reported cases in the United States, United Kingdom, and Australia (**Figure 2B**). The N gene does not yet have successful mutations that have attained reference sequence status, but R203K, G204R, A220V, D3L, and S235F all reached a frequency of 50% or higher during the study period. A220V emerged in August of 2020 and reached a frequency higher than 50% in October of 2020 (**Figure 2D**). Mutations G204R, R203K, and A220V in the N gene are reported at high frequency in the United Kingdom, but were not been detected in South America, Central, East, and South African regions (**Figure 2C**). We also observe the emergence of several mutations in S and N present in the B.1.1.7 variant, which now have a frequency higher than 65% (**Figures 2C,D** and **Table 2**). The analyses of these data illustrate the localization of the most prevalent mutations to date, which appear to be mostly present in Western high-income countries. This, however, is undoubtable attributable to overall more intense testing and sequencing.

Localization and Molecular Interactions of Prevalent S Protein Mutations

Here we illustrate the molecular interactions and spatial localization of the mutations in the S protein. PyMOL was used to model structures of the S protein binding to ACE2, and we analyze the possible effects of specific mutations at given positions in the protein. **Figure 3A** presents an overview of the physical positions of S protein mutations. The A222V mutation has no reported effects on protein stability, neutralizing antibody escape, and affinity for hACE2 (**Table 1**). The substitution from A to V results in a low steric clash between neighboring residues (**Figure 3B**). The S477N substitution in the RBD enables increased stability during hACE2–RBD interactions (**Figure 3C** and **Table 1**). In the N-terminal domain (NTD), mutation L18F leads to a steric clash between neighboring residues (**Figure 3D**). However, this does not appear to impact the stability of the protein given that no such effects have been reported so far (**Table 1**). In the closed conformation, D614 makes an ionic bond with K854 in the S2 subunit of another S protein monomer (**Figure 3F**; Benton et al., 2020). In the open conformation, D614 (or G614) does not make interactions or display steric clashes with neighboring residues (**Figure 3E**).

The Emergence of the B.1.1.7 Variant in the United Kingdom

A new variant was discovered in late 2020 in the United Kingdom that displayed increased affinity to hACE2, virulence, and Nabs escape capabilities (**Figure 4**; Starr et al., 2020; Greaney et al., 2021; Haas et al., 2021; Wang et al., 2021). Here we further investigated the B.1.1.7 variant by looking at S protein mutations of this variant in complex with Nabs and hACE2. We mapped

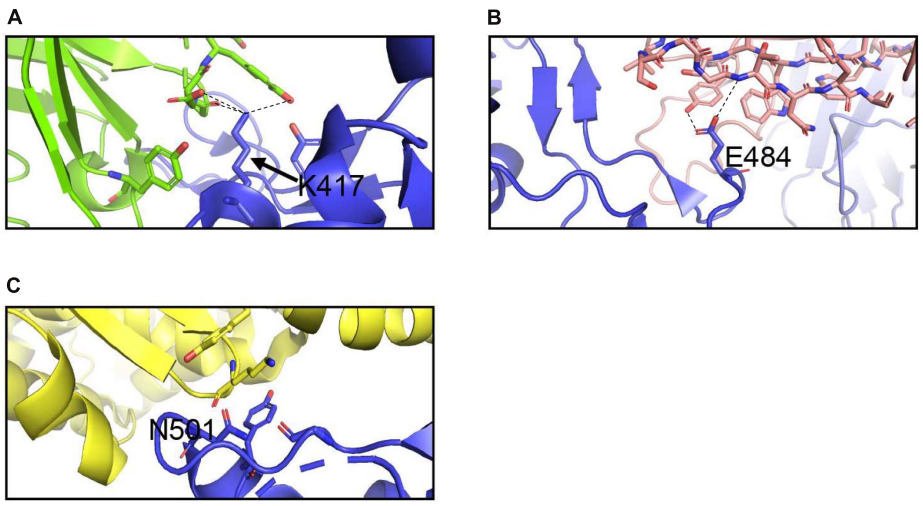


FIGURE 6 | Interactions of K417, E484, and N501 of the S protein with neutralizing antibodies and hACE2. **(A)** Interaction of K417 (blue) with C102 Nab (green) residues. **(B)** Interaction of E484 (blue) with C121 Nab (light pink). **(C)** Interaction of N501 (blue) with hACE2 (yellow) Dashes lines indicate interactions between residues. The graphs were generated using PyMOL.

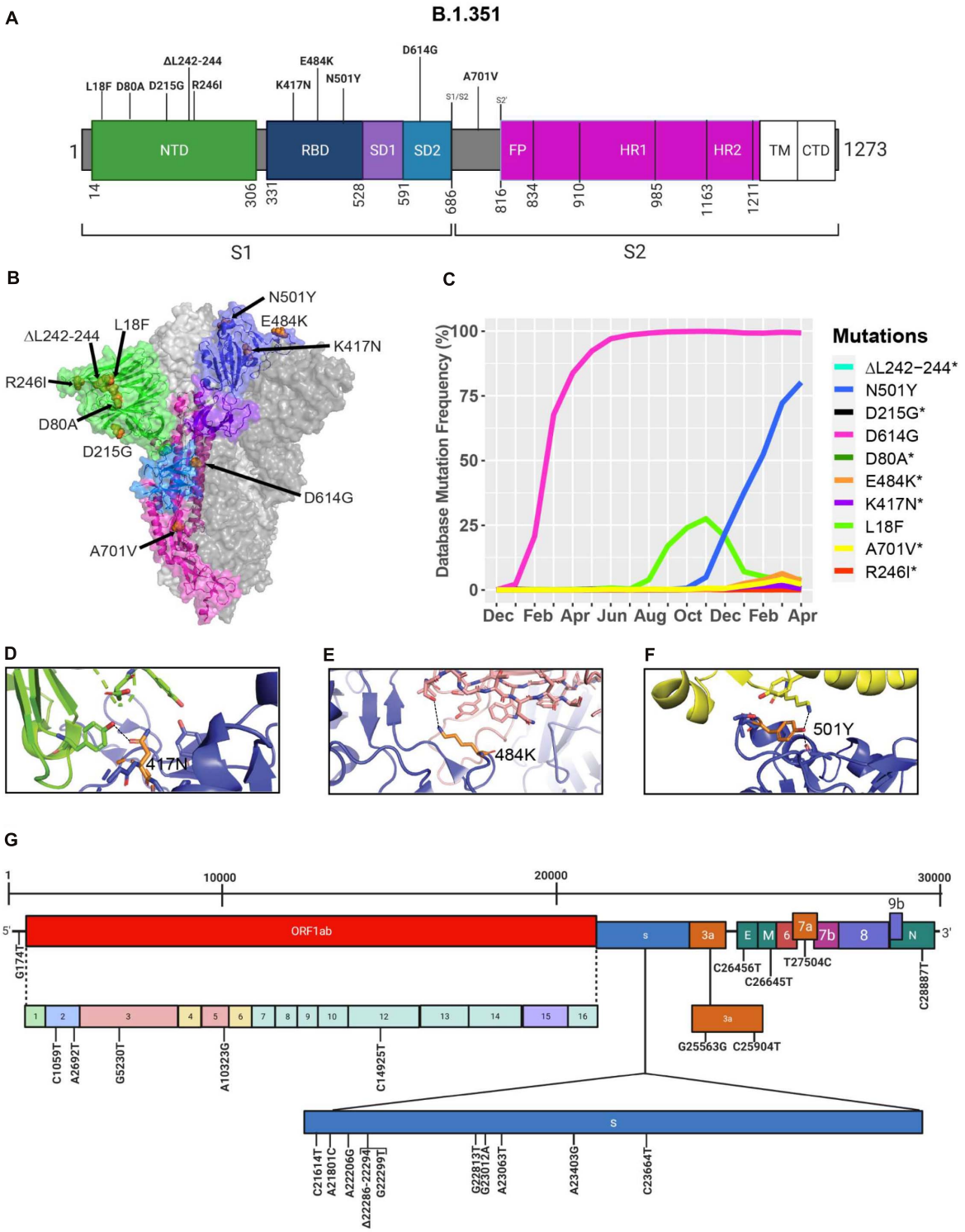


FIGURE 7 | Analysis of mutations in the B.1.351 variant. **(A)** Mutation map of the spike protein of B.1.351. **(B)** Structural representation of spike with ACE2. B.1.352 S protein mutations are presented in orange. NTD (green), RBD (blue), SD1 (purple), SD2 (light blue), and S2 (magenta) are illustrated. The other S protein monomers are illustrated in gray and white. **(C)** Frequency of the mutations in the S protein B.1.351 variant from December 2019 to April 30, 2021. Interaction of **(D)** 417N with C102 Nab (green), **(E)** 484K with C121 Nab (light pink), and **(F)** 501Y with hACE2 (yellow). The mutant residues are illustrated in orange, and the dashed lines represent interactions with adjacent residues. **(G)** Genome of the SARS-CoV-2 B.1.351 variant with identified nucleotide substitutions or deletions. Graphs were generated using Biorender, PyMOL, and RStudio. *Overlapping curves.

the localization of mutations with available Cryo-EM structures of the S protein and assessed the frequency of the variant by interrogating the GISAID database. There are nine mutations in the S protein out of the total 24 mutations in the B.1.1.7 SARS-CoV-2 genome (**Figures 5A,E**). Mutations in the S protein of the B.1.1.7 variant, except D614G, emerged in October of 2020 and reached a worldwide frequency of 70% to 80% in April 2021 (**Figures 4**, **5C** and **Table 2**). N501Y, found in the RBD, can interact with K353 in hACE2 (**Figures 5B,D**, **6C**). The N501Y mutation is associated with an increased affinity to hACE2, along with an increase in infectivity and virulence (**Table 2**). **Figure 5E** illustrates the whole genome of SARS-CoV-2 with all nucleotide substitutions of the B.1.1.7 variant. Substitutions C913T, C5986T, C14676T, C15279T, C16176T in ORF1ab, and T26801C in M protein are all synonymous mutations. Also, the C27972T mutation has a frequency of 75% and produces a premature stop codon in NS8 that inactivates the protein (Q27stop) without obvious consequences (**Figure 5E** and **Table 2**) (Pereira, 2020). These results allow us to better understand the frequencies, localization, and interactions of mutations in the S protein of the B.1.1.7 variant. For viruses, both synonymous and non-synonymous mutations are of importance. Synonymous mutations that do not change the amino acid sequence of proteins can still exercise very important functions. They can play a key role in docking sites for RNA-binding proteins, transcription factors, and primers, or partake in important secondary structures of the viral RNA like functional loops and folds.

The Emergence of the B.1.351 Variant in South Africa

During the spread of the B.1.1.7 variant in the United Kingdom, another variant was emerging in South Africa, known as B.1.351 (O'Toole et al., 2021; Tegally et al., 2021). The GISAID database was used to identify mutations and Cryo-EM structures of the S protein to model the effects of point mutations (**Figure 7**). Most of the mutations in the S protein of the B.1.351 variant are localized in the S1 subunit, with only A701V in the S2 subunit. Additionally, three mutations reside in the RBD,

TABLE 3 | Synonymous, non-synonymous mutations and deletions in the B.1.351 (South Africa) variant.

Variant	Protein (gene)	Genome nucleotide mutation Tegally et al., 2021	S or NS	AA mutation	Domain	Frequency* (%)	Effect
B.1.351	*5′-UTR*	G174T	S			N/A	N/A
	NSP2	C1059T	NS	T85I		13.95	N/A
	(ORF1ab)	A2692T	S			N/A	N/A
	NSP3 (ORF1ab)	G5230T	NS	K837N		1.41	N/A
	NSP5 (ORF1ab)	A10323G	NS	K90R		3.07	N/A
	NSP12 (ORF1ab)	C14925T	S			N/A	N/A
	S	C21614T	NS	L18F	NTD	4.11	N/A
		A21801C	NS	D80A	NTD	0.50	N/A
		A22206G	NS	D215G	NTD	0.51	N/A
		22286-22294del	NS	ΔL242/ΔA243/ ΔL244	NTD	0.51	Antibody escape Weisblum et al., 2020
		G22299T	NS	R246I	NTD	0.004	Antibody escape Weisblum et al., 2020
		G22813T	NS	K417N	RBD	0.51	Antibody escape Wang et al., 2021
		G23012A	NS	E484K	RBD	3.51	Antibody escape Barnes et al., 2020; Covid-19 Genomic UK Consortium, 2021
		A23063T	NS	N501Y	RBD	80.16	Increases affinity for hACE2 Wu et al., 2017
		A23403G	NS	D614G	SD2	99.29	Moderate increase in transmissbility Li Q. et al., 2020; Covid-19 Genomic UK Consortium, 2021
		G23664T	NS	A701V	S1/S2 – S2′	2.15	N/A
	NS3	G25563T	NS	Q57H		14.36	N/A
	(ORF3a)	C25904T	NS	S171L		1.77	N/A
	E	C26456T	NS	P71L		1.42	N/A
	M	C26645T	S				N/A
	NS7a (ORF7a)	T27504C	S				N/A
	N	C28887T	NS	T205I		7.00	N/A

**Mutation frequency as of April 30, 2021.*

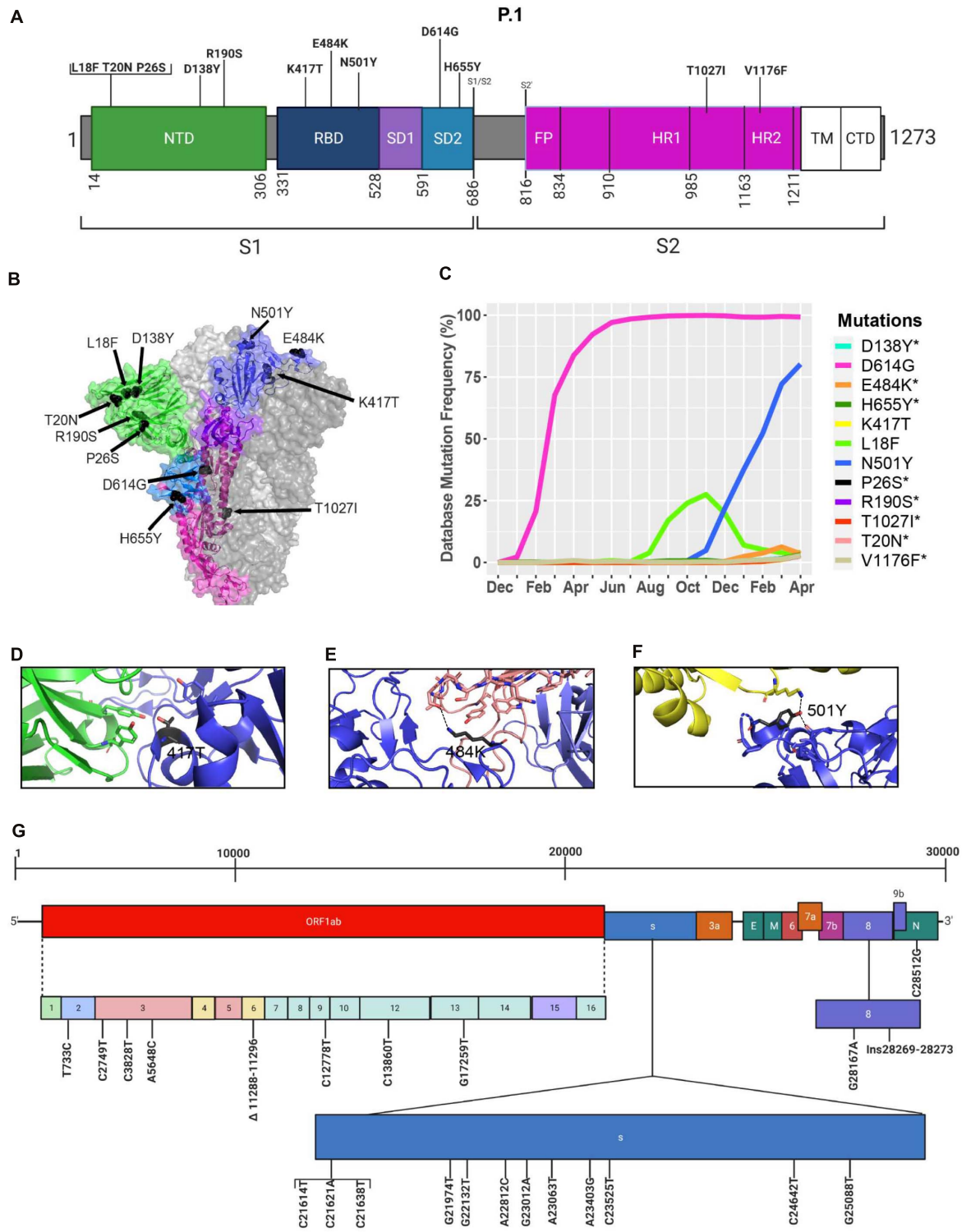


FIGURE 8 | Analysis of mutations in the P.1 variant. **(A)** Mutation map of the spike protein of P.1. **(B)** Structural representation of spike with ACE2. P.1 S protein mutations are presented in black. NTD (green), RBD (blue), SD1 (purple), SD2 (light blue), and S2 (magenta) are illustrated. The other S protein monomers are illustrated in gray and white. **(C)** Frequency of P.1 variant S protein mutations from December 2019 to April 30, 2021. Interaction of **(D)** 417T with C102 Nab (green), **(E)** 484K with C121 Nab (light pink), and **(F)** 501Y with hACE2 (yellow). The mutations are colored in black and interaction with adjacent residues are demonstrated by dashed lines. **(G)** Genome of the SARS-CoV-2 P.1 variant with identified nucleotides substitution, deletions, and insertions. Figures were generated using Biorender, PyMOL, and RStudio. *Overlapping curves.

among which two of them are not found in the B.1.1.7 variant (K417N, E484K) (**Figures 7A,B**). However, this variant contains D614G and N501Y, which are also seen in the B.1.1.7 variant (**Figure 5A**). Furthermore, many of the mutations found in the B.1.351 variant have worldwide frequencies lower than 5%. The exception is D614G, L18F, and N501Y (**Figure 7C** and **Table 3**). In comparison to B.1.1.7 and P.1 variants, the B.1.351 variant has not reached a frequency higher than 2% as of April 2021 (**Figure 4**). By using the mutagenesis tool of PyMOL, we modeled the S protein in complex with hACE2, and with C103 and C121 Nabs, which are human recombinant class I and II neutralizing antibodies, respectively (Barnes et al., 2020). Our *in silico* mutagenesis modeling predicts that B.1.351 mutations in the RBD favor a loss of interactions with C102 and C121 Nabs. Diminished interactions between RBD and C102 are predicted when the K417 is mutated to Asn producing Nabs escape capabilities (**Figures 7D**, **6A**). Diminished interactions with RBD are also predicted with C121 when the E484K mutation is present

TABLE 4 | Synonymous, non-synonymous, and deletions in the P.1 (Brazil) variant.

Variant	Protein (gene)	Genome nucleotide mutation Faria et al., 2021	S or NS	AA mutation	Domain	Frequency* (%)	Effect
P.1	*NSP2*	T733C	S			N/A	N/A
	(ORF1ab)	C2749T	S			N/A	N/A
	NSP3	C3828T	NS	S370L		3.09	N/A
	(ORF1ab)	A5648C	NS	K977Q		3.00	N/A
	NSP6 (ORF1ab)	11288-11296del	NS	ΔS106/ ΔG107/ΔF108		87.61	N/A
	NSP9 (ORF1ab)	C12778T	S			N/A	N/A
	NSP12 (ORF1ab)	C13860T	S			N/A	N/A
	NSP13 (ORF1ab)	G17259T	NS	E341D		2.97	N/A
	S	C21614T	NS	L18F	NTD	4.11	N/A
		C21621A	NS	T20N	NTD	2.77	N/A
		C21638T	NS	P26S	NTD	3.01	N/A
		G21974T	NS	D138Y	NTD	2.95	N/A
		G22132T	NS	R190S	NTD	2.71	N/A
		A22812C	NS	K417T	RBD	2.88	Antibody escape Garcia-Beltran et al., 2021
		G23012A	NS	E484K	RBD	3.51	Antibody escape Barnes et al., 2020; Covid-19 Genomic UK Consortium, 2021
		A23063T	NS	N501Y	RBD	80.16	Increase affinity for hACE2 Covid-19 Genomic UK Consortium, 2021
		A23403G	NS	D614G	SD2	99.29	Moderate increase in transmissibility Li Q. et al., 2020; Covid-19 Genomic UK Consortium, 2021
		C23525T	NS	H655Y	SD2	3.16	N/A
		C24642T	NS	T1027I	S2	2.79	N/A
		G25088T	NS	V1176F	S2	2.79	N/A
	NS8	G28167A	NS	E92K		2.96	N/A
	(ORF8)	Ins28269-28273	S			N/A	N/A
	N	C28512G	NS	P80R		2.94	N/A

**Mutation frequency as of April 30, 2021.*

TABLE 5 | Efficacy of vaccines against SARS-CoV-2 variants.

Vaccine	hCoV-19/Wuhan/WIV-4/2019	B.1.1.7	B.1.351	P.1
Pfizer-BioNTech	95.0% (Polack et al., 2020)	95.3% (Haas et al., 2021; Abu-Raddad et al., 2021)	75.0% (Abu-Raddad et al., 2021)	N/A
Moderna	94.1% (Baden et al., 2020)	N/A	N/A	N/A
Oxford-AstraZeneca	70.4% (Voysey et al., 2021)	70.4% (Emary et al., 2021)	10% (Madhi et al., 2021)	N/A
Johnson & Johnson	66% (Sadoff et al., 2021)	N/A	52.0% and 64.0% (14 and 28 days) (Sadoff et al., 2021)	66% (Sadoff et al., 2021)
Novavax	96% (Shinde et al., 2021)	86% (Shinde et al., 2021)	51% (Shinde et al., 2021)	

(**Figures 7E**, **6B**). As previously mentioned, N501Y is also found in B.1.351, and our modeling predicts that it will have similar effects as those observed with the B.1.1.7 variant (**Figure 7F**). **Figure 7G** illustrates the position of the nucleotide substitutions of the B.1.351 variant.

The Emergence of the P.1 Variant in Brazil

Similar to the B.1.351 variant, the P.1 variant harbors the N501Y and E484K mutations, but position 417 of the S protein displays a threonine (T) instead of a lysine (K) residue (Faria et al., 2021; **Figure 8**). Similar to the United Kingdom and South African variants, we present in **Figure 8C** mutations in the S protein that are characteristic of the P.1 variant and their frequencies from December 2019 to April 30, 2021. The P.1 variant harbors amino acid substitutions L18F, T20N, P26S, D138Y, and R190S in the NTD of the S protein. H655Y and T1027I, V1176F are in the subdomain 2 (SD2) of S1 and in the S2 subunit, respectively (**Figures 8A, B** and **Table 4**). The V1176F mutation is not shown because the structure is unresolved in this region. Overall, P.1-specific mutations have currently worldwide frequencies less than 5% (**Figure 8C** and **Table 4**). D614G, L18F, and N501Y are not specific to P1. Similar to B.1.351, the P.1 variant has a low frequency representing less than 5% of global variants as of April 2021 (**Figure 4**). Here, we also modeled P.1 mutations to investigate interaction alterations with known recombinant neutralizing antibodies (Barnes et al., 2020). The K417T mutation reduces interactions with neighboring residues in the C102 Nab, and therefore, the model predicts, as with K417N in B.1.1.7, an increased ability to escape neutralization (**Figure 8D**). Also, the P.1 variant has E484K and N501Y mutations in the RBD. Modeling predicts that they will have the same effect reported for the B.1.1.7 and B.1.351 variants (**Figures 8E,F**). In **Figure 8G**, we indicate the positions of synonymous, non-synonymous, and deletions in the SARS-CoV-2 P.1 genome.

DISCUSSION

The emergence of new genetic variants that are more transmissible, virulent, and resistant to antibody neutralization has highlighted the importance of studying the function of mutations in the viral genome. The number of sequenced viral genomes uploaded to the GISAID database grew rapidly from 131,417 at the end of September 2020 to 451,913 by January 30, 2021 (Shu and McCauley, 2017). GISAID is a formidable tool for tracking the emergence of mutations, identifying the geographic region where it emerged, and tracking its spread around the globe. Given that the risk mutations and new variants pose for neutralizing antibody therapy and vaccines for SARS-CoV-2, it is crucial to continuously monitor the susceptibility of these variants to neutralization by humoral and cellular immune responses either induced through natural exposure to the reference strain or induced by vaccination (Barnes et al., 2020; Greaney et al., 2021; Wang et al., 2021). Recent reports on the efficacy of the various vaccines have shown that these are variable and often diminished against variants (**Tables 5**, **6**).

TABLE 6 | Resistance of SARS-CoV-2 variants to vaccine-induced antibody neutralization.

Vaccine	Fold reduction in neutralization		
	B.1.1.7	**B.1.351**	**P.1**
Pfizer-BioNTech	[a]2.1 (Garcia-Beltran et al., 2021)	[a](34.5–42.4) (Garcia-Beltran et al., 2021)	[a]6.7 (Garcia-Beltran et al., 2021)
Moderna	[a]2.3 (Garcia-Beltran et al., 2021)	[a](19.2–27.7)(Garcia-Beltran et al., 2021)	[a]4.5 (Garcia-Beltran et al., 2021)
Oxford-AstraZeneca	[b]2.5 (Supasa et al., 2021)	[b]9.0 (Zhou et al., 2021)	[b]2.6 (Dejnirattisai et al., 2021)
Johnson & Johnson	N/A	N/A	N/A
Novavax	[a](0.85 to 20) (Shen et al., 2021b)	[a]13.1 (Shen et al., 2021a)	N/A

[a]*Pseudotyped virus neutralization assay.*
[b]*Live virus neutralization assay.*

Most available vaccines generally remain near fully efficacious against the B.1.1.7 variant (Emary et al., 2021; Haas et al., 2021; Wu et al., 2021). However, the Oxford-AstraZeneca vaccine has displayed compromised efficacy against the B.1.351 variant with only 10% vaccine efficacy (Madhi et al., 2021). Preliminary data with the Pfizer-BioNTech and Moderna mRNA vaccines also show a reduction in efficacy against B.1.351 (Abu-Raddad et al., 2021; Liu et al., 2021; Wu et al., 2021). Furthermore, antibodies induced by the Pfizer-BioNTech and Moderna vaccines appear to display a 6.7-fold and 4.5-fold decrease in neutralization efficacy against the P.1 variant (**Table 6**; Garcia-Beltran et al., 2021). The recently approved Johnson & Johnson adenovirus-based vaccine only requires a single dose, in comparison with the two for the mRNA vaccines and Oxford-AstraZeneca vaccine, and has an efficacy of 66% against the original Wuhan reference strain, 52.0% to 64.0% against the B.1.351 variant and 66% against the P.1 variant (Sadoff et al., 2021). The Novavax vaccine also shows reduced efficacy against the variants with 86% and 51% against the B.1.1.7 and P.1 variants, respectively. Nevertheless, humoral responses are only one component of the adaptive immune response. T-cell responses have not been studied in detail against these variants at this time and still provide robust contribution to overall protection against severe COVID-19 disease.

Studying the effect of individual mutations in a viral protein may not always reflect their true impact on virus features if they appear in combination with other mutations. Epistasis is the combinatory effect of two or more mutations in a genome (Phillips, 2008). Epistasis has previously been intensely studied in the surface protein hemagglutinin (HA) of influenza viruses, and positive epistasis was identified in several regions of the HA receptor-binding domain (Wu et al., 2017). In relation to the S protein of SARS-CoV-2, epistatic mutations could, for instance, allow the S protein to adopt a specific conformation when all the mutations are present, thereby producing a unique folding of the protein and enabling new features. A recent study demonstrated the impact of antigenicity and infectivity of D614G SARS-CoV-2 variants with a combination of different mutations occurring in

the S protein (Li Q. et al., 2020). The study shows that D614G alone only mildly increases infectivity, but in combination with different other mutations in the S protein, together, these can either more intensely increase or decrease viral infectivity. Similar findings have been reported regarding sensitivity to Nabs. D614G alone has undetectable effects on Nabs escape. However, the combination of D614G with other mutations in S can enable Nabs escape (Li Q. et al., 2020). This data suggests that the continuous emergence of epistatic mutations in SARS-CoV-2 will likely be involved in further altering the properties of the virus, including transmissibility, pathogenicity, stability, and Nab resistance.

Our analyses have highlighted that several new fast-spreading mutations had frequency trajectories that eventually plunged or stabilized at low frequencies. Only the D614G in the S protein and the P323L in the RdRp have maintained their presence in the consensus sequence (**Figures 1**, **2**). Analyses of the GISAID database also reveal which countries upload the most sequences to the database and are therefore carrying out the most testing and sequencing. The global frequency of mutations and variants in the database is therefore biased to represent the genetic landscape of the virus in the countries doing the most testing, sequencing, and data sharing. Emergence of new variants may therefore go undetected until they leave their point of origin and enter countries with high testing and sequencing rates. This delayed notification constitutes a major obstacle in identifying and preventing the spread of nefarious variants that are potentially resistant to current vaccines and neutralizing antibody therapy. Another caveat in analyzing sequences in the GISAID database is that consensus sequences are uploaded, but subsequences and quasispecies are generally not included. Therefore, it is possible that mutations in sub-variants may be missed that could contribute to virus replication, pathogenicity, and spread (Rocheleau et al., 2021). A global approach to analyzing both transmitted variants and non-transmitted sub-variants and quasispecies could provide a better understanding of the effects of SARS-CoV-2 mutations.

In conclusion, our metadata analysis of emerging mutations has highlighted the natural fluctuations in mutation prevalence. We also illustrate how mutations sometimes need to co-emerge in order to create a favorable outcome for virus propagation. Tracking mutations and the evolution of the SARS-CoV-2 genome is critical for the development and deployment of effective treatments and vaccines. Thus, it is the responsibility of all countries and governing jurisdictions to increase testing and sequencing, and uploading SARS-CoV-2 genomes to open access databases in real-time. This will result in more accurate information to inform policy and decision makers about interventions required to blunt the global transmission of the virus and ensure that vaccines remain effective against all circulating variants.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.gisaid.org.

AUTHOR CONTRIBUTIONS

NC performed the analysis and drafted the manuscript. WZ contributed to study design, provided guidance, and drafted the manuscript. M-AL designed the study, provided guidance, and drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by a COVID-19 Rapid Response grant to M-AL by the Canadian Institute of Health Research (CIHR) and by a grant supplement by the Canadian Immunity Task Force (CITF).

ACKNOWLEDGMENTS

The authors wish to thank Dr. Sean Li at Health Canada for helpful comments on our manuscript. M-AL holds a Canada Research Chair in Molecular Virology and Intrinsic Immunity.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.676314/full#supplementary-material

REFERENCES

Abu-Raddad, L. J., Chemaitelly, H., and Butt, A. A. (2021). Effectiveness of the BNT162b2 Covid-19 vaccine against the B.1.1.7 and B.1.351 variants. *N. Engl. J. Med.* doi: 10.1056/NEJMc2104974

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., et al. (2020). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* 384, 403–416.

Barnes, C. O., Jette, C. A., Abernathy, M. E., Dam, K.-M. A., Esswein, S. R., Gristick, H. B., et al. (2020). SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* 588, 682–687.

Benton, D. J., Wrobel, A. G., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., et al. (2020). Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* 588, 327–330. doi: 10.1038/s41586-020-2772-0

Chu, H., Chan, J. F.-W., Yuen, T. T.-T., Shuai, H., Yuan, S., Wang, Y., et al. (2020). Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. *Lancet Microbe* 1, e14–e23.

Covid-19 Genomic UK Consortium (2021). *COG-UK Report on SARS-CoV-2 Spike Mutations of Interest in the UK*. London: Covid-19 Genomic UK Consortium Available online at: https://www.cogconsortium.uk/wp-content/uploads/2021/01/Report-2_COG-UK_SARS-CoV-2-Mutations.pdf (accessed May 13, 2021).

Dejnirattisai, W., Zhou, D., Supasa, P., Liu, C., Mentzer, A. J., Ginn, H. M., et al. (2021). Antibody evasion by the Brazilian P.1 strain of SARS-CoV-2. *bioRxiv* [preprint] doi: 10.1101/2021.03.12.435194

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 20, 533–534. doi: 10.1016/s1473-3099(20)30120-1

Drake, J. W. (1993). Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. U.S.A.* 90, 4171. doi: 10.1073/pnas.90.9.4171

Duffy, S. (2018). Why are RNA virus mutation rates so damn high? *PLoS Biol.* 16:e3000003. doi: 10.1371/journal.pbio.3000003

Eckerle, L. D., Becker, M. M., Halpin, R. A., Li, K., Venter, E., Lu, X., et al. (2010). Infidelity of SARS-CoV Nsp14-Exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.* 6:e1000896. doi: 10.1371/journal.ppat.1000896

Emary, K. R. W., Golubchik, T., Aley, P. K., Ariani, C. V., Angus, B., Bibi, S., et al. (2021). Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory analysis of a randomised controlled trial. *Lancet* 397, 1351–1362.

Faria, N. R., Claro, I. M., Candido, D., Moyses Franco, L. A., Andrade, P. S., Coletti, T. M., et al. (2021). *Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in Manaus: Preliminary Findings. Virological*. Available online at: https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586 (accessed January 23, 2021).

Garcia-Beltran, W. F., Lam, E. C., St. Denis, K., Nitido, A. D., Garcia, Z. H., Hauser, B. M., et al. (2021). Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* 184, 2372–2383.e9. doi: 10.1016/j.cell.2021.03.013

Greaney, A. J., Starr, T. N., Gilchuk, P., Zost, S. J., Binshtein, E., Loes, A. N., et al. (2021). Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* 29, 44–57.e9.

Gu, H., Chen, Q., Yang, G., He, L., Fan, H., Deng, Y.-Q., et al. (2020). Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* 369:1603.

Haas, E. J., Angulo, F. J., McLaughlin, J. M., Anis, E., Singer, S. R., Khan, F., et al. (2021). Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet* 397, 1819–1829. doi: 10.1016/S0140-6736(21)00947-8

Kannan, S. R., Spratt, A. N., Quinn, T. P., Heng, X., Lorson, C. L., Sönnerborg, A., et al. (2020). Infectivity of SARS-CoV-2: there is something more than D614G? *J. Neuroimmune Pharmacol.* 15, 574–577. doi: 10.1007/s11481-020-09954-3

Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell* 181, 914–921.e10.

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. *Cell* 182, 812–827.e19.

Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., et al. (2020). The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 182, 1284–1294.e9.

Li, X., Giorgi, E. E., Marichannegowda, M. H., Foley, B., Xiao, C., Kong, X.-P., et al. (2020). Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci. Adv.* 6:eabb9153. doi: 10.1126/sciadv.abb9153

Liu, Y., Liu, J., Xia, H., Zhang, X., Fontes-Garfias, C. R., Swanson, K. A., et al. (2021). Neutralizing activity of BNT162b2-elicited serum — preliminary report. *N. Engl. J. Med.* 384, 1466–1468. doi: 10.1056/NEJMc2102017

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., et al. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. *Proc. Natl. Acad. Sci. U.S.A.* 112:9436. doi: 10.1073/pnas.1508686112

Madhi, S. A., Baillie, V., Cutland, C. L., Voysey, M., Koen, A. L., Fairlie, L., et al. (2021). Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B.1.351 variant. *N. Engl. J. Med.* 384, 1885-1898. doi: 10.1056/NEJMoa2102214

Minskaia, E., Hertzig, T., Gorbalenya, A. E., Campanacci, V., Cambillau, C., Canard, B., et al. (2006). Discovery of an RNA virus 3′→5′ exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 103:5108. doi: 10.1073/pnas.0508200103

O'Toole, A., Hill, V., Pybus, O. G., Watts, A., Bogoch, I. I., Khan, K., et al. (2021). *Tracking the international spread of SARS-CoV-2 lineage B.1.1.7 and B.1.351/501Y-V2*. Available online at: https://virological.org/t/tracking-the-international-spread-of-sars-cov-2-lineages-b-1-1-7-and-b-1-351-501y-v2/592 (accessed May 13, 2021).

Peck, K. M., and Lauring, A. S. (2018). Complexities of viral mutation rates. *J. Virol.* 92:e01031-17.

Pereira, F. (2020). Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. *Infect. Genet. Evol.* 85, 104525–104525. doi: 10.1016/j.meegid.2020.104525

Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9, 855–867. doi: 10.1038/nrg2452

Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., et al. (2020). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592, 116–121. doi: 10.1038/s41586-020-2895-3

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* 383, 2603–2615.

Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., et al. (2020). *Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations*. Available online at: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed May 13, 2021).

Rausch, J. W., Capoferri, A. A., Katusiime, M. G., Patro, S. C., and Kearney, M. F. (2020). Low genetic diversity may be an Achilles heel of SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* 117:24614. doi: 10.1073/pnas.2017726117

Rocheleau, L., Laroche, G., Fu, K., Stewart, C. M., Mohamud, A. O., Côté, M., et al. (2021). Identification of a high-frequency intra-host SARS-CoV-2 spike variant with enhanced cytopathic and fusogenic effect. *bioRxiv* [preprint] doi: 10.1101/2020.12.03.409714

Sadoff, J., Gray, G., Vandebosch, A., Cárdenas, V., Shukarev, G., Grinsztejn, B., et al. (2021). Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *N. Engl. J. Med.* 384, 2187–2201. doi: 10.1056/NEJMoa2101544

Shen, X., Tang, H., McDanal, C., Wagh, K., Fischer, W., Theiler, J., et al. (2021a). SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. *Cell Host Microbe* 29, 529–539.e3.

Shen, X., Tang, H., Pajon, R., Smith, G., Glenn, G. M., Shi, W., et al. (2021b). Neutralization of SARS-CoV-2 Variants B.1.429 and B.1.351. *N. Engl. J. Med.*

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. doi: 10.1126/science.abb7015

Shinde, V., Bhikha, S., Hoosain, Z., Archary, M., Bhorat, Q., Fairlie, L., et al. (2021). Efficacy of NVX-CoV2373 Covid-19 vaccine against the B.1.351 variant. *N. Engl. J. Med.* 384, 1899-1909.

Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* 22:30494.

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., et al. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 182, 1295–1310.e20.

Supasa, P., Zhou, D., Dejnirattisai, W., Liu, C., Mentzer, A. J., Ginn, H. M., et al. (2021). Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* 184, 2201–2211.e7.

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.

Ugurel, O. M., Mutlu, O., Sariyer, E., Kocer, S., Ugurel, E., Inci, T. G., et al. (2020). Evaluation of the potency of FDA-approved drugs on wild type and mutant SARS-CoV-2 helicase (Nsp13). *Int. J. Biol. Macromol.* 163, 1687–1696. doi: 10.1016/j.ijbiomac.2020.09.138

Voysey, M., Clemens, S. A. C., Madhi, S. A., Weckx, L. Y., Folegatti, P. M., Aley, P. K., et al. (2021). Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* 397, 99–111.

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. *Cell* 181, 281–292.e6.

Wang, P., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., et al. (2021). Increased resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7 to antibody neutralization. *bioRxiv* [preprint] doi: 10.1101/2021.01.25.428137

Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J. C., et al. (2020). Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *ELife* 9:e61312.

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.

Wu, K., Werner, A. P., Koch, M., Choi, A., Narayanan, E., Stewart-Jones, G. B. E., et al. (2021). Serum neutralizing activity elicited by mRNA-1273 vaccine — Preliminary Report. *N. Engl. J. Med.* 384, 1468–1470. doi: 10.1056/NEJMc2102179

Wu, N. C., Xie, J., Zheng, T., Nycholat, C. M., Grande, G., Paulson, J. C., et al. (2017). Diversity of functionally permissive sequences in the receptor-binding site of influenza hemagglutinin. *Cell Host Microbe* 21, 742–753.e8.

Wu, S., Tian, C., Liu, P., Guo, D., Zheng, W., Huang, X., et al. (2020). Effects of SARS-CoV-2 mutations on protein structures and intraviral protein–protein interactions. *J. Med. Virol.* 93, 2132–2140. doi: 10.1002/jmv.26597

Yi, C., Sun, X., Ye, J., Ding, L., Liu, M., Yang, Z., et al. (2020). Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell. Mol. Immunol.* 17, 621–630. doi: 10.1038/s41423-020-0458-z

Zahradník, J., Marciano, S., Shemesh, M., Zoler, E., Chiaravalli, J., Meyer, B., et al. (2021). SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv* [preprint] doi: 10.1101/2021.01.06.425392

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., et al. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* 11:6013.

Zhou, D., Dejnirattisai, W., Supasa, P., Liu, C., Mentzer, A. J., Ginn, H. M., et al. (2021). Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* 184, 2348–2361.e6.

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

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 July 2021
doi: 10.3389/fmicb.2021.683152

Gene Segment Interactions Can Drive the Emergence of Dominant Yet Suboptimal Gene Constellations During Influenza Virus Reassortment

Sanja Trifkovic[1†], Brad Gilbertson[1], Emily Fairmaid[1], Joanna Cobbin[1†], Steven Rockman[1,2] and Lorena E. Brown[1,3]*

[1] The Department of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia, [2] Seqirus, Parkville, VIC, Australia, [3] Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education, Hokkaido University, Sapporo, Japan

A segmented genome enables influenza virus to undergo reassortment when two viruses infect the same cell. Although reassortment is involved in the creation of pandemic influenza strains and is routinely used to produce influenza vaccines, our understanding of the factors that drive the emergence of dominant gene constellations during this process is incomplete. Recently, we defined a spectrum of interactions between the gene segments of the A/Udorn/307/72 (H3N2) (Udorn) strain that occur within virus particles, a major interaction being between the NA and PB1 gene segments. In addition, we showed that the Udorn PB1 is preferentially incorporated into reassortant viruses that express the Udorn NA. Here we use an influenza vaccine seed production model where eggs are coinfected with Udorn and the high yielding A/Puerto Rico/8/34 (H1N1) (PR8) virus and track viral genotypes through the reassortment process under antibody selective pressure to determine the impact of Udorn NA-PB1 co-selection. We discovered that 86% of the reassortants contained the PB1 from the Udorn parent after the initial co-infection and this bias towards Udorn PB1 was maintained after two further passages. Included in these were certain gene constellations containing Udorn HA, NA, and PB1 that confered low replicative fitness yet rapidly became dominant at the expense of more fit progeny, even when co-infection ratios of the two viruses favoured PR8. Fitness was not compromised, however, in the corresponding reassortants that also contained Udorn NP. Of particular note is the observation that relatively unfit reassortants could still fulfil the role of vaccine seed candidates as they provided high haemagglutinin (HA) antigen yields through co-production of non-infectious particles and/or by more HA molecules per virion. Our data illustrate the dynamics and complexity of reassortment and highlight how major gene segment interactions formed during packaging, in addition to antibody pressure, initially restrict the reassortant viruses that are formed.

Keywords: influenza virus, reassortment, pandemics, gene segment interactions, vaccine production

OPEN ACCESS

Edited by:
Kai Huang,
University of Texas Medical Branch at Galveston, United States

Reviewed by:
Hongquan Wan,
United States Food and Drug Administration, United States
Gloria Consuelo Ramirez-Nieto,
National University of Colombia, Colombia

****Correspondence:***
Lorena E. Brown
lorena@unimelb.edu.au

†Present address:
Sanja Trifkovic,
St Jude Children's Research Hospital, Memphis, TN, United States
Joanna Cobbin,
School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW, Australia

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *20 March 2021*
Accepted: *23 June 2021*
Published: *14 July 2021*

Citation:
Trifkovic S, Gilbertson B, Fairmaid E, Cobbin J, Rockman S and Brown LE (2021) Gene Segment Interactions Can Drive the Emergence of Dominant Yet Suboptimal Gene Constellations During Influenza Virus Reassortment. Front. Microbiol. 12:683152. doi: 10.3389/fmicb.2021.683152

INTRODUCTION

Reassortment, or the swapping of gene segments, is a major mechanism of influenza virus evolution. After co-infection of a single cell by two or more influenza viruses, each viral genome is replicated and various combinations of individual gene segments, in the form of viral ribonucleoprotein complexes (vRNPs), are co-packaged into progeny virions. Since influenza has eight vRNPs, co-infection with two viruses could theoretically yield up to 256 (2^8) different gene segment constellations. Although reassortment can occur at high frequency (Ghedin et al., 2005; Lu et al., 2014; Steel and Lowen, 2014), studies have shown this to be a non-random process (Lubeck et al., 1979; Rabadan et al., 2008; Varich et al., 2008; Marshall et al., 2013) with progeny populations restricted due to incompatibilities between gene products (Li et al., 2008, 2010; Song et al., 2011), preferential co-packaging of gene segments (Gavazzi et al., 2013b; Cobbin et al., 2014; Gilbertson et al., 2016), and certain genes reassorting at higher frequencies (Downie, 2004; Nelson et al., 2012; Schrauwen et al., 2013; Lu et al., 2014).

Using cross-linking of gene segments *in virio* together with deep sequencing of digested products, we recently uncovered an extensive network of inter-segment interactions believed to be used by the virus to package its genome (Dadonaite et al., 2019). Contrary to existing dogma, these interactions were not restricted to the previously defined "packaging sequences" at the ends of the segments but occurred throughout their entire length (Cobbin et al., 2014; Dadonaite et al., 2019). The pattern of interactions differed markedly between viruses of different subtypes and to a lesser extent between strains of the one subtype (Dadonaite et al., 2019). Importantly, the observed interactions were numerous, with hundreds being detected in the population of virus particles as a whole. These findings led us to propose that the ability of the influenza virus genome to utilise different sets of interactions to package one of each of the eight gene segments provides sufficient flexibility to allow reassortment to occur between different influenza viruses. Importantly, of this suite of interactions, some were found at very high frequency, indicating these were likely present in the majority of virus particles. One such high-frequency interaction occurred between the NA and PB1 genes in the early H3N2 virus A/Udorn/307/72 (Udorn) (Cobbin et al., 2014; Gilbertson et al., 2016) at nucleotides NA 512-550:PB1 2004-2037 (3′–5′), equivalent to NA 917-955:PB1 305-338 (5′–3′) (Dadonaite et al., 2019). This NA:PB1 interaction was also maintained in a reverse engineered virus containing the NA and PB1 genes from Udorn and the remaining genes from the H1N1 virus A/Puerto Rico/8/34 (PR8) where the pattern of interactions were found to be essentially inherited from both parent viruses (Dadonaite et al., 2019). This supported the notion that the stronger interactions are preferentially maintained and shape the gene constellations of resulting reassortant progeny.

Our initial interest in the NA:PB1 gene segment interaction developed from the retrospective analysis of the gene constellations of H3N2 influenza vaccine seed candidate viruses (Cobbin et al., 2013). These are produced by reassortment with the highly egg-adapted PR8 virus to enable the creation of viruses displaying greater H3N2 surface antigen yields in eggs through incorporation of non-surface antigen genes from the PR8 parent (Kilbourne and Murphy, 1960). This classical reassortment process consists of an initial co-infection step in eggs, followed by antibody selection for reassortant viruses with seasonal haemagglutinin (HA) and neuraminidase (NA) surface antigens, and finally cloning by limit dilution to isolate dominant viruses, some of which will display high haemagglutination titres suitable for vaccine seeds. By genotyping past H3N2 vaccine seed candidates, we showed that, unlike other non-surface antigen genes, the PB1 gene of the seasonal H3N2 virus was present in high frequency in these vaccine seeds. Furthermore, the Udorn virus exemplified this process, with 75% of the progeny virus expressing Udorn PB1 following reassortment with PR8 in eggs (Cobbin et al., 2013). This occurred despite the fact that the resulting reassortant progeny expressing Udorn HA in addition to Udorn NA and PB1 were significantly less fit than viruses that contained the Udorn HA and NA genes together with the PB1 gene from the H1N1 strain (Cobbin et al., 2014).

Here we used the vaccine seed model, where antibody drives selection of viruses with Udorn HA and NA, and gene segment interactions potentially drive co-selection of Udorn NA and PB1 with the aim of better understanding the dynamics of co-selection of the seasonal NA and PB1 genes and the effects of the co-infection ratio of Udorn and PR8 viruses on the outcome. This model also allowed us to examine why the viruses that have gene constellations resulting in inferior replicative fitness, exemplified by those expressing Udorn HA, NA, and PB1 genes, could ever be chosen on the basis of high antigen yield as we had observed in our retrospective analysis of vaccine seed strains. Our study illustrates the dynamics and complexity of classical reassortment and the impact of gene co-selection in this process. We also present data that highlights the effect of genotype on phenotype in relation to antigen expression.

MATERIALS AND METHODS

Influenza A Viruses

The highly egg-adapted A/Puerto Rico/8/34 (PR8; H1N1) virus, currently used as a reassortment partner for the production of H3N2 vaccine seeds, and A/Udorn/307/72 (Udorn; H3N2) as a model seasonal isolate, were used in this study. An eight-plasmid DNA transfection system (Hoffmann et al., 2000) was used to generate reverse engineered (rg) virus. Specific gene constellations are referred to using the following standard nomenclature e.g., PR8(Ud-HA,NA,PB1,NP) referring to an isolate containing Udorn HA, NA, PB1, and NP genes with the remaining genes from PR8. All viruses were propagated in 10-day old embryonated hen's eggs at 35°C for 2 days, then allantoic fluid was harvested and stored at −80°C.

Classical Reassortment

This was performed using a modification of the method devised by Kilbourne (1969) (**Figure 1A**). Ten-day-old embryonated hen's eggs were co-infected with PR8 and Udorn viruses (**Figure 1B**) and incubated at 35°C in a moist environment for 24 h. As in seed strain production, different ratios of the two

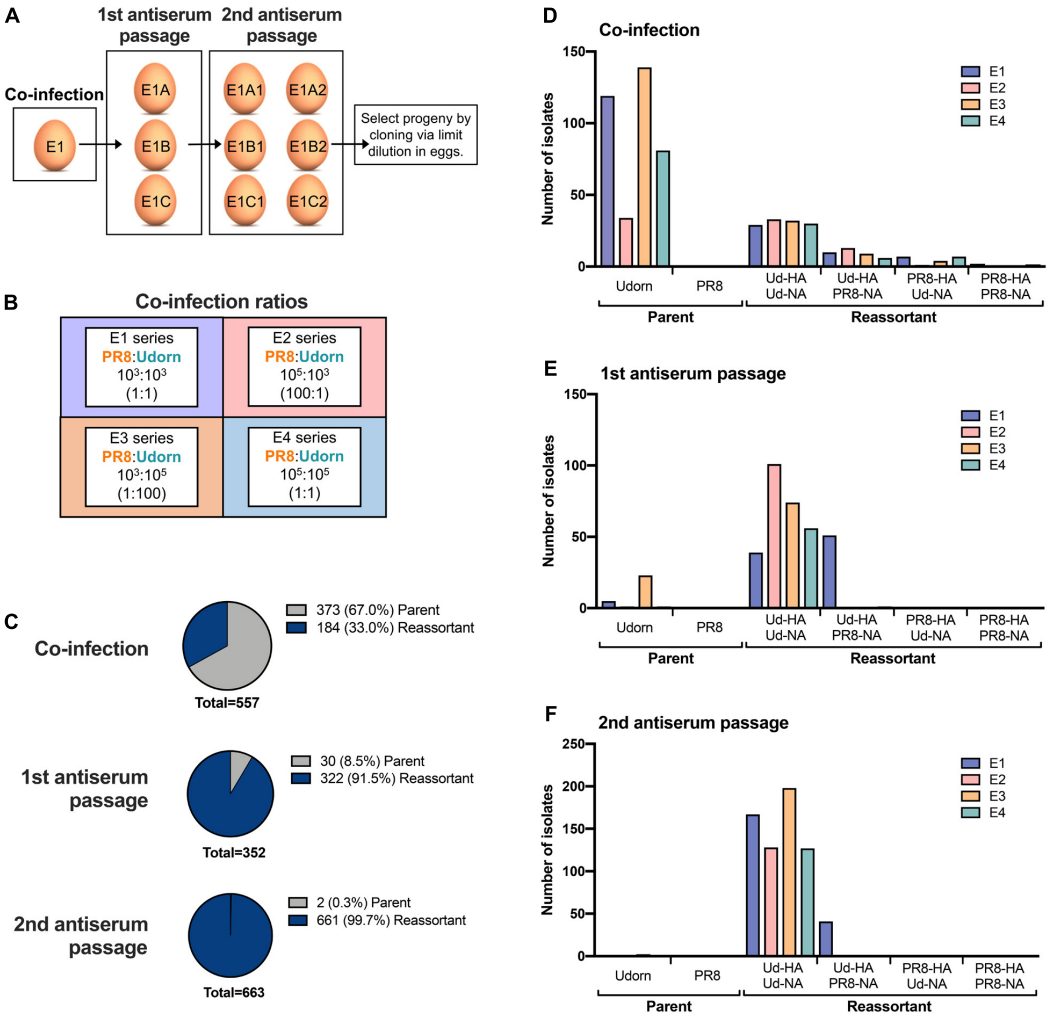


FIGURE 1 | The frequency of parent versus reassortant genotypes isolated throughout the classical reassortment process. **(A)** Schematic overview of the classical reassortment process used for the generation of seasonal influenza vaccine seeds. Ten-day-old embryonated hen's eggs were co-infected with PR8 and Udorn viruses. The allantoic fluid from these eggs was then harvested and passaged into three eggs (e.g., E1A, B, and C) in the presence of antiserum against the PR8 surface glycoproteins. The allantoic fluid from each of these eggs was then further passaged into two eggs (e.g., E1A1 and E1A2). Viruses from these were cloned by limit dilution. **(B)** The experiment was carried out four times at different co-infection ratios (series E1-E4). The co-infection ratios and amounts of infectious PR8 and Udorn virus (PFU) used to inoculate the first eggs of each series are indicated. **(C)** The ratio of reassortant genotypes to the Udorn parent at each stage of the classical reassortment process. Viruses in the allantoic fluid were isolated by plaque formation in MDCK cells. For the co-infected eggs, plaquing was performed in the presence of anti-PR8 antiserum. Individual plaques were picked into 0.05% Triton X-100 and gene specific RT-PCR performed to determine the origin of each gene. **(D–F)** The number of PR8 and Udorn parent genotypes recovered as well as the number of viruses containing different combinations of HA and NA genes present on reassortant viruses in **(D)** the initial coinfection, **(E)** the first antiserum passage, and **(F)** the second antiserum passage.

viruses were examined (**Figure 1B**). Allantoic fluid from the co-infected eggs was harvested and passaged twice in eggs in the presence of antibodies to neutralise the PR8 parent virus. These antibodies were in the form of purified gamma globulin from polyclonal sheep anti-PR8 antiserum that had been trypsin- and periodate-treated to remove non-specific inhibitors. For the first antiserum passage, eggs were inoculated with 0.2 ml virus-infected allantoic fluid from the co-infected eggs (10^{-3} dilution) followed by 0.2 ml of the anti-PR8 antibody preparation after 1 h. After 48 h the allantoic fluid was harvested and genotyped to detect the presence of the Udorn HA and NA genes. Virus-infected allantoic fluid with both Udorn HA and NA genes present was subjected to a second egg passage, this time being pre-incubated with the antiserum (the same amount as for the first passage) for 1 h prior to inoculation of the mixture into eggs. After 48 h the allantoic fluid was harvested and genotyped to determine the origin of the HA, NA, and M genes. The dominant progeny from classical reassortment were isolated by cloning via limit dilution. Briefly, serial dilutions of the allantoic fluid were inoculated into five eggs each and after 48 h the allantoic fluid was harvested. A haemagglutination assay was performed to detect the highest dilution containing virus for subsequent genotyping. If required, a further round of cloning by limit dilution was performed to obtain a pure population.

Virus Quantitation

Assays to quantitate haemagglutination (Fazekas De St and Webster, 1966) (as haemagglutinating units, HAU) were performed in microtitre plates with 1% chicken red blood cells and infectious virus yields (plaque-forming units, PFU) were determined by plaque assay on confluent MDCK cells monolayers (Tannock et al., 1984). To establish the relative HA yields of reassortant viruses, a standardised infection was performed by inoculating 10-day embryonated eggs ($n = 5$) with 100 PFU of virus injected into the allantoic cavity. After 2 days at 35°C the eggs were chilled overnight at 4°C before harvesting the allantoic fluid.

Gene Specific RT-PCR for the Identification of Viral Genes

The origin of viral genes present in samples were determined by gene-specific reverse transcription polymerase chain reaction (RT-PCR) using a SensiFast Probe No-ROX one-step reverse transcription kit (Bioline, Meridian Biosciences, OH, United States). Each 20 μl reaction contained 5 μl of virus sample in 0.05% Triton-X 100, 10 μl of 2x SensiFast Probe No-ROX one-step master mix, 0.2 μl reverse transcriptase, 0.4 μl RiboSafe RNase inhibitor mix, 0.8 μl of each 10 μM forward and reverse primer and 0.08 μl of each 25 μM gene specific probe. The reverse transcription, amplification, and detection were performed in a BioRad CFX96 PCR System. The reaction conditions, primers (Geneworks, Adelaide, SA, Australia) and probe [Integrated DNA Technologies (IDT), Coralville, IA, United States] sequences are available on request.

Viral Replicative Fitness

Viral replication kinetics were determined by infecting MDCK cells using a multiplicity of infection (MOI) of 0.001 PFU/cell. After a 1-h adsorption period the inoculum was removed, cells were washed and incubated in media supplemented with 1 μg/ml TPCK (designated time = 0 h). Cell culture supernatants were harvested at various time points and stored at −80°C until required. Infectious viral titres were determined by plaque assay as above.

Mini Genome Assay

A β-lactamase (BLA) reporter assay (Cavrois et al., 2002) was used to compare the activities of viral polymerase complexes identified in the dominant reassortant viruses as previously described (Cobbin et al., 2013). Briefly, 293T cells were transfected with 2 ng each of plasmids expressing the three influenza virus polymerase genes (PB2, PB1, and PA) and the nucleoprotein (NP) gene, together with 2 ng of a plasmid encoding BLA (provided by CSL Ltd.). These pHW2000 plasmids were those used for genetic engineering of influenza viruses in which viral cDNA is inserted into a bicistronic expression system (CMV and RNA polymerase 1 promotors) (Hoffmann et al., 2000). For the BLA-expressing plasmid the reporter gene was cloned between the 5′ and 3′ non-coding regions of the H1 HA gene to provide specificity for the influenza polymerase complex. Background levels of BLA produced by direct transcription from the CMV promotor was assessed in cells transfected with the reporter gene alone. After incubation at 37°C and 5% CO_2 for 24 h, cells were lysed and LyticBLazer^{TM}-FRET B/G substrate (Life Technologies) added. The BLA cleavage of the green substrate to the blue product was measured by optical density every 15 min for a 2-h period using a CLARIOstar (BMG Labtech, Ortenberg, Germany) fluorescence reader with excitation at 405 nm and blue and green emission detected at 445 and 520 nm, respectively. Specific BLA activity and thus relative polymerase activity was calculated as follows: (445 nm/520 nm ratio of the sample)/(445 nm/520 nm ratio of BLA plasmid alone).

Electron Microscopy

Transmission electron microscopy was performed at CSL Ltd. by Ross Hamilton using a method modified from that of Hayat and Miller (1990). Allantoic fluid was transferred onto Formar-coated copper TEM grids (Athene), which were inverted onto 2% agar plates to remove excess liquid. Grids were negatively stained using 2% sodium phosphotungstate (PTA) pH 7.0, excess stain was blotted away using filter paper (Whatman) and allowed to air-dry. Negatively-stained samples were examined using a Philips CM10 transmission electron microscope running iTEM software (EMSIS GmbH).

Quantification of Viral RNA and Viral mRNA/cRNA

To quantitate viral RNA production in MDCK cells, total RNA from infected cells was extracted using an RNeasy Mini Kit (Qiagen). Viral vRNA and mRNA were detected by polarity-specific quantitative RT-PCR using the SensiFast Probe No-ROX one-step kit (Bioline). Each 20 μl reaction contained 5 μl of RNA, 10 μl of 2x SensiFast Probe No-ROX one-step master mix, 0.2 μl Reverse Transcriptase, 0.4 μl RiboSafe RNase inhibitor mix and 0.08 μl of 25 μM gene specific probe. Detection of vRNA or mRNA/cRNA was facilitated by addition of 0.8 μl of 10 μM gene-specific forward or reverse primer, respectively. The RT reaction was incubated at 45°C for 10 min. Following this step, 0.8 μl of 10 μM of the opposite primer was added. In all qRT-PCR assays, serially diluted plasmids of corresponding influenza genes with known copy number were used as standards for quantification.

Viral Protein Quantification by Slot Blot Analysis

The relative protein content of virus in allantoic fluid was determined by staining for viral proteins in native conformation using a slot blot apparatus. Viral samples were adsorbed to nitrocellulose using a vacuum pump. The nitrocellulose strips were blocked in PBS with casein for 30 min at room temperature and then probed for 1 h with monoclonal antibodies (mAb) recognising H3 HA [clone 36/2 (Brown et al., 1990)] or IAV M1matrix protein (clone MCA401, BioRad). Strips were washed and incubated for 1 h with a horse radish peroxidase (HRP)-conjugated secondary immunoglobulin (DAKO). The strips were washed and immersed in citrate-EDTA before addition of the TMB substrate for 1 hour. Strips were then washed, dried overnight and digitised. The signal intensities for the viral HA

and M1 protein bands were determined using ImageJ analysis software (NIH, Bethesda, MD, United States).

Viral Protein Quantification by Flow Cytometry

The relative viral protein content within infected MDCK cells was determined by flow cytometry. Infected MDCK monolayers were harvested 6 h after 1-h absorption period. Cells were fixed, permeabilised and incubated with either clone 36/2 recognising the HA or clone MCA401 recognising M1 for 30 min. Cells were washed and incubated with an Alexa Fluor 488-conjugated rabbit anti-mouse immunoglobulin (Life Technologies) for 30 min. Data were acquired using a FACSCanto II and analysed with FlowJo X version 10 (Tree Star, Ashland, OR, United States).

Statistical Analysis

All of the data were analysed using GraphPad Prism version 8.4.2 (GraphPad Software Inc., La Jolla, CA, United States). Unless otherwise stated, the data were analysed for statistical significance by either a one-way analysis of variance (ANOVA) with a Tukey post-test or a two-way ANOVA with a Sidak post-test. A probability (p) value of <0.05 was considered to be statistically significant.

RESULTS

Classical Reassortment Generates a Diverse Range of Gene Constellations

To examine the spectrum of reassortants generated through the initial stages of the classical reassortment process (**Figure 1A**), the model seasonal strain Udorn (H3N2) was reassorted with the highly egg-adapted PR8 strain (H1N1) in four different co-infection conditions representing independent experiments (**Figure 1B**). Throughout the process, allantoic fluid from individual eggs was screened to detect haemagglutination and the presence of the Udorn HA and NA genes. Viruses were then isolated by plaque assay and the origin of each gene determined by gene-specific RT-PCR.

After the initial co-infection step, the progeny viruses were isolated by plaque formation in the presence of anti-PR8 antisera, which effectively prevented residual PR8 parent virus or viruses with PR8 surface antigens from dominating the plaque assay (**Figures 1C,D**). Of the remaining viruses present after co-infection, cumulative data across the four co-infection conditions revealed that 33% of the non-neutralised viruses were reassortants and 67% were the Udorn parent (**Figure 1C**). Consistent with what is observed for a seasonal influenza virus, the Udorn parent was represented least in the E2 series where the co-infection ratio favoured PR8 over Udorn by 100-fold (**Figure 1D**). The percentage of reassortants continued to increase throughout the process, with 91.5% of the viruses being reassortants after the first antibody passage and 99.7% after the second antibody passage (**Figures 1C,E,F**), indicating strong selective pressure for reassortment.

Figure 2A illustrates the complexity of the different genotypes present in the resulting reassortant pool (detailed in the **Supplementary Table 1**). After the initial co-infection, 63 discrete gene constellations were present and after two additional rounds of replication and antibody selection, this reduced to 39 distinct genotypes. Depending on the co-infection ratio and input dose, differences were observed in the gene constellations detected or their frequencies in each reassortment series (**Figure 2** and **Supplementary Table 1**). In most cases, the dominant reassortant detected in the first and second antiserum passages was not identical. In the E4 series, the dominant reassortant of the second antiserum passage represented between 40 and 53% of the total reassortants isolated and yet only represented approximately 10% of the reassortants in the first antiserum passage. When considering reassortants incorporating the PB1 gene from the Udorn parent (**Figure 2B**), it was clear that these dominated after the initial co-infection and after the second round of antiserum selection for the Udorn HA and NA. Overall, 85.5% (cumulative data from all series) contained the Udorn PB1 after the initial co-infection and 88.5% after two antiserum passages. In the E3 series 100% of the reassortants contained the Udorn PB1 after the first antiserum passage. Even in the E2 series where the co-infection ratio favoured PR8 over Udorn by 100-fold, 90% of isolates contained the Udorn PB1 after the second antiserum passage.

The cumulative frequency of other non-HA and NA genes in the reassortants in each series at the different stages is shown for comparison in **Figure 3**. The pattern observed with Udorn PB1 was closely reflected by that of the Udorn NP, which dominated after both the initial co-infection and second antiserum passage, present in 81 and 81.3% of genotypes, respectively (**Figure 3D**). In contrast, the PA and M genes of the PR8 virus dominated by the second antiserum passage, present in 84.5 and 99.5%, respectively of the gene constellations isolated (**Figures 3C,E**). Only the PB2 and NS genes appeared to be greatly influenced by the initial coinfection ratios, with the E3 series having the respective Udorn genes dominating (**Figures 3A,F**). Three reassortant genotypes, PR8(Ud-HA,NA,PB2,PB1,NP,NS), PR8(Ud-HA,NA,PB2,PB1,NP), and PR8(Ud-HA,NA,PB1,NP), all containing Udorn PB1 and NP in addition to the HA and NA genes, occurred most frequently in the final reassortant populations and were detected in all the reassortment series.

Reassortants That Produce High Haemagglutination Titres in Eggs Do Not Always Have a High Yield of Infectious Virus

Use of limit dilution to obtain the most abundant reassortants from each egg after the second antiserum passage yielded 11 different viruses, 9 of which incorporated Udorn NA and PB1 genes. Although PR8(Ud-HA,NA), which is expected to be the high HA yielding vaccine seed virus in our system, was present in some eggs after the second passage, it was not the dominant species in these eggs (**Supplementary Table 1**) and not isolated by limit dilution. The surrogate virus, rg PR8(Ud-HA,NA), was therefore made by reverse genetics as a comparator. Eight of the 11 reassortants had significantly

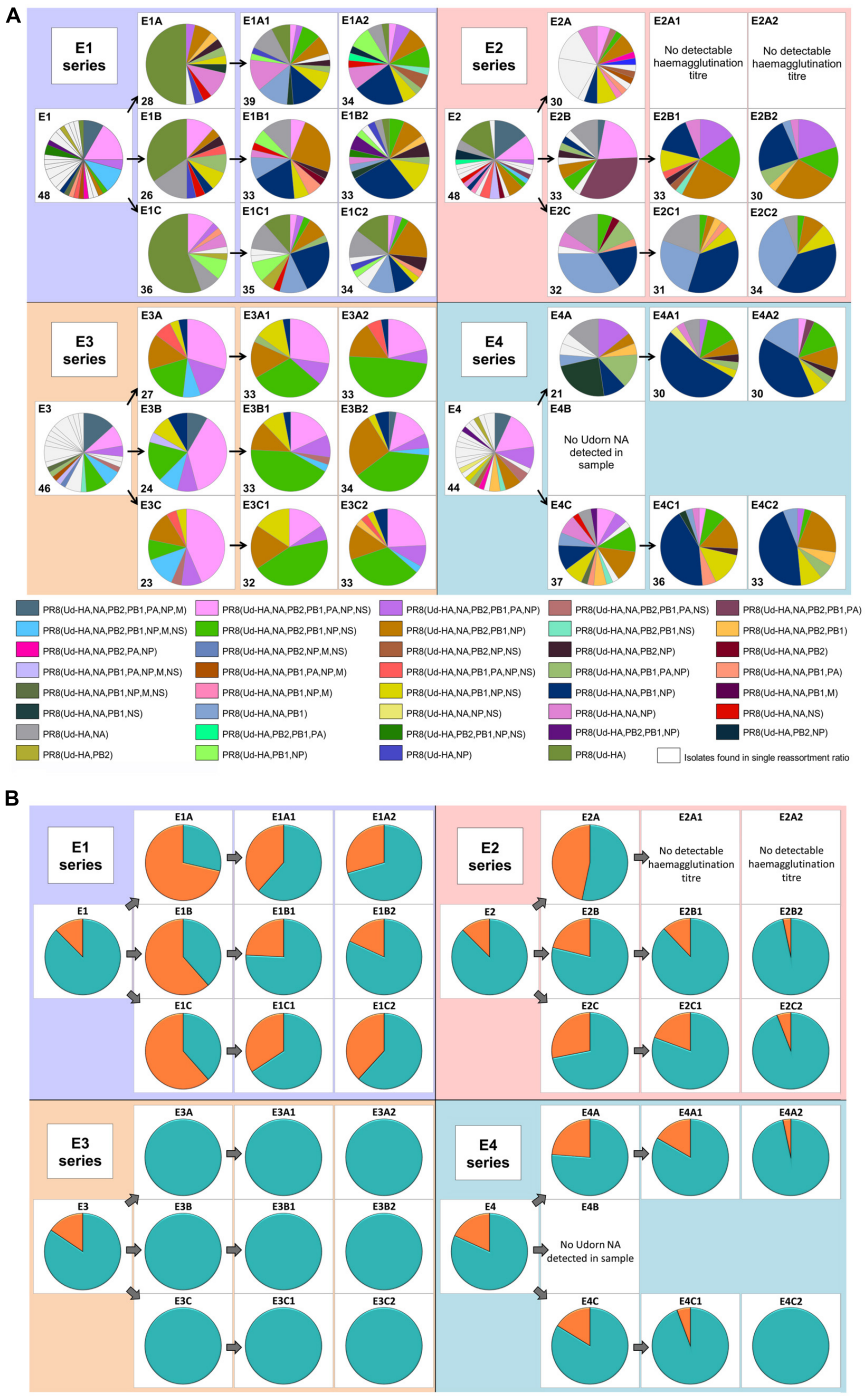


FIGURE 2 | The diversity of reassortant viruses isolated during the reassortment process. Viruses from the allantoic fluid of each egg were isolated by plaque assay. For the co-infected eggs, plaquing was performed in the presence of PR8 antiserum. The genotype of virus in individual plaques was determined using gene specific RT-PCR. Each of the four larger panels represents eggs that originated from the different co-infection ratios (**Figure 1B**). The pie charts represent the viruses isolated in the individual eggs corresponding to the co-infection and antiserum passages (**Figure 1A**). The dominant gene constellations present in the individual eggs are represented according to the colours in the key. The nomenclature used indicates the Udorn genes on the PR8 background, e.g., PR8(Ud-HA) is a virus with the HA gene from Udorn and the remaining seven genes from PR8. White segments represent gene constellations that were only present in that particular egg. No data is available for eggs E2A1 and E2A2 as the allantoic fluids did not have detectable haemagglutination titres and were not examined further. Initial screening of egg E4B showed that Udorn NA was not present so complete genotypes were not analysed. The number of individual plaques analysed for each egg is indicated in the bottom left of each pie chart panel. **(B)** The frequency of the Udorn PB1 gene (teal) and PR8 PB1 gene (orange) in the genotypes of the individual virus plaques shown in panel **(A)**.

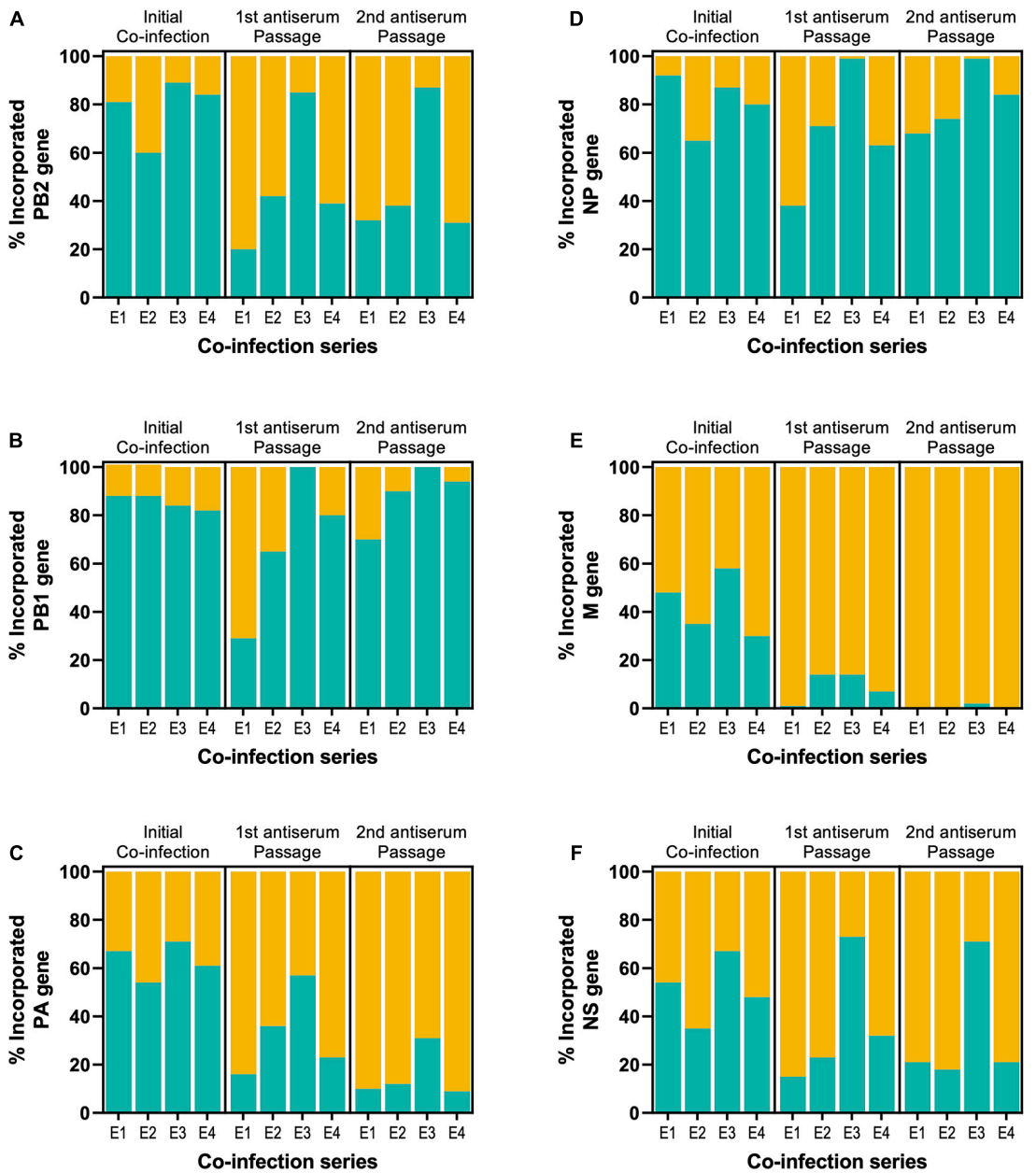


FIGURE 3 | Tracking the frequency of the six internal Udorn (teal) and PR8 (orange) genes throughout the reassortment process. The source of individual genes in the reassortants shown in **Figure 2** is summarised for each series. Data are expressed as a percentage of the total viruses isolated in that series at each stage of the reassortment process, with replicate eggs pooled for the first (two eggs) and second (four eggs) antisera passages. The origin of the **(A)** PB2, **(B)** PB1, **(C)** PA, **(D)** NP, **(E)** M, and **(F)** NS genes were determined to be either from the PR8 parent (orange) or the Udorn parent (teal).

higher haemagglutination titres than the parent Udorn virus (**Figure 4A**). Four, namely R9 PR8(Ud-HA,NA,PB1), R4 PR8(Ud-HA,NA,PB2,PB1), R8 PR8(Ud-HA,NA,PB1,NP), and R3 PR8(Ud-HA,NA,PB2,PB1,NP) were not significantly different to PR8 virus, illustrating the power of reassortment to enhance the yield of HA from seasonal isolates. To establish whether a greater replicative capacity was responsible for selection of these reassortants, infectious viral titres were determined by plaque assay in MDCK cells (**Figure 4B**). Compared to Udorn virus, all but two reassortants had higher infectious virus titres ($p < 0.0001$; one-way ANOVA), with six reassortants having equivalent titres to PR8 virus ($p > 0.05$; one-way ANOVA). The rgPR8(Ud-HA,NA) virus also had an infectious virus titre that was comparable to PR8 virus ($p > 0.05$; one-way ANOVA) and significantly greater than that of Udorn virus and five of the reassortants ($p < 0.0001$; one-way ANOVA). These data indicate that the absence of PR8(Ud-HA,NA) virus from the final dominant reassortant progeny was not due

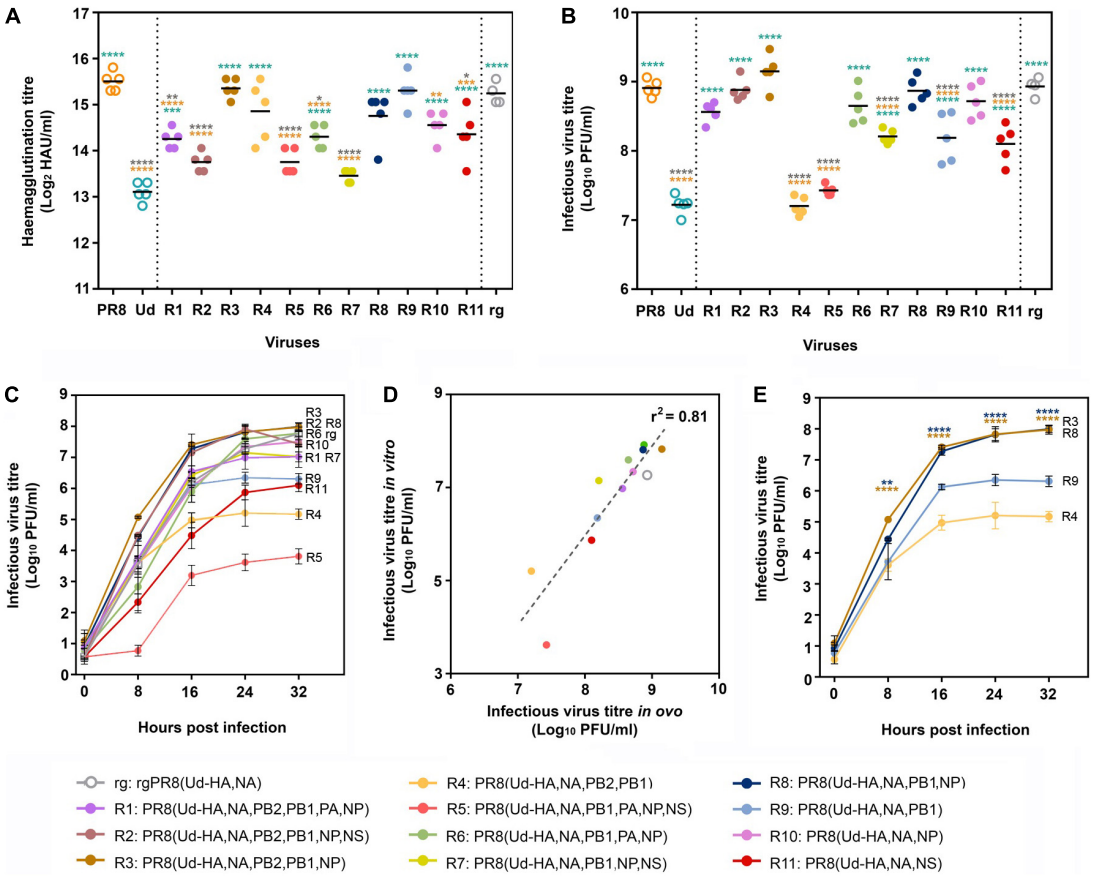


FIGURE 4 | Haemagglutination titres and replicative fitness of dominant reassortant progeny isolated through classical reassortment. **(A,B)** Eggs were infected with a standard dose (100 PFU) of one of the 11 reassortant viruses (R1–R11) isolated after limit dilution ($n = 5$), the parental PR8 and Udorn (Ud) viruses ($n = 5$) or the reverse genetics-derived virus rgPR8(Ud-HA,NA) ($n = 4$). Virus genotypes are shown in the key by number and colour. Three days post infection, allantoic fluid was harvested and **(A)** haemagglutination titres and **(B)** infectious viral titres determined. Each symbol represents the titre of virus from an individual egg and the line represents the geometric mean. Statistical significance was determined by one-way ANOVA with Tukey's multiple comparisons test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$) for panel **(A)** ($F_{13,55} = 23.39$) and **(B)** ($F_{13,55} = 61.78$). Orange, teal, and grey asterisk indicate comparisons of viruses with PR8, Udorn, and rgPR8(Ud-HA,NA), respectively. **(C)** MDCK cells were infected with the panel of reassortant viruses and rgPR8(Ud-HA,NA) at an MOI of 0.001. At the specified time points post infection, virus titres in the supernatants were determined by plaque assay on confluent MDCK cell monolayers. The data represents the mean and standard deviation of three individual experiments. **(D)** Infectious yields *in vitro* **(C)** at 24 hr as a function of the infectious viral titres *in ovo* **(B)**. The dotted line represents a linear regression analysis ($p < 0.0001$, $r^2 = 0.81$). **(E)** The data from (C) pertaining only to the reassortants with haemagglutination titres equivalent to PR8 *in ovo* **(A)** are shown for clarity. The data represents the mean and standard deviation of three individual experiments. Statistical significance was determined by two-way ANOVA with Sidak's multiple comparisons test ($F_{3,8} = 97.28$, **$p < 0.01$, ****$p < 0.0001$). Brown asterisk indicate comparisons between PR8(Ud-HA,NA,PB2,PB1,NP) with PR8(Ud-HA,NA,PB2,PB1) and blue asterisk indicate comparisons between PR8(Ud-HA,NA,PB1,NP) with PR8(Ud-HA,NA,PB1).

to a reduced replicative capacity. Of the four reassortants with haemagglutination titres similar to that of PR8 virus, R8 PR8(Ud-HA,NA,PB1,NP) and R3 PR8(Ud-HA,NA,PB2,PB1,NP) viruses also replicated to the same extent as the PR8 virus ($p > 0.05$; one-way ANOVA) yet the two other reassortants, R9 PR8(Ud-HA,NA,PB1) and R4 PR8(Ud-HA,NA,PB2,PB1), did not ($p < 0.0001$; one-way ANOVA). These data indicate that the enhanced haemagglutination phenotype for some reassortants must be due to mechanisms other than the acquisition of greater replication capacity.

To confirm the phenotypes of the viruses observed in eggs, MDCK monolayers were infected with a standardised dose of the different viruses and viral loads within cultures determined over time (**Figure 4C**). Infectious virus yields at the plateau of the replication curve in MDCK cells mirrored the hierarchy found in embryonated eggs as shown by regression analysis of the titres in MDCK cells at 24 h post infection versus those in eggs ($p < 0.0001$, $r^2 = 0.81$; **Figure 4D**). Of the four reassortants with haemagglutination titres not significantly different to PR8 virus, PR8(Ud-HA,NA,PB1) had an approximately 1.5 log lower plateau titre than the corresponding virus with the Udorn NP, PR8(Ud-HA,NA,PB1,NP) ($p > 0.0001$; two-way ANOVA; **Figure 4E**). Likewise, PR8(Ud-HA,NA,PB2,PB1) had a 2.5 log lower titre compared to PR8(Ud-HA,NA,PB2,PB1,NP) ($p > 0.0001$; two-way ANOVA; **Figure 4E**). Therefore, the addition of the Udorn NP dramatically improved the infectious virus yield obtained when the Udorn HA, NA and PB1 were present in a virion, despite no difference in haemagglutination capacity.

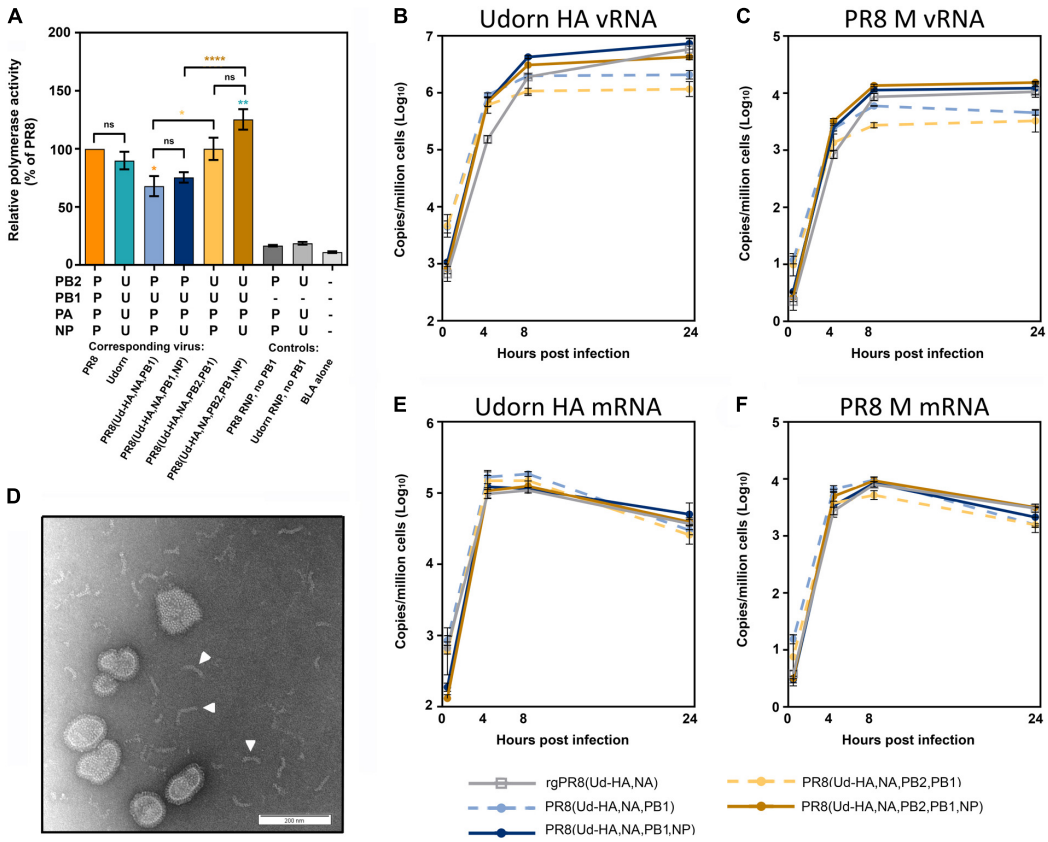


FIGURE 5 | The relative polymerase activity and viral RNA production of the reassortants with high haemagglutination titres. **(A)** A β-lactamase (BLA) reporter assay was performed in HEK 293T cells that were transfected with the pCAGGS-BLA reporter gene and four pHW2000 plasmids coding for the indicated PB2, PB1, PA, and NP genes, which correspond to the RNPs of the indicated viruses or controls without the PB1-encoding plasmid. The relative polymerase activities were normalised to the PR8 RNP activity and each bar represents the mean and standard error of four experiments at the half max of each curve. Statistical significance was determined by one-way ANOVA with Tukey's multiple comparison test ($F_{8,27} = 49.28$, ns $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ****$p < 0.0001$). Orange, teal, brown, and yellow asterisk indicate comparisons between the RNP complexes with $P_{PB2}P_{PB1}P_{PA}P_{NP}$, $U_{PB2}U_{PB1}U_{PA}U_{NP}$, $U_{PB2}U_{PB1}P_{PA}U_{NP}$ and $U_{PB2}U_{PB1}P_{PA}P_{NP}$, respectively. **(B,C,E,F)** MDCK cells were infected with the indicated virus at an MOI of 3 and total RNA extracted from 1×10^6 cells at 0, 4, 8, and 24 h after a 1 h virus absorption period. The copy numbers of **(B)** Udorn HA vRNA, **(C)** PR8 M vRNA, **(E)** Udorn HA mRNA/cRNA, and **(F)** PR8 M mRNA/cRNA were assessed by quantitative RT-PCR and the data represents the mean and standard deviation ($n = 3$) and is representative of four experiments. **(D)** The presence of free RNP complexes in the allantoic fluid of PR8(Ud-HA,NA,PB1,NP) and PR8(Ud-HA,NA,PB2,PB1,NP) were visualised by negative staining transmission electron microscopy. A representative image is shown with some of the free RNP complexes indicated by a white arrowhead.

Reduced Replicative Capacities of PR8(Ud-HA,NA,PB2,PB1) and PR8(Ud-HA,NA,PB1) Viruses Are Not Due to Low Polymerase Activity

It has been proposed that differences in polymerase activity can account for differences in viral replication (Li et al., 2008; Octaviani et al., 2011; Nakazono et al., 2012; Hara et al., 2013). To investigate this, a BLA reporter assay was performed in cells co-transfected with the different combinations of viral polymerase and NP genes represented in the viruses under study (**Figure 5A**). Although infection with PR8(Ud-HA,NA,PB1) and PR8(Ud-HA,NA,PB2,PB1) viruses had resulted in lower infectious virus yields compared to the corresponding viruses with Udorn NP (**Figure 4E**), their respective RNP complexes ($P_{PB2}U_{PB1}P_{PA}P_{NP}$ and $U_{PB2}U_{PB1}P_{PA}P_{NP}$), demonstrated no significant difference in activity ($p > 0.05$; one-way ANOVA; **Figure 5A**) to complexes with the Udorn NP ($P_{PB2}U_{PB1}P_{PA}U_{NP}$ and $U_{PB2}U_{PB1}P_{PA}U_{NP}$). Substitution of PR8 PB2 with Udorn PB2 significantly increased the polymerase activity of $P_{PB2}U_{PB1}P_{PA}P_{NP}$ ($p < 0.01$, one-way ANOVA) and $P_{PB2}U_{PB1}P_{PA}U_{NP}$ ($p < 0.001$, one-way ANOVA), yet no increase in infectious titres was observed for the corresponding viruses both *in vitro* and *in vivo*. These data demonstrate that the viral polymerase activity, as measured by the BLA assay, does not correlate with the efficiency of replication of viruses with these different gene constellations.

The Addition of the Udorn NP to PR8(Ud-HA,NA,PB1) and PR8(Ud-HA,NA,PB2,PB1) Results in Increased vRNA Production

The amount of Udorn HA vRNA (**Figure 5B**) and PR8 M vRNA (**Figure 5C**), which are common to all the reassortants,

were assessed in infected MDCK cells via quantitative RT-PCR. The presence of the Udorn PB1± the Udorn PB2 in a PR8(Ud-HA,NA) background, resulted in significantly reduced levels of HA and M vRNA at 24 h post-infection compared to the rgPR8(Ud-HA,NA) virus ($p < 0.001$ and $p < 0.0001$, respectively; two-way ANOVA). Upon the inclusion of the Udorn NP to the PR8(Ud-HA,NA,PB1) and PR8(Ud-HA,NA,PB2,PB1) backgrounds, the HA and M vRNA levels were restored to the levels seen in rgPR8(Ud-HA,NA) at 24 h post infection. The addition of the Udorn NP to PR8(Ud-HA,NA,PB1) virus increased the amount of HA vRNA at the 8 and 24 h time points ($p < 0.05$, $p < 0.0001$ respectively; two-way ANOVA) and also the levels of M vRNA at the 8 and 24 h time points ($p < 0.01$, $p < 0.0001$ respectively; two-way ANOVA). Similarly, PR8(Ud-HA,NA,PB2,PB1,NP) displayed significantly higher levels of HA vRNA than PR8(Ud-HA,NA,PB2,PB1) at the 8 and 24 h time points ($p < 0.001$, $p < 0.0001$ respectively; two-way ANOVA) and M vRNA at the 4, 8, and 24 h time points ($p < 0.001$, $p < 0.0001$, $p < 0.0001$, respectively; two-way ANOVA). These data suggest that the higher levels of vRNA produced by viruses with Udorn NP, in addition to Udorn HA, NA, PB1 ± PB2, may contribute to the greater infectious particle yields of these viruses compared to the corresponding reassortants with PR8 NP. That said, we observed by electron microscopy of the virus-containing allantoic fluid preparations of PR8(Ud-HA,NA,PB1,NP) and PR8(Ud-HA,NA,PB2,PB1,NP) that free RNPs were present in these (**Figure 5D**) but not in the corresponding reassortants with PR8 NP, suggesting the possibility that the greater levels of vRNA produced by these Udorn NP-containing viruses may not be all incorporated into virions.

High Haemagglutination Titres Did Not Result From Increased HA Protein in the Virion Due to Increased Transcription or Translation of HA in Infected Cells

It was possible that the lower levels of HA and M vRNA production observed for PR8(Ud-HA,NA,PB2,PB1) and PR8(Ud-HA,NA,PB1) compared to the corresponding viruses with Udorn NP, might be due to the preferential production of viral mRNA over vRNA. To examine this, the level of Udorn HA and PR8 M viral mRNA transcription was quantified in infected MDCK cells by gene-specific qRT-PCR (**Figures 5E,F**). Comparison between rgPR8(Ud-HA,NA) and the four reassortant viruses showed no difference in viral mRNA production for either the Udorn HA or PR8 M genes at any time point ($p > 0.05$, two-way ANOVA).

To examine whether selective protein modulation was occurring in our system, the amount of HA present in infected MDCK cells was assessed by flow cytometry 6 h post infection (**Figure 6A**). Matrix protein was also analysed to determine whether any potential modulation of Udorn HA translation was specific for the HA (**Figure 6B**). The data showed a trend towards slightly more HA produced in cells infected with rgPR8(Ud-HA,NA) than with PR8(Ud-HA,NA,PB2,PB1) and PR8(Ud-HA,NA,PB2,PB1,NP) ($p < 0.05$; one-way ANOVA). However, the levels of M1 protein expression between infected cells differed considerably with rgPR8(Ud-HA,NA)-infected cells producing more of this protein than the reassortants ($p < 0.0001$; one-way ANOVA). When the ratio of HA staining to M1 staining was calculated for each allantoic fluid and the means of the different reassortant groups expressed relative to rgPR8(Ud-HA,NA) (**Figure 6** key), the fold difference was less than 1.3 suggesting only minor, if any, effects on protein expression specific for HA in infected cells.

An Increased Number of Non-infectious Particles and a Higher HA Density Contributes to the High HAU

To further examine what contributes to the high haemagglutination titres of the reassortants we required a measure of relative particle number. Usually this is done by some measurement of the RNA, however, we had observed that the Udorn NP-containing reassortants had free RNPs (**Figure 5F**) that would make such measurement misleading. Instead, an analysis of HA protein relative to M1 protein content of the different viruses was performed. As virion size and morphology were similar between the viruses when observed via electron microscopy (**Figure 7A**), we made the assumption that the amount of matrix protein (M1) present in the viral particles in the different preparations would be approximately equivalent on average and correlate directly with particle number in the preparation.

The amounts of HA and M1 in each of the reassortants were determined by the binding of specific monoclonal antibodies using a slot blot assay (**Figures 7B,C**). Compared to rgPR8(Ud-HA,NA), less infectious virus was required to provide the same amount of HA staining for each of the reassortants (**Figure 7B**), particularly for PR8(Ud-HA,NA,PB2,PB1) where only about one-hundredth of the infectious virus was required (**Table 1**). With the exception of PR8(Ud-HA,NA,PB2,PB1,NP), less infectious virus was also required to provide the same amount of PR8 M1 staining compared to rgPR8(Ud-HA,NA) (**Figure 7C**). By dividing the relative PFU/M1 unit by the relative PFU/HA unit, the resultant HA/M1 ratio indicates the relative amount of HA protein present per arbitrary amount of M1 protein, which, given the assumption of a correlation between M1 and virions, reflects the relative abundance of HA per virion (**Table 1**). The 3.9-fold greater HA protein density per PR8(Ud-HA,NA,PB1) particle compared to rgPR8(Ud-HA,NA) supports the findings of Cobbin et al. (2013), who also identified a four-fold increase in HA per viral particle for rgPR8(Ud-HA,NA,PB1) by western blotting. The substitution of Udorn NP for PR8 NP in the PR8(Ud-HA,NA,PB1) virus resulted in an 8-fold decrease in HA density from 3.9 to 0.5 HA/M1. The PR8(Ud-HA,NA,PB2,PB1) virus was similar to rgPR8(Ud-HA,NA) with 1.2 HA/M1 and, in this case, the presence of the Udorn NP led to an increase in HA density to 4 HA/M1. Thus, changes in the gene constellation of the virus can dramatically alter the amount of HA on the virion surface.

Data from **Table 1**, taken in conjunction with the infectious virus titres from the allantoic fluid of eggs infected with a standard amount of each of the viruses (**Figure 4B**), allows an estimation of how many total virus particles are predicted to

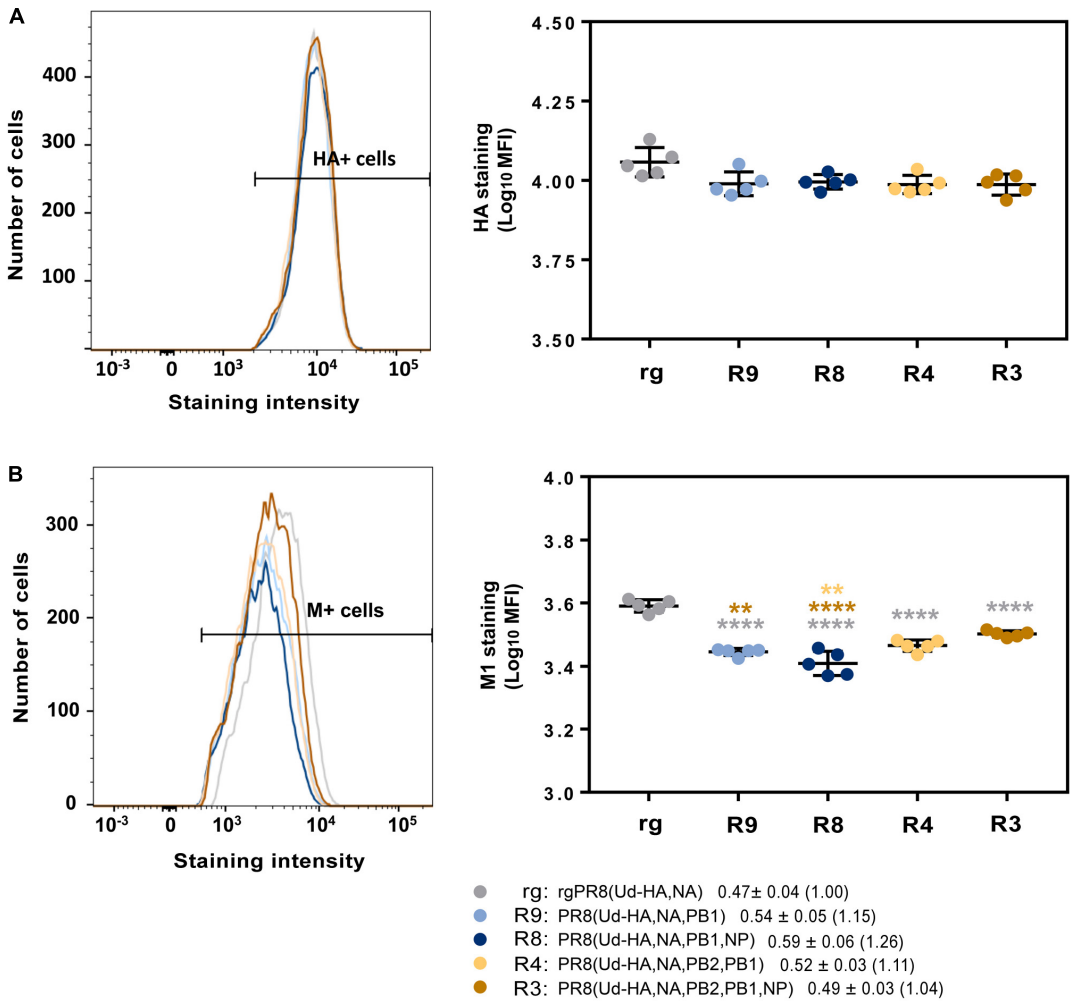


FIGURE 6 | Flow cytometric analysis of Udorn HA and PR8 M1 protein production in infected cells. MDCK cells (1×10^6) were infected with different viruses at an MOI of 3 and protein production assessed 6 h after a 1 h virus absorption period. Cells were stained with either **(A)** anti-Udorn HA (AF647) or **(B)** anti-M1 (MCA401) followed by a secondary antibody coupled to a fluorescent dye for the detection of proteins. Left panels show example flow cytometry histograms of stained cells with the fluorescence range for positive cells indicated; right panels show individual data for the mean fluorescent intensity (MFI) from five replicate cultures and is representative of three experiments. Statistical significance was determined by one-way ANOVA with Sidak's multiple comparison test (**$p < 0.01$, ****$p < 0.0001$) for (A) ($F_{4,20} = 3.24$) and **(B)** ($F_{4,20} = 45.59$). Grey, brown and yellow asterisks indicate comparison of viruses with rgPR8(Ud-HA,NA), PR8(Ud-HA,NA,PB2,PB1,NP), and PR8(Ud-HA,NA,PB2,PB1) respectively. Values in the key represent the mean and SD of the ratios of HA staining to M1 staining for each individual allantoic fluid within each group with the fold difference compared to rgPR8(Ud-HA,NA) in brackets.

be present and the relative total yield of HA from these eggs (**Table 2**). As an example of these calculations and the underlying assumptions, for a given amount of M1 protein, which we equate to a given amount of virus particles, PR8(Ud-HA,NA,PB1) has only half the number of infectious particles as rgPR8(Ud-HA,NA) (**Table 1**, relative PFU/M1 unit). To provide a baseline to calculate relative particle numbers of the reassortants we assume that all the virions in the reverse engineered rgPR8(Ud-HA,NA) are infectious. So for a given amount of infectious PR8(Ud-HA,NA,PB1) virions there is an equivalent number of non-infectious virions contributing to the amount of HA in the egg. Dividing the observed titre of infectious virions by the relative PFU/M1 unit ratio for each virus (**Table 1**) provides the predicted total number of virions in the egg in M1 units/ml (**Table 2**, predicted total virions). In this way, the PR8(Ud-HA,NA,PB1) virus is predicted to have a total virion count of 5.6×10^8 M1 units/ml. As each particle had 3.9 times the amount of HA than did rgPR8(Ud-HA,NA) (**Table 1** relative HA/M1), which we designate as having 1 arbitrary unit of HA per virion, the predicted total yield of HA in the PR8(Ud-HA,NA,PB1)-infected eggs is 2.2×10^9 HA units/ml (M1 units/ml × HA/M1) which is 2.5-fold greater than rgPR8(Ud-HA,NA) (**Table 2**, relative HA yield). Likewise, PR8(Ud-HA,NA,PB1,NP) had a 1.5-fold greater overall yield of HA, PR8(Ud-HA,NA,PB2,PB1) had 2.3-fold and PR8(Ud-HA,NA,PB2,PB1,NP) had 5.3-fold greater HA yield than did rgPR8(Ud-HA,NA) (**Table 2**).

The relationship between HA density per particle, the number of virus particles and the capacity to haemagglutinate chicken

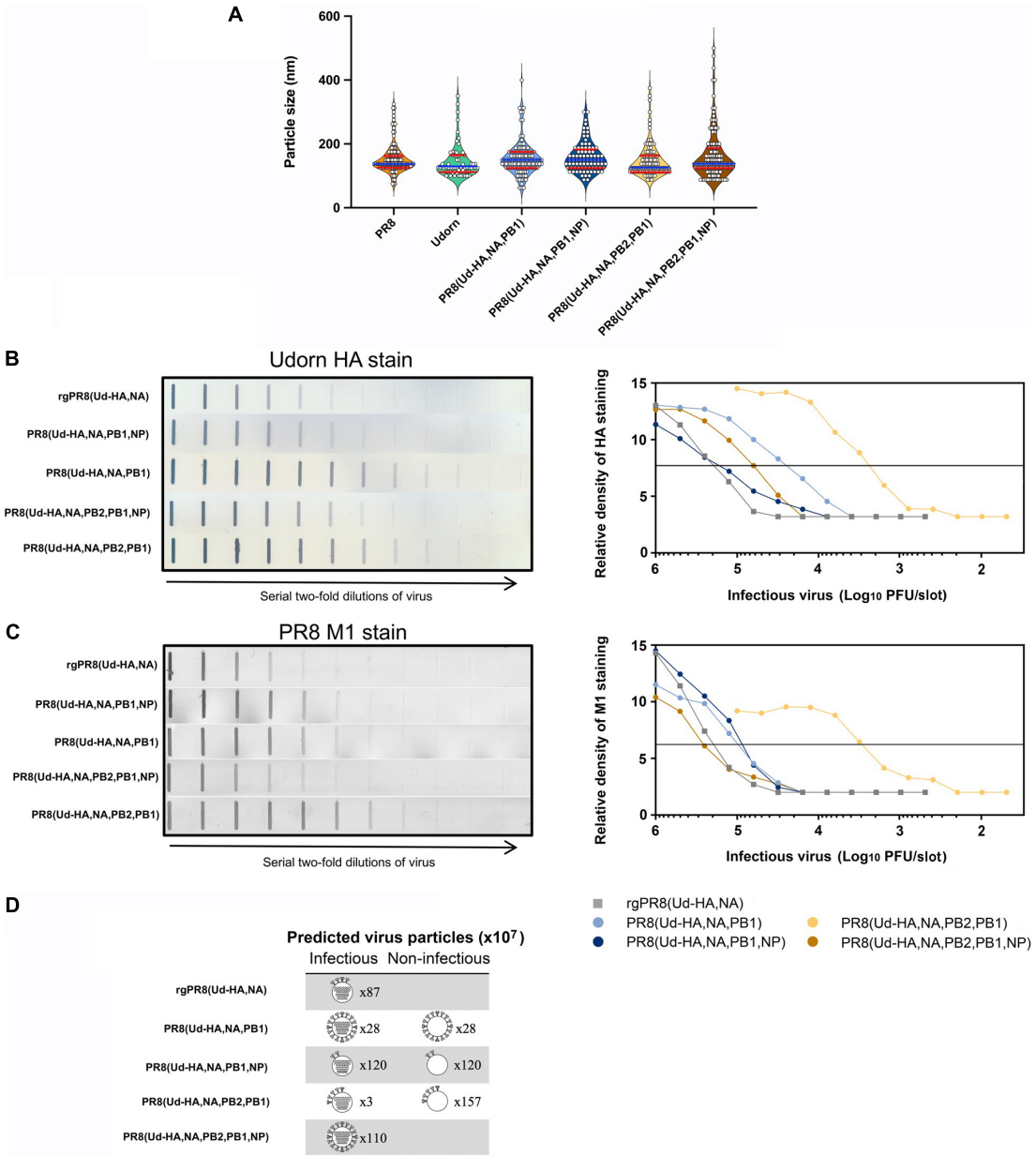


FIGURE 7 | The structure and HA protein content of reassortant viruses. **(A)** Violin plots of particle sizes visualised by transmission electron microscopy of infected allantoic fluid. White filled circles represent individual measurements along the longest axis of 60–130 distinctive viral particles from one to three printed images per virus preparation converted to nm. The median size is shown as a blue line with quartiles shown as red lines. Analysis by one-way ANOVA shows no difference in the mean particle size (p = 0.27) **(B,C)** Allantoic fluid containing infectious virus was diluted to 10^7 PFU/ml, or 10^6 PFU/ml for PR8(Ud-HA,NA,PB2,PB1), and two-fold serial dilutions were performed. Each dilution (100 μl) was transferred to a nitrocellulose membrane and the **(B)** Udorn HA protein or **(C)** PR8 M1 protein was detected with a mouse monoclonal antibody and visualised by secondary staining with rabbit anti-mouse HRP and the addition of substrate (left panels). Densitometry analysis was performed on the bands obtained in each of the strips and the peak heights were plotted against the amount of infectious virus in that sample (right panels). The horizontal line is used to determine the number of PFU required to provide an arbitrary amount of HA or M1 staining. The data is representative of two experiments. **(D)** Depiction of the relative amount of HA and its distribution between infectious and non-infectious virions. Infectious particles are depicted with gene segments; non-infectious particles are depicted without gene segments. The predicted number of particles ($\times 10^7$) is indicated beside each virion (**Table 2**). For ease of representation one arbitrary unit of HA protein (**Table 1**) is depicted as four HA spikes on the virion. The distribution of the HA protein was assumed to be equal between the infectious and non-infectious particles.

erythrocytes is yet to be completely understood. Nevertheless, together these data provide a possible explanation of how high HA yields, predicted from HAU titres, could be achieved by PR8(Ud-HA,NA,PB1) and PR8(Ud-HA,NA,PB2,PB1) despite their poor replicative capacity. This work highlights how different mechanisms operate to achieve high HA yields (**Figure 7D**)

TABLE 1 | Amount of infectious virus required to achieve an equivalent level of HA or M1 staining and relative abundance of HA per virion.

Virus	PFU/HA unit[a]	Relative PFU/HA unit[b]	PFU/M1 unit[c]	Relative PFU/M1 unit[d]	Relative HA/M1[e]
rgPR8(Ud-HA,NA)	1.8×10^5	1.00	2.0×10^5	1.00	1.0
PR8(Ud-HA,NA,PB1)	2.3×10^4	0.13	9.8×10^4	0.50	3.9
PR8(Ud-HA,NA,PB1,NP)	1.5×10^5	0.80	8.6×10^4	0.44	0.5
PR8(Ud-HA,NA,PB2,PB1)	2.3×10^3	0.012	2.9×10^3	0.015	1.2
PR8(Ud-HA,NA,PB2,PB1,NP)	6.0×10^4	0.33	2.6×10^5	1.33	4.0

[a] *From slot blot analysis, the infectious virus (PFU) corresponding to an arbitrarily level (referred to as 1 unit) of HA staining (read at the line indicated in* ***Figure 6B*** *graph).*
[b] *PFU/HA unit of reassortant expressed relative to rgPR8(Ud-HA,NA).*
[c] *From slot blot analysis, the infectious virus (PFU) corresponding to an arbitrarily level (referred to as 1 unit) of M1 staining (read at the line indicated in* ***Figure 6C*** *graph).*
[d] *PFU/M1 unit of reassortant expressed relative to rgPR8(Ud-HA,NA).*
[e] *Calculated as PFU/M1 units divided by PFU/HA units and corresponds to the relative amount of HA per virion assuming that the amount of M1 is proportional to the total number of virions and is present in similar amounts on average in infectious and non-infectious virus.*

TABLE 2 | Predicted HA distribution on infectious and non-infectious particles.

Virus	Infectious virus titre (PFU/ml)[a]	Predicted total virions (M1 units/ml)[b]	Predicted total HA yield (HA units/ml)[c]	Relative HA yield[d]
rgPR8(Ud-HA,NA)	$8.7 \pm 1.8 \times 10^8$	$8.7 \pm 1.8 \times 10^8$	$8.7 \pm 1.8 \times 10^8$	1.0
PR8(Ud-HA,NA,PB1)	$2.8 \pm 1.8 \times 10^8$	$5.6 \pm 3.6 \times 10^8$	$2.2 \pm 1.4 \times 10^9$	2.5
PR8(Ud-HA,NA,PB1,NP)	$1.1 \pm 0.5 \times 10^9$	$2.4 \pm 0.1 \times 10^9$	$1.3 \pm 0.6 \times 10^9$	1.4
PR8(Ud-HA,NA,PB2,PB1)	$2.4 \pm 0.7 \times 10^7$	$1.6 \pm 0.5 \times 10^9$	$2.0 \pm 0.6 \times 10^9$	2.2
PR8(Ud-HA,NA,PB2,PB1,NP)	$1.5 \pm 0.2 \times 10^9$	$1.1 \pm 0.1 \times 10^9$	$4.6 \pm 0.5 \times 10^9$	5.2

[a] *Infectious titres determined by plaque assay on allantoic fluid from eggs infected with a standard dose of virus (****Figure 3B****). Mean ± SD.*
[b] *Observed infectious virus titre (PFU/ml) divided by relative PFU/M1 unit (****Table 1****). Prediction of total virions assumes all rgPR8(Ud-HA,NA) virions are infectious.*
[c] *M1 units/ml multiplied by the relative HA/M1 (****Table 1****). Assumes rgPR8(Ud-HA,NA) has one arbitrary unit of HA per virion.*
[d] *Predicted yield of HA relative to rgPR8(Ud-HA,NA).*

in different reassortants. A higher HA density per virion was the main contributing factor for PR8(Ud-HA,NA,PB1) and PR8(Ud-HA,NA,PB2,PB1,NP) whereas a greater number of non-infectious particles allowed for the high HA yields of PR8(Ud-HA,NA,PB2,PB1) and PR8(Ud-HA,NA,PB1,NP).

DISCUSSION

In this study we utilised a model of vaccine seed production with the aim of assessing the influence of selective pressures on viral reassortment over multiple passages. Contrary to current thinking, we show that replicative fitness and antibody resistance are not the only determinants dictating which reassortant progeny will come to dominate in this system. We show data compatible with gene co-selection, resulting from dominant gene segment interactions, as a powerful force for shaping the viruses that are formed and we document the dynamics of the rise and fall of different constellations across the stages of the process. Though counter intuitive, we show that gene co-selection can lead to less fit viruses being positively selected to the point that they are isolated by limit dilution as dominant progeny. We also asked how these dominant but less fit viruses may display high antigen yields such that they can be chosen as vaccine seed candidates and reveal different phenotypes that are likely to account for this.

Generally, reassortment of H3N2 seasonal viruses with PR8 is thought to increase the yields of HA protein by generating reassortants that replicate to high titres (Gerdil, 2003) by virtue of their acquisition of a full complement of genes encoding PR8 non-surface antigens. However, from the four reassortants that expressed high haemagglutination titres equivalent to PR8, only PR8(Ud-HA,NA,PB1,NP) and PR8(Ud-HA,NA,PB2,PB1,NP) displayed high infectious virus yields that likely contributed to the high haemagglutination titres achieved. Comparison between the gene constellations of the four final "vaccine seed candidates" indicated that the presence of the Udorn NP significantly improved infectious virus yields, suggesting some incompatibility of Udorn HA, NA or PB1 with PR8 NP. Viral infectious yields are often associated with the function of the polymerase complex and the relative activity is believed to determine the replicative capacity of a virus (Li et al., 2008; Octaviani et al., 2011; Nakazono et al., 2012; Hara et al., 2013). However, the addition of the Udorn NP had no impact on polymerase activity in the reporter assay, demonstrating that polymerase activity is not necessarily indicative of the replicative ability of a virus. Despite no difference in polymerase activity, *in vitro* analysis of RNA production demonstrated that viruses with the Udorn NP had significantly increased levels of vRNA, but not viral mRNA, compared to the corresponding viruses with PR8 NP. The NP protein encapsidates vRNA and complementary cRNA but not mRNA and interacts directly with PB1 and PB2 (Biswas et al., 1998). In addition, regions in the NP have been identified as important for vRNA production (Mena et al., 1999; Li et al., 2009; Davis et al., 2017), for selective modulation of NA expression (Brooke et al., 2014) and for efficient packaging (Brooke et al., 2014; Moreira et al., 2016; Bolte et al., 2019). Although virus strain differences in these functions have not yet been investigated, it is possible that a mismatch of NP with PB1

or NA may impair correct packaging of segments resulting in the over-abundance of non-infectious particles in some reassortant genotypes, such as PR8(Ud-HA,NA,PB2,PB1) but not in the corresponding virus with Udorn NP. However it occurs, the restoration of infectious yields by the Udorn NP is possibly contributed to by the increased amounts of vRNA available for packaging into progeny virions to generate a greater number of infectious particles.

In this study, we provide a possible explanation for how dominant gene constellations could provide a high HA yielding phenotype. We show this may be as a result of either a high replicative capacity, the concomitant production of a large number of non-infectious particles, a greater density of HA per virion or a combination of these factors. PR8(Ud-HA,NA,PB1) and PR8(Ud-HA,NA,PB2,PB1,NP) had an overall increase of HA/particle compared to rgPR8(Ud-HA,NA) whereas PR8(Ud-HA,NA,PB1,NP) and PR8(Ud-HA,NA,PB2,PB1) achieved high haemagglutination titres mainly through an increase in the number of total viral particles. Although the calculations used to determine the relative numbers of infectious and non-infectious particles were based on a number of underlying assumptions, the observations on which these assumptions were made, such as infectious yield and the HA and M1 protein content, were all determined experimentally. Therefore, although the actual numbers are notional, relative differences between reassortants stem from empirical data. Previously in our laboratory it was demonstrated that the Udorn PB1 could selectively modulate Udorn HA protein production in cells (Cobbin et al., 2013). This as yet undefined mechanism, thought to operate post-transcriptionally (Cobbin et al., 2013), may be operating here to explain how PR8(Ud-HA,NA,PB1) and PR8(Ud-HA,NA,PB2,PB1) achieved high haemagglutination titres despite their reduced replication kinetics.

Until recently, viral replicative fitness was considered the main factor to drive the emergence of dominant gene constellations (Kimble et al., 2014; Steel and Lowen, 2014). Viruses with lower infectious yields would likely be outcompeted by genotypes with higher replication kinetics, which would then dominate the population. However, in this study, the isolation of six dominant reassortant gene constellations with infectious yields lower than rgPR8(Ud-HA,NA), but not PR8(Ud-HA,NA) itself, demonstrates that viral fitness doesn't always dictate dominance, at least not in the initial stages of the selection process. Interactions between viral gene segments are known to be important during the assembly and packaging of the eight RNPs into progeny virions (Fournier et al., 2012; Essere et al., 2013; Gavazzi et al., 2013a; Le Sage et al., 2020) and the data shown here enforce our previous observation of the preferential co-packaging of the Udorn PB1 and NA gene segments during progeny virion formation (Cobbin et al., 2014; Gilbertson et al., 2016). We show that the co-selection of the Udorn PB1 gene with the Udorn NA gene is likely responsible for the prevalence of the Udorn PB1 in all stages of the reassortment process and in the final dominant reassortants in our model system and the increased replicative fitness observed upon the inclusion of the Udorn NP in the presence of the Udorn PB1, HA and NA, likely drove the increased prevalence of the Udorn NP. However, we need to entertain the possibility that the Udorn PB1-NA co-selection relationship is reinforced by other co-selection relationships between the PR8 genes that might remove them from the available "packaging pool." As we are neutralising reassortants expressing PR8 HA and or NA with antisera, genes that co-select strongly with PR8 HA or NA genes will remain unidentified in this study. The fact that the reassortant PR8(Ud-HA,NA) was not isolated as one of the dominant progeny, despite its high replicative capacity and high HA yield, attests to the strength of co-selection in our system and its influence on the availability of genes that would otherwise create highly fit viruses.

The Udorn PB1-NA co-selection relationship is in accord with our retrospective analysis of past H3N2 vaccine seed strains, where the seasonal PB1 was present at a higher frequency compared to the other non-HA and NA genes from the seasonal virus parent (Cobbin et al., 2013). The co-selection of internal genes with the HA or NA genes also has direct implications for the generation of pandemic strains. In the event of reassortment between human and avian IAVs, progeny reassortant viruses expressing the surface glycoproteins of the human strain would be inhibited by pre-existing antibodies within the human population, allowing reassortants with the avian HA and potentially also NA, to dominate. As the HA and NA genes may have co-selection relationships with other internal gene segments, this can dictate which avian internal genes are also carried through into the human-infecting strain, shaping the phenotype of the emergent virus and influencing its impact on the human population. For example, co-selection of an avian PB1 would allow expression of a full-length and inflammatory PB1-F2 (McAuley et al., 2010, 2017) which has been shown to be a driver of severe secondary bacterial infection (McAuley et al., 2007).

Our study provides new information on the drivers of influenza virus reassortment and the factors that may influence the phenotype of dominant progeny. The eventual move from classical reassortment to reverse engineering for vaccine seed generation will require this understanding so that gene constellations that provide greatest HA and NA protein content can be produced. In addition, a greater understanding of the factors that dictate gene constellations and their corresponding phenotypes likely to arise by reassortment between influenza viruses of human and other reservoir species, will help in prediction of the likelihood and impact of future pandemics.

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 author.

AUTHOR CONTRIBUTIONS

ST performed all experiments with the exception of the haemagglutination assays, which were performed by EM, and genotyping of certain limit dilution viruses by JC. ST analysed

the experiments and wrote the draft manuscript. LB, BG, and SR supervised the work, further analysed the data, and contributed to the writing of the submitted manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by a National Health and Medical Research Council of Australia Program grant ID1071916 to LB. ST was supported by an Australian Postgraduate Award.

ACKNOWLEDGMENTS

We thank Ross Hamilton from CSL Ltd., for providing the electron microscopic images.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.683152/full#supplementary-material

REFERENCES

Biswas, S. K., Boutz, P. L., and Nayak, D. P. (1998). Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. *J. Virol.* 72, 5493–5501. doi: 10.1128/JVI.72.7.5493-5501

Bolte, H., Rosu, M. E., Hagelauer, E., García-Sastre, A., and Schwemmle, M. (2019). Packaging of the influenza virus genome is governed by a plastic network of RNA- and nucleoprotein-mediated interactions. *J. Virol.* 93:e01861-18. doi: 10.1128/JVI.01861-18

Brooke, C. B., Ince, W. L., Wei, J., Bennink, J. R., and Yewdell, J. W. (2014). Influenza A virus nucleoprotein selectively decreases neuraminidase gene-segment packaging while enhancing viral fitness and transmissibility. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16854–16859. doi: 10.1073/pnas.1415396111

Brown, L. E., Murray, J. M., White, D. O., and Jackson, D. C. (1990). An analysis of the properties of monoclonal antibodies directed to epitopes on influenza virus hemagglutinin. *Arch. Virol.* 114, 1–26. doi: 10.1007/BF01311008

Cavrois, M., De Noronha, C., and Greene, W. C. (2002). A sensitive and specific enzyme-based assay detecting HIV-1 virion fusion in primary T lymphocytes. *Nat. Biotechnol.* 20, 1151–1154. doi: 10.1038/nbt745

Cobbin, J. C. A., Ong, C., Verity, E., Gilbertson, B. P., Rockman, S. P., and Brown, L. E. (2014). Influenza virus PB1 and neuraminidase gene segments can cosegregate during vaccine reassortment driven by interactions in the PB1 coding region. *J. Virol.* 88, 8971–8980. doi: 10.1128/JVI.01022-14

Cobbin, J. C. A., Verity, E. E., Gilbertson, B. P., Rockman, S. P., and Brown, L. E. (2013). The source of the PB1 gene in influenza vaccine reassortants selectively alters HA content of the resulting seed virus. *J. Virol.* 10, 5577–5585. doi: 10.1128/JVI.02856-12

Dadonaite, B., Gilbertson, B., Knight, M. L., Trifkovic, S., Rockman, S., Laederach, A., et al. (2019). The structure of the influenza A virus genome. *Nat. Microbiol.* 4, 1781–1789. doi: 10.1038/s41564-019-0513-7

Davis, A. M., Ramirez, J., and Newcomb, L. L. (2017). Identification of influenza A nucleoprotein body domain residues essential for viral RNA expression expose antiviral target. *Virol. J.* 14, 22–34. doi: 10.1186/s12985-017-0694-8

Downie, J. C. (2004). "Reassortment of influenza A virus genes linked to PB1 polymerase gene," in *Options for the Control of Influenza V*, ed. Y. Kawaoka (Elsevier), 714–718.

Essere, B., Yver, M., Gavazzi, C., Terrier, O., Isel, C., Fournier, E., et al. (2013). Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc. Natl. Acad. Sci. U.S.A.* 110, E3840–E3848. doi: 10.1073/pnas.1308649110

Fazekas De St, G., and Webster, R. G. (1966). Disquisitions of original antigenic sin. I. Evidence in man. *J. Exp. Med.* 124, 331–345. doi: 10.1084/jem.124.3.331

Fournier, E., Moules, V., Essere, B., Paillart, J. C., Sirbat, J. D., Cavalier, A., et al. (2012). Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine* 30, 7359–7367. doi: 10.1016/j.vaccine.2012.09.079

Gavazzi, C., Isel, C., Fournier, E., Moules, V., Cavalier, A., Thomas, D., et al. (2013a). An in vitro network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. *Nucleic Acids Res.* 41, 1241–1254. doi: 10.1093/nar/gks1181

Gavazzi, C., Yver, M., Isel, C., Smyth, R. P., Rosa-Calatrava, M., Lina, B., et al. (2013b). A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16604–16609. doi: 10.1073/pnas.1314419110

Gerdil, C. (2003). The annual production cycle for influenza vaccine. *Vaccine* 21, 1776–1779. doi: 10.1016/s0264-410x(03)00071-9

Ghedin, E., Sengamalay, N. A., Shumway, M., Zaborsky, J., Feldblyum, T., Subbu, V., et al. (2005). Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 437, 1162–1166. doi: 10.1038/nature04239

Gilbertson, B., Zheng, T., Gerber, M., Printz-Schweigert, A., Ong, C., Marquet, R., et al. (2016). Influenza NA and PB1 gene segments interact during the formation of viral progeny: localization of the binding region within the PB1 gene. *Viruses* 8, 238–255. doi: 10.3390/v8080238

Hara, K., Nakazono, Y., Kashiwagi, T., Hamada, N., and Watanabe, H. (2013). Co-incorporation of the PB2 and PA polymerase subunits from human H3N2 influenza virus is a critical determinant of the replication of reassortant ribonucleoprotein complexes. *J. Gen. Virol.* 94, 2406–2416. doi: 10.1099/vir.0.053959-0

Hayat, M. A., and Miller, S. E. (1990). *Negative Staining*. New York, NY: McGraw-Hill Publishing Company.

Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G., and Webster, R. G. (2000). A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6108–6113. doi: 10.1073/pnas.100133697

Kilbourne, E. D. (1969). Future influenza vaccines and the use of genetic recombinants. *Bull. World Health Organ.* 41, 643–645.

Kilbourne, E. D., and Murphy, J. S. (1960). Genetic studies of influenza viruses .1. Viral morphology and growth capacity as exchangeable genetic traits – rapid in ovo adaptation of early passage asian strain isolates by combination with PR8. *J. Exp. Med.* 111, 387–406. doi: 10.1084/jem.111.3.387

Kimble, J. B., Angel, M., Wan, H., Sutton, T. C., Finch, C., and Perez, D. R. (2014). Alternative reassortment events leading to transmissible H9N1 influenza viruses in the ferret model. *J. Virol.* 88, 66–71. doi: 10.1128/JVI.02677-13

Le Sage, V., Kanarek, J. P., Snyder, D. J., Cooper, V. S., Lakdawala, S. S., and Lee, N. (2020). Mapping of influenza virus RNA-RNA interactions reveals a flexible network. *Cell Rep.* 31:107823. doi: 10.1016/j.celrep.2020.107823

Li, C., Hatta, M., Nidom, C. A., Muramoto, Y., Watanabe, S., Neumann, G., et al. (2010). Reassortment between avian H5N1 and human H3N2 influenza viruses creates hybrid viruses with substantial virulence. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4687–4692. doi: 10.1073/pnas.0912807107

Li, C., Hatta, M., Watanabe, S., Neumann, G., and Kawaoka, Y. (2008). Compatibility among polymerase subunit proteins is a restricting factor in reassortment between equine H7N7 and human H3N2 influenza viruses. *J. Virol.* 82, 11880–11888. doi: 10.1128/JVI.01445-08

Li, Z., Watanabe, T., Hatta, M., Watanabe, S., Nanbo, A., Ozawa, M., et al. (2009). Mutational analysis of conserved amino acids in the influenza A virus nucleoprotein. *J. Virol.* 83, 4153–4162. doi: 10.1128/JVI.02642-08

Lu, L., Lycett, S. J., and Leigh Brown, A. J. (2014). Reassortment patterns of avian influenza virus internal segments among different subtypes. *BMC Evol. Biol.* 14:16. doi: 10.1186/1471-2148-14-16

Lubeck, M. D., Palese, P., and Schulman, J. L. (1979). Nonranom association of parental genes in influenza A virus recombinants. *Virology* 95, 269–274. doi: 10.1016/0042-6822(79)90430-6

Marshall, N., Priyamvada, L., Ende, Z., Steel, J., and Lowen, A. C. (2013). Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. *PLoS Pathog.* 9:e1003421. doi: 10.1371/journal.ppat.1003421

McAuley, J., Deng, Y.-M., Gilbertson, B., Mackenzie-Kludas, C., Barr, I., and Brown, L. (2017). Rapid evolution of the PB1-F2 virulence protein expressed by human seasonal H3N2 influenza viruses reduces inflammatory responses to infection. *Virol. J.* 14:162. doi: 10.1186/s12985-017-0827-0

McAuley, J. L., Chipuk, J. E., Boyd, K. L., Van De Velde, N., Green, D. R., and Mccullers, J. A. (2010). PB1-F2 proteins from H5N1 and 20th century pandemic influenza viruses cause immunopathology. *PLoS Pathog.* 6:e1001014. doi: 10.1371/journal.ppat.1001014

McAuley, J. L., Hornung, F., Boyd, K. L., Smith, A. M., Mckeon, R., Bennink, J., et al. (2007). Expression of the 1918 influenza A virus PB1-F2 enhances the pathogenesis of viral and secondary bacterial pneumonia. *Cell Host Microbe* 2, 240–249. doi: 10.1016/j.chom.2007.09.001

Mena, I., Jambrina, E., Albo, C., Perales, B., Ortín, J., Arrese, M., et al. (1999). Mutational analysis of influenza A virus nucleoprotein: identification of mutations that affect RNA replication. *J. Virol.* 73, 1186–1194. doi: 10.1128/JVI.73.2.1186-1194.1999

Moreira, E. A., Weber, A., Bolte, H., Kolesnikova, L., Giese, S., Lakdawala, S., et al. (2016). A conserved influenza A virus nucleoprotein code controls specific viral genome packaging. *Nat. Commun.* 7:12861. doi: 10.1038/ncomms12861

Nakazono, Y., Hara, K., Kashiwagi, T., Hamada, N., and Watanabe, H. (2012). The RNA polymerase PB2 subunit of influenza A/HongKong/156/1997 (H5N1) restrict the replication of reassortant ribonucleoprotein complexes. *PLoS One* 7:e32634. doi: 10.1371/journal.pone.0032634

Nelson, M. I., Detmer, S. E., Wentworth, D. E., Tan, Y., Schwartzbard, A., Halpin, R. A., et al. (2012). Genomic reassortment of influenza A virus in North American swine, 1998–2011. *J. Gen. Virol.* 93, 2584–2589. doi: 10.1099/vir.0.045930-0

Octaviani, C. P., Goto, H., and Kawaoka, Y. (2011). Reassortment between seasonal H1N1 and pandemic (H1N1) 2009 influenza viruses is restricted by limited compatibility among polymerase subunits. *J. Virol.* 85, 8449–8452. doi: 10.1128/JVI.05054-11

Rabadan, R., Levine, A. J., and Krasnitz, M. (2008). Non-random reassortment in human influenza A viruses. *Influenza Other Resp. Viruses* 2, 9–22. doi: 10.1111/j.1750-2659.2007.00030.x

Schrauwen, E. J. A., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., Fouchier, R. A. M., and Herfst, S. (2013). Reassortment between avian H5N1 and human influenza viruses is mainly restricted to the matrix and neuraminidase gene segments. *PLoS One* 8:e59889. doi: 10.1371/journal.pone.0059889

Song, M. S., Pascua, P. N., Lee, J. H., Baek, Y. H., Park, K. J., Kwon, H. I., et al. (2011). Virulence and genetic compatibility of polymerase reassortant viruses derived from the pandemic (H1N1) 2009 influenza virus and circulating influenza A viruses. *J. Virol.* 85, 6275–6286. doi: 10.1128/JVI.02125-10

Steel, J., and Lowen, A. C. (2014). "Influenza A Virus Reassortment," in *Influenza Pathogenesis and Control*, Vol. I, eds R. W. Compans and M. B. A. Oldstone (Switzerland: Springer International Publishing), 377–401.

Tannock, G. A., Paul, J. A., and Barry, R. D. (1984). Relative immunogenicity of the cold-adapted influenza virus A/Ann Arbor/6/60 (A/AA/6/60-ca), recombinants of A/AA/6/60-ca, and parental strains with similar surface antigens. *Infect. Immun.* 43, 457–462. doi: 10.1128/IAI.43.2.457-462.1984

Varich, N. L., Gitelman, A. K., Shilov, A. A., Smirnov, Y. A., and Kaverin, N. V. (2008). Deviation from the random distribution pattern of influenza A virus gene segments in reassortants produced under non-selective conditions. *Arch. Virol.* 153, 1149–1154. doi: 10.1007/s00705-008-0070-5

Conflict of Interest: SR is an employee of the influenza vaccine manufacturing company Seqirus.

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership