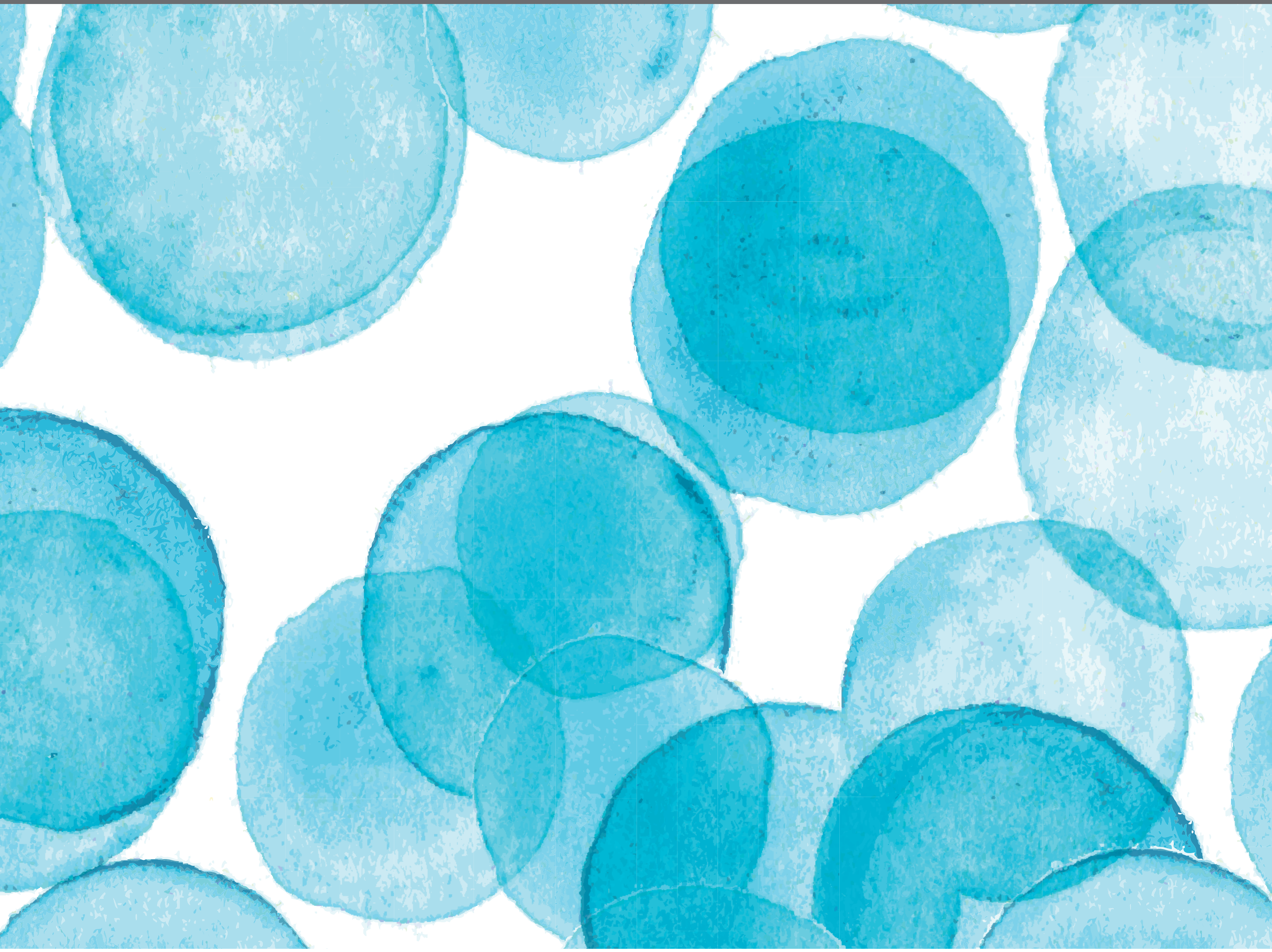
EPIDEMIOLOGY OF INFLUENZA AND OTHER RESPIRATORY PATHOGENS DURING THE CORONAVIRUS COVID-19 PANDEMIC

EDITED BY: Clyde Dapat, Ian George Barr, Hassan Zaraket and
Hitoshi Oshitani

PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-529-4
DOI 10.3389/978-2-88976-529-4

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

EPIDEMIOLOGY OF INFLUENZA AND OTHER RESPIRATORY PATHOGENS DURING THE CORONAVIRUS COVID-19 PANDEMIC

Topic Editors:
Clyde Dapat, Tohoku University Graduate School of Medicine, Japan
Ian George Barr, WHO Collaborating Centre for Reference and Research on Influenza (VIDRL), Australia
Hassan Zaraket, American University of Beirut, Lebanon
Hitoshi Oshitani, Tohoku University Graduate School of Medicine, Japan

Citation: Dapat, C., Barr, I. G., Zaraket, H., Oshitani, H., eds. (2022). Epidemiology of Influenza and Other Respiratory Pathogens During the Coronavirus COVID-19 Pandemic. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-529-4

Table of Contents

05 ***Analysis of Indian SARS-CoV-2 Genomes Reveals Prevalence of D614G Mutation in Spike Protein Predicting an Increase in Interaction With TMPRSS2 and Virus Infectivity***

Sunil Raghav, Arup Ghosh, Jyotirmayee Turuk, Sugandh Kumar, Atimukta Jha, Swati Madhulika, Manasi Priyadarshini, Viplov K. Biswas, P. Sushree Shyamli, Bharati Singh, Neha Singh, Deepika Singh, Ankita Datey, Kiran Avula, Shuchi Smita, Jyotsnamayee Sabat, Debdutta Bhattacharya, Jaya Singh Kshatri, Dileep Vasudevan, Amol Suryawanshi, Rupesh Dash, Shantibhushan Senapati, Tushar K. Beuria, Rajeeb Swain, Soma Chattopadhyay, Gulam Hussain Syed, Anshuman Dixit, Punit Prasad, Odisha COVID-19 Study Group, ILS COVID-19 Team, Sanghamitra Pati and Ajay Parida

18 ***Clinical and Laboratory Profile of Hospitalized Symptomatic COVID-19 Patients: Case Series Study From the First COVID-19 Center in the UAE***

Suad Hannawi, Haifa Hannawi, Kashif Bin Naeem, Noha Mousaad Elemam, Mahmood Y. Hachim, Ibrahim. Y. Hachim, Abdulla Salah Darwish and Issa Al Salmi

28 ***Sex-Disaggregated Data on Clinical Characteristics and Outcomes of Hospitalized Patients With COVID-19: A Retrospective Study***

Mengdie Wang, Nan Jiang, Changjun Li, Jing Wang, Heping Yang, Li Liu, Xiangping Tan, Zhenyuan Chen, Yanhong Gong, Xiaoxv Yin, Qiao Zong, Nian Xiong and Guopeng Zhang

36 ***Dynamics of SARS-CoV2 Infection and Multi-Drug Resistant Bacteria Superinfection in Patients With Assisted Mechanical Ventilation***

Annarita Mazzariol, Anna Benini, Ilaria Unali, Riccardo Nocini, Marcello Smania, Anna Bertoncelli, Francesco De Sanctis, Stefano Ugel, Katia Donadello, Enrico Polati and Davide Gibellini

46 ***Research on Influencing Factors and Classification of Patients With Mild and Severe COVID-19 Symptoms***

Xiaoping Chen, Lihui Zheng, Shupei Ye, Mengxin Xu, YanLing Li, KeXin Lv, Haipeng Zhu, Yusheng Jie and Yao-Qing Chen

60 ***The Burden of Respiratory Viruses and Their Prevalence in Different Geographical Regions of India: 1970–2020***

Rushabh Waghmode, Sushama Jadhav and Vijay Nema

72 ***Clinical and Immunological Characteristics of Patients With Adenovirus Infection at Different Altitude Areas in Tibet, China***

Bowen Wang, Mengjia Peng, Li Yang, Guokai Li, Jie Yang, Ciren Yundan, Xiaohua Zeng, Qianqi Wei, Qi Han, Chang Liu, Ke Ding, Kaige Peng and Wen Kang

86 ***Genomic Epidemiology Reveals Multiple Introductions of Severe Acute Respiratory Syndrome Coronavirus 2 in Niigata City, Japan, Between February and May 2020***

Keita Wagatsuma, Ryosuke Sato, Satoru Yamazaki, Masako Iwaya, Yoshiki Takahashi, Akiko Nojima, Mitsuru Oseki, Takashi Abe, Wint Wint Phyu, Tsutomu Tamura, Tsuyoshi Sekizuka, Makoto Kuroda, Haruki H. Matsumoto and Reiko Saito

101 ***Phylogenetic and Spatiotemporal Analyses of the Complete Genome Sequences of Avian Coronavirus Infectious Bronchitis Virus in China During 1985–2020: Revealing Coexistence of Multiple Transmission Chains and the Origin of LX4-Type Virus***

Wensheng Fan, Jiming Chen, Yu Zhang, Qiaomu Deng, Lanping Wei, Changrun Zhao, Di Lv, Liting Lin, Bingsha Zhang, Tianchao Wei, Teng Huang, Ping Wei and Meilan Mo

117 ***Rapid and Accurate Detection of SARS Coronavirus 2 by Nanopore Amplicon Sequencing***

Xiao-xiao Li, Chao Li, Peng-cheng Du, Shao-yun Li, Le Yu, Zhi-qiang Zhao, Ting-ting Liu, Cong-kai Zhang, Sen-chao Zhang, Yu Zhuang, Chao-ran Dong and Qing-gang Ge

frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 November 2020
doi: 10.3389/fmicb.2020.594928

OPEN ACCESS

Edited by:
Clyde Dapat,
Tohoku University Graduate School of Medicine, Japan

Reviewed by:
Arunachalam Ramaiah,
University of California, Irvine, United States
Subir Sarker,
La Trobe University, Australia

***Correspondence:**
Sunil Raghav
sunilraghav@ils.res.in;
raghuvanshi2010@yahoo.co.in
Ajay Parida
ajayparida@ils.res.in;
drajayparida@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: *14 August 2020*
Accepted: *21 October 2020*
Published: *23 November 2020*

Citation:
Raghav S, Ghosh A, Turuk J, Kumar S, Jha A, Madhulika S, Priyadarshini M, Biswas VK, Shyamli PS, Singh B, Singh N, Singh D, Datey A, Avula K, Smita S, Sabat J, Bhattacharya D, Kshatri JS, Vasudevan D, Suryawanshi A, Dash R, Senapati S, Beuria TK, Swain R, Chattopadhyay S, Syed GH, Dixit A, Prasad P, Odisha COVID-19 Study Group, ILS COVID-19 Team, Pati S and Parida A (2020) Analysis of Indian SARS-CoV-2 Genomes Reveals Prevalence of D614G Mutation in Spike Protein Predicting an Increase in Interaction With TMPRSS2 and Virus Infectivity. Front. Microbiol. 11:594928. doi: 10.3389/fmicb.2020.594928

Analysis of Indian SARS-CoV-2 Genomes Reveals Prevalence of D614G Mutation in Spike Protein Predicting an Increase in Interaction With TMPRSS2 and Virus Infectivity

Sunil Raghav[1]*†, Arup Ghosh[1†], Jyotirmayee Turuk[2], Sugandh Kumar[1], Atimukta Jha[1], Swati Madhulika[1], Manasi Priyadarshini[1], Viplov K. Biswas[1], P. Sushree Shyamli[1], Bharati Singh[1], Neha Singh[1], Deepika Singh[1], Ankita Datey[1], Kiran Avula[1], Shuchi Smita[1], Jyotsnamayee Sabat[2], Debdutta Bhattacharya[2], Jaya Singh Kshatri[2], Dileep Vasudevan[1], Amol Suryawanshi[1], Rupesh Dash[1], Shantibhushan Senapati[1], Tushar K. Beuria[1], Rajeeb Swain[1], Soma Chattopadhyay[1], Gulam Hussain Syed[1], Anshuman Dixit[1], Punit Prasad[1], Odisha COVID-19 Study Group[1], ILS COVID-19 Team[1], Sanghamitra Pati[2] and Ajay Parida[1]*

[1] Institute of Life Sciences (ILS), Bhubaneswar, India, [2] Regional Medical Research Centre (RMRC), Bhubaneswar, India

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, has emerged as a global pandemic worldwide. In this study, we used ARTIC primers–based amplicon sequencing to profile 225 SARS-CoV-2 genomes from India. Phylogenetic analysis of 202 high-quality assemblies identified the presence of all the five reported clades 19A, 19B, 20A, 20B, and 20C in the population. The analyses revealed Europe and Southeast Asia as two major routes for introduction of the disease in India followed by local transmission. Interestingly, the19B clade was found to be more prevalent in our sequenced genomes (17%) compared to other genomes reported so far from India. Haplotype network analysis showed evolution of 19A and 19B clades in parallel from predominantly Gujarat state in India, suggesting it to be one of the major routes of disease transmission in India during the months of March and April, whereas 20B and 20C appeared to evolve from 20A. At the same time, 20A and 20B clades depicted prevalence of four common mutations 241 C > T in 5′ UTR, P4715L, F942F along with D614G in the Spike protein. D614G mutation has been reported to increase virus shedding and infectivity. Our molecular modeling and docking analysis identified that D614G mutation resulted in enhanced affinity of Spike S1–S2 hinge region with TMPRSS2 protease, possibly the reason for increased shedding of S1 domain in G614 as compared to D614. Moreover, we also observed an increased concordance of G614 mutation with the viral load, as evident from decreased Ct value of Spike and the ORF1ab gene.

Keywords: SARS-CoV-2, COVID-19, phylogeny, India, D614G, viral RNA sequencing, protein-protein interaction

INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a betacoronavirus belonging to the Coronaviridae family. The first occurrence of this novel coronavirus was observed in Wuhan, China, in late December 2019, which later spread globally via human-to-human contact transmission (Lu et al., 2020). According to the World Health Organization (WHO) weekly epidemiological update until September 14, 2020, the virus has spread to more than 180 countries with 28.6 million total confirmed cases and 0.9 million deaths worldwide (World Health Organization, 2020). The first occurrence of a coronavirus case in India was observed in mid-January, but the number of cases started to increase from the first week of March. According to the WHO, the number of cases in India has reached 4.7 million with 78,500 deceased (World Health Organization, 2020).

Genomic studies performed to understand the origin of SARS-CoV-2, a positive single-stranded RNA virus, unraveled that it has a zoonotic origin and is transmitted to humans from bats via Malayan pangolins (Zhang T. et al., 2020). The nucleotide sequence of SARS-CoV-2 is ∼79% similar to SARS-CoV-1 and about 50% with MERS-CoV (Middle East respiratory syndrome coronavirus) (Lu et al., 2020). Approximately 30-kb genome of SARS-CoV-2 features a cap structure in $5'$ and $3'$ poly (A) like other members of the coronavirus family. A major portion of the genome is covered by two ORFs (ORF1a and ORF1b), which code for 15 non-structural proteins, including crucial proteins required for viral replication, such as viral proteases nsp3, nsp5, and nsp12, also known as RNA-dependent RNA polymerase (RdRP) (Kim et al., 2020). The genomic RNA (gRNA) also codes for structural proteins like Spike protein (S), nucleocapsid protein (N), membrane protein (M), and envelope protein (E), which are required for packaging of the virus and at least accessory proteins, but their ORFs are still not experimentally validated (Kim et al., 2020).

The entry of viral particles in the human body occurs through binding with angiotensin I–converting enzyme 2 (hACE2) receptor present on lung epithelial cells (Hoffmann et al., 2020). The predominant infection site is the respiratory tract due to the route of infection. Other than the lungs, it is known to infect other organs of the body, such as the kidney, liver, and intestine, as ACE2 expression is found to be quite high. After the viral entry into host cells, to initiate viral replication, negative-sense RNA intermediates are first synthesized by RdRP activity; these templates are then utilized for synthesis of gRNA and subgenomic RNAs (Kim et al., 2020). Because of the low fidelity of RdRP, mutations are incorporated with high frequency in gRNA, and such mutations are often known to increase the pathogenicity and fitness of the virus (Graepel et al., 2017; Ferron et al., 2018; Korber et al., 2020).

Recently, a predominant mutation, i.e., D614G in Spike protein, has been identified in virus strains sequenced from European population (Korber et al., 2020). There are several reports depicting that D614G mutation in Spike protein is associated with enhanced infectivity and spread of the virus owing to increased interaction with ACE2 receptor present on host cells (Korber et al., 2020). It has also been indicated that this mutation is present on the S2 domain of the Spike protein that is important for cleavage by TMPRSS2 enzyme for cleavage of S1 to facilitate the fusion of the viral Spike with the host cell membrane (Korber et al., 2020; Zhang L. et al., 2020). It is indeed interesting to understand further at a molecular level the changes in the protein structure induced by this mutation and its functional association with the infectivity and disease severity.

At the same time, it has been reported that D614G mutation co-occurs with three more mutations, i.e., 241 in UTR, 3307, and 14408 (Korber et al., 2020). The functional significance of these co-occurring mutations in evolution and virus selectivity is interesting to understand.

In the present study, we have sequenced 225 COVID-19 isolates from patient samples from the state of Odisha, those migrated from the 13 most affected Indian states as a part of DBT's PAN-India 1000 SARS-CoV-2 RNA genome sequencing consortium, using an amplicon sequencing–based methodology. The travel history of these patients was collected along with their symptomatic and asymptomatic behavior. We performed phylogenetic analysis from sequenced data to understand the genetic diversity and evolution of SARS-CoV-2 in the Indian subcontinent. From the sequenced isolates, we have identified 247 single-nucleotide variants; most of them are observed in ORF1ab, Spike and nucleocapsid protein coding region. Moreover, we have analyzed the D614G mutations in the samples to obtain information on its evolution in Indian population. Protein-modeling analysis of D614G mutation was carried out to identify the impact on structural changes at protein levels. We further performed protein–protein docking simulation to predict the impact of D614G mutation on the interaction between of wild-type and mutated Spike protein with TMPRSS2 enzyme to assess its impact on binding and perhaps viral infectivity.

MATERIALS AND METHODS

Sample Collection

All hospitalized and quarantined patients (March 2020 to June 2020), based on their clinical symptoms (fever or respiratory symptoms) or travel history, were preliminarily involved in this study. We received throat swabs in viral transport media samples of these patients used for SARS−CoV−2 detection. Patients with missing or with negative SARS−CoV−2 test results were excluded from this study based on Ct values obtained by quantitative polymerase chain reaction (qPCR) of isolated RNA. All patients involved in this study were residents of Odisha, India, during the outbreak period of COVID−19. The samples were collected and processed as per the guidelines of the Institutional Ethics and Biosafety Committee. Institutional Biosafety Committee (IBSC) approval (IBSC file no. V-122-MISC/2007-08/01) was taken before processing the samples in BSL3 laboratory.

Viral Load Detection

RNA isolation for all the 248 human patients was performed using QIAamp Viral RNA Mini Kit (Qiagen, cat. no. 52906). The isolated RNA was subjected to qPCR for determining viral load by

Ct values. For qPCR, we performed one-step multiplex real-time PCR using TaqPath^{TM} 1-Step Multiplex Master Mix (Thermo Fisher Scientific, cat. no. A28526), targeting three different gene-specific primer and probe sets—envelope glycoprotein Spike (S), nucleocapsid (N), and open reading frame 1 (ORF1).

Viral RNA Library Preparation and Sequencing

We prepared amplicon libraries for viral genome sequencing using QIAseq FX DNA Library Kit and QIAseq SARS-CoV-2 Primer Panel (Qiagen, cat. no. 180475, cat. no. 333896) as instructed by the manufacturer's manual, and the library was subsequently sequenced using Illumina platform. The adapter sequence used for each sample was compatible with Illumina sequencing instrument with 96-sample configurations (Qiaseq unique dual Y-adapter kit). The average insert length was in the 250–500 bp range. Prepared libraries were then pooled as a batch of 96 samples and sequenced using Illumina NextSeq 550 platform in 150 × 2 layout.

Raw Data Preprocessing

Quality of the sequenced files was checked using FastQC tool (0.11.9) (Andrews, 2010), followed by removal of low quality bases (–nextseq-trim, Q < 20), Illumina Universal adapter sequence and reads with less than 30-bp length using Cutadapt (2.10) (Martin, 2011). To access the quantity of host genomic DNA and other contaminants, Kraken (2.0.9-beta) (Wood et al., 2019) was used, and the reports were summarized using Krona (2.7.1) (Ondov et al., 2011). All the files were then aligned to human genome (assembly version GRCh38) using HISAT2 (2.2.0) (Kim et al., 2015), and unmapped reads were extracted using SAMTOOLS (1.10) (Li et al., 2009) and converted to FASTQ format using BEDTOOLS (2.29.2) (Quinlan and Hall, 2010) bamToFastq option.

Alignment With Viral Genome

The unmapped reads were then aligned to SARS-CoV-2 reference assembly (NCBI accession NC_045512) using HISAT2 (2.2.0) (Kim et al., 2015). Amplicon primes from the aligned file were removed using iVar (1.2.2) (Grubaugh et al., 2019) guided by Artic Network V3 primer scheme[1]. The aligned files were then deduplicated using Picard Tools (2.18.7)[2]. Alignment quality was checked using SAMTOOLS (1.10) (Li et al., 2009) flagstat option.

Consensus Sequence Generation and Variant Calling

A consensus sequence for each isolate was generated using Bcftools (1.10) (Li, 2011) and SEQTK[3]. After generating a reference-based consensus sequence, we selected 202 genomes with less than 5% N's and more than 10× coverage for phylogenetic and mutation analysis. Single-nucleotide variants were called and filtered (QUAL > 40 and DP > 20) using Bcftools (1.10) (Li, 2011). Effects of the filtered variants were annotated using SnpEff (4.5) (Cingolani et al., 2012). All of the consensus sequences were deposited in GISAID (Shu and McCauley, 2017); accession IDs are provided in **Supplementary Table 2**.

Phylogenetic Analysis

Phylogenetic tree analyses of all the samples were performed using SARS-CoV-2 analysis protocol standards and tools provided by Nextstrain (Hadfield et al., 2018) pipeline. First, all the sequences are aligned against the WH01 reference genome using Augur wrapper of MAFFT (Katoh and Standley, 2013), and low-quality variant sites are masked from the alignment. The initial maximum likelihood tree was generated by the IQTREE2 tool (Nguyen et al., 2015) with 1,000 bootstraps. We have provided the maximum likelihood tree with bootstrap percentage marked for branches with ≥ 60 support and clade information in **Supplementary Figure 3**. Further refinement of the tree was done using the Augur refine command, and the tree was rooted using the reference sequence with timeline information incorporation using TimeTree (Kumar et al., 2017). To finalize the tree for Nextstrain auspice visualization the tree was annotated using ancestral traits, clades, nucleotide mutation and amino acid mutation. The resulting tree was visualized using an Auspice instance, and the visualization was refined using the ggtree R package.

We have used Nextstrain year-letter clade nomenclature that started with 19A and 19B branched by C8782T and T28144C nucleotide changes and was initially prevalent in Asia during initial outbreak. Later, 20A emerged in European outbreak from 19A parents having C3037T, C14408T, and A23403G as distinctive features. 20B emerged as a distinct clade in Europe with three consecutive mutations, e.g., G28881A, G28882A, and G28883C. Further the 20C emerged as a North America–specific clade with C1059T and G25563T nucleotide changes.

Haplotype Network Analysis

For haplotype network analysis, we took a total of 287 (China 15, Germany 23, Italy 25, Saudi Arabia 23, Singapore 14, and South Korea) SARS-CoV-2 whole-genome sequences with less than 1% N and with collection date of March, April, and May from GISAID database. The selected samples are then aligned to the WH01 reference genome using MAFFT (Katoh and Standley, 2013). After filtering aligned sequences, we used POPART (Leigh and Bryant, 2015) software to generate haplotype network using a median joining method with 2,000 iterations.

Modeling of the Protein Structures

The sequences of SARS-CoV-2 proteins (NPS3, NSP4b, NSP6, nucleocapsid, and Spike) were retrieved from NCBI. As most of the proteins do not have a three-dimensional (3D) structure in protein data bank (PDB), they were modeled using Modeler 9.21 (Webb and Sali, 2016). The suitable templates for modeling of the proteins (**Supplementary Table 1**) were selected by DELTA-BLAST (Boratyn et al., 2012) against the PDB proteins. One hundred models were generated for each of the proteins. The best model was selected based on the lowest DOPE score (Shen and

[1] https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V3

[2] https://broadinstitute.github.io/picard/

[3] https://github.com/lh3/seqtk

Sali, 2006). The D614G mutant of the Spike protein was generated by Modeler 9.21. The loop in Spike protein (670–690) was refined using loop modeling procedure in Modeler by generating 100 loop models. The model with the lowest dope score was finally chosen for both the mutant and wild-type protein. Similarly, the host transmembrane serine protease 2 (TMPRSS2) was also modeled using Modeler. The PROCHECK (Laskowski et al., 1993) server was used to assess the stereochemistry of the generated models.

Protein–Protein Docking

The standalone version of HADDOCK2.2 (van Zundert et al., 2016) was used to perform protein–protein docking with SARS-CoV-2 Spike protein with human TMPRSS2. The docking was conducted by the restraining of the receptor (Spike) and ligand (TMPRSS2) residues known to be at the interface Spike-TMPRSS2 interface. Specifically, residues within 5 Å of the reported cleavage site (Arg685, Ser686) (Hoffmann et al., 2020) of the Spike and catalytic triad of TMPRSS2 (H296, D345, and S441) and binding residue D435 were restrained to be at the docking interface. A total of 100 docking poses were generated and ranked based on the HADDOCK2.2 docking score (van Zundert et al., 2016), which is composed of van der Waals energy, electrostatic energy, restraints energy, etc. The best-ranked docking pose was visualized using Pymol.

Statistical Analysis and Plotting

All the statistical analysis and plots were generated in R (3.6.1) statistical programming language using ggplot2, dplyr, reshape2, lubridate, ggsci, and ggpubr package available from CRAN and Bioconductor[4] repository.

RESULTS

Demographics, Clinical Status, and Travel History

The average age of the 225 patients was 30.98 ± 11.79 years with age range 1–75 years and median age of 30 years. The overall gender ratio of male-to-female was 201:24 with median age of male patients 31.75 and 24.5 years for female patients (**Supplementary Figures 1A,B**). Almost every female patient in this study reported no strong symptoms during sample collection, whereas in case of the male patients, we found that the numbers of symptomatic cases were almost the same with the number of asymptomatic cases (**Supplementary Figure 1C**). For the samples in our study, we did not observe any fatality in the patient group to our knowledge.

The majority of the patients (89%) disclosed their travel history during the sample collection procedure. We found that the majority of the patients included in the study were found to migrate from Gujarat, West Bengal, Tamil Nadu, Maharashtra, Delhi, Kerala, and Andhra Pradesh state in India (**Supplementary Figure 1D**). As most of the patients did not have any direct foreign travel (n = 1) history, the primary source of infection is local contacts in their workplaces, and this helped us to understand the COVID-19 strain diversity in the most affected states of India (**Supplementary Figure 1D**).

Phylogenetic Analysis

The phylogenetic analysis was carried out with 202 high-quality (<5% N's) SARS-CoV-2 sequences that revealed presence of four major clades, i.e., 19A (n = 39), 19B (n = 36), 20A (n = 73), 20B (n = 60), and one minor clade 20C (n = 4) (**Figure 1A**). To understand the abundance of clades with time, we plotted cumulative counts of clades against the sample collection date and observed the introduction of clade 20A occurred in April 2020 with parallel emergence of 19A, 19B, and 20B in May (**Figure 1B**). After May, no new occurrence of clade 19A or 19B was observed in our data, but in mid-June, we started to observe emergence of 20C clade (**Figure 1B**). When we overlaid the clinical status information (asymptomatic vs. symptomatic), we observed that isolates from clade 20A, 20B, and 20C depicted a higher number of symptomatic patients (**Figure 1C**). Although without the information about predisposition of complication in patients, it is hard to establish an association between the mutations and the clinical manifestations.

To further understand the transmission of the virus in our patients, we performed phylogenetic analysis of our dataset with 1,042 high-coverage Indian SARS-CoV-2 whole-genome sequences (N < 1%) obtained from GISAID on July 12, 2020 (**Supplementary Figure 1F**). From the transmission map generated using Nextstrain, we observed that most of the sequences in 20A clades share sequence similarity with samples from Gujarat, which has very high prevalence of this clade (**Supplementary Figure 1E**). The 19A clade is very prevalent in Delhi and Telangana region and Tamil Nadu, but the 20B clade is only prevalent in the southern parts of India (**Supplementary Figure 1E**). Both 20A and 20B clades predominantly contain a mutation in protein coding sequence 23403A > G, and the mutation is traced back to the West European region in late January (Korber et al., 2020). The missense mutation in Spike protein coding gene causes a change in 614 D > G position of Spike protein and reported to increase the shedding of S1 subunit of the protein, which leads to the increased infectivity (Korber et al., 2020).

Mutation Analysis

After filtering out low-quality genomes with <5% N's in the assembly and at least 10 × average coverage, we observed a total of 247 single-nucleotide variants from 202 SARS-CoV-2 isolates. Of these variants, 156 variants observed only in single isolates, 25 variants were classified as common variants with occurrence of more than 5% and 19 variants as rare with 2–5% occurrence (**Supplementary Table 1**). Among the common variants, the most frequent mutations are 23403 A > G (D614G, S gene), 241 C > T (5′ UTR), 14408 C > T (P4715L, RdRP gene), 3037 C > T (F942F, NSP3), 28881 G > A (R203K, N gene), 28882 G > A (R203R, N gene), 28883 G > C (G204R, N gene), and 28144 T > C (L84S, ORF8) with presence in more than 15% of all of our sequenced genomes (**Supplementary Table 1**). Plotting mutation diversity (>2%) in a clade-wise manner, we

[4] https://CRAN.R-project.org/package=tidyvers

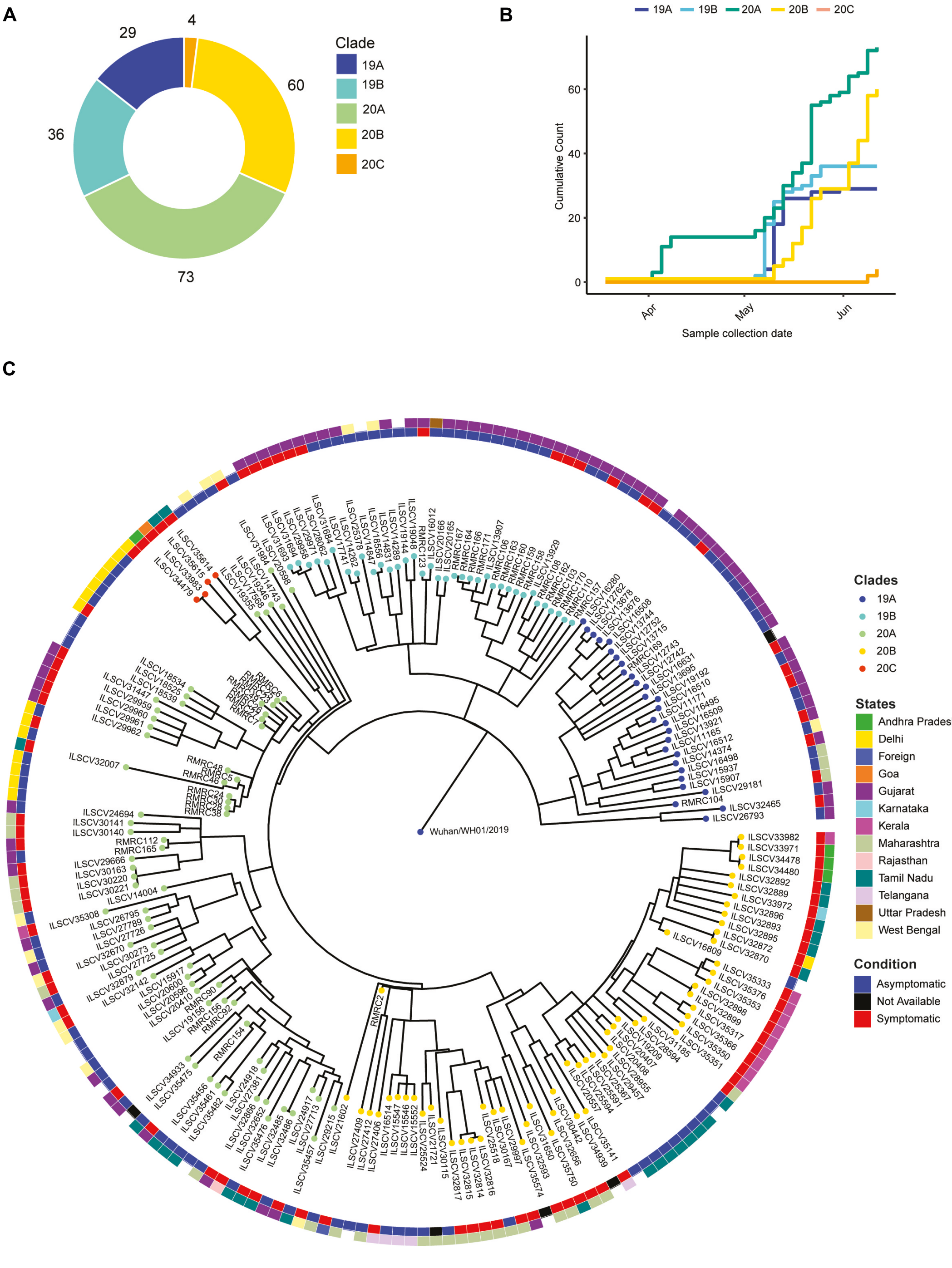


FIGURE 1 | Phylogenetic analysis of the SARS-Cov genomes and their distribution into different Nextstrain defined new clades. **(A)** A donut chart representing the sequenced sample (*n* = 202) distribution across the clades (clade nomenclature obtained using Nextstrain). **(B)** Cumulative count of clades plotted against sample collection date showing abundance of clades with time. **(C)** Time tree (1,000 bootstraps) of the sequenced samples (*n* = 202) generated using Nextstrain time-tree pipeline overlaid with clinical status (condition, inner circle) of the patients during sample collection, place of migration (state, outer circle), and clade information (clades).

observed distinct mutation signatures in different clades. The isolates that were grouped in 19A clade depicted prevalence of mostly ORF1ab mutations with one distinct N gene C > T mutation at 28311 position (**Figure 2C**). In clade 19B samples, we observed a very distinct ORF8 T > C mutation at 28144 position, two N gene mutations at positions 28326 and 28878 with some ORF1ab mutation in lower frequency (**Figure 2C**). Clades 20A and 20B have almost similar mutation profile with

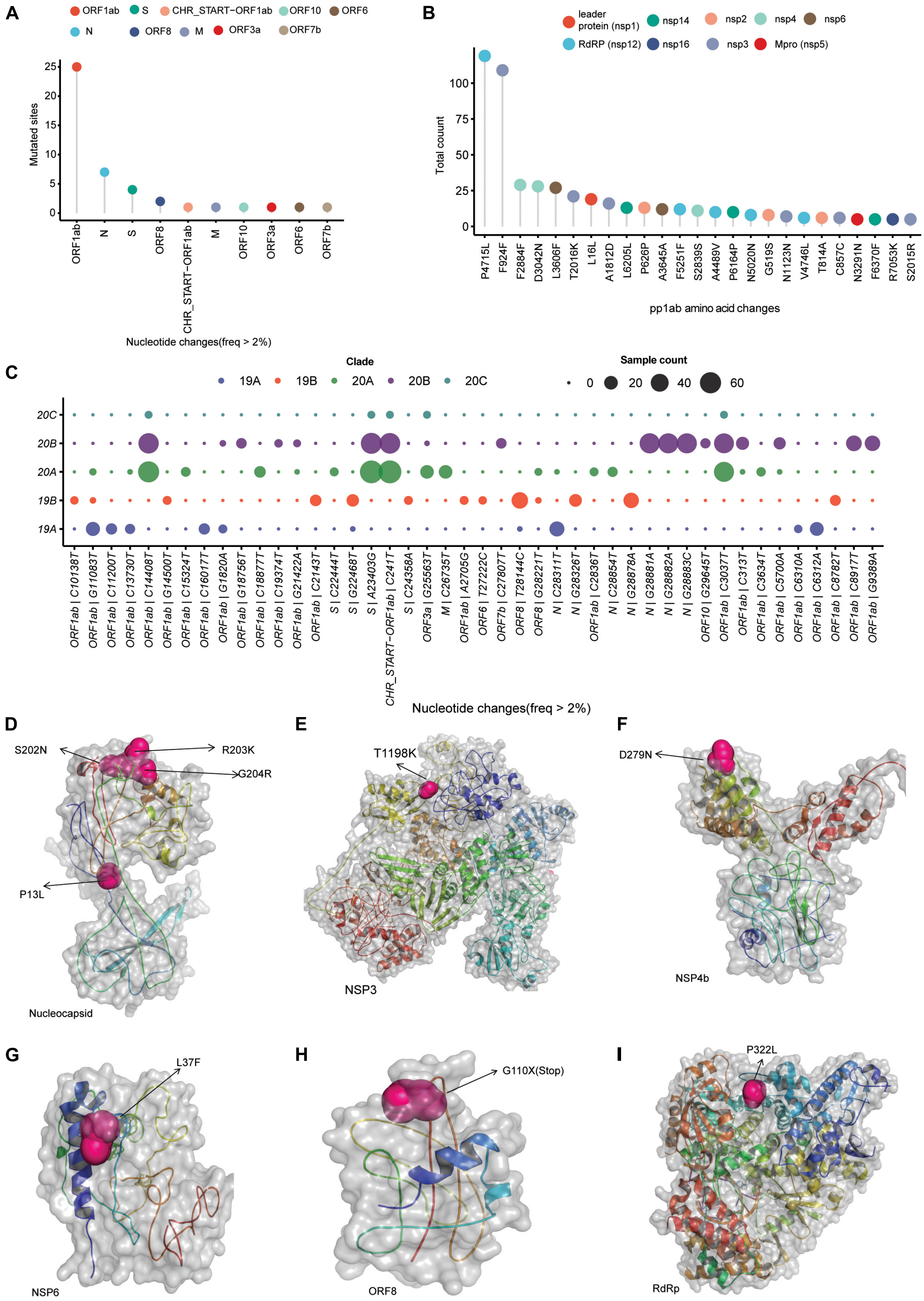


FIGURE 2 | SARS-CoV-2 clade distribution and their prevalent mutation profiles. **(A)** Dot plot representing the number of single-nucleotide mutation (occurred in more than 2% of the samples) present in different genomic segments of SARS-CoV-2 genome. **(B)** The ORF1ab region codes for a polypeptide are later cleaved to several mature peptides. The dot plot represents the amino acid changes (location of amino acid acids as per location in polypeptide sequence) in the mature peptides of ORF1ab. **(C)** Clade-wise occurrence of nucleotide mutations with presence in more than 2% of sequenced samples (*n* = 202). Color of the dots represents the clade and size of the dots represents number of the samples showing presence of the single-nucleotide variant. **(D–I)** The mutation sites on the modeled structures of the SARS-CoV-2 proteins. The mutation site(s) of the NSP3, NSP4b, NSP6, RdRP, and nucleocapsid proteins are marked as sphere, while the rest of the structure is shown in cartoon representation.

major mutated positions 241 C > T mutation in leader sequence, ORF1ab 3037 C > T, ORF1ab 14408 C > T (RdRP), and S gene 23403 A > G mutations. The very distinct characteristic that we observed for 20B clade is three consecutive N gene mutations at positions 28,881, 28,882, and 28,883 (**Figure 2C**). Of these three mutations, two are missense mutations resulting in change of protein sequence. In clade 20A, we observed two mutations at ORF3a and M protein coding gene at position 25,563 G > T, 26,735 C > T but in less frequencies in comparison with other mutated sites (**Figure 2C**). Overall, to conclude, we summarized all the mutated sites present in all the samples and observed ORF1ab is the most mutated region, followed by N gene, S gene, and ORF8 in SARS-CoV-2 samples from India (**Figure 2A**). Among the ORF1ab mutations, 4,715 P > L change in nsp12, also known as RdRP (the viral RNA dependent RNA polymerase), was the most common one followed by a synonymous change (F924F) in nsp3 protein (**Figure 2B**).

To understand the impact of identified prominent missense mutations (present on > 10% of the samples) on the annotated viral proteins in our sequenced population, we used protein crystal structures and protein models as crystal structures were not available. The PROCHECK (Laskowski et al., 1993) results showed that the generated models have acceptable stereochemistry. First, we looked into the location of the mutated amino acid with respect to its functional domains as any change in the functional domain has high probability to perturb the protein function. The sites of mutation on the modeled protein structures (nucleocapsid, NPS3, NSP4b, NSP6, ORF8, and RdRP) are marked to indicate their location (**Figures 2D–I**).

Haplotype Network Analysis

First, we created a median-joining haplotype network to look for transmission within India, and we observed two major branches (**Supplementary Figure 2A**). When we colored the sequences with clade information, the branches represented 19A and 20A, which later furcate into 19B, 20B, and 20C (**Supplementary Figure 2A**). When we overlaid migration information in the network, we observed the majority of 19A and 19B isolates migrated from Gujarat (**Figure 3A**). Migration of isolates identified in newly prevalent clades occurred from the southern path of India (**Figure 3A**). To understand the transmission source of SARS-CoV-2 infection in India, we constructed haplotype network using our sequencing data combined with genome sequences obtained from GISAID (China 15, Germany 23, Italy 25, Saudi Arabia 23, Singapore 14, and South Korea). From the haplotype network, we observed distinct clusters of genome sequences that were grouped in four major nodes. A large group of sequences clustered in two major haplotype clusters, one with genome sequences from China, Singapore, and South Korea and the other one with Italy, Saudi Arabia, and Germany with 2- to 4-nucleotide substitutions (**Figure 3B**). From the collection date, we observed that 20A clade, which is prevalent in Europe, became abundant in Odisha from April, representing the top cluster, whereas the bottom cluster represents samples belonging to clades 19A and 19B having a common origin in Southeast Asia (**Figure 3B**).

Effect of D614G on Viral Load

To understand the prevalence and effect of G614 in our sequencing dataset, we plotted week-wise count (based on collection date) of D614 and G614 and observed that the occurrence of the mutation has been observed in early March (12th week), and the cumulative frequency increased over time (**Figure 4A**). We also assessed the frequency of D614G mutation from Covid19 Beacon database (CSIRO and CSIR-IGIB, access date: February 8, 2020) in global and all the SARS-CoV-2 sequence published from India, and the occurrence of G614 in Indian genomes is 76.31% where the global frequency is 43.8% (**Supplementary Figure 2C**). Every sample we sequenced as a part of the study was also checked for the levels of ORF1 and S gene using qPCR. The Ct obtained is also a direct indicator of viral load (lesser the Ct, higher the viral load) in the individual. When we plotted the Ct values of all sequenced samples, we observed that except week 21, the Ct values of the isolates having G614 mutation are less in comparison to isolates having D614 (**Figures 4B,C**). We also had Ct values (ORF1, S gene) of 637 positive isolates available as a partner institute of the COVID19 surveillance program in Odisha, India. Plotting the data against the date of sample collection for testing, we observed a sharp and significant ($p < 0.05$) decline (median ∼5 Ct change) in the Ct values (**Figure 4E**) from April 2020 to May 2020, and there was median ∼1 Ct change of S gene, as well as ORF1ab (**Figures 4D–F**) from the month of May to June 2020. In the case of ORF1 expression, we also observed a sharp change between April and May, but there is almost no change between May and June 2020 (**Supplementary Figure 2A**).

Molecular Modeling Depicted Enhanced Interaction of D614G-Mutated Spike Protein With TMPRSS2 Protease

As reported in earlier publications, we also noticed D614G as a highly prevalent mutation in highly transmitted and evolved strains belonging to clades 20A and 20B in the Indian scenario; therefore, we carried out molecular modeling analysis. It was observed that the wild-type (D614) and mutated form (G614) of Spike protein showed a slightly different arrangement of structural elements near the mutation site, which is present in hinge region linking S1 and S2 domain, i.e., S1 furin cleavage site (**Figure 4G**). The S2 domain of the Spike protein is reported to interact with TMPRSS2 protease necessary for shedding of S1 domain for viral entry inside the host cells by facilitating the merging of virus with the host cell membrane. The proteolytic cleavage of Spike protein by TMPRSS2 results in the shedding of S1 domain, which is one of the key steps of the virus infection in host cells. Cryo–electron microscopy structures indicated that side chains of D614 protomer and T859 of the neighboring protomer form a hydrogen bond in between bringing together S1 domain with S2 (Lu et al., 2020; Walls et al., 2020). This substitution could modulate the glycosylation at N616 site as well, perturbing the interaction between the neighboring protomer. Our Spike protein model depicted this hydrogen bonding between D614-T859 of S1 and S2 domain

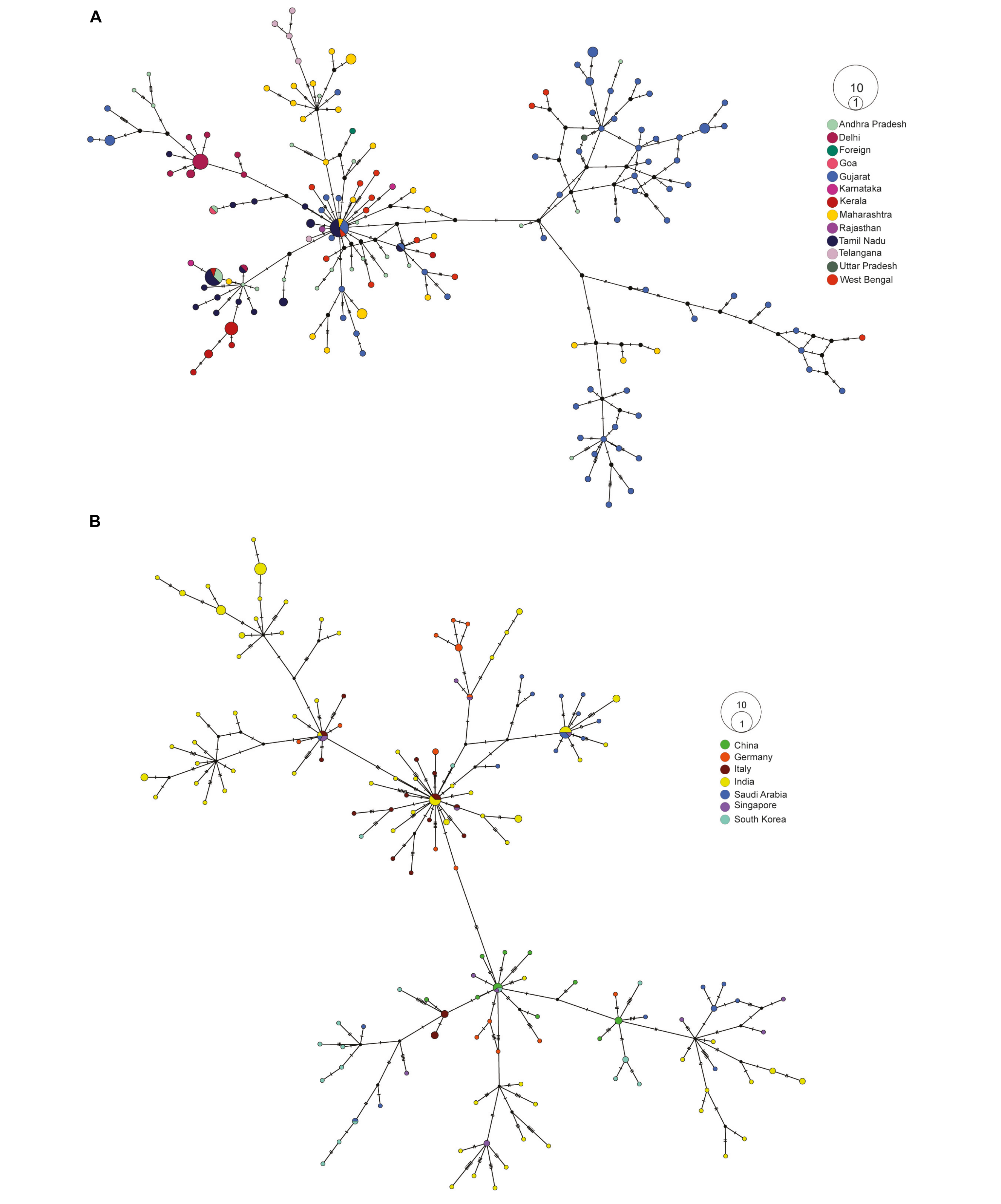


FIGURE 3 | Haplotype network analysis of SARS-CoV-2 sequences. **(A)** Haplotype network of 202 SARS-CoV-2 whole-genome sequences from our dataset colored by their respective place of migration. **(B)** Haplotype network of 100 high-coverage SARS-CoV-2 genomes obtained from GISAID (China 15, Germany 23, Italy 25, Saudi Arabia 23, Singapore 14, South Korea 17) combined with 170 samples sequenced from Odisha with less than <5% N's present in consensus sequence.

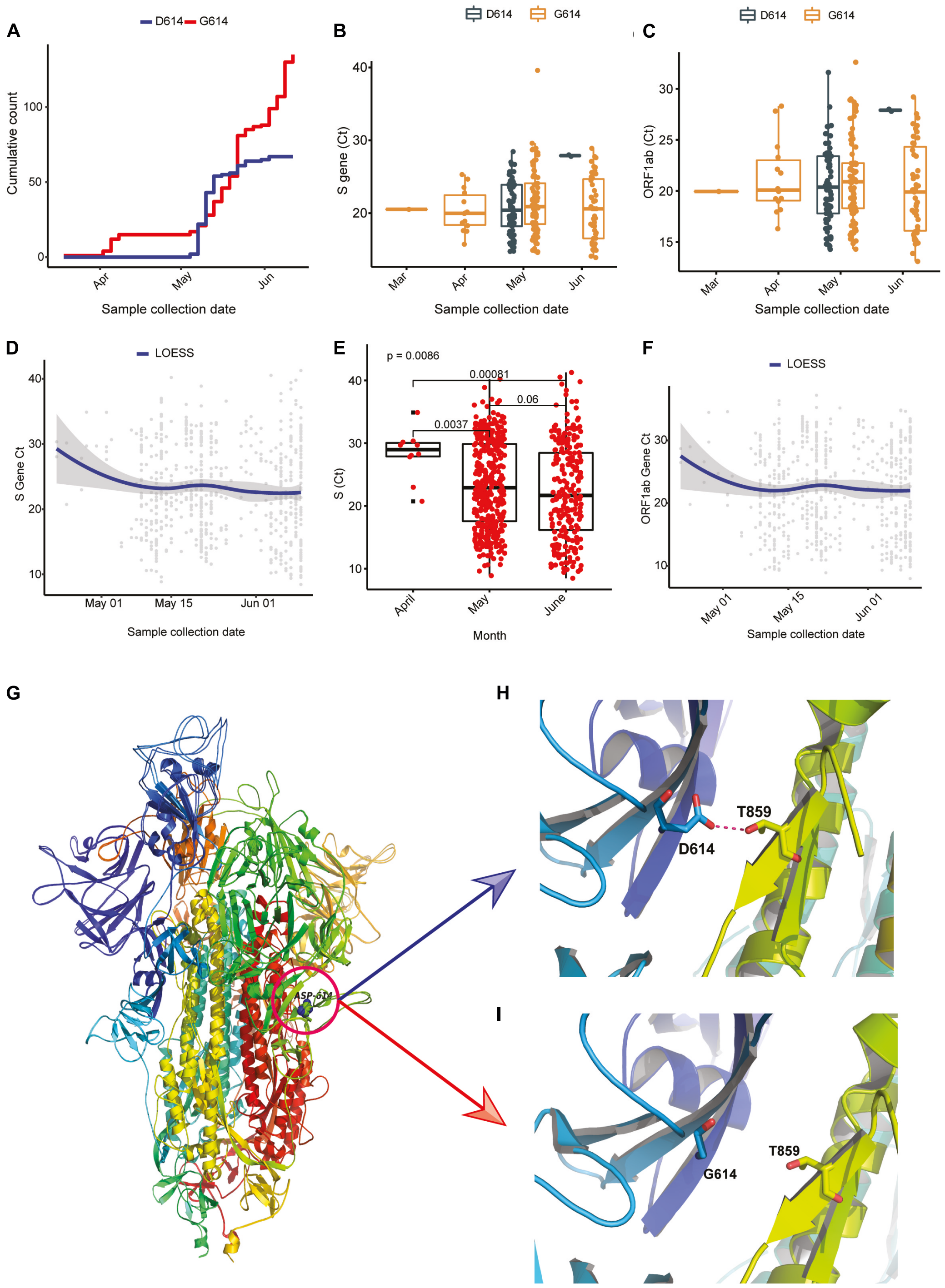


FIGURE 4 | D614G in Spike gene increases infectivity portrayed by Ct values as a surrogate for viral load. **(A)** Cumulative count of the occurrence of D and G in 614 position of Spike protein in sequenced genomes (*n* = 202). **(B,C)** Ct value distribution of S gene and ORF1ab for the sequenced genomes (*n* = 202). **(D–F)** Ct value distribution of S gene and ORF1ab in all the positive samples tested at Institute of Life Sciences until June 17, 2020. **(G–I)** The superimposed 3D structures G614 mutant and wild-type Spike protein. **(G)** The mutant site is highlighted with a circle at 614 position. **(H)** The hydrogen bond (D614-T859) shown as dotted line between Spike S1 and S2 domain in wild type. **(I)** The hydrogen bond is lost as a result of D614G mutation.

(**Figure 4H**). The mutation of D614 to G614 eliminates this side-chain hydrogen bonding between S1 and S2 domain (**Figure 4I**), leading to increased main-chain flexibility enabling a more favorable orientation of Q613, possibly facilitating cleavage by TMPRSS2 by perturbing its affinity with the S1-furin cleavage site. It has also been proposed that D614 forms an intrasalt bridge with R646, which makes the conformation unfavorable for S1 association with S2 domain (Zhang L. et al., 2020). Interestingly, the protein docking analysis depicted better hydrogen bonding interactions between the Spike protein cleavage sites (Arg685, Ser686) with the catalytic triad of TMPRSS2 in mutant condition as compared to wild-type. In the case of mutant Spike protein, the Arg682 and residues at primary cleavage site of Spike protein (Arg685 and Ser686) formed six hydrogen bonding interactions with Glu299, Lys300, Asp338, and Gln438 residues of TMPRSS2 (**Figures 5A,B**), whereas in the D614 wild-type form there were five hydrogen bonds observed between the cleavage site of Spike protein S2 domain and TMPRSS2 (**Figures 5C,D**). The binding energy was observed to be better for the G614 mutant (-143.03 kcal/mol) as compared to that of the wild type (−113.67 kcal/mol), indicating better binding of TMPRSS2 with the mutated Spike protein (**Figure 5E**).

DISCUSSION

The pattern of COVID-19 pandemic spread in India is quite different from all over the world in terms of slow transmission

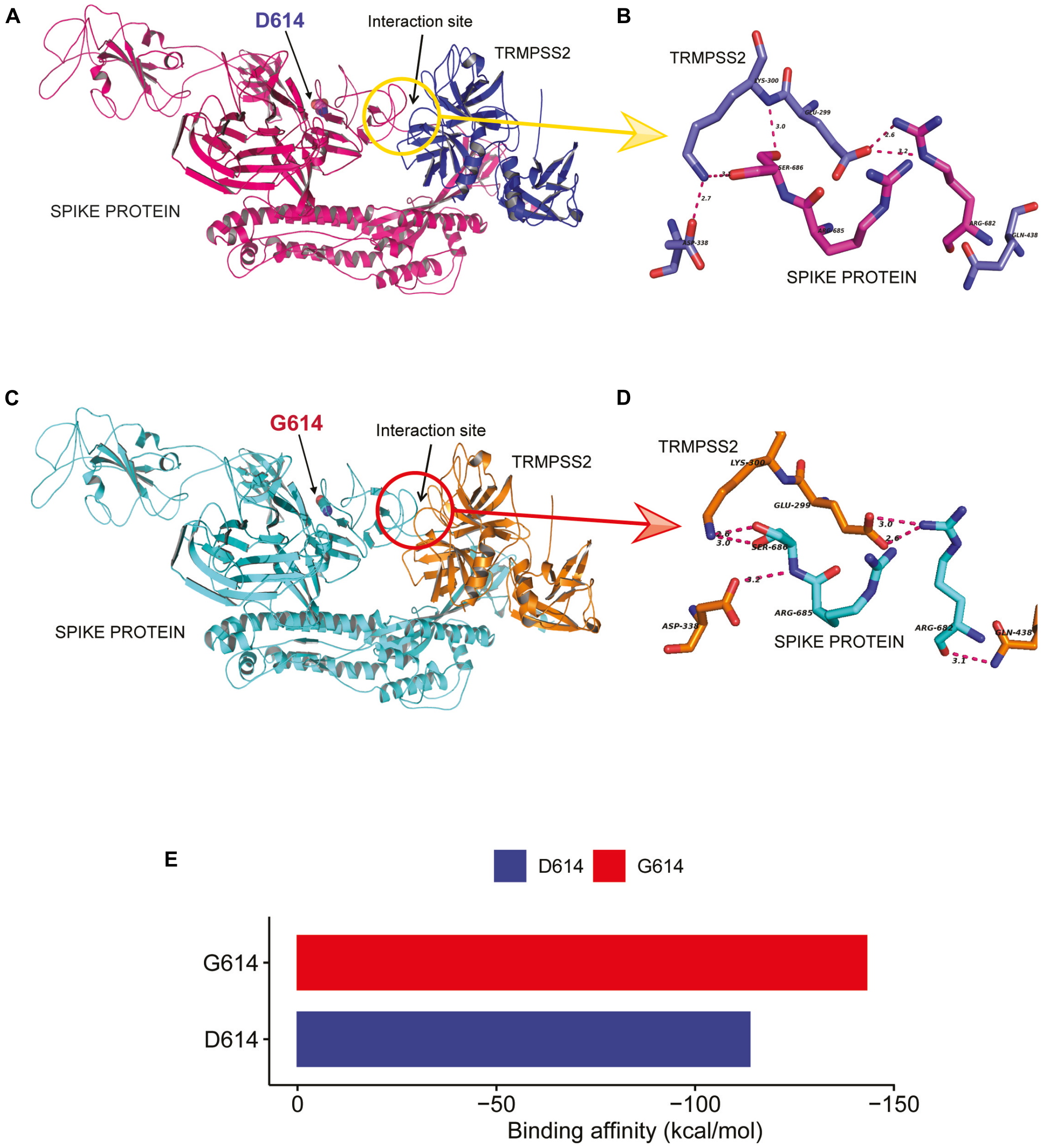


FIGURE 5 | D614G change in Spike protein enhanced TMPRSS2 protease interaction that might be responsible for increased virus infectivity. **(A–D)** The docking study of TMPRSS2 with the wild-type (D614) and mutant (G614) Spike protein. The interaction site and the mutation position (614) is marked with an arrow. The hydrogen bond interactions are shown in pink dotted lines with distance marked in Å. **(A)** The overview of the docking site location on WT Spike protein, **(B)** the interactions between TMPRSS2 and wild type, **(C)** the overview of the docking site location on WT Spike protein, **(D)** the interactions between TMPRSS2 and mutant Spike protein, and **(E)** the average binding energy (kcal/mol) values for the top poses selected from five different clusters.

and lower mortality rates in the beginning. Therefore, it was extremely important to understand the transmission dynamics of SARS-CoV-2 and its evolution during different phases of disease in India. The mutations that accumulate in the virus genome with time act as a molecular clock that can provide insight into emergence and evolution of the virus. These analyses could be helpful to prevent or control the transmission of the virus. To track the COVID-19 outbreak and understand the genomic clades of SARS-CoV-2 prevalent in India as compared to rest of the world, we performed the genome sequencing of oropharyngeal and nasopharyngeal swabs samples collected from individuals migrated from different regions of India to the state of Odisha after the reported incidences of COVID-19 pandemic in India.

For SARS-CoV-2 genome sequencing, we did amplicon-based sequencing for 225 SARS-CoV-2 genomes to capture the major clade diversity in India, of which we selected 202 SARS-CoV-2 sequences based on their assembly coverage (<5% gap) for further detailed analysis. We selected migrant groups from different parts of the country so that diversity of COVID-19 prevalent in the country could be captured in the phylogenetic analysis to understand disease transmission and virus evolution with time. The sequenced samples represent migratory populations from North, East, West, and Southern parts of India. Our analysis depicted that all the genomes analyzed in our study were grouped into four major clades, 19A, 19B, 20A, and 20B, according to the new Nextstrain clade nomenclature. Clade 19A is the Wuhan clade from China. Interestingly, we were able to capture occurrence of a rare clade with very less occurrence in India, i.e., 19B with 17% (n = 36) in our samples. This clade was found to be prevalent in East and Southeast Asia during the early outbreak of the pathogen. This confirmed that the foundation of the source of early outbreak infection in India also came from Southeast Asian countries. The migration information for these early collected samples were not well defined; therefore, it was difficult for us to pinpoint exactly which Southeast Asian country the transmission of 19B started. The mutation analysis depicted that both 19A and 19B clades almost evolved in parallel as prevalent mutations in both the clades are highly variable. On the other side, the phylogenetic analysis showed that clades 20A and 20B evolved quite rapidly in the Indian population and are a major source of disease transmission in the country, whereas the 20C strain is rarely detected and appeared to be less adapted or somehow contracted by contact tracing at early stages of infection. The haplotype network construction also pointed to the later strains belonging to 20A; 20B clades originated from Western Europe and transmitted directly or via Saudi Arabia are mostly prevalent in the southern and western part of India. In the clade, we also observed a very less frequent clade 20C, prevalent only in a handful of the Middle Eastern countries making up a marginal portion (n = 4) of our sequenced samples. This suggests the requirement of constant monitoring of SARS-CoV-2 with sequencing technology to understand the source of infection and design prevention mechanisms such as strategic lookdowns and region specific travel restrictions.

The fitness of the virus strain and its transmission depend on the adaptive mutations that it acquires with time. From initial observation, we have seen a 10.34% occurrence of C6312A, which has been associated with an India-specific clade called I/A3i (Banu et al., 2020). The mutation is dominant only India as the global frequency of the mutation is around 1.2% according to the Covid19 Beacon database. We found that four common variants, i.e., 241 C > T in the UTR region, 3,037 C > T in NSP3 gene, 14,408 C > T in the NSP12, and 23,403 A > G in S gene coevolved mostly in the 20A and 20B clades. As 20A and 20B clade frequency increased with time in the population, which indicates that these strains have some selective advantage with time for increased transmission. It has been reported that the leader sequence present in the UTR region of positive strand RNA viruses like SARS-CoV is important for the replication and strand switching to generate negative strands (Kim et al., 2020). This mutation in leader sequence is 20 nucleotides upstream of translation start site of ORF1ab gene. Therefore, it might be providing an advantage to virus in preventing stem-loop generation required during strand switching by RdRP, which needs further experimental evaluation. At the same time, several reports documented that 23,403 A > G (D614G) missense mutation in the Spike protein enhanced the infectivity rate of the virus. One of the reports showed that the shedding of S1 domain of Spike protein changes due to change in the hydrogen bonding between S1 and S2 domains. We observed in our protein modeling analysis that TMPRSS2 binding to Spike protein is enhanced by this mutation of aspartic acid to glycine (D614G) as it resulted in increased hydrogen bonding interactions. This change enhances the interaction of TMPRSS2 with the S2 domain, which is important for the cleavage of the S1 domain and virus entry into cells by facilitating its entry into the host cells. The overall primary *in silico* docking study showed that mutant Spike protein has a greater number of hydrogen bonds with TMPRSS2 at the cleavage site as compared to the wild type, resulting in better docking energy. Overall analysis indicates that the breakage of a hydrogen bond as a result of the mutation may facilitate greater cleavage of the mutant Spike protein as compared to the wild type. All the sequenced genomes were submitted in the GISAID database.

ODISHA COVID-19 STUDY GROUP (AUTHOR NAMES ARE ARRANGED IN ALPHABETICAL MANNER)

Arvind Kumar Singh, Baijayantimala Mishra, Banajini Parida, Binod Kumar Patro, D. P. Dogra, Dasarathi Das, Deepa Prasad, Dhaneswari Jena, Dharitri Mohapatra, Dinesh Prasad Sahu, Durga Madhab Satapathy, Durgesh Prasad Sahoo, Jayanta Panda, Jaya Singh Khatri, Kaushik Mishra, Manoranjan Satpathy, Nirupama Chaini, Roma Rattan, Sadhu Panda, Sangeeta Das, Somen Kumar Pradhan, Srikanta Kanungo, Sriprasad Mohanty, Subrata Kumar Palo.

ILS COVID-19 GROUP (AUTHOR NAMES ARE ARRANGED IN ALPHABETICAL MANNER)

Aditi Chatterjee, Adyasha Mishra, Ajit Kumar Singh, Amrita Ray, Ankita Datey, Aliva Minz, Ashish Yadav, Auromira Khuntia, Anshuman Dixit, Debyashreeta Barik, Deepak Singh, Eshna Laha, Hiren G. Dodia, Jeky Chawla, Kautilya Jena, Kaushik Sen, Niyati Das, Omprakash Shriwas, P. M. Vaishali, Parej Nath, Paritosh Nath, Prabhudutta Mamidi, Priyanka Mohapatra, Rahul Das, Rina Yadav, Sachikanta Rout, Saikat De, Sanchari Chatterjee, Sandhya Suranjika, Satyaranjan Sahoo, Shamima Ansari, Shifu Aggarwal, Shiva Pradhan, Sivaram Krishna, Sneha Dutta, Soumendu Mahapatra, Soumyajit Gosh, Subhabrata Barik, Sudhir Boral, Supriya Suman Keshry, Swatismita Priyadarshini, Tsheten Sherpa.

DATA AVAILABILITY STATEMENT

In the data availability statement the name of the Online repository is GISAID (https://www.gisaid.org/) accession ids are available in **Supplementary table 2**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Human Ethics Committee, Institute of Life Sciences. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AP, SP, SR, and JT planned and designed the study. AJ, SM, MP, VB, PS, BS, NS, DS, AD, SK, KA, SSm, and JS did the experiments. SC, GS, RD, SSe, RS, TB, and PP coordinated sampling and COVID-19 testing analysis. AG and SR did the genomic data analysis. SK and AD performed protein modeling and docking analysis. All authors have read and approved the manuscript.

FUNDING

Authors would like to acknowledge Department of Biotechnology, India for providing the funding and approvals under the umbrella of DBT's PAN-India 1000 SARS-CoV-2 RNA genome sequencing consortium, ILS intramural grant for supporting for sequencing resources and Odisha state authorities to provide the clinical samples for testing from where the virus RNA were isolated for sequencing analysis. PS and SM were supported by the DST inspire fellowship.

ACKNOWLEDGMENTS

We acknowledge GISAID, the genomic data sharing platform for influenza viruses and the contributors sequencing data used in this study. We gratefully acknowledge the efforts of volunteers who helped in coordinating the testing and related activities that supported implementation of this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.594928/full#supplementary-material

Supplementary Figure 1 | Patient demographics, transmission map and diversity of COVID-19 in India. **(A–D)** Represents the distribution of gender, age, clinical status and geographical location of sampled patients ($n = 225$). **(E–F)** Transmission map and Timetree of the 1,042 high coverage Indian SARS-CoV-2 whole genome sequences ($N < 1\%$) obtained from GISAID on 12th July, 2020.

Supplementary Figure 2 | ORF1ab Ct values in tested positive samples ($n = 637$) and haplotype network of sequenced data colored by clade. **(A)** ORF1ab Ct values in tested positive samples ($n = 637$) binned in their respective collection months. **(B)** Haplotype network of 202 SARS-CoV-2 whole genome sequences from our dataset colored by their respective clade. **(C)** Comparison of D614G mutation in global and Indian SARS-CoV-2 sequences.

Supplementary Figure 3 | Maximum likelihood tree of SARS-CoV-2 genomes ($n = 202$). **(A)** Maximum likelihood tree of SARS-CoV-2 genomes with 1,000 bootstraps and branch support values with =60 threshold.

Supplementary Table 1 | Information about patients, List of single nucleotide mutations and the annotated genes, corresponding amino acid changes and its predicted impact on protein function.

Supplementary Table 2 | Accession ids of GISAID submissions.

REFERENCES

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Banu, S., Jolly, B., Mukherjee, P., Singh, P., Khan, S., Zaveri, L., et al. (2020). A distinct phylogenetic cluster of Indian SARS-CoV-2 isolates. *Open Forum Infect. Dis.* 7, ofaa434. doi: 10.1093/ofid/ofaa434

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., Madden, T. L., et al. (2012). Domain enhanced lookup time accelerated BLAST. *Biol. Direct.* 7:12. doi: 10.1186/1745-6150-7-12

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Ferron, F., Subissi, L., Silveira, De Morais, A. T., Le, N. T. T., Sevajol, M., et al. (2018). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. U.S.A.* 115, E162–E171.

Graepel, K. W., Lu, X., Case, J. B., and Sexton, N. R. (2017). Proofreading-deficient coronaviruses adapt for increased fitness over long-term passage without reversion of exoribonuclease-inactivating mutations. *mBio* 8:e01503-17. doi: 10.1128/mBio.01503-17

Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., et al. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* 20:8.

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

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

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

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., Chang, H., et al. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell* 181, 914.e10–921.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.e19–827.e19.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34, 1812–1819. doi: 10.1093/molbev/msx116

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291. doi: 10.1107/s0021889892009944

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

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/s0140-6736(20)30251-8

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10. doi: 10.14806/ej.17.1.200

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

Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12:385.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Shen, M. Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15, 2507–2524. doi: 10.1110/ps.062416606

Shu, Y., and McCauley, J. (2017). GISAID: global initiative on sharing all influenza data - from vision to reality. *Eur. Surveill* 22:30494.

van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., et al. (2016). The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* 428, 720–725. doi: 10.1016/j.jmb.2015.09.014

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., Veesler, D., et al. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281.e6–292.e6.

Webb, B., and Sali, A. (2016). comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* 54, 561–5637.

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20:257.

World Health Organization (2020). *Coronavirus Disease Coronavirus Disease (COVID-19) Spreads*, Mbbs LR, D LBSM, Neurosurgery MS. Vol. 75. 95–97. Available online at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200423-sitrep-94-covid-19.pdf (accessed September 28, 2020).

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Rangarajan, E. S., Izard, T., et al. (2020). The D614G mutation in the SARS-CoV-2 Spike protein reduces S1 shedding and increases infectivity. *bioRxiv* [Preprint]. doi: 10.1101/2020.06.12.148726

Zhang, T., Wu, Q., and Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* 30, 1346.e2–1351.e2.

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 Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 26 February 2021
doi: 10.3389/fcimb.2021.632965

Clinical and Laboratory Profile of Hospitalized Symptomatic COVID-19 Patients: Case Series Study From the First COVID-19 Center in the UAE

Suad Hannawi[1*†], *Haifa Hannawi*[1,2], *Kashif Bin Naeem*[1], *Noha Mousaad Elemam*[3,4], *Mahmood Y. Hachim*[5†], *Ibrahim. Y. Hachim*[3,4†], *Abdulla Salah Darwish*[6] and *Issa Al Salmi*[7,8†]

[1] Department of Medicine, Ministry of Health and Prevention, Dubai, United Arab Emirates, [2] Department of Reserach, Ministry of Health and Prevention, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates, [3] Sharjah Institute for Medical Research, College of Medicine, University of Sharjah, Sharjah, United Arab Emirates, [4] Clinical Sciences Department, College of Medicine, University of Sharjah, Sharjah, United Arab Emirates, [5] College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates, [6] Gulff Medical School, Ajman, Gulf Medical University, Ajman, United Arab Emirates, [7] Department of Medicine, Oman Medical Specialty Board, The Royal Hospital, Muscat, Oman, [8] The Research Section, Oman Medical Speciality Board, The Royal Hospital, Muscat, Oman

OPEN ACCESS

Edited by:
Hassan Zaraket,
American University of Beirut, Lebanon

Reviewed by:
Salah Awaidy,
Ministry of Health, Oman
Riti Sharan,
Texas Biomedical Research Institute,
United States

***Correspondence:**
Suad Hannawi
suad1@ausdoctors.net

†ORCID:
Suad Hannawi
orcid.org/0000-0002-5689-6710
Mahmood Y. Hachim
orcid.org/0000-0002-7096-0731
Ibrahim Y. Hachim
orcid.org/0000-0001-6438-9455
Issa Al Salmi
orcid.org/0000-0002-3443-5972

Specialty section:
This article was submitted to
Clinical Microbiology,
a section of the journal
Frontiers in Cellular
and Infection Microbiology

Received: 24 November 2020
Accepted: 25 January 2021
Published: 26 February 2021

Citation:
Hannawi S, Hannawi H, Naeem KB, Elemam NM, Hachim MY, Hachim IY, Darwish AF and Al Salmi I (2021) Clinical and Laboratory Profile of Hospitalized Symptomatic COVID-19 Patients: Case Series Study From the First COVID-19 Center in the UAE. Front. Cell. Infect. Microbiol. 11:632965. doi: 10.3389/fcimb.2021.632965

Introduction: COVID-19 is raising with a second wave threatening many countries. Therefore, it is important to understand COVID-19 characteristics across different countries.

Methods: This is a cross-sectional descriptive study of 525 hospitalized symptomatic COVID-19 patients, from the central federal hospital in Dubai-UAE during period of March to August 2020.

Results: UAE's COVID-19 patients were relatively young; mean (SD) of the age 49(15) years, 130 (25%) were older than 60 and 4 (<1%) were younger than 18 years old. Majority were male(47; 78%). The mean (SD) BMI was 29 (6) kg/m^2. While the source of contracting COVID-19 was not known in 369 (70%) of patients, 29 (6%) reported travel to overseas-country and 127 (24%) reported contact with another COVID-19 case/s. At least one comorbidity was present in 284 (54%) of patients and 241 (46%) had none. The most common comorbidities were diabetes (177; 34%) and hypertension (166; 32%). The mean (SD) of symptoms duration was 6 (3) days. The most common symptoms at hospitalization were fever (340; 65%), cough (296; 56%), and shortness of breath (SOB) (243; 46%). Most of the laboratory values were within normal range, but (184; 35%) of patients had lymphopenia, 43 (8%) had neutrophilia, and 116 (22%) had prolong international normalized ratio (INR), and 317 (60%) had high D-dimer. Chest x ray findings of consolidation was present in 334 (64%) of patients and CT scan ground glass appearance was present in 354 (68%). Acute cardiac injury occurred in 124 (24%), acute kidney injury in 111 (21%), liver injury in 101 (19%), ARDS in 155 (30%), acidosis in 118 (22%), and septic shock in 93 (18%). Consequently, 150 (29%) required ICU admission with 103 (20%) needed mechanical ventilation.

Conclusions: The study demonstrated the special profile of COVID-19 in UAE. Patients were young with diabetes and/or hypertension and associated with severe infection as shown by various clinical and laboratory data necessitating ICU admission.

Keywords: COVID-19, acute respiratory distress syndrome, clinical characteristics, laboratory features, organ failure, mechanical ventilation

INTRODUCTION

Corona virus disease -2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first detected in Wuhan-China on December 31st, 2019 (Wang C et al., 2020). Shortly after, the World Health Organization (WHO). declared COVID-19 as a global pandemic on March 2020 (Mahase, 2020). As on January 19th, 2021 the total global number of affected cases was 96,073,719 million and 2,051,377 deaths Available at: https://www.worldometers.info/coronavirus/. In UAE A Chinese family on holiday in the UAE were the first people in the country to be given positive coronavirus diagnoses on January 23rd, 2020 (Zaatar,), Thereafter, the cases increased to reach 63,819 as on August 14th 2020 (The Supreme Council for National Security,).

SARS-CoV-2 is highly contagious (Mohapatra et al., 2020) and one of way to control the pandemic is through preventive methods, sensitive diagnostic approaches, and using accessible drugs, in a view of understanding the characteristics of COVID-19 among different population (Lotfi et al., 2020).

Although the majority of the patients have mild symptoms, COVID-19 can be fatal and require intensive medical care (Sheleme et al., 2020). COVID-19 severity and fatality have been found to be associated with many host factors, including age, gender, race, ethnicity and presence of other comorbidities (Biswas and Mudi, 2020). Moreover, it had been reported that COVID-19 mortality rates differ between countries and between geographical areas within the same country with genetic factors and climates differences as an etiology for this variation (Biswas and Mudi, 2020; Al-Tawfiq et al., 2020). Report showed that Asian individuals had an increased risk of infection and worse clinical outcomes and higher mortality (Pan et al., 2020). Therefore, this study aimed at studying the clinical and the laboratory characteristics of symptomatic hospitalized COVID-19 patients that had been admitted to the main federal hospital in Dubai, United Arab Emirates (UAE). Understanding COVID-19 profile in each country will help in containing the disease and to set proper guidelines and strategies to deal with any subsequent waves.

METHODS

Study Design

This clinical observational study included 525 hospitalized symptomatic confirmed COVID-19 patients admitted to Al Kuwait-Dubai hospital (AKH), Dubai, United Arab Emirates (UAE) between February until August, 2020. The AKH is the only federal hospital operated by the Ministry of health and Prevention (MOHAP) in Dubai, and it is the first hospital in the UAE that had been evacuated from all other cases and had been declared as a COVID-19 center. Up to the date of writing this manuscript, AKH is still function as a COVID-19 center.

Data Collection

COVID-19 had been confirmed using polymerase chain reaction (PCR) from nasal swab sample. The laboratory used Sacace Real Time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) test to diagnose COVID-19 disease. The test was performed on patients' nasopharyngeal swabs. RNA was extracted using SaMag Viral Nucleic Acid Extraction system. Extracted RNA was amplified using BGI-Real Time Fluorescent RT-PCR kit for the detection of COVID-19. Suspected cases were admitted to a different specialized hospital devoted for unconfirmed cases.

MOHAP hospitals use electronic medical file system where all clinical, laboratory, radiological and management data are collected prospectively.

Collected data included (1) demography; age, gender, body mass index (BMI); calculated as weight(KG)/height(m^2), (2) source of infection; travel to overseas, contact with another COVID-19 patient, or unknown source (identified as either no travel history and no contact with another known positive case of COVID-19), (3) smoking status (current or ex-smoker), (4) comorbidities; diabetes mellitus (DM), hypertension, cardiovascular disease (CVD), chronic lung disease, chronic kidney disease (CKD), stroke/transient ischemic attack (TIA), cancer, and any other documented comorbidities (5) COVID-19 symptoms; (A) duration of symptoms (days), (B) upper respiratory tract infection symptoms; headache, fever, fatigue, myalgia, rhinorrhea, and sore throat (C) lower respiratory tract infection symptoms; cough, shortness of breath (SOB), sputum production, and hemoptysis, (D) other symptoms; ageusia, anosmia, anorexia, nausea, vomiting, diarrhea, and confusion, (6) radiological features; chest X-ray consolidation and computed tomography (CT) ground glass appearance, (7) clinical complications such as (A) presence of organ failure; (i) acute cardiac injury (defined as at least one documented elevated high-sensitivity troponin-I) (ii) acute kidney injury; defined as a raise in the serum creatinine of ≥26.5 μmol/liter within 48 h.; or increase in serum creatinine ≥1.5 times than the baseline that occurred within the preceding 7 days; or if there were reduction

Abbreviations: ARDS, Acute Respiratory Distress Syndrome; ALT, Alanine Transaminase; AST, Aspartate Aminotransferase; BMI, Body Mass Index; COVID-19, Corona Virus Disease-2019; CRP, C-Reactive Protein; CVD, Cardiovascular Disease; DM, Diabetes Mellitus; eGFRigure, Estimated Glomerular Filtration Rate; ICU, Intensive Care Unit; INR, International Normalized Ratio; IL-6, Interlukin-6; LDH, Lactate Dehydrogenase; LRTI, Lower Respiratory Tract Infection; SOB, Shortness of Breath; URTI, Upper Respiratory Tract Infection; WCC, White Cell Count.

in the urine volume <0.5 ml/kg/hr. for 6 consecutive h, and (iii) acute liver injury; defined as presence of high alanine transaminase (ALT) and/or high aspartate aminotransferase (AST) by more than 5 times the upper limit of normal range, (iv) acute respiratory distress syndrome (ARDS); using Berlin definition (Force et al., 2012), acidosis, (5) septic shock; defined as per the Third International Consensus Definitions for Sepsis and Septic Shock (Singer et al., 2016), (6) need for intensive care unit (ICU) medical care, (7) need for mechanical ventilation, and (8) death rate.

We followed the Chinese CDC report in categorizing the COVID-19 clinical manifestations to 3 levels of severity; (1) Mild disease; if there were non-pneumonia or if there were mild pneumonia, (2) Severe if there were dyspnea, respiratory rate ≥ 30/min, blood oxygen saturation (SpO2) ≤ 93%, PaO2/FiO2 ratio [the ratio between the partial pressure of oxygen (PaO2) and the fraction of inspired oxygen (FiO2)] < 300, and/or lung infiltrates > 50% within 24 to 48 h, and (3) Critical, if there were respiratory failure, septic shock, and/or multiple organ dysfunction/failure (Wu and McGoogan, 2020).

The medications used in treating COVID-19 during their admission are as follow: Chloroquine/Hydroxychloroquine, Lopinavir/Ritonavir, antibiotics, Steroid, Interferon, Tocilizumab, and anti-fungal.

The laboratory tests were collected and include: (1) blood rheology; hemoglobin (Hb) (normal reference range; NR: male: 12.0-15.5 g/dl, and female: 13.5 -17.5 g/dl), total white cell count (WCC) (NR: 4,000 -11,000 x10(3)/mcl), lymphocyte count (NR: 1,000 and 4,800 x10(3)/mcl), neutrophil count (NR: 1.5-4 x10 (3)/mcl), and platelet count (NR: 150,000 - 400,000 x10(3)/mcl), (2) inflammatory markers; ferritin (NR: males: 26-388 ng/mL, and female: 8-252 ng/mL), C-reactive protein (CRP) (NR: 0.0-5.0 mg/liter), procalcitonin (NR < 0.1 ng/mL), lactate dehydrogenase (LDH) (NR: Male: 85 – 227 U/liter, and Female: 81 – 234 U/liter) and lactic acid (NR: 0.4 – 2.0 mmol/liter), (3)glucose status; glycosylated hemoglobin; HbA1C (NR: 4.8 - 6.0%), (3) liver function test; ALT (NR: males: 16 – 63 IU/liter, and females:14 – 59 IU/liter), AST (NR: 15 – 37 U/liter), albumin (34 – 50 g/liter), bilirubin (NR: 0-3 μmol/liter), and alkaline phosphatase level (NR: 46-116 U/liter) (4) renal function test, included, blood urea (NR: 0.0-8.3 mmole/liter), and creatinine (NR: 44-133 μmole/liter). Estimated glomerular filtration rate (eGFR) (NR: 90 to 120 mL/min/1.73 m^2) had been calculated using Modification of Diet in Renal Disease (MDRD) equation, 186 × (SCr mg/dl)-1.154 × (age)−0203 × 0.702 [if Female] x 1.212[if Black], (5) Coagulation profile included International normalized ratio (INR) (NR <1.1), and D-dimer (>55 mg/liter considered positive) (6) electrolytes included sodium (Na) level (NR: 135-145 mEq/liter) and potassium (K) level (NR: 3.6 to 5.2 mmol/liter), and (7) indicator of cardiac injury; troponin level (NR: 0 to 60.4 ng/liter).

Further, laboratory parameters had been categorized based on the upper and lower limits of the normal reference ranges to lymphopenia if lymphocyte count was < 1 x10^3/mcL, neutrophilia if neutrophil counts were >14 x10(3)/mcL, high ALT if were > 63 IU/liter for male and >59 IU/liter for female, high AST if were > 37 U/liter, high LDH if were > 227 IU/liter for males and > 234 IU/liter for females, high ferritin if were > 388 ng/ml for males and were > 252 ng/ml for female, high D-dimer if were > 0.5 mg/liter, high troponin if were > 60 ng/liter, high lactic acid if were >2.0 mmol/liter, and high procalcitonin if were > 0.1 μg/liter.

Statistical Analysis

Data had been presented as mean and standard deviation for continuous variables and frequency (number and percentage; %) for categorical variables. To assess the differences between COVID-19 patients needed ICU admission vs. no ICU admission Student's t-test was used for the continuous variables and Chi-square test was used for the categorical variables. P value <0.05 had been considered significant. All the analysis had been carried out using STATA 9/SE statistical software (Stata Corp, College Station, Texas, USA).

Ethical Approval

The study was approved by the Scientific Research Committee. Approval Number. MOHAP/DXB-REC/MMM/NO.44/2020 and certify that the study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments ethical standards.

RESULTS

Demography

The basic clinical and epidemiological features of this cohort is shown in **Table 1**. While the majority of the patients were from the indicant subcontinent; 332 (63%), the UAE citizens made 53 (10%) and the other Arabs were 75 (14%). The mean (SD) of the age was 49 ± 15 years, with only 130 (25%) of the patient were older than 60 years old, and 4 (0.76%) were younger than 18

TABLE 1 | Demography of 525 symptomatic COVID-19 patients.

Variables	
Nationalities	
United Arab Emirates	53 (10%)
Other Arabs	75 (14%)
Iranian	10 (2%)
Indian Subcontinent	332 (63%)
South East Asia	35 (7%), out of which 2 (0.4%) were Chinese
African	5 (1%)
European	13 (2.5%)
South American	2 (0.4%)
Age*	49 (15)
Age > 60 yrs.	130 (25%)
Age <18 yrs.	4 (0.76%)
Gender (Females: Males)	118: (22%): 407 (78%)
BMI (kg/m^2)*	29 (6)
Smokers (current or ex-smoker)	25 (5%)
Symptoms duration (days)	6 (3)*
Source of COVID-19 infection	
Travel history	29 (6%)
Contact history	127 (24%)
No known source for COVID-19	369 (70%)

Data presented as number (%) or as mean (SD; standard deviation), >; more than, <; less than.*

years old. Male composed the majority of the patients with ratio of female to male was 118: (22%) to 407 (78%). The mean (SD) of the BMI was 29 ± 6 kg/m^2 for all COVID 19 patients.

The source of infection transmission was unknown for the majority of the patients 369 (70%) with only 29 (6%) and 127 (24%) had history of travel to foreign overseas country and history of contact with another COVID-19 patient, respectively. Smoking history; either current smoking or history of smoking presented in 25 (5%) of patients. At presentation to the hospital the mean (SD) of COVID-19 symptoms duration was 6 ± 3 days.

Comorbidities

While 284 (54%) of the patients had at least one comorbidity, 241 (45%) had none. The most common comorbidities were DM (177; 34%) and hypertension (166; 32%). CVD was present in 27 (5%) of patients. Small percentage of patients had thyroid dysfunction (16; 3%), chronic lung disease (14; 3%), and chronic kidney disease (13; 2%), as shown in **Table 2**.

Clinical Symptoms

The most common symptoms were fever (340; 65%), cough (296; 56%), and SOB (243; 46%). The cough was mainly dry with only 10 (2%) had sputum production with no hemoptysis. Small percentage of patients had myalgia (62; 12%), sore throat (44; 8%), headache (30; 6%) and fatigue (34; 6%). Less than and/or equal to 5% of patients had anorexia (14; 3%), rhinorrhea (18; 3%), nausea (11; 2%), vomiting (10; 2%), and diarrhea (27; 5%). 2 (0.4%) of patients had anosmia, and 4(1%) had ageusia. None of the patients had confusion, as shown in **Table 3** and in **Figure 1**.

Laboratory Parameters and Radiological Features

While most of the blood rheology parameters values (Hb, WCC, and platelet count) were within normal range at time of first encounter to the hospital,184 (35%) of the patients had lymphopenia and 43 (8%) had neutrophilia. The coagulation profile showed 116 (22%) of the patients had prolonged INR, and 317 (60%) had high D-dimer.

TABLE 2 | Comorbidities of 525 symptomatic COVID-19 patients.

Comorbidities	No		Yes	
	n	%	n	%
Presence of any comorbidities	241	45.90	284	54
Diabetes mellitus	348	66	177	34
Hypertension	359	68	166	32
Cardiovascular disease	498	95	27	5
Thyroid dysfunction	509	97	16	3
Chronic lung disease	511	97	14	3
Chronic kidney disease	512	98	13	2
Stroke	521	99.24	4	0.76
Cancer	522	99.43	3	0.57
Epilepsy	524	99.8	1	0.2
Multiple sclerosis	524	99.8	1	0.2
Autoimmune rheumatic diseases	523	99.6	2	0.4
Thalassemia	524	99.8	1	0.2

Data presented as n; number and %; percentage.

TABLE 3 | Symptoms of 525 symptomatic COVID-19 patients.

Symptoms	NO		YES	
	n	%	n	%
Fever	185	35	340	65
Cough	229	44	296	56
Fatigue	491	94	34	6
Anorexia	511	97	14	3
Shortness of breath	282	54	243	46
Sputum production	515	98	10	2
Myalgia	463	88	62	12
Headache	495	94	30	6
Confusion	0	0	0	0
Rhinorrhea	507	97	18	3
Sore throat	481	92	44	8
Hemoptysis	0	0	0	0
Vomiting	515	98	10	2
Diarrhea	497	95	27	5
Nausea	514	98	11	2
Anosmia	523	99.6	2	0.4
Ageusia	521	99	4	1

Data presented as n; number and %; percentage.

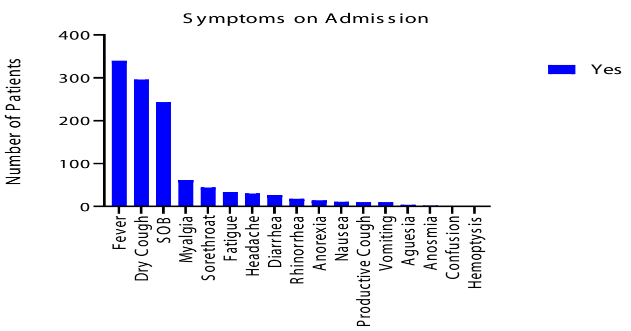


FIGURE 1 | The distribution of the 525 symptomatic COVID-19 patients on admission.

Radiological investigations at admission showed bilateral chest x ray consolidation was present in 334 (64%) of patients and CT scan ground glass appearance was present in 354 (68%), as shown in **Table 4** and **Figure 2**.

Clinical Complications and Outcome

Despite that most of renal and liver function tests were within normal range, 111 (21%) had acute kidney injury and 101 (19%) had liver injury. Also, 124 (24%) of COVID-19 patients had acute cardiac injury. In addition, ARDS developed in 155 (30%), acidosis in 118 (22%) and septic shock in 93 (18%) of patients. As a result, 150 (29%) required ICU admission with 103 (20%) needed mechanical ventilation. Eventually, 93 (18%) of the patients deceased, as shown in **Table 5** and **Figures 3** and **4**.

Medications Used in the Treatment of COVID-19 Patients

Drugs used to treat patients as per National (UAE) Guidelines for Clinical Management and Treatment of COVID-19. 20 April

TABLE 4 | Laboratory features of 525 symptomatic COVID-19 patients.

	mean	SD
Complete blood count		
WCC [×10(3)/mcl]	8.44	4.05
Neutrophils count, [×10(3)/mcl]	6.39	3.88
Neutrophilia >14 [×10(3)/mcl], n(%)*	43 (8%)	
Lymphocytes count, [×10(3)/mcl]	1.36	0.74
Lymphocytes <1 [×10(3)/mcl] n (%)*	184 (35%)	
Hb (g/dl)	13	2
Platelet,[×10(3)/mcl]	250	102
Coagulation profile		
INR	1.07	0.14
Prolong INR*	116 (22%)	
D-dimer (mg/liter)	2.67	5.97
High D-dimer, n(%),(mg/liter)*	317 (60%)	
Electrolytes and renal profile		
Sodium (Na) (mmole/liter)	136.29	4.72
Potassium (K) (mmole/liter)	4.06	0.64
Urea (mmole/liter)	6.27	5.67
Creatinine (μmole/liter)	117.63	317.32
eGFR (ml/min)	86.00	32.16
Liver function test		
Serum bilirubin (μmol/liter)	16.03	39.21
ALT (IU/liter)	67.76	120.38
AST (U/liter)	58.85	116.11
High ALT/AST (IU/liter/U/liter), n(%)*	281 (54%)	
Alkaline phosphatase (IU/liter)	88.34	54.14
Albumin (g/liter)	30.77	7.21
Inflammatory markers		
CRP mg/liter	82.73	97.53
High CRP>3mg/liter, n(%)*	447 (85%)	
Procalcitonin (ug/liter)	0.82	4.03
High procalcitonin, n(%),(ug/liter)*	206 (39%)	
LDH (IU/liter)	417.834	323.13
High LDH, n(%),(IU/L)*	393 (75%)	
Lactic acid (mmole/liter)	1.62	0.89
High lactic acid, mmole/liter, n(%)*	49 (9%)	
High ferritin, n(%),(mcg/liter)*	291 (55%)	
Ferritin (mcg/liter)	1049.50	1429.78
Other parameters		
Troponin (ng/liter)	278.62	2744.61
High troponin (ng/liter)*	69 (13%)	
HbA1C (%)	7.08	2.41

Data presented as mean and SD; or as n; number (%; percentage), standard deviation, WCC; white cell count, Hb; hemoglobin, INR; International normalized ratio, eGFR; estimated glomerular filtration rate, ALT; alanine transaminase, AST; aspartate aminotransferase, CRP; C-reactive protein, LDH; lactate dehydrogenase, HbA1C; glycosylated hemoglobin.*

2020, version 3.1. The guidelines recommend different classes of drugs according to the clinical status. For laboratory-confirmed COVID-19 patients without pneumonia, hydroxychloroquine/chloroquine with lopinavir-ritonavir is recommended. In the presence of radiological evidence of pneumonia, favipiravir and remdesivir is considered additionally. In case of critical illness, tocilizumab and interferon-alfa 2 is considered additionally. Steroids are used in the presence of significant hypoxia with high inflammatory markers. Antibiotics are added when a super-added bacterial infection is suspected either clinically or positive cultures with high lab markers like CRP and procalcitonin. Antifungals are used if clinical suspicion of fungal infection or positive cultures.

This study showed that Chloroquine or hydroxychloroquine were the most commonly used medications in 493 (94%) of patients, followed by anti-viral; lopinavir/ritonavir in 380 (72%). Another anti-viral treatment used was Favipiravir which was used in 181 (34%) of patients. Anti-biotics had been prescribed in 372 (71%) of patients. Steroid used in over half of the patient; 298 (57%). Interferon, Tocilizumab (IL-6 inhibitor) and anti-fungal medications had been used in a smaller number of cases; 137 (26%), 105 (20%), and 59 (11%), respectively, as shown in **Table 6**.

COVID-19 Patients Needed ICU Care vs. COVID-19 Patients Did Not Need ICU Care

Among all the COVID-19 patients 150 (29%) required ICU admission without any prediction for any nationality to be admitted to the ICU. COVID-19 patients who needed ICU admission were older in age 55 ± 13 vs. 46 ± 15 years ($p < 0.001$), with 53 (35%) were older than 60 years compared to 77 (21%) of patients who did not required ICU care. Although male gender was the majority of COVID-19 patient but the percentage was higher among the group that required ICU care; 139 (93%) vs. 268 (21%), $p < 0.001$. More, they were more diabetic; 79 (53%) vs. 98 (26%) and hypertensive; 74 (49%) vs. 92 (25%) (p value < 0.001 for both) (**Tables 7** and **8**).

DISCUSSION

To the best of our knowledge, this is the largest study from our region of the Gulf Cooperation Council (GCC) to examine both the clinical and laboratory features of admitted COVID 19 patients. It showed that patients were young, mostly male and overweight or obese. Majority were teetotal and their infection source was unknow in majority of patients with average symptoms duration of 6 days. More than 50% had either diabetes or hypertension or both as a medical comorbidity. Majority had typical LRTI symptoms and signs at presentation. Also, radiological findings were very common including ground glass appearance. Multi organ failure was noted and mostly include AKI, ALI, and ACI with almost a third needed ICU management. Management strategies were instituted as experience progressed with majority received including Chloroquine or hydroxychloroquine, anti-viral, antibiotics, steroids, and various biological medications.

The excess number of males in our sample is ongoing with what had been reported of increase vulnerability of men to COVID-19. Bwire et al. attributed this to the different ability of each gender to fight SARS-2-CoV-2 as a result of biological differences in the immune systems between men and women (Bwire, 2020). As well, it had been reported that females are more resistant to infections than men, and this is mediated by several factors including sex hormones and high expression of coronavirus receptors (ACE 2) in men (Bwire, 2020). More, life style differences between men and women such as smoking and drinking that are more common among men. As well, women have more responsible attitude toward the Covid-19 pandemic than men. Others reported that as men show higher mortality from diseases including heart disease and diabetes, then these diseases which are known to show sex-specific occurrence could be contributing factors for the sex-biased mortality from COVID-19 (Pradhan and Olsson, 2020).

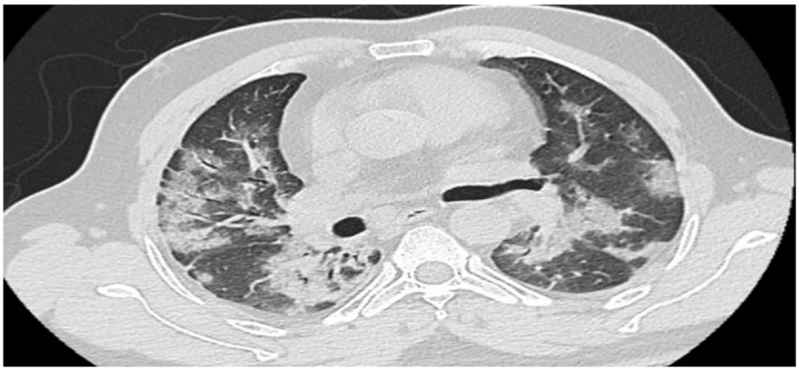


FIGURE 2 | Bilateral multilobar ground-glass opacification and consolidations, preferentially peripheral and subpleural.

TABLE 5 | Clinical complications and clinical outcomes of 525 symptomatic COVID-19 patients.

Clinical complications and outcomes	No		YES	
	n	%	n	%
Severity				
Mild to moderate			181	34
Sever			197	38
Critical			147	28
Radiological features				
Bilateral chest X-ray consolidation	191	36	334	64
CT ground glass appearance	171	33	354	68
Organ failures and complications				
Acute cardiac injury	401	76	124	24
Acute kidney injury	414	79	111	21
Liver injury	424	81	101	19
Acidosis	407	78	118	22
Septic shock	432	82	93	18
ARDS	370	70	155	30
Mechanical ventilation	422	80	103	20
ICU admission	375	71	150	29
Death	432	82	93	18

Data presented as n; number and %; percentage, CT; computed tomography, ARDS; acute respiratory distress syndrome, ICU; intensive care unit.

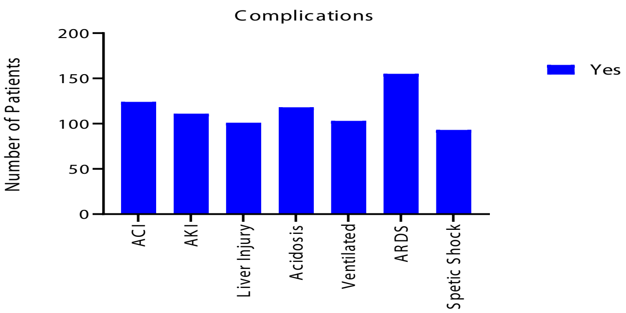


FIGURE 3 | Clinical complications of the 525 symptomatic COVID-19 patients.

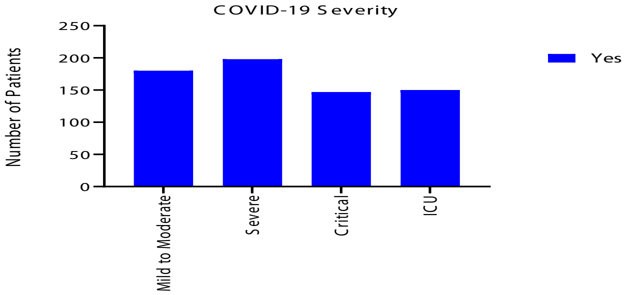


FIGURE 4 | Clinical severity of the 525 symptomatic COVID-19 patients.

TABLE 6 | Treatment of 525 symptomatic COVID-19 patients.

Medications	No		YES	
	n	%	n	%
Chloroquine/hydroxychloroquine	32	6	493	94
Lopinavir/ritonavir	145	28	380	72
Favipiravir	344	66	181	34
Anti-biotics	153	29	372	71
steroid	227	43	298	57
Interferon	388	74	137	26
Tocilizumab (Interlukin-6; IL-6 inhibitor)	420	80	105	20
Anti-fungal	466	89	59	11

Data presented as n; number and %; percentage.

Despite older age had been related to worse clinical outcomes in COVID-19 (Liu Y et al., 2020), the average age of our COVID-19 showed that they are relatively young (37.0 ± 13.0 years) compared to what had been reported in Oman (Khamis et al., 2020) and 41 years in Kuwait (Singh, 2020). As well, there were only 4 (0.76%) of our study patients below the age of 18 years which is close to what had been observed from the first cases reported in Wuhan where they reported no clinical cases in children below 15 years of age (Li et al.,

TABLE 7 | Nationalities of the 525 symptomatic COVID-19 patients and admission to the intensive care unit (ICU).

	Total number of cases	ICU care (no)	ICU care (yes)
UAE nationals	53	42 (79%)	11 (21%)
Other Arabs	75	55 (73%)	20 (27%)
Iranian	10	8 (80%)	2 (20%)
Indian subcontinents	332	226 (68%)	106 (32%)
Chinese	2	2 (100%)	0 (0%)
African	5	5 (100%)	0 (0%)
European	13	12 (92%)	1 (8%)
South East Asian	33	23 (70%)	10 (30%)
South America	2	2 (100%)	0 (0%)

Data presented as number (%). P value 0.24.

TABLE 8 | Age, gender, body mass index, symptoms duration and comorbidities of 150 symptomatic COVID-19 patients needed Intensive care unit (ICU) admission vs. 375 did not need ICU admission.

	ICU care (no) (n = 375)	ICU care (yes) (n = 150)	P value
Age (years)*	46 (15)	55 (13)	<0.001
Age > 60 years	77 (21%)	53 (35%)	<0.001
Gender (male: female)	268 (71%): 107 (29%)	139 (93%):11 (7%)	<0.001
BMI (kg/m2)*	29 (6)	29 (5)	0.878
Symptoms duration*	5.6 (3)	5.7 (3)	0.735
Comorbidities			
Cardiovascular disease	15 (4%)	12 (8%)	0.061
Diabetes	98 (26%)	79 (53%)	<0.001
Hypertension	92 (25%)	74 (49%)	<0.001
Stroke	3 (0.8%)	1 (0.7%)	0.874
Cancer	2 (0.5%)	1 (0.7%)	0.855
Chronic lung disease	9 (2%)	5 (3%)	0.549
Chronic kidney disease	8 (2%)	5 (3%)	0.424
Smoking history	17 (5%)	8 (5%)	0.697
Thyroid disease	12 (3%)	4 (3%)	0.748

Data presented as number (%) or as mean (SD; standard deviation).*

2020). Further, other reports showed that COVID-19 appears to be less severe in children (Rezaei, 2020). UAE is a young nation with majority of its population between age 25 to 54 years (Statistic center D,).

A systemic review demonstrated that proportion of male patients ranged from 29.0% to 77.0 (Fu et al., 2020). Other studies showed that there were no significant gender differences in prevalence of COVID-19 (Li et al., 2020), but men are more at risk for worse COVID-19 outcomes and death, independent of their age (Jin et al., 2020). In our study the majority of cases were male; 407 (78%). This could be explained by having in our study symptomatic COVID-19 patients who had been admitted to the hospital, therefore, our sample sized represent the worst sector of patients who required medical care for their disease.

Among our patients 5% had a history of smoking; either current or past. This is going with what had been reported in systemic review that the proportion of COVID-19 patients who were current smokers ranged from 0.0% to 18.0% (median 7.2%; nine studies) (Fu et al., 2020). Further, smoking is culturally not accepted in our region and is faced by a stigma in a conservative culture (Saravanan et al., 2019; Ayalon, 2019). Or, because of the stigma patients were not truthful revealing their smoking status.

The mean BMI of COVID-19 patients were 29 ± 6 kg/m^2, which lies within the overweight class as per the WHO classification (Pi-Sunyer, 2000). Population and patients with high BMI have moderate to high risk of medical complications with COVID-19 (Malik et al., 2020), with increased adiposity destabilizes the pulmonary function and contribute to viral pathogenesis (Dorner et al., 2010).

The source of infection was known in about quarter of our COVID-19 patients while the majority of our symptomatic patients had no known source of infection. The unknown source of infection in the majority of our patient is in line with our previous finding (under review) that highlighted the adopted policy of paying attention to patients with travel or contact history resulted in unintentional negligence of other COVID-19 patients who had no history of travel or history of contact with another COVID-19 case. This strategy contributed to a worse clinical outcome among the COVID-19 patients with no obvious source of contracting the disease.

Additionally, the unknown source of infection can be explained by the transmission efficiency of SARS-CoV-2 that had proved to be high (Fauci et al., 2020). An infected individual can release aerosols and droplets containing SARS-CoV-2 by coughing, sneezing, speech and breathing (Chao et al., 2009; Tang et al., 2011). More, Aerosols (< 10-µm diameter) and droplets (> 10-µm diameter) can promote infection through deposition on surfaces and subsequent hand-to-mouth/nose/eye transfer, and through inhalation (van Doremalen et al., 2020).

Previous publications from other parts of the world revealed that pre-existing conditions (e.g., cardiovascular, pulmonary, and renal diseases) render a person more vulnerable to more severe COVID-19 infection (Zhou et al., 2020). Among our patient the most common comorbidities were DM and hypertension. Indeed, DM, hypertension and obesity are the most prevalent comorbidities in the UAE and its neighboring countries (Alharbi et al., 2014; Tailakh et al., 2014).

The most common symptoms among our patients were fever (65%), cough (56%), and SOB (46%). This is contradicting the results of a systemic review which showed fever, sore throat, and muscle soreness or fatigue as the most common symptoms (Sun et al., 2020). This might raise the alert that the manifestations of COVID-19 vary across different countries and different nations. Therefore, in our region more attention might be needed to be paid to LRTI symptoms than to the URTI symptoms.

In the early stage of COVID-19, decreased lymphocyte count had been reported and had been demonstrated as a negative prognostic factor (Cascella et al., 2020). The presence of lymphopenia had been seen in more than third of our patients; precisely in 35%. The relatively high procalcitonin level in 39% of our patients might indicate the presence of secondary bacterial infection. Presence of high CRP in 85%, high LDH in 75%, and high ferritin in 55% going with the inflammatory status of the disease. As well, presence of the high ALT, AST, and troponin level is aligned with the presence of the underlying organ dysfunction.

Although at early stage of COVID-19, the chest x ray might be normal, the presence of bilateral changes in 334 (64%) of our COVID-19 patients could be because of presence of patients to the hospital after 6 days of disease onset. And hence the CT scan ground glass appearance was seen in 354 (68%) of the patients. Yet, the CT changes are less than what had been reported in other studies where it was found to be 96.6% (Sun et al., 2020).

Up to the time of writing this manuscript, there is no specific antiviral treatment recommended for COVID-19. Yet, several medications have been proposed such as Lopinavir/Ritonavir (Bimonte et al., 2020) that we used in 94% of our patients. Despite that Lopinavir/Ritonavir showed some benefit is some studies, other demonstrated no benefit with Lopinavir/Ritonavir treatment compared to standard care (Cao et al., 2020). Favipiravir is another anti-viral agent that demonstrated efficacy, improve the discharging rate and decrease the mortality rate of COVID-19 patients (Wang Z et al., 2020), and had been used in 34% of our patients. Chloroquine and hydroxychloroquine were proposed as immunomodulatory therapy in COVID-19 and it was the most commonly used treatment in our patients; 94%. Chloroquine/hydroxychloroquine had been used massively in our patients despite the conflict results about its efficacy, with some studies support its use (Liu J et al., 2020) and other studies antagonize it use in COVID-19 (Kim et al., 2020). Corticosteroids is another drug that introduced as COVID-19 therapy; corticosteroid dexamethasone was suggested to have anti-inflammatory and immunosuppressive roles (Selvaraj et al., 2020; Theoharides and Conti, 2020). Steroid had been used in about in 57% of our patients.

Antibiotic is another medication that had been used in 71% of our patients based. The excess use of antibiotics in our patients is a reflection of anxiety, fear, and uncertainty surrounded the pandemic and the absence of anti-viral medication with proven efficacy. As well, AKH was the first hospital in UAE to be declared as a COVID-19 center, therefore, employee of AKH were at the frontline of fighting the first reported cases of COVID-19 in the UAE. More, fever and cough were the most common symptoms in our patients and the presence of the radiological infiltrates presented in more than 60% of the patients. Fever, cough and radiological infiltrate are hallmarks of bacterial community-acquired pneumonia which requires antibiotic treatment (Huttner et al., 2020), Hence, this contributed to excess use of antibiotics and some other medications in our COVID-19 patients.

The excess use of antibiotics is supported by the fact that literature does not indicate that antibiotics are effective in treating COVID-19 (Zhou et al., 2020), and the incidence of bacterial coinfections appears low among COVID-19 (Rawson et al., 2020). Rawson et al. reported that although among their COVID-19 patients only 8% experienced a bacterial or fungal coinfection, 72% received antibiotics. a percentage that is similar to what we found in our COVID-19 patients (Rawson et al., 2020).

Older age, hypertension and diabetes were the risk factor for ICU admission is aligned with what had published previously that showed that the same risk factors more common among ICU patients (Hachim et al., 2020) and it contribute to the development of acute cardiac injury among COVID-19 patients (Naeem et al., 2020).

Strength and Limitations

Although the study is a single center and it lack the social determinants of the patients but it included the first cases affected by COVID-19 in the UAE and the sample size is fairly good with wide range clinical and laboratory data.

CONCLUSION

Our patients are younger and mainly male, DM and hypertension were the most common comorbidities. The clinical symptoms are mainly of LRTI and there were considerable percentage of clinical complications and organ dysfunction. The treatment showed excess use of antibiotics.

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.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Scientific Research Committee of MOHAP, approval letter No. MOHAP/DXB-REC/MMM/NO.44/2020. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors have contributed equally to the manuscript. They all contributed in the planning, data collection, analysis, drafting and writing the article. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank all our patients for their braveness and patience. We thank all the medical workers for their great work and sacrifices during this crisis. In addition, we extend our gratitude to the Research Ethics Committee of MOHAP.

REFERENCES

Available at: https://www.worldometers.info/coronavirus/ (Accessed January 19th 2021).

Alharbi, N. S., Almutari, R., Jones, S., Al-Daghri, N., Khunti, K., and de Lusignan, S. (2014). Trends in the prevalence of type 2 diabetes mellitus and obesity in the Arabian Gulf States: systematic review and meta-analysis. *Diabetes Res. Clin. Pract.* 106 (2), e30–e33.

Al-Tawfiq, J. A., Leonardi, R., Fasoli, G., and Rigamonti, D. (2020). Prevalence and fatality rates of COVID-19: What are the reasons for the wide variations worldwide? *Travel Med. Infect. Dis.* 35, 101711.

Ayalon, L. (2019). Perceived Discrimination and Stigma in the Context of the Long-Term Care Insurance Law from the Perspectives of Arabs and the Jews in the North of Israel. *Int. J. Environ. Res. Public Health* 16 (19), 3511. doi: 10.3390/ijerph16193511

Bimonte, S., Crispo, A., Amore, A., Celentano, E., Cuomo, A., and Cascella, M. (2020). Potential Antiviral Drugs for SARS-Cov-2 Treatment: Preclinical Findings and Ongoing Clinical Research. *Vivo* 34 (3 Suppl), 1597–1602.

Biswas, S. K., and Mudi, S. R. (2020). Genetic variation in SARS-CoV-2 may explain variable severity of COVID-19. *Med. Hypotheses* 143, 109877.

Bwire, G. M. (2020). Coronavirus: Why Men are More Vulnerable to Covid-19 Than Women? *SN Compr. Clin. Med.* 1–3. doi: 10.1007/s42399-020-00341-w

Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., et al. (2020). A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N. Engl. J. Med.* 382 (19), 1787–1799.

Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., and Di Napoli, R. (2020). *Features, Evaluation and Treatment Coronavirus (COVID-19)* (Treasure Island (FL: StatPearls).

Chao, C. Y. H., Wan, M. P., Morawska, L., Johnson, G. R., Ristovski, Z. D., Hargreaves, M., et al. (2009). Characterization of expiration air jets and droplet size distributions immediately at the mouth opening. *J. Aerosol. Sci.* 40 (2), 122–133.

Dorner, T. E., Schwarz, F., Kranz, A., Freidl, W., Rieder, A., and Gisinger, C. (2010). Body mass index and the risk of infections in institutionalised geriatric patients. *Br. J. Nutr.* 103 (12), 1830–1835.

Fauci, A. S., Lane, H. C., and Redfield, R. R. (2020). Covid-19 - Navigating the Uncharted. *N. Engl. J. Med.* 382 (13), 1268–1269.

Force, A. D. T., Ranieri, V. M., Rubenfeld, G. D., Thompson, B. T., Ferguson, N. D., Caldwell, E., et al. (2012). Acute respiratory distress syndrome: the Berlin Definition. *JAMA* 307 (23), 2526–2533.

Fu, L., Wang, B., Yuan, T., Chen, X., Ao, Y., Fitzpatrick, T., et al. (2020). Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. *J. Infect.* 80 (6), 656–665.

Hachim, M. Y., Hachim, I. Y., Naeem, K. B., Hannawi, H., Salmi, I. A., and Hannawi, S. (2020). D-dimer, Troponin, and Urea Level at Presentation With COVID-19 can Predict ICU Admission: A Single Centered Study. *Front. Med. (Lausanne)* 7, 585003.

Huttner, B. D., Catho, G., Pano-Pardo, J. R., Pulcini, C., and Schouten, J. (2020). COVID-19: don't neglect antimicrobial stewardship principles! *Clin. Microbiol. Infect.* 26 (7), 808–810.

Jin, J. M., Bai, P., He, W., Wu, F., Liu, X. F., Han, D. M., et al. (2020). Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Front. Public Health* 8, 152.

Khamis, F., Al Rashidi, B., Al-Zakwani, I., Al Wahaibi, A. H., and Al Awaidy, S. T. (2020). Epidemiology of COVID-19 Infection in Oman: Analysis of the First 1304 Cases. *Oman Med. J.* 35 (3), e145.

Kim, A. H. J., Sparks, J. A., Liew, J. W., Putman, M. S., Berenbaum, F., Duarte-Garcia, A., et al. (2020). A Rush to Judgment? Rapid Reporting and Dissemination of Results and Its Consequences Regarding the Use of Hydroxychloroquine for COVID-19. *Ann. Intern. Med.* 172 (12), 819–821.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N. Engl. J. Med.* 382 (13), 1199–1207.

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.

Liu, Y., Mao, B., Liang, S., Yang, J. W., Lu, H. W., Chai, Y. H., et al. (2020). Association between age and clinical characteristics and outcomes of COVID-19. *Eur. Respir. J.* 55 (5).

Lotfi, M., Hamblin, M. R., and Rezaei, N. (2020). COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clin. Chim. Acta* 508, 254–266.

Mahase, E. (2020). Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. *BMJ* 368, m1036.

Malik, V. S., Ravindra, K., Attri, S. V., Bhadada, S. K., and Singh, M. (2020). Higher body mass index is an important risk factor in COVID-19 patients: a systematic review and meta-analysis. *Environ. Sci. Pollut. Res. Int.* 27 (33), 42115–42123. doi: 10.1007/s11356-020-10132-4

Mohapatra, R. K., Pintilie, L., Kandi, V., Sarangi, A. K., Das, D., Sahu, R., et al. (2020). The recent challenges of highly contagious COVID-19, causing respiratory infections: Symptoms, diagnosis, transmission, possible vaccines, animal models, and immunotherapy. *Chem. Biol. Drug Des.* 96 (5), 1187–1208.

Naeem, K. B., Hachim, M. Y., Hachim, I. Y., Chkhis, A., Quadros, R., Hannawi, H., et al. (2020). Acute cardiac injury is associated with adverse outcomes, including mortality in COVID-19 patients. A single-center experience. *Saudi Med. J.* 41 (11), 1204–1210.

Pan, D., Sze, S., Minhas, J. S., Bangash, M. N., Pareek, N., Divall, P., et al. (2020). The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. *EClinicalMedicine* 23, 100404.

Pi-Sunyer, F. X. (2000). Obesity: criteria and classification. *Proc. Nutr. Soc.* 59 (4), 505–509.

Pradhan, A., and Olsson, P. E. (2020). Sex differences in severity and mortality from COVID-19: are males more vulnerable? *Biol. Sex Differ.* 11 (1), 53.

Rawson, T. M., Moore, L. S. P., Zhu, N., Ranganathan, N., Skolimowska, K., Gilchrist, M., et al. (2020). Bacterial and Fungal Coinfection in Individuals With Coronavirus: A Rapid Review To Support COVID-19 Antimicrobial Prescribing. *Clin. Infect. Dis.* 71 (9), 2459–2468.

Rezaei, N. (2020). COVID-19 affects healthy pediatricians more than pediatric patients. *Infect. Control Hosp. Epidemiol.* 1.

Saravanan, C., Attlee, A., and Sulaiman, N. (2019). A Cross Sectional Study on Knowledge, Beliefs and Psychosocial Predictors of Shisha Smoking among University Students in Sharjah, United Arab Emirates. *Asian Pac. J. Cancer Prev.* 20 (3), 903–909.

Selvaraj, V., Dapaah-Afriyie, K., Finn, A., and Flanigan, T. P. (2020). Short-Term Dexamethasone in Sars-CoV-2 Patients. *R. I. Med. J. (2013)* 103 (6), 39–43.

Sheleme, T., Bekele, F., and Ayela, T. (2020). Clinical Presentation of Patients Infected with Coronavirus Disease 19: A Systematic Review. *Infect. Dis. (Auckl).* 13, 1178633720952076.

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315 (8), 801–810.

Singh, A. K. (2020). COVID-19 experience in Kuwait: A high prevalence of asymptomatic cases and increased mortality in smokers. *EClinicalMedicine* 24, 100462.

Statistic center D. Available at: https://www.dsc.gov.ae/en-us/Pages/Population.aspx (Accessed August 7th, 2020).

Sun, P., Qie, S., Liu, Z., Ren, J., Li, K., and Xi, J. (2020). Clinical characteristics of hospitalized patients with SARS-CoV-2 infection: A single arm meta-analysis. *J. Med. Virol.* 92 (6), 612–617.

Tailakh, A., Evangelista, L. S., Mentes, J. C., Pike, N. A., Phillips, L. R., and Morisky, D. E. (2014). Hypertension prevalence, awareness, and control in Arab countries: a systematic review. *Nurs. Health Sci.* 16 (1), 126–130.

Tang, J. W., Nicolle, A. D., Pantelic, J., Jiang, M., Sekhr, C., Cheong, D. K., et al. (2011). Qualitative real-time schlieren and shadowgraph imaging of human exhaled airflows: an aid to aerosol infection control. *PLoS One* 6 (6), e21392.

The Supreme Council for National Security *National emergency Crises and Disasters Management Authority*. Available at: https://covid19.ncema.gov.ae/en (Accessed August 14th 2020).

Theoharides, T. C., and Conti, P. (2020). Dexamethasone for COVID-19? Not so fast. *J. Biol. Regul. Homeost. Agents.* 34 (3).

van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., et al. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *N. Engl. J. Med.* 382 (16), 1564–1567.

Wang, C., Horby, P. W., Hayden, F. G., and Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *Lancet* 395 (10223), 470–473.

Wang, Z., Yang, B., Li, Q., Wen, L., and Zhang, R. (2020). Clinical Features of 69 Cases With Coronavirus Disease 2019 in Wuhan, China. *Clin. Infect. Dis.* 71 (15), 769–777.

Wu, Z., and McGoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*.

Zaatar, M. T. *Charting the UAE"s battle against COVID-19*. Available at: https://www.natureasia.com/en/nmiddleeast/article/10.1038/nmiddleeast.2020.84 (Accessed January 18th, 2021. 2020).

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395 (10229), 1054–1062.

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.

ORIGINAL RESEARCH
published: 26 May 2021
doi: 10.3389/fcimb.2021.680422

Sex-Disaggregated Data on Clinical Characteristics and Outcomes of Hospitalized Patients With COVID-19: A Retrospective Study

Mengdie Wang[1†], *Nan Jiang*[2†], *Changjun Li*[3†], *Jing Wang*[2], *Heping Yang*[4], *Li Liu*[5], *Xiangping Tan*[6], *Zhenyuan Chen*[2], *Yanhong Gong*[2], *Xiaoxv Yin*[2], *Qiao Zong*[2], *Nian Xiong*[1] and *Guopeng Zhang*[7*]

[1] Department of Neurology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2] Department of Social Medicine and Health Management, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3] Department of Neurology, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [4] School of Nursing, Wuchang University of Technology, Wuhan, China, [5] Office of Academic Research, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [6] Lichuan Center for Disease Control and Prevention, Lichuan, China, [7] Department of Nuclear medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:
Hassan Zaraket,
American University of Beirut, Lebanon

Reviewed by:
Kei Yamamoto,
National Center For Global Health and Medicine, Japan
Liriye Kurtovic,
Burnet Institute, Australia

***Correspondence:**
Guopeng Zhang
zhgp101505@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Clinical Microbiology, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 15 March 2021
Accepted: 11 May 2021
Published: 26 May 2021

Citation:
Wang M, Jiang N, Li C, Wang J, Yang H, Liu L, Tan X, Chen Z, Gong Y, Yin X, Zong Q, Xiong N and Zhang G (2021) Sex-Disaggregated Data on Clinical Characteristics and Outcomes of Hospitalized Patients With COVID-19: A Retrospective Study. Front. Cell. Infect. Microbiol. 11:680422. doi: 10.3389/fcimb.2021.680422

Background: Sex and gender are crucial variables in coronavirus disease 2019 (COVID-19). We sought to provide information on differences in clinical characteristics and outcomes between male and female patients and to explore the effect of estrogen in disease outcomes in patients with COVID-19.

Method: In this retrospective, multi-center study, we included all confirmed cases of COVID-19 admitted to four hospitals in Hubei province, China from Dec 31, 2019 to Mar 31, 2020. Cases were confirmed by real-time RT-PCR and were analyzed for demographic, clinical, laboratory and radiographic parameters. Random-effect logistic regression analysis was used to assess the association between sex and disease outcomes.

Results: A total of 2501 hospitalized patients with COVID-19 were included in the present study. The clinical manifestations of male and female patients with COVID-19 were similar, while male patients have more comorbidities than female patients. In terms of laboratory findings, compared with female patients, male patients were more likely to have lymphopenia, thrombocytopenia, inflammatory response, hypoproteinemia, and extrapulmonary organ damage. Random-effect logistic regression analysis indicated that male patients were more likely to progress into severe type, and prone to ARDS, secondary bacterial infection, and death than females. However, there was no significant difference in disease outcomes between postmenopausal and premenopausal females after propensity score matching (PSM) by age.

Conclusions: Male patients, especially those age-matched with postmenopausal females, are more likely to have poor outcomes. Sex-specific differences in clinical characteristics and outcomes do exist in patients with COVID-19, but estrogen may not be the primary cause. Further studies are needed to explore the causes of the differences in disease outcomes between the sexes.

Keywords: COVID-19, SARS-CoV-2, sex, menopause, estrogen, China

INTRODUCTION

Coronavirus disease 2019 (COVID-19) has been reported in 223 countries and regions, with a cumulative total of 131,837,512 confirmed cases and 2,862,664 deaths as of April 7, 2021, according to the World Health Organization. COVID-19 pandemic has posed a serious threat to global health and economy. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the pathogen that causes COVID-19 *via* using angiotensin converting enzyme 2 (ACE2) as a receptor (Zhao et al., 2020). It is a member of the beta coronavirus genus along with severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) (DiMaio et al., 2020). The typical clinical manifestations of patients with COVID-19 are fever, dry cough, fatigue, and in severe cases, dyspnea (Guan et al., 2020; Hu et al., 2021). In previous epidemics of coronavirus-induced diseases, infected populations have shown gender differences in clinical outcomes. Studies related to SARS-CoV showed that males were more susceptible to infection and had a significantly higher case fatality rate than females (Karlberg et al., 2004; Leung et al., 2004; Leong et al., 2006). Among patients infected by MERS-CoV, morbidity and mortality were higher in males than females (Alghamdi et al., 2014). Similar to these respiratory diseases caused by coronaviruses, sex and gender are crucial variables in COVID-19 (Channappanavar et al., 2017; Kadel and Kovats, 2018; Guan et al., 2020). Epidemiologic data suggested that males may be more susceptible to COVID-19 and have higher clinical severity and mortality, especially older males with chronic illnesses (Epidemiology Working Group for NCIP Epidemic Response and Chinese Center for Disease Control and Prevention, 2020; Guan et al., 2020; Onder et al., 2020). Such sex-specific differences are probably attributed to sex hormones, different copy numbers of immune response genes on X chromosomes, and the presence of disease susceptibility genes in females and males (Sue, 2017; Kadel and Kovats, 2018; Schurz et al., 2019). Behavioral and cultural factors may also be involved (Suen et al., 2019; Cai, 2020).

Sex- and gender- specific epidemiologic observations would help deepen our understanding towards COVID-19 and make sex- or gender- specific recommendations (Bhopal, 2020; Wenham et al., 2020). Although current studies have reported the impact of sex on patients with COVID-19, studies based on Chinese populations are scarce and limited in small sample sizes (Sha et al., 2021). Estrogen has been reported to play a crucial role in disease outcomes in COVID-19 patients, attributed to its ability to reduce inflammatory IL-6, IL-8 and TNF-α levels (Alwani et al., 2021). However, this conclusion is not consistent (Alwani et al., 2021; Sha et al., 2021). This study aimed to provide information on differences in clinical characteristics and outcomes between male and female patients and explore the effect of estrogen on disease outcomes.

MATERIALS AND METHODS

Study Design and Participants

This multi-center retrospective study analyzed information on hospitalized patients with COVID-19 admitted to four hospitals (the Central Hospital of Wuhan, Wuhan Red Cross Hospital, the Central Hospital of Enshi Tujia and Miao Autonomous Prefecture and Lichuan People's Hospital) in Hubei Province, China. All these hospitals are government-appointed hospitals dedicated to the treatment of COVID-19. The diagnosis of COVID-19 was confirmed according to the WHO interim guidance and the Diagnosis and Treatment Protocol for Coronavirus Pneumonia (trial version 7) released by National Health Commission of China (National Health Commission of the People's Republic of China; World Health Organization). A total of 2501 patients with COVID-19 admitted to these hospitals from 31st December 2019 to 31st March 2020 were enrolled. This study was approved by the Medical Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology (Number: 2020IECA252) and complied with the principles of the Declaration of Helsinki. The requirement for informed consent from patients was waived by the ethics committee due to the retrospective nature of the study. The data are anonymous, and all authors could only use the anonymized data for statistical analysis, with no direct interaction with patients or patient samples.

Data Collection

The demographic characteristics, medical history, laboratory findings, chest computed tomography (CT) on admission, and outcome data were extracted from electronic medical records. Laboratory assessments comprised complete blood count, coagulation test, biochemical test (including liver and renal function, cardiac enzymes), and infection-related indices. The laboratory findings presented in the study were collected at hospital admission of COVID-19 patients. The primary outcome variables of the study were disease severity and in-hospital mortality. The secondary outcomes were complications, which included shock, acute respiratory distress syndrome (ARDS), acute cardiac injury, acute kidney injury (AKI), secondary bacterial infection, and urinary tract infection. ARDS and shock were defined in term of the interim guidance of WHO for novel coronavirus (World Health Organization). Cardiac injury was diagnosed when serum levels of cardiac biomarkers (e.g., high-sensitivity troponin I, hs-TNI) were above the 99th percentile upper reference limit (Shi et al., 2020). Acute kidney injury was identified according to the kidney disease: Improving Global Outcomes definition (Kellum et al., 2012). Secondary bacterial infection was confirmed if the patients had symptoms or signs of nosocomial pneumonia or bacteremia, and a positive culture of a new pathogen from a lower respiratory tract specimen or from blood samples taken after admission (Garner et al., 1988). Urinary tract infection was determined by an abnormally elevated leukocyte count in the urine (Long and Koyfman, 2018). The severity of COVID-19 (severe *vs.* non-severe) was assessed at admission according to the Diagnosis and Treatment Protocol for Coronavirus Pneumonia (trial version 7) released by National Health Commission of China (National Health Commission of the People's Republic of China). The clinical outcomes were categorized into discharges and mortality and monitored up to Apr 5, 2020.

To explore the effects of estrogen, female patients with COVID-19 were divided into two groups (premenopausal and postmenopausal female patients) according to whether they are menopausal, and male ones were classed into two groups based on their age (male patients <50 years and ≥50 years), which refer to the median age of female menopause.

Statistical Analysis

In this study, continuous variables were represented by median and interquartile range (IQR), and categorical variables were presented as frequencies and percentage (%). Comparison of parameters between two groups were conducted with the Wilcoxon-Mann-Whitney-Test for continuous variables. Pearson's χ^2 test or Fisher's exact tests were used for categorical variables. The risk of outcomes of interest was calculated by multivariable logistic regression. Hospital was modeled as a random effect in the random effect logistic regression. Adjusted odd ratios (ORs) and 95% confidence intervals (CIs) were calculated for different groups. Multivariate analyses were all adjusted for age and comorbidities (hypertension, diabetes, coronary heart disease, cerebrovascular disease, COPD, malignancy, chronic liver disease, and chronic kidney disease). Propensity score-matched analysis was used to balance the age between premenopausal and postmenopausal females. Two cohorts were matched at a ratio of 1:1 with a caliper width of 0.2. For all comparisons, differences were tested using two-tailed tests and p-values less than 0.05 were considered statistically significant. Statistical analysis was performed using SAS version 9.4.

RESULTS

Clinical Characteristics and Radiographic Parameters of Female and Male Patients

A total of 2501 hospitalized patients with COVID-19 were included in the analysis. Of these, 1305 were females and 1196 were males. The features of the study population are shown in **Table 1**. The median age was 56 years (IQR, 40-67 years) for females and 59 years (IQR, 44-69 years) for males. The clinical manifestations of male and female patients were similar. Fever and dry cough were the major common symptoms, whereas myalgia, diarrhea, and vomiting were rare. Fever was present more in male patients (795[66.5%] *vs.* 806[61.8%], *P*=0.0142). Diarrhea was present more in female patients (75[5.7%] *vs.* 45 [3.8%], *P*=0.0204). Male patients had more comorbidities, including hypertension (457[38.2%] *vs.* 418[32.0%], *P*=0.0012), coronary heart disease (145[12.1%] *vs.* 90[6.9%], *P*<0.0001), cerebrovascular disease (107[8.9%] *vs.* 62[4.8%], *P*<0.0001), COPD (109[9.1%] *vs.* 64[4.9%], *P*<0.0001), and chronic kidney disease (87[7.3%] *vs.* 50[3.8%], *P*=0.0002).

Supplementary Table 1 further showed the subgroup analysis of clinical characteristics. The prevalence of coronary heart disease, cerebrovascular disease, COPD, and chronic kidney disease among postmenopausal female patients were significantly lower than that of their age-matched male patients. There was a significant difference in the prevalence of hypertension between premenopausal female and their age-matched male patients. Females in postmenopausal group reported significantly higher prevalence of hypertension, diabetes, coronary heart disease, cerebrovascular disease, COPD, malignancy, chronic liver disease, and chronic kidney diseases compared to the premenopausal group. Propensity score matching (PSM) was performed to avoid age interference in the association between estrogen and disease outcomes. After PSM, 74 females in premenopausal group were matched at a 1:1 ratio to 74 females in postmenopausal group. There was no significant difference between the two groups in terms of clinical characteristics (**Supplementary Table 2**).

Of 2259 patients who underwent chest CT on admission, 165 (13.9%) of females and 145(13.5%) of males showed ground-glass opacity, 162(13.6%) of females and 114(10.6%) of males showed unilateral pneumonia, and 861(72.5%) of females and 812(75.8%) of males showed bilateral pneumonia. There was no difference between the female and male groups (*P*=0.0810) (**Table 1**).

Laboratory Findings of Female and Male Patients

In terms of laboratory findings on admission, male patients had more lymphopenia (455[38.2%] *vs.* 410[31.8%], *P*=0.0008) and thrombocytopenia (145[12.2%] *vs.* 87[6.7%], *P*<0.0001) as compared with female patients. Male patients also showed a higher inflammatory response (C-reactive protein: 710[61.1%] *vs.* 624[49.5%], *P*<0.0001; procalcitonin: 62[8.3%] *vs.* 28[3.3%], *P*<0.0001; interleukin-6: 123[45.2%] *vs.* 99[30.9%], *P*=0.0003) and were more prone to have hypoproteinemia (albumin: 568 [48.0%] *vs.* 544[42.4%], *P*=0.0054) and extrapulmonary organ damage, such as cardiac injury (creatine kinase: 193[18.4%] *vs.* 103[9.0%], *P*<0.0001; high-sensitivity troponin I: 119[15.6%] *vs.* 99[12.0%], *P*=0.0394; myohemoglobin: 77[19.3%] *vs.* 48[11.7%], *P*=0.0028) and liver injury (alanine aminotransferase: 258 [21.8%] *vs.* 152[11.9%], *P*<0.0001; aspartate aminotransferase: 173[17.7%] *vs.* 130[12.3%], *P*=0.0006; total bilirubin: 95[8.1%] *vs.* 42[3.3%], *P*<0.0001) (**Table 2**). The absolute values of the median and interquartile range (IQR) of laboratory findings were also shown in **Supplementary Table 3**. Normal ranges of each laboratory findings were listed in **Supplementary Table 4**.

TABLE 1 | Clinical characteristics on admission and outcomes in male and female patients with COVID-19[a].

Variables	Female (n = 1305)	Male (n = 1196)	*P* value
Age, median (IQR), year	56(40-67)	59(44-69)	0.0002[b]
Days from symptom onset to admission, median (IQR), (Missing=20)	9(5-15)	9(5-15)	0.2708[b]
Length of stay, median (IQR)	16(11-25)	16(11-26)	0.5095[b]
Comorbidities			
Any	588(45.1)	646(54.0)	<0.0001
Hypertension	418(32.0)	457(38.2)	0.0012
Diabetes	204(15.6)	213(17.8)	0.1445
Coronary heart disease	90(6.9)	145(12.1)	<0.0001
Cerebrovascular disease	62(4.8)	107(8.9)	<0.0001
COPD	64(4.9)	109(9.1)	<0.0001
Malignancy	90(6.9)	67(5.6)	0.1824
Chronic liver disease	78(6.0)	89(7.4)	0.1428
Chronic kidney disease	50(3.8)	87(7.3)	0.0002
Signs and symptoms			
Fever	806(61.8)	795(66.5)	0.0142
Dry cough	661(50.7)	562(47.0)	0.0673
Shortness of breath	284(21.8)	255(21.3)	0.7886
Fatigue	239(18.3)	224(18.7)	0.7896
Chest stuffiness	233(17.9)	217(18.1)	0.8507
Expectoration	184(14.1)	171(14.3)	0.8873
Anorexia	144(11.0)	120(10.0)	0.4157
Myalgia	71(5.4)	63(5.3)	0.8478
Diarrhea	75(5.7)	45(3.8)	0.0204
Vomiting	26(2.0)	26(2.2)	0.7506
Chest computed tomography findings (Missing=242)			0.0810
Ground-glass opacity	165(13.9)	145(13.5)	
Local patchy shadowing	162(13.6)	114(10.6)	
Bilateral patchy shadowing	861(72.5)	812(75.8)	
Complications			
Shock	13(1.0)	25(2.1)	0.0255
Acute respiratory distress syndrome	99(7.6)	135(11.3)	0.0015
Acute cardiac injury (Missing=913)	127(15.4)	165(21.6)	0.0015
Acute kidney injury (Missing=31)	369(28.7)	271(22.9)	0.0009
Secondary infection (Missing=568)	303(30.2)	370(39.8)	<0.0001
Urinary tract infection	116(22.9)	75(17.4)	0.0365
Disease severity			
Severe	562(43.1)	656(54.9)	<0.0001
Clinical outcome			<0.0001
Discharged	1238(94.9)	1068(89.3)	
Died	67(5.1)	128(10.7)	

COPD, Chronic obstructive pulmonary disease.

[a] *Unless otherwise indicated, values shown are n(%).*

[b] *These P values are associated with Wilcoxon-Mann-Whitney-Test; all other P values are associated with χ^2 tests.*

Subgroup analysis of laboratory findings indicated that postmenopausal females had a lower incidence of lymphopenia, thrombocytopenia, and hypoproteinemia than age-matched males. The serum markers indicated that postmenopausal females are less likely to have elevated inflammatory response, disturbed coagulation function, and extrapulmonary organ damage, including cardiac injury and liver injury, than age-matched males. Premenopausal female showed weaker inflammatory response and liver injury than age-matched male patients (**Supplementary Table 5**).

Complications and Outcomes of Female and Male Patients

Compared to female patients, male patients had a higher proportion of severe cases (656[54.9%] *vs.* 562[43.1%], *P* <0.0001) and deaths (128[10.7%] *vs.* 67[5.1%], *P*<0.0001). During hospitalization, male patients were more likely to have shock (25[2.1%] *vs.* 13[1.0%], *P*=0.0255), ARDS (135[11.3] *vs.* 99 [7.6%], *P*=0.0015), acute cardiac injury (165[21.6%] *vs.* 127 [15.4%], *P*=0.0015), secondary bacterial infections (370[39.8%] *vs.* 303 [30.2%], *P*<0.0001). Female patients were more prone to have acute kidney injury (369[28.7%] *vs.* 271[22.9%], *P*=0.0009) and urinary tract infection (116[22.9%] *vs.* 75[17.4%], *P*=0.0365) (**Table 1**).

Multivariable logistic regression analysis indicated that males were more likely to progress into severe type (OR=1.46; 95%CI: 1.24-1.73) and prone to ARDS (OR=1.37; 95%CI: 1.03-1.83), secondary bacterial infections (OR=1.39; 95%CI: 1.14-1.69) and death (OR=1.90; 95%CI: 1.36-2.66), but had a lower probability of acute kidney injury (OR=0.57; 95%CI: 0.47-0.70) and urinary tract infection (OR=0.60; 95%CI: 0.43-0.85) (**Table 3**). After further subgrouping the patients by disease severity, the

TABLE 2 | Laboratory findings in female and male patients with COVID-19 on admission.

Variables	Female (n = 1305)	Male (n = 1196)	*P* value[a]
Hematologic			
Blood leukocyte count $>10\times10^{9}$/L	94(7.3)	105(8.8)	0.1630
Lymphocyte count $<1.1\times10^{9}$/L	410(31.8)	455(38.2)	0.0008
Neutrophil count $>6.3\times10^{9}$/L	151(11.7)	154(12.9)	0.3533
Platelet count $<125\times10^{9}$/L	87(6.7)	145(12.2)	<0.0001
Biochemical			
Hemoglobin <110 g/L	292(22.6)	269(22.6)	0.9675
Alanine aminotransferase >50U/L	152(11.9)	258(21.8)	<0.0001
Aspartate aminotransferase >40U/L	130(12.3)	173(17.7)	0.0006
Lactate dehydrogenase >225U/L	247(24.4)	242(25.9)	0.4458
Total bilirubin >21μmol/L	42(3.3)	95(8.1)	<0.0001
Albumin <35g/L	544(42.4)	568(48.0)	0.0054
Blood urea >8.2mmol/L	95(7.4)	139(11.7)	0.0002
Creatinine >133μmol/L	208(16.2)	186(15.7)	0.7394
Creatine kinase >190U/L	103(9.0)	193(18.4)	<0.0001
Creatine kinase-MB >6.73ng/ml	10(4.3)	14(6.3)	0.3424
High-sensitivity troponin I >0.014ng/ml (99th percentile)	99(12.0)	119(15.6)	0.0394
Myohemoglobin >75ng/ml	48(11.7)	77(19.3)	0.0028
Brain natriuretic peptide >100pg/ml	157(30.2)	169(35.7)	0.0670
Coagulation test			
Prothrombin time >15s	146(12.0)	228(20.0)	<0.0001
Activated partial thromboplastin time >40s	51(4.2)	58(5.1)	0.2939
D-dimer >1ug/ml	476(38.7)	457(40.5)	0.3773
Infection-related indices			
C-reactive protein >5 mg/L	624(49.5)	710(61.1)	<0.0001
Procalcitonin ≥0.5 ng/ml	28(3.3)	62(8.3)	<0.0001
Interleukin-6 >7pg/ml	99(30.9)	123(45.2)	0.0003

[a] *All p-values are associated with $\chi2$ tests.*

incidence of secondary bacterial infections was also significantly higher in male patients with severe type of COVID-19 than in females (**Supplementary Table 6**).

In the subgroup analysis, male patients age-matched with postmenopausal females were more likely to progress into severe type (OR=1.37; 95%CI: 1.11-1.68), and prone to ARDS (OR=1.36; 95%CI: 1.00-1.83), secondary bacterial infection (OR=1.67; 95%CI: 1.31-2.12), and death (OR=1.89; 95%CI: 1.34-2.67) than postmenopausal females. But they had lower risk of acute renal injury (OR=0.59; 95%CI: 0.46-0.75) and urinary tract infection (OR=0.43; 95%CI: 0.29-0.66). Male patients age-matched with premenopausal females were more likely to progress into severe type (OR=1.65; 95%CI: 1.23-2.22), but were less likely to prone to renal injury (OR=0.54; 95% CI: 0.37-0.77) than premenopausal females. The postmenopausal females were more likely to have urinary tract infection (OR=2.33; 95%CI: 1.14-4.75), but had a lower probability of secondary bacterial infections (OR=0.46; 95%CI: 0.30-0.70) compared with premenopausal female patients (**Table 3**). However, these differences were not observed after PSM between premenopausal and postmenopausal females by age (**Supplementary Table 7**).

TABLE 3 | Multivariable logistic regression analysis of associations of groups with outcomes[a].

Variables	Male *vs.* Female[b]	Postmenopausal females *vs.* Premenopausal females[b]	Males ≥50 years *vs.* Postmenopausal females[b]	Males <50 years *vs.* Premenopausal females[b]
Complications				
Shock	1.68(0.81-3.47)	0.66(0.06-6.98)	1.75(0.82-3.73)	0.91(0.06-15.00)
Acute respiratory distress syndrome	1.37(1.03-1.83)*	1.82(0.83-4.01)	1.36(1.00-1.83)*	1.42(0.61-3.34)
Acute cardiac injury (Missing=913)	1.25(0.94-1.68)	0.87(0.43-1.74)	1.36(0.99-1.86)	0.74(0.34-1.64)
Acute kidney injury (Missing=31)	0.57(0.47-0.70)***	0.91(0.61-1.36)	0.59(0.46-0.75)***	0.54(0.37-0.77)**
Secondary infection (Missing=568)	1.39(1.14-1.69)**	0.46(0.30-0.70)**	1.67(1.31-2.12)***	0.97(0.69-1.37)
Urinary tract infection	0.60(0.43-0.85)**	2.33(1.14-4.75)*	0.43(0.29-0.66)***	1.35(0.72-2.51)
Disease severity				
Severe	1.46(1.24-1.73)***	1.09(0.77-1.56)	1.37(1.11-1.68)**	1.65(1.23-2.22)**
Clinical outcome				
Died	1.90(1.36-2.66)**	2.05(0.59-7.15)	1.89(1.34-2.67)**	1.46(0.33-6.46)

****P < 0.0001, **P < 0.01, *P < 0.05.*

[a] *Adjusted for age and comorbidities including hypertension, diabetes, coronary heart disease, cerebrovascular disease, chronic obstructive pulmonary disease, malignancy, chronic liver disease, and chronic kidney disease. Hospital was modeled as a random effect in the multivariable logistic regression.*

[b] *The reference.*

DISCUSSION

In the present study, we reported detailed sex-disaggregated data on COVID-19 and explored the effect of estrogen in the disease outcomes. We found differences in clinical characteristics and outcomes do exist between male and female patients. These sex-dependent differences were more pronounced in males who were age-matched to postmenopausal. Sex is a crucial variable in the prognosis of COVID-19.

Significant sex difference in severe disease and death of COVID-19 were observed in this study. Male patients were more likely to be severe cases and had a significantly higher proportion of death than females, which is consistent with the findings of SARS-CoV (Channappanavar et al., 2017) and the MERS-CoV (Matsuyama et al., 2016). This may be related to the different respond of females and males to many virus infections (Kadel and Kovats, 2018). Females usually have stronger innate and adaptive immune responses and are relatively resistant to viral infections. Instead, males generate less robust immune responses and are more susceptible to infection (Bouman et al., 2005; Rettew et al., 2008; Klein and Flanagan, 2016). The X chromosome, sex hormones and differential expression of disease susceptibility genes between sexes may be involved in these sex-specific differences following virus infections (Schurz et al., 2019). The X chromosome contains a large number of immune-related genes. Therefore, females can clear pathogens faster and induce vaccine effectiveness greater than males (Schurz et al., 2019). Moreover, as a functional receptor for SARS-CoV-2, ACE2 plays an important role in the pathogenesis. Plasma concentrations of ACE2, have been found higher in male than in female with heart failure (Sama et al., 2020). This may be the addition reason for the differences in virus loads, tissue damage and outcomes between the sexes. In a mouse model of SARS-CoV infection, oophorectomy or estrogen receptor antagonist treatment increased mortality in female mice (Channappanavar et al., 2017). However, there was no significant difference in disease severity and mortality between postmenopausal and premenopausal females after eliminating the confounding of age in this study. A retrospective study in patients with COVID-19 in China also reported that estrogen might not be directly related to the lower mortality in females (Sha et al., 2021). Due to the small sample size after PSM, large sample investigations are still needed to verify this conclusion.

Male patients were also more prone to severe complications. Besides more severe inflammatory immune response aforementioned, more pre-existing diseases (i.e., hypertension, coronary heart disease, cerebrovascular disease, COPD, and chronic kidney disease) in male patients may be an additional reason. Current evidence from Wuhan, China has shown that the COVID-19 mortality rate is comorbidity-dependent, and the crude case mortality rate increased to 10.5%, 7.3% and 6.3% in patients with cardiovascular disease, diabetes mellitus or hypertension, respectively (Epidemiology Working Group for NCIP Epidemic Response and Chinese Center for Disease Control and Prevention, 2020). The presence of these comorbidities may put male patients in weaker immune functions and further aggravated hyperinflammatory state after SARS-CoV-2 infection, leading to more severe multiple organ damage. Pre-existing diseases usually cause organs damage, which are also more likely to deteriorate when SARS-CoV-2 infection occurs. Furthermore, effects of comorbidities on ACE2 expression and activities should be considered (Sama et al., 2020). We also found secondary infection was more common in male patients than in females, and the sex difference remained significant in severe cases. The results of a recent study also showed that males in severe illness were more likely to develop secondary infection (Su et al., 2020). It is suggested that the incidence of secondary infection in severe cases should be closely monitored by clinicians and reported as a complication, especially in male patients.

Previous studies showed that AKI was more common in male patients compared to females (Su et al., 2020; Vahidy et al., 2021). However, our study found that the incidence of AKI was higher in females than males. The reason may be due to the fact that on the one hand, gastrointestinal symptoms, especially diarrhea, are more common in females, and differences in care for dehydration will affect the frequency of AKI. Therefore, the higher incidence of AKI in females may be related to pre-renal acute kidney injury. On the other hand, urinary tract infections were more common in females, which may also contribute to the higher incidence of AKI in females than males in this study. Furthermore, since the incidence of AKI may vary depending on the diagnostic criteria (Luo et al., 2014), this may also be one reason why the present findings differ from other studies. Clinical trials with large samples are still needed to explore the causes of the different incidence of AKI in male and female patients with COVID-19.

The subgroup results showed that males that age-matched with postmenopausal females were more common in severe illness and prone to death. Data from Global Health 5050 also indicated that the overall case fatality ratio in males is indeed higher than females (Global Health 50/50). Thus, combined with recent studies (Chen et al., 2020; Dangis et al., 2020; Guan et al., 2020), we believed that sex is a crucial variable in the prognosis. Male sex, especially those aged over 50 years old, is related with the severe disease and death from SARS-Cov-2 infection.

Our study has several limitations. First, some laboratory tests (i.e., high-sensitivity troponin I, N-terminal pro-brain natriuretic peptide, creatinine, and cytokine level measurements) were not done in all the patients, and missing data might lead to bias of results. Second, due to all patients included in this study being adults, it was hard to assess whether sex-specific differences in clinical characteristics and disease outcomes also existed among younger age groups. Future studies should pay more attention to COVID-19 patients younger than 18 years, such as adolescents, to fill these gaps.

CONCLUSIONS

In conclusion, collecting sex-disaggregated data is essential to understanding the feature of COVID-19, the risk factors of poor prognosis, and developing the strategy of treatment. Here, we reported detailed sex-disaggregated data on SARS-CoV-2 infection and confirmed that sex is a crucial variable in the clinical characteristics and outcomes. A comprehensive management plan with a sex perspective is necessary. Males, especially those age-matched with postmenopausal females, require additional prevention, surveillance, or earlier intensive intervention. Estrogen may not be the primary reason to the sex-specific differences in disease outcomes.

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.

ETHICS STATEMENT

This study was approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MW, NJ, CL, and GZ were responsible for the conception, design, and writing of the manuscript. CL, HY, LL, XT, YG, and XY were responsible for the acquisition of data and literature research. MW, NJ, JW, ZC, QZ, NX were responsible for the analysis and interpretation of data. All authors contributed to the article and approved the submitted version. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

FUNDING

This study was funded by the Fundamental Research Funds for the Central Universities, Huazhong University of Science and Technology (No. 2020kfyXGYJ073).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.680422/full#supplementary-material

REFERENCES

Alghamdi, I. G., Hussain, I. I., Almalki, S. S., Alghamdi, M. S., Alghamdi, M. M., and El-Sheemy, M. A. (2014). The Pattern of Middle East Respiratory Syndrome Coronavirus in Saudi Arabia: A Descriptive Epidemiological Analysis of Data From the Saudi Ministry of Health. *Int. J. Gen. Med.* 7, 417–423. doi: 10.2147/IJGM.S67061

Alwani, M., Yassin, A., Al-Zoubi, R. M., Aboumarzouk, O. M., Nettleship, J., Kelly, D., et al. (2021). Sex-Based Differences in Severity and Mortality in COVID-19. *Rev. Med. Virol.*, 1–11. doi: 10.1002/rmv.2223

Bhopal, R. (2020). Covid-19 Worldwide: We Need Precise Data by Age Group and Sex Urgently. *BMJ* 369, m1366. doi: 10.1136/bmj.m1366

Bouman, A., Heineman, M. J., and Faas, M. M. (2005). Sex Hormones and the Immune Response in Humans. *Hum. Reprod. Update* 11, 411–423. doi: 10.1093/humupd/dmi008

Cai, H. (2020). Sex Difference and Smoking Predisposition in Patients With COVID-19. *Lancet Respir. Med.* 8, e20. doi: 10.1016/S2213-2600(20)30117-X

Channappanavar, R., Fett, C., Mack, M., Ten Eyck, P. P., Meyerholz, D. K., and Perlman, S. (2017). Sex-Based Differences in Susceptibility to Severe Acute Respiratory Syndrome Coronavirus Infection. *J. Immunol.* 198, 4046–4053. doi: 10.4049/jimmunol.1601896

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study. *Lancet* 395, 507–513. doi: 10.1016/S0140-6736(20)30211-7

Dangis, A., De Brucker, N., Heremans, A., Gillis, M., Frans, J., Demeyere, A., et al. (2020). Impact of Gender on Extent of Lung Injury in COVID-19. *Clin. Radiol.* 75, 554–556. doi: 10.1016/j.crad.2020.04.005

DiMaio, D., Enquist, L. W., and Dermody, T. S. (2020). A New Coronavirus Emerges, This Time Causing a Pandemic. *Annu. Rev. Virol.* 7, iii–iiv. doi: 10.1146/annurev-vi-07-042020-100001

Epidemiology Working Group for NCIP Epidemic Response and Chinese Center for Disease Control and Prevention. (2020). The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) in China. *Zhonghua. Liu. Xing. Bing. Xue. Za. Zhi.* 41, 145–151. doi: 10.3760/cma.j.issn.0254-6450.2020.02.003

Garner, J. S., Jarvis, W. R., Emori, T. G., Horan, T. C., and Hughes, J. M. (1988). CDC Definitions for Nosocomial Infections, 1988. *Am. J. Infect. Control* 16, 128–140. doi: 10.1016/0196-6553(88)90053-3

Global Health 50/50. (2020). *Sex, Gender and COVID-19: Overview and Resources 2020*. Available at: http://globalhealth5050.org/covid19 (Accessed May 19, 2020).

Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X., et al. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. doi: 10.1056/NEJMoa2002032

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

Kadel, S., and Kovats, S. (2018). Sex Hormones Regulate Innate Immune Cells and Promote Sex Differences in Respiratory Virus Infection. *Front. Immunol.* 9:1653. doi: 10.3389/fimmu.2018.01653

Karlberg, J., Chong, D. S., and Lai, W. Y. (2004). Do Men Have a Higher Case Fatality Rate of Severe Acute Respiratory Syndrome Than Women Do? *Am. J. Epidemiol.* 159, 229–231. doi: 10.1093/aje/kwh056

Kellum, J. A., Lameire, N., Aspelin, P., Barsoum, R. S., Burdmann, E. A., Goldstein, S. L., et al. (2012). Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int. Suppl.* 2, 1–138. doi: 10.1038/kisup.2012.1

Klein, S. L., and Flanagan, K. L. (2016). Sex Differences in Immune Responses. *Nat. Rev. Immunol.* 16, 626. doi: 10.1038/nri.2016.90

Leong, H. N., Earnest, A., Lim, H. H., Chin, C. F., Tan, C., Puhaindran, M. E., et al. (2006). SARS in Singapore–Predictors of Disease Severity. *Ann. Acad. Med. Singap.* 35, 326–331.

Leung, G. M., Hedley, A. J., Ho, L. M., Chau, P., Wong, I. O., Thach, T. Q., et al. (2004). The Epidemiology of Severe Acute Respiratory Syndrome in the 2003 Hong Kong Epidemic: An Analysis of All 1755 Patients. *Ann. Intern. Med.* 141, 662–673. doi: 10.7326/0003-4819-141-9-200411020-00006

Long, B., and Koyfman, A. (2018). The Emergency Department Diagnosis and Management of Urinary Tract Infection. *Emerg. Med. Clin. North Am.* 36, 685–710. doi: 10.1016/j.emc.2018.06.003

Luo, X., Jiang, L., Du, B., Wen, Y., Wang, M., and Xi, X. (2014). Beijing Acute Kidney Injury Trial (BAKIT) Workgroup. A Comparison of Different

Diagnostic Criteria of Acute Kidney Injury in Critically Ill Patients. *Crit. Care* 18, R144. doi: 10.1186/cc13977

Matsuyama, R., Nishiura, H., Kutsuna, S., Hayakawa, K., and Ohmagari, N. (2016). Clinical Determinants of the Severity of Middle East Respiratory Syndrome (MERS): A Systematic Review and Meta-Analysis. *BMC Public Health* 16, 1203. doi: 10.1186/s12889-016-3881-4

National Health Commission of the People's Republic of China. (2020). *New Coronavirus Pneumonia Prevention and Control Program (Version 7.0)*. Available at: http://www.gov.cn/zhengce/zhengceku/2020-03/04/content_5486705.htm (Accessed May 23, 2020).

Onder, G., Rezza, G., and Brusaferro, S. (2020). Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA* 323, 1775–1776. doi: 10.1001/jama.2020.4683

Rettew, J. A., Huet-Hudson, Y. M., and Marriott, I. (2008). Testosterone Reduces Macrophage Expression in the Mouse of Toll-Like Receptor 4, A Trigger for Inflammation and Innate Immunity. *Biol. Reprod.* 78, 432–437. doi: 10.1095/biolreprod.107.063545

Sama, I. E., Ravera, A., Santema, B. T., van Goor, H., Ter Maaten, J. M., Cleland, J. G. F., et al. (2020). Circulating Plasma Concentrations of Angiotensin-Converting Enzyme 2 in Men and Women With Heart Failure and Effects of Renin-Angiotensin-Aldosterone Inhibitors. *Eur. Heart J.* 41, 1810–1817. doi: 10.1093/eurheartj/ehaa373

Schurz, H., Salie, M., Tromp, G., Hoal, E. G., Kinnear, C. J., and Möller, M. (2019). The X Chromosome and Sex-Specific Effects in Infectious Disease Susceptibility. *Hum. Genomics* 13:2. doi: 10.1186/s40246-018-0185-z

Sha, J., Qie, G., Yao, Q., Sun, W., Wang, C., Zhang, Z., et al. (2021). Sex Differences on Clinical Characteristics, Severity, and Mortality in Adult Patients With Covid-19: A Multicentre Retrospective Study. *Front. Med. (Lausanne)* 8:607059. doi: 10.3389/fmed.2021.607059

Shi, S., Qin, M., Shen, B., Cai, Y., Liu, T., Yang, F., et al. (2020). Association of Cardiac Injury With Mortality in Hospitalized Patients With Covid-19 in Wuhan, China. *JAMA Cardiol.* 5, 802–810. doi: 10.1001/jamacardio.2020.0950

Sue, K. (2017). The Science Behind "Man Flu". *BMJ* 359, j5560. doi: 10.1136/bmj.j5560

Suen, L. K. P., So, Z. Y. Y., Yeung, S. K. W., Lo, K. Y. K., and Lam, S. C. (2019). Epidemiological Investigation on Hand Hygiene Knowledge and Behaviour: A Cross-Sectional Study on Gender Disparity. *BMC Public Health* 19, 401. doi: 10.1186/s12889-019-6705-5

Su, W., Qiu, Z., Zhou, L., Hou, J., Wang, Y., Huang, F., et al. (2020). Sex Differences in Clinical Characteristics and Risk Factors for Mortality Among Severe Patients With COVID-19: A Retrospective Study. *Aging* 12, 18833–18843. doi: 10.18632/aging.103793

Vahidy, F. S., Pan, A. P., Ahnstedt, H., Munshi, Y., Choi, H. A., Tiruneh, Y., et al. (2021). Sex Differences in Susceptibility, Severity, and Outcomes of Coronavirus Disease 2019: Cross-sectional Analysis From a Diverse US Metropolitan Area. *PLoS One* 16, e0245556. doi: 10.1371/journal.pone.0245556

Wenham, C., Smith, J., and Morgan, R. (2020). Covid-19: The Gendered Impacts of the Outbreak. *Lancet* 395, 846–848. doi: 10.1016/S0140-6736(20)30526-2

World Health Organization. (2020). *Coronavirus Disease 2019 (Covid-19) Situation Report-165*. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200703-covid-19-sitrep-165.pdf?sfvrsn=b27a772e_2 (Accessed July 15, 2020).

World Health Organization. (2020). *Laboratory Testing for 2019 Novel Coronavirus (2019-nCoV) in Suspected Human Cases Interim Guidance*. Available at: https://www.who.int/publications-detail/laboratory-testing-for-2019-novel-coronavirus-in-suspected-human-cases-20200117 (Accessed May 20, 2020).

World Health Organization. (2020). *Novel Coronavirus-China*. Available at: http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ (Accessed May 15, 2020).

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

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.

ORIGINAL RESEARCH
published: 12 August 2021
doi: 10.3389/fcimb.2021.683409

Dynamics of SARS-CoV2 Infection and Multi-Drug Resistant Bacteria Superinfection in Patients With Assisted Mechanical Ventilation

Annarita Mazzariol[1*], *Anna Benini*[2], *Ilaria Unali*[1], *Riccardo Nocini*[3], *Marcello Smania*[4], *Anna Bertoncelli*[1], *Francesco De Sanctis*[5], *Stefano Ugel*[5], *Katia Donadello*[6], *Enrico Polati*[6] and *Davide Gibellini*[1]

[1] Microbiology Section, Department of Diagnostics and Public Health, University of Verona, Verona, Italy, [2] Pharmacology Section, Department of Diagnostics and Public Health, University of Verona, Verona, Italy, [3] Department of Otolaryngology-Head and Neck Surgery, University Hospital of Verona, Verona, Italy, [4] Unità Operativa Complessa of Microbiology, University and Hospital Trust of Verona, Verona, Italy, [5] Immunology Section, Department of Medicine, University and Hospital Trust of Verona, Verona, Italy, [6] Intensive Care Unit, Department of Surgery, Dentistry, Maternity and Infant, University and Hospital Trust of Verona, Verona, Italy

OPEN ACCESS

Edited by:
Ian George Barr,
WHO Collaborating Centre for Reference and Research on Influenza (VIDRL), Australia

Reviewed by:
Tonyia Eaves-Pyles,
University of Texas Medical Branch at Galveston, United States
Benjamin M. Liu,
Johns Hopkins University, United States

****Correspondence:***
Annarita Mazzariol
annarita.mazzariol@univr.it

Specialty section:
This article was submitted to Clinical Microbiology, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *24 March 2021*
Accepted: *22 July 2021*
Published: *12 August 2021*

Citation:
Mazzariol A, Benini A, Unali I, Nocini R, Smania M, Bertoncelli A, De Sanctis F, Ugel S, Donadello K, Polati E and Gibellini D (2021) Dynamics of SARS-CoV2 Infection and Multi-Drug Resistant Bacteria Superinfection in Patients With Assisted Mechanical Ventilation. Front. Cell. Infect. Microbiol. 11:683409. doi: 10.3389/fcimb.2021.683409

Objective: To investigate the presence of bacteria and fungi in bronchial aspirate (BA) samples from 43 mechanically ventilated patients with severe COVID-19 disease.

Methods: Detection of SARS-CoV-2 was performed using Allplex 2019-nCoV assay kits. Isolation and characterisation of bacteria and fungi were carried out in BA specimens treated with 1X dithiothreitol 1% for 30 min at room temperature, using standard culture procedures.

Results: Bacterial and/or fungal superinfection was detected in 25 out of 43 mechanically ventilated patients, generally after 7 days of hospitalisation in an intensive care unit (ICU). Microbial colonisation (colony forming units (CFU) <1000 colonies/ml) in BA samples was observed in 11 out of 43 patients, whereas only 7 patients did not show any signs of bacterial or fungal growth. *Pseudomonas aeruginosa* was identified in 17 patients. Interestingly, 11 out of these 17 isolates also showed carbapenem resistance. The molecular analysis demonstrated that resistance to carbapenems was primarily related to OprD mutation or deletion. *Klebsiella pneumoniae* was the second most isolated pathogen found in 13 samples, of which 8 were carbapenemase-producer strains.

Conclusion: These data demonstrate the detection of bacterial superinfection and antimicrobial resistance in severe SARS-CoV-2-infected patients and suggest that bacteria may play an important role in COVID-19 evolution. A prospective study is needed to verify the incidence of bacterial and fungal infections and their influence on the health outcomes of COVID-19 patients.

Keywords: SARS-Cov-2, bronchial aspirate samples, mechanically ventilated patients, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, bacterial superinfection, antimicrobial resistance

INTRODUCTION

In the current era of multi-drug-resistant bacteria, a new viral pandemic due to the SARS-CoV-2 virus began near the end of 2019. SARS-CoV-2 is the cause of coronavirus disease 2019 (COVID-19), which had infected 113 472 187 people and resulted in 2 520 653 deaths worldwide as of 02 March 2021 (WHO website). Coronaviruses are an important viral family composed of seven viruses: four of these (HKU1CoV, OC43CoV, NL63CoV, and 229ECoV) are related to mild respiratory diseases, whereas the other three (SARS-CoV, SARS-CoV-2, and MERS-CoV) are associated with severe pneumonia. In particular, SARS-CoV induced severe acute respiratory syndrome (SARS) outbreaks in 2002 and 2003, with 10% mortality (Drosten et al., 2003; Ksiazek et al., 2003), while MERS-CoV caused severe respiratory disease outbreaks in 2012 in the Middle East, with 36% mortality (Zaki et al., 2012).

SARS-CoV-2 was first isolated in several cases of pneumonia of unknown origin in Wuhan, China; three months later, WHO declared a SARS-CoV-2 pandemic. The origin of SARS-CoV-2 is not fully understood, but the high genome sequence homology with bat and pangolin coronaviruses suggests a possible spillover from bats to humans through pangolins, with the acquisition of an essential basic amino acid sequence in the viral S protein (Letko et al., 2020). The spread of SARS-CoV-2 infection occurred quickly and the clinical impact was dramatic, especially in comorbid patients >65 years old. Notwithstanding the viral damage to pulmonary structures, SARS-CoV-2-induced alteration of the immune response has also been observed in critically ill COVID-19 patients (Pedersen and Ho, 2020; Bost et al., 2021).

In this paper, we investigated the co-occurrence of bacterial superinfections in severe COVID-19 patients. It is well known that the patients at the greatest risk for multi-drug-resistant infections are those who are already more vulnerable to viral lung infections (McCullers, 2014). This association has severe implications for global health, as bacterial co-infection and/or superinfection leads to both increased morbidity and mortality (Smith and McCullers, 2014; Morris et al., 2017). While the impacts of co-infection with bacteria and COVID-19 have not been well studied, people who are already immunocompromised by one organism are usually much more susceptible to a secondary infection and increased mortality. Interestingly, superinfection is also considered an important risk factor for COVID-19-related mortality (Mirzaei et al., 2020; Zhang et al., 2020).

Several antivirals and antimicrobials have been suggested for the management of COVID-19, with contrasting and unclear results (Morris et al., 2017; Rawson et al., 2020). Select antimicrobial therapy appears to be effective in the treatment of suspected or confirmed bacterial respiratory co-infection or superinfection; this therapy may be empiric or targeted and be conducted in patients presenting to the hospital or for the management of nosocomial infections. Recently, Cox *et al.* (Cox et al., 2020) highlighted the importance in the analysis and management of COVID-19 patients' co-infections of determining whether these induce disease progression and appropriately tailoring antimicrobial stewardship. It is vital that we understand whether the COVID-19 pandemic may have contributed to an increase in antimicrobial resistance, since this is a growing problem with implications for both global health and the world economy (NoAuthor, 2020).

In order to pinpoint the dynamics of bacterial and/or fungal superinfection in SARS-CoV-2 patients, we investigated the presence of bacteria and fungi in bronchial aspirate (BA) samples of mechanically ventilated patients with severe COVID-19 disease.

MATERIALS AND METHODS

Sample Population and Data Collection

We analysed BA samples obtained during routine clinical specimen collection from 43 hospitalised patients with confirmed SARS-CoV-2 infection who were subjected to mechanical ventilation (MV) between March and April 2020. BA samples were collected as part of the normal diagnostics routine from 43 patients with clinical symptoms of SARS-CoV-2 infection that required mechanical ventilation.

Viral Infection Determination

The samples were transferred using viral SARS-Cov-2 extraction kits, and amplification was performed using Allplex 2019-nCoV assay kits (Seegene Inc., Seoul, South Korea). This multiplex real-time RT-PCR assay is based on simultaneous amplification of three viral target genes (E, N, and RdRP).

At admission, multiplex Allplex respiratory panel assays (Seegene Inc.) for the detection of 17 viruses (influenza virus A [H1, H3] and B; RSV A and B; adenovirus; enterovirus; rhinovirus; coronavirus [OC43, 229E, NL63]; bocavirus, metapneumovirus, and parainfluenza virus [PIV 1, 2, 3, 4]) were performed using nasopharyngeal swab samples.

Bacterial Cultures

The BA samples were cultured to determine the presence of bacterial or fungal co-infections/superinfections. We considered superinfection a bacterial/fungal infection acquired after 48 h after admission of a COVID-19 patient (Garcia-Vidal et al., 2021). Samples appearing viscous were treated with 1X dithiothreitol 1% (Sigma, St Louis, USA), a reducing mucolytic agent, for 30 min at room temperature. Twenty microliters (20 µl) of treated samples were streaked out on various agar media—namely, blood agar, chocolate agar, Columbia Nalidixic acid agar, McConkey agar, mannitol salt agar, and Sabouraud dextrose agar—in order to identify all pathogens. Samples were incubated at 37°C for 18–24 h, and semi-quantitative analysis of bacterial and fungal growth was performed. Bacterial/fungal load of 10^5–10^6 colony forming units (CFU)/ml was considered as infection, while bacterial/fungal load lower than 10^3 CFU/ml was considered as coloniser.

Bacteria Identification and Antimicrobial Susceptibility Tests

All isolated strains were identified using a VITEK MS MALDI-TOF system (BioMérieux, Marcy L'Etoile, France). Antimicrobial susceptibility was determined using an agar diffusion test, and the results were interpreted following the EUCAST 2019 criteria https://eucast.org/clinical_breakpoints/.

Resistance Tests

The presence of carbapenemase enzymes was investigated using i) the NG-Test CARBA 5 (NG-Biotech, Guipry, France), an immunochromatographic assay that uses a multiplex lateral flow immunoassay (LFIA) to detect NDM, KPC, VIM, IMP, and OXA-48-like enzymes, and ii) specific PCR to determine the presence of the *bla*$_{KPC}$ gene (Mazzariol et al., 2012), with the amplicons sequenced (MWG Operons, Ebersberg, Germany) as indicated in the subsequent sections. OprD porin that is involved in the carbapenem resistance of *P. aeruginosa* strains was also investigated. Deletions or mutations of the *oprD* gene were analysed by PCR (Ocampo-Sosa et al., 2012) and sequenced.

RESULTS

Cohort Characteristics and Infection Data

We studied 43 severe COVID-19 patients. All patients were hospitalised between March and April 2020 and exhibited positive detection of viral genes through a SARS-CoV-2 RT-PCR assay. The average age of the patients was 61.4 years, and 21% of the patients were female.

Analysis of Microbiological Samples

All patients were followed for SARS-CoV-2 infection during the clinical course of the disease. The RT-PCR determination of three major viral target genes (E, N, and RdRP) demonstrated the presence of viral infection through BA samples. The data in **Figure 1** show when bacterial colonization/infection started after admission to the intensive care unit (ICU) and demonstrate that positive test results were obtainable from patients for a variable number of days. Bacteria were typically detected 5–15 days after hospital admission. **Figure 2** represents the relationship between the presence of SARS-CoV-2 and bacterial load. It is noteworthy that while all patients started with viral infections, the Ct value of viral gene amplification decreased in all patients until it was no longer detectable. Meanwhile, in the BA samples of these patients, increasing loads of *P. aeruginosa* and/or *K. pneumoniae* were recovered. When the bacteria reached a high load, the virus was usually no longer detectable, at least in the BA samples; the bacteria took over and constituted the major actors in the evolution of the infectious process.

Interestingly, bacterial and/or fungal superinfection was detected in 25 out of 43 patients (58.13%), generally after 7 days in the ICU with MV. Microbial colonisation (CFU <1000 colonies/ml) in BA samples was observed in 11 out of the 43 patients, whereas only seven patients did not show any signs of bacterial or fungal growth. **Table 1** reports the bacterium and fungus species isolated from BA samples, with the semi-quantitative load and the detection of SARS-CoV-2.

Analysis of the bacteria and fungi involved in these superinfections showed that the most frequently isolated bacterial species was *P. aeruginosa*, which was found in 17 out of 43 patients (39.53%), all of whom had bacterial superinfections (68% of superinfected patients). Eleven out of 17 isolates were carbapenem-resistant *P. aeruginosa*, considered to be multi-drug-resistant strains. Nine of these 11 carbapenem-resistant *P. aeruginosa* strains were detected in sequential BA samples with a high load, ranging from 10^5 to 10^6 CFU/ml. The remaining six (i.e., non-carbapenem-resistant) *P. aeruginosa* isolates produced antibiotic-susceptible strains. However, only one of these susceptible *P. aeruginosa* strains played a role as an infective agent; the other five isolates can be considered simply as coloniser strains. It is noteworthy that in one patient (number 7, **Table 2**), *P. aeruginosa* infection started with a susceptible strain that subsequently acquired resistance to antibiotics.

We also investigated the characteristics of antibiotic resistance in *P. aeruginosa* strains. Resistance to carbapenems in multi-drug-resistant *P. aeruginosa* isolates was related to *oprD* mutation or deletion (**Table 2**), not the detection of carbapenemases. Five out of 11 carbapenem-resistant *P. aeruginosa* strains involved in infection showed *oprD* gene deletion and produced a negative result following *oprD*-specific PCR amplification. The other six resistant strains showed successful *oprD* gene amplification; however, after sequencing, the gene displayed insertions, deletions, or mutations with amino acid substitution, with deleterious consequences for OprD expression and synthesis. As expected, all six susceptible strains presented an *oprD* gene encoding for a functional porin that facilitates carbapenem entry into the cell.

As reported in **Table 2**, the second most isolated pathogen was *K. pneumoniae*, detected in samples from 13 out of 43 patients (30.23%) and representing 52% of the superinfected patients (13 out 25). The phenotypic distribution of the 13 K. *pneumoniae* strains consisted of two susceptible strains, three ESBL-producer strains, and eight carbapenemase-producer strains. All the carbapenemase-producer strains harboured a *bla*$_{KPC}$ gene and expression was confirmed by CARBA 5 immunochromatographic tests. Both ESBL and carbapenemase producers were multi-drug resistant, with few therapeutic choices. In addition, 8 out of the 13 (61.53%) *K. pneumoniae* isolates exhibited high loads and were isolated in multiple BA samples from these patients, while 6 were KPC producers.

Nine patients were infected with both *P. aeruginosa* and *K. pneumoniae*; in four of these individuals, high loads of both strains were detected. In these high-load patients, the *P. aeruginosa* strains always showed a resistant pattern, while the *K. pneumoniae* strains exhibited variable patterns of susceptibility—namely, one susceptible, one ESBL producer, and two KPC producers. Other bacteria that were isolated and play a role as an infectious agent were *Stenotrophomonas maltophilia* and *Escherichia coli*, which were each detected only in one case. High loads of fungi were detected in just one patient, together with *E. coli* superinfection; otherwise, fungi were found only as colonisers at very low loads (less than 10^3 CFU/ml).

DISCUSSION

The SARS-CoV-2 virus is the etiological agent of COVID-19. This airborne infection can present as a SARS with the need for MV. WHO reported mortality rates ranging from less than 0.1% to over 25% in different countries, while reports from Johns Hopkins

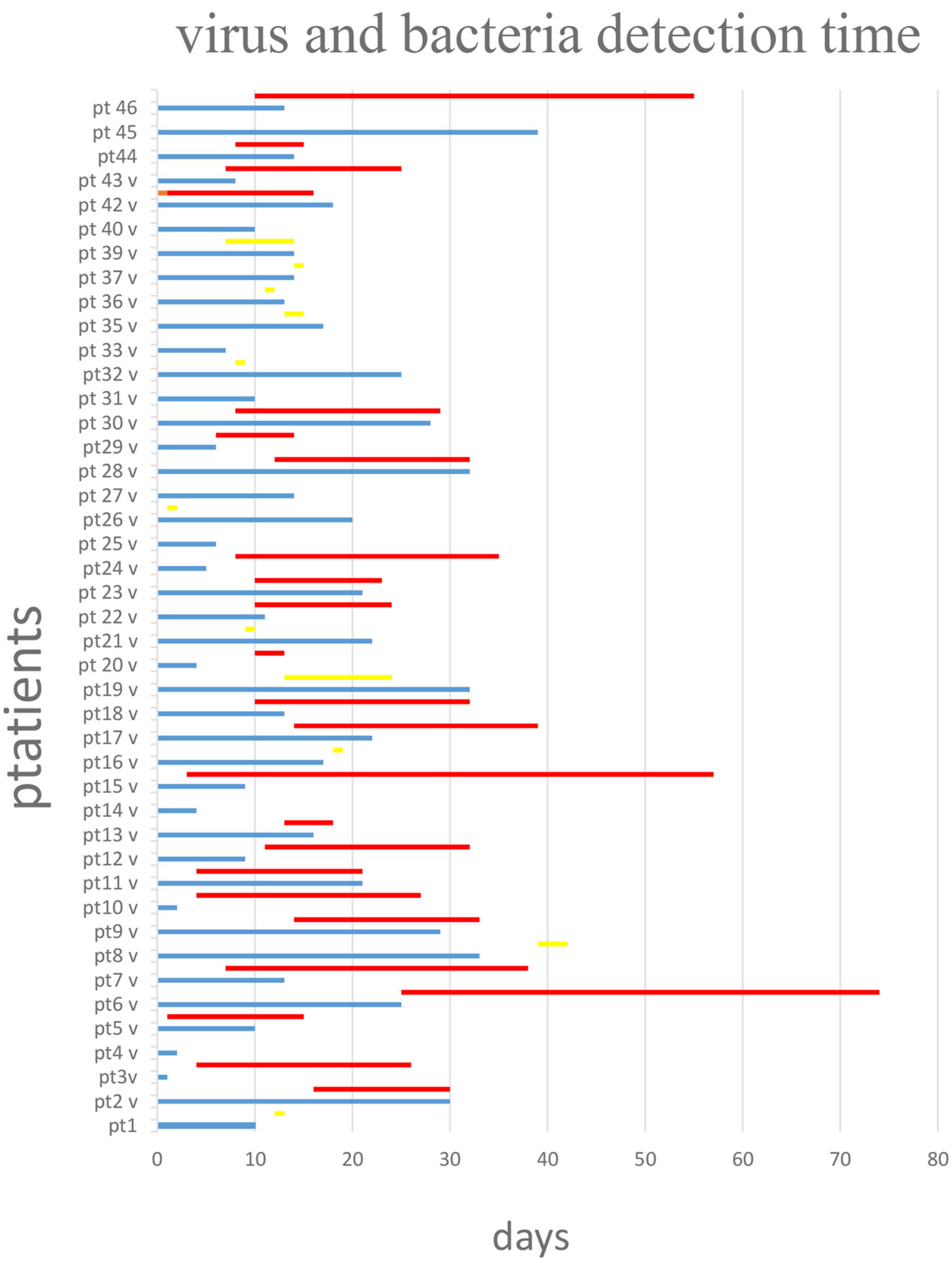


FIGURE 1 | Virus (blue lines) and bacteria (red line if high load, yellow line if in low load) detection over the time starting from day 0 that represent the admission in intensive care unit. Virus and bacteria were detected in the BA of the 43 patients with COVID-19 disease. Blue line: days of virus detection. Red line: days of bacteria detection in high load. Yellow line: days of bacteria detection in low load.

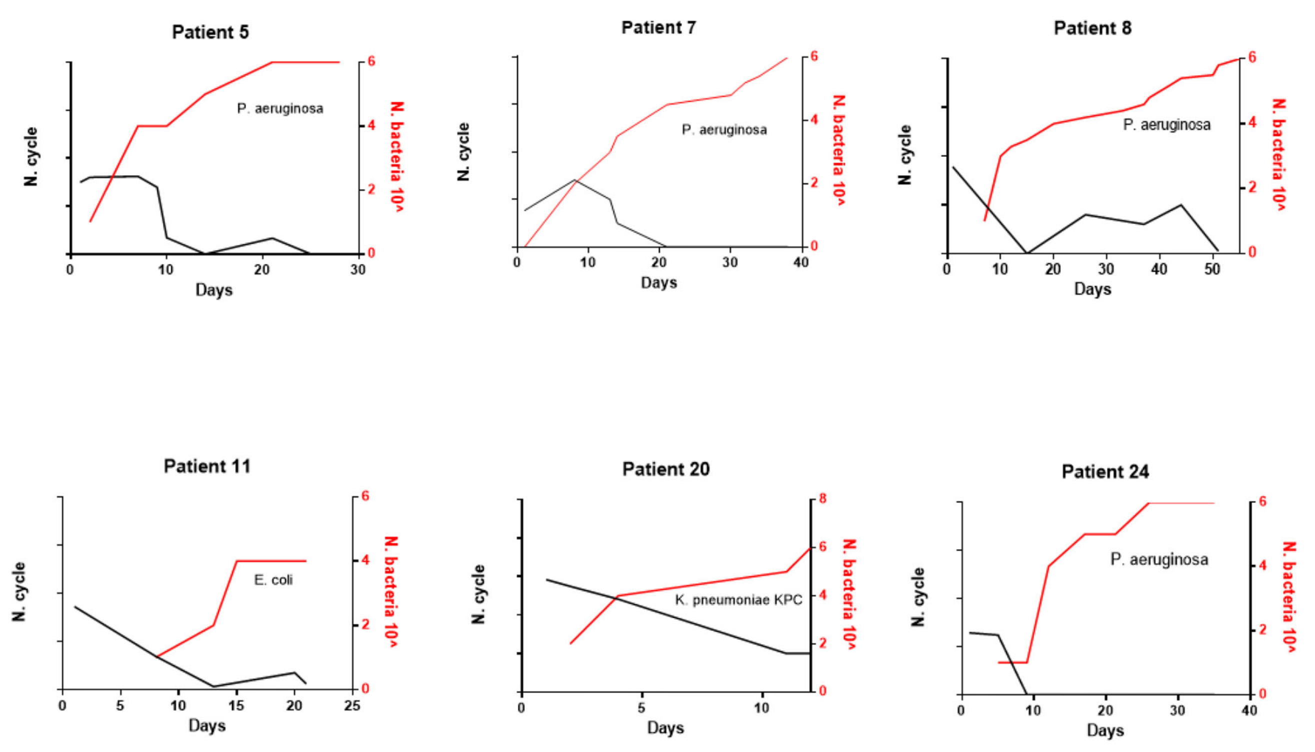


FIGURE 2 | Evolution of SARS-CoV-2 c/t detection (black line) and bacteria load (red line) over the time of hospitalization in six representative patients.

University estimated a mortality rate of 3–4% for COVID-19 patients. Pedersen *et al.* (Pedersen and Ho, 2020) reported studies showing increased levels of IL-6, IL-10 and TNF-α in SARS-CoV-2-infected individuals. The concentrations of these inflammatory cytokines, which are also involved in other systemic inflammatory syndromes such as cytokine storms, declined during patients' recovery from SARS-CoV-2 infection (Liu and Hill, 2020; Liu et al., 2021). Interestingly, bacterial superinfection plays a very important role in influenza virus infection and other respiratory virus infections. A review by McCullers (Bost et al., 2021) reported the complex interaction between bacterial and influenza viruses in pulmonary pathogenesis, indicating that bacterial superinfections might promote severe disease and mortality.

Cox et al. (2020) underlined the importance of the accurate management of co-infections and the evaluation of related antimicrobial resistance. To elucidate this aspect, we collected microbiologic data from 43 severe COVID-19 patients. Diagnosis was performed using real-time PCR, and all patients showed SARS-CoV-2 target amplification at early Ct values. All patients were admitted to the ICU and required MV. BA samples were analysed for SARS-CoV-2, and some were also cultured to investigate the presence of bacteria and fungi. Over half (58.13%) of patients showed the presence of high loads of bacteria. These data might indicate an important role of bacterial secondary infection. The occurrence of bacterial co-infections/ superinfection reported in COVID-19 patients ranges from 5.1% (Chen et al., 2020) to 29.8% (Zhang et al., 2020). Interestingly, as shown in **Figure 2**, SARS-CoV-2 Ct values always progressively declined until becoming undetectable, while bacterial load simultaneously increased.

P. aeruginosa, which was isolated from 68% of patients showing bacterial infection, played a very important role in superinfection. High loads of carbapenem-resistant *P. aeruginosa* were detected in 10 out of 11 patients. These data are intriguing as they increase our knowledge about SARS-CoV-2 bacterial superinfections. Indeed, Mirzaei *et al.* (Mirzaei et al., 2020) reported the known microorganisms that co-infected and/ or were responsible for pneumonia in SARS-CoV-2-infected patients. They listed atypical pneumonia agents such as *Mycoplasma pneumoniae* and *Legionella pneumophila* that were not found in our patients, along with bacteria such as *Staphylococcus aureus*, *K. pneumoniae*, *Enterobacter cloacae*, and *Acinetobacter baumannii*. This discrepancy between *Mirzaei et al.*'s findings and our study could be related to differences in epidemiology. The second most frequently detected microorganism was *K. pneumoniae*, another classic pathogen responsible for nosocomial pneumonia. In this case as well, multi-drug-resistant microorganisms were predominant. Eight out of 13 strains were KPC-producer strains that are endemic in Italy (Giani et al., 2013). Six patients exhibited high loads of these strains.

Both multi-drug-resistant *P. aeruginosa* and *K. pneumoniae* showed carbapenem resistance, although this was supported by different mechanisms. Our investigation showed that KPC production by *K. pneumoniae* and *oprD* gene deletion or mutation in the case of *P. aeruginosa* results in a non-functional

TABLE 1 | SARS-CoV-2 and bacteria detection for each patient with superinfection are reported with the SARS-CoV-2 gene N c/t and the bacterial identification and load at the corresponding day of hospitalization.

Patient	Days of hospitalization	SARS-CoV-2 gene N c/t	Bacteria	Bacterial load
2	1	23.31		
	9	29.83		
	16	33.78		
	21		*E. aerogenes*	+++
	23	36.22	*P. aeruginosa*	+
	30	32.41	*P. aeruginosa*	+++
			K. pneumoniae	+
3	20	23.70	*P. aeruginosa*	+++
	26		*P. aeruginosa*	+++
4	2	29.39	*K. pneumoniae*	+
5	1	25.02	*P. aeruginosa*	+
			K. pneumoniae	+
	7	23.79	*P. aeruginosa*	+
	8	–	*E. faecium*	+
	9	26.07		
	15	NR	*P. aeruginosa*	++
	21	26.68		
6	1	21.91		
	12	33.50		
	14	33.36		
	18	34.40		
	25	39.31		
	26	NR	*S. maltophylia*	+++
	28	NR	*S. maltopylia*	+++
	30	NR	*S. maltophylia*	+
	32	NR	*S. maltophylia*	+++
	35	NR	*E. faecalis*	+
	46	NR	*S. maltophylia*	+
	54	NR	*S. maltophylia*	+++
	58	NR	*S. maltophylia*	+
			K. pneumoniae	+
	74	NR	*P. aeruginosa*	+++
7	2	32.27	*P. aeruginosa*	+
	13	25.85		
	14		*P. aeruginosa*	+
			K. pneumoniae	++
	21		*P. aeruginosa*	+++
			K. pneumoniae	+
	30		*P. aeruginosa*	+
	31		*P. aeruginosa*	++
	32		*P. aeruginosa*	+
			C. albicans	++
	34		*P. aeruginosa*	+
	38		*P. aeruginosa*	+++
8	2	22.10		
	10		*Yeasts*	+
	12	15.07		
	15		*Yeasts*	++
	26			
	33	26.81		
	37	25.21		
	38	33.96		
	40	34.15	*P. aeruginosa*	++
	42		*P. aeruginosa*	++
	44	28.15		
	50	38.23		
	51	39.44		
9	8	28.90		
	15	33.35	*E. aerogenes*	+
	23	39.79	*K. pneumoniae*	+++
	29	39.25	*K. pneumoniae*	+
	36	36.34		

(Continued)

TABLE 1 | Continued

Patient	Days of hospitalization	SARS-CoV-2 gene N c/t	Bacteria	Bacterial load
	37		*P. aeruginosa*	++
			K. pneumoniae	+
			E. aerogenes	rco
10	1	24.54		
	5		*E. cloacae*	++
			P. mirabilis	+
			Yeasts	+
	19		*P. aeruginosa*	+
	26		*P. aeruginosa*	+
	27		*P. aeruginosa*	+++
12	1	26.71		
	2	29.08		
	5	31.06		
	8	28.88		
	16		*K. pneumoniae*	++
	22	38.53	*K. pneumoniae*	+
	27		*K. pneumoniae*	+
	29		*P. aeruginosa*	+
	31		*K. pneumoniae*	+
			P. aeruginosa	+
13	1	26.23	*K. pneumoniae*	+
	15	36.13	*K. pneumoniae*	+
	16	38.86		
	17		*K. pneumoniae*	+++
15	1	28.69		
	2	23.55		
	4		*E. aerogenes*	+
			P. aeruginosa	++
	6	36.39		
	7		*P. aeruginosa*	+++
	11		*P. aeruginosa*	+
	18		*P. aeruginosa*	+
	33	35.80	*P. aeruginosa*	+
	38		*P. aeruginosa*	++
	41		*P. aeruginosa*	+++
	45	37.91		
	46		*P. aeruginosa*	+
	57		*P. aeruginosa*	+
17	1	16.54		
	8	30.70		
	15	38.92	*P. aeruginosa*	+++
	22	38.36	*P. aeruginosa*	++
	25		*P. aeruginosa*	+
			K. pneumoniae	+
	38		*P. aeruginosa*	
			K. pneumoniae	
	39		*P. aeruginosa*	
			K. pneumoniae	
19	1	26.96		
	10	31.51		
	15		*P. aeruginosa*	+
	17	32.14		
	22		*P. aeruginosa*	+
	25	38.23	*P. aeruginosa*	+
	26			
	31	38.21		
	32	36.66		
20	1	20.77		
	2	22.54		
	4	25.85		
	11		*K. pneumoniae*	++
	12		*K. pneumoniae*	+++
22	1	24.40		

(Continued)

TABLE 1 | Continued

Patient	Days of hospitalization	SARS-CoV-2 gene N c/t	Bacteria	Bacterial load
	11	33.13	*P. aeruginosa*	+++
	14		*P. aeruginosa*	+
	24	16.84	*P. aeruginosa*	++
24	1	27.12		
	5	27.58		
	10		*P. aeruginosa*	+
	13		*P. aeruginosa*	+
	20		*P. aeruginosa*	+
	24		*P. aeruginosa*	++
	33		*P. aeruginosa*	+++
	35		*P. aeruginosa*	+++
28	1	18.14		
	6	18.20		
	13	23.80	*K. pneumoniae*	+
			C. albicans	++
	14	26.92		
	18	25.62	*K. pneumoniae*	+
	25	23.65		
	32	38.11	*K. pneumoniae*	++
	34	39.12		
35	1	19.01		
	5	27.98		
	7	24.43		
	10	24		
	14	30.5	*P. aeruginosa*	+
	17	30.39	*K. pneumoniae*	+
39	1	20.33		
	6	22.33		
	8	30.06	*P. aeruginosa*	+
	11	30.11		
	14	30.19	*P. aeruginosa*	+++
	19	34.66		
	20	34.42		
42	1	25.05	*E. coli*	+++
	2	28.31		
	4			
	9	27.76	*E. coli*	+++
	12	28.98		
	15	33.63	*E. coli*	++
	18	32.88		
43	1	27.27		
	3	30.32		
	8	28.67	*P. aeruginosa*	+
	10			
	13		*P. aeruginosa*	+
	19	38.48	*P. aeruginosa*	+
	25		*P. aeruginosa*	++
46	1	24.85	*H. parainfluenzae*	+
			Streptococcus spp	+
	11	24.83		
	13	26.44		
	24		*P. aeruginosa*	+
	36		*P. aeruginosa*	+
	45		*P. aeruginosa*	+
	46		*P. aeruginosa*	++
	47		*P. aeruginosa*	+++

+ : 10^{1}-10^{2}-10^{3}
+: 10^{4}
++: 10^{5}
+++10^{6}

TABLE 2 | Antimicrobial profile, mechanisms of resistance, and load of the two main pathogens recovered in superinfected patients.

Patient N.	*P. aeruginosa*						*K. pneumoniae*						Outcome
	Antimicrobial profile				Resistance mechanisms	Load	Antimicrobial profile				Resistance mechanisms	Load	
	MPM	CAZ	CIP	AK			MPM	CAZ	CIP	AK			
2	R	S	S	S	OprD MUT	***	S	R	R	S	ESBL	*	Other ICU
3	R	S	R	S	OprD DEL	***	S	R	R	S	ESBL	*	DEAD
4	–	–	–	–	–	–	S	S	S	S	–	*	Pneumology
5	S	S	S	S	–	**	S	S	S	S	–	*	Infectioous disease
6	S	S	S	S	–	***	R	R	R	S	KPC	*	Pneumology
7	S→R	S→R	S→R	S	OprD DEL	***	S	S	S	S	–	**	DEAD
9	R	S	S	S	OprD MUT	**	R	R	R	S	KPC	***	DEAD
10	R	S	S	S	OprD MUT	*	–	–	–	–	–	–	Pneumology
12	R	S	S	–	OprD MUT	*	S	I	S	S	ESBL	**	DEAD
13	–	–	–	–	–	–	R	R	R	S	KPC	***	Other ICU
15	R	R	S	S	OprD MUT	***	–	–	–	–	–	–	Pneumology
17	R	R	S	S	OprD DEL	***	R	R	R	R	KPC	*	Pneumology
19	S	S	S	S	–	*	–	–	–	–	–	–	DEAD
20	–	–	–	–	–	–	R	R	R	I	KPC	***	Other ICU
22	S	S	S	S	–	***	–	–	–	–	–	–	Pneumology
24	R	R	R	S	OprD DEL	**	–	–	–	–	–	–	Other ICU
28	–	–	–	–	–	–	R	R	R	R	KPC	***	Other ICU
35	S	S	S	S	–	***	R	R	R	S	KPC	***	Medicine
39	S	R	S	S	–	***	–	–	–	–	–	–	Pneumology
43	R	S→R	S	S	OprD DEL	**	–	–	–	–	–	–	Pneumology
46	R	R	S	S	OprD MUT	***	R	R	R	S	KPC	*	Pneumology

**: 10^1; 10^2; 10^3*

: 10^4; **: 10^5; *: 10^6*

MPM, Meropenem; CAZ, ceftazidime; CIP, Ciprofloxacin; AK, Amikacin; R, Resistant; S, Susceptible; OPR D MUT, mutations present in the Porin D of P. aeruginosa affecting meropenem resistance; OPR D DEL, deletion of the porin D gene in P. aeruginosa affecting meropenem resistance; ESBL, Extended Spectrum Beta-Lactamase; KPC, Klebsiella pneumoniae carbapenemase; ICU, intensive care unit.

outer membrane channel are responsible of carbapenem resistance. These data support the idea that we are faced with hypermutable strains of *P. aeruginosa*. This hypothesis must be investigated further, together with the possible genetic relatedness of isolates.

This work analysed the impact of SARS-CoV-2 infection on antimicrobial resistance. Concerns about antimicrobial resistance did not disappear during the pandemic; rather, it has become a pivotal concern in the management of SARS-CoV-2-infected patients. The use of antibiotics to treat pneumonia caused by COVID-19 must take into account the fact superinfection could, in many cases, be sustained by multi-drug-resistant microorganisms with a substantial impact on therapeutic choices. This is even more important when patients are admitted to the ICU, a ward where antimicrobial resistance can develop and spread easily, especially if there is normally a high incidence of resistance. In our study, we documented a 58.13% incidence of superinfection in SARS-CoV-2-infected patients admitted to the ICU. Most of the patients in our study were infected with carbapenem-resistant *P. aeruginosa*, followed by multi-drug-resistant *K. pneumoniae*.

Rhee et al. (2020) recently reported clinical evidence suggesting increased mortality related to inadequate empiric antibiotic treatment. In this scenario, the microbiology laboratory would play a pivotal role both in identifying bacterial superinfections and determining resistance patterns and the underlying mechanisms of resistance, so as to enable health care providers to make the appropriate choices concerning antimicrobial treatments.

CONCLUSIONS

These data suggest that bacteria may play an important role in COVID-19 evolution and that antimicrobial resistance might negatively impact outcomes for affected patients. A prospective study will be necessary to verify the incidence of bacterial and fungal infections and their influence on COVID-19 outcomes.

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

AM and DG: conceptualization. IU, RN, and MS: investigation. AM, ABen, and IU: formal analysis. AM and ABen: data curation. ABen, FDS, SU, KD, and EP: resource. AM, ABen, and DG: writing original draft. AM, FS, SU, KD, EP, and DG: writing review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Cariverona Foundation, ENACT project VIRO-COVID.

REFERENCES

No Author. (2020). Antimicrobial Resistance in the Age of COVID-19. *Nat. Microbiol.* 5, 779. doi: 10.1038/s41564-020-0739-4

http://www.eucast.org/clinical breakpoints/.

Bost, P., De Sanctis, F., Canè, S., Ugel, S., Donadello, K., Castellucci, M., et al. (2021). Deciphering the State of Immune Silence in Fatal COVID-19 Patients. *Nat. Commun.* 12 (1), 1428. doi: 10.1038/s41467-021-21702-6

Chen, N., Zhou, M., Dong X Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study. *Lancet* 395, 507–513. doi: 10.1016/S0140-6736(20)30211-7

Cox, M. J., Loman, N., Bogaert, D., and O'Grady, J. (2020). Co-Infections: Potentially Lethal and Unexplored in COVID-19. *Lancet Microbe* 1, e11. doi: 10.1016/S2666-5247(20)30009-4

Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., et al. (2003). Identification of a Novel Coronavirus in Patients With Severe Acute Respiratory Syndrome. *N Engl. J. Med.* 348, 1967–1 76. doi: 10.1056/NEJMoa030747

Garcia-Vidal, C., Sanjuan, G., Moreno-García, E., Puerta-Alcalde, P., Garcia-Pouton, N., Chumbita, M., et al. (2021). Incidence of Co-Infections and Superinfections in Hospitalized Patients With COVID-19: A Retrospective Cohort Study. *Clin. Microbiol. Infect.* 27 (1), 83–88. doi: 10.1016/j.cmi.2020.07.041

Giani, T., Pini, B., Arena, F., Conte, V., Bracco, S., Migliavacca, R., et al. (2013). Epidemic Diffusion of KPC Carbapenemase-Producing *Klebsiella Pneumoniae* in Italy: Results of the First Countrywide Survey, 15 may to 30 June 2011. *Euro Surveill.* 18, 20489. doi: 10.2807/ese.18.22.20489-en

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., et al. (2003). A Novel Coronavirus Associated With Severe Acute Respiratory Syndrome. *N Engl. J. Med.* 348, 1953–1966. doi: 10.1056/NEJMoa030781

Letko, M., Seifert, S. N., Olival, K. J., and Plowright, R. K. (2020). Bat-Borne Virus Diversity, Spill Over and Emergence. *Nat. Rev. Microbiol.* 18, 461–471. doi: 10.1038/s41579-020-0394-z

Liu, B. M., and Hill, H. R. (2020). Role of Host Immune and Inflammatory Responses in COVID-19 Cases With Underlying Primary Immunodeficiency: A Review. *J. Interferon Cytokine Res.* 40 (12), 549–554. doi: 10.1089/jir.2020.0210

Liu, B. M., Martins, T. B., Peterson, L. K., and Hill, H. R. (2021). Clinical Significance of Measuring Serum Cytokine Levels as Inflammatory Biomarkers in Adult and Pediatric COVID-19 Cases: A Review. *Cytokine* 142:155478. doi: 10.1016/j.cyto.2021.155478

Mazzariol, A., Lo Cascio, G., Ballarini, P., Ligozzi, M., Soldani, F., Fontana, R., et al. (2012). Rapid Molecular Technique Analysis of a KPC-3-Producing Klebsiella Pneumoniae Outbreak in an Italian Surgery Unit. *J. Chemother.* 24, 93– 96. doi: 10.1179/1120009X12Z.00000000020

McCullers, J. A. (2014). The Co-Pathogenesis of Influenza Viruses With Bacteria in the Lung. *Nat. Rev. Microbiol.* 12, 252–262. doi: 10.1038/nrmicro3231

Mirzaei, R., Goodarzi, P., Asadi, M., Soltani, A., Aljanabi, H. A. A., Jeda, A. S., et al. (2020). Bacterial Co-Infections With SARS-CoV-2. *IUBMB Life* 72, 2097–2111. doi: 10.1002/iub.2356

ACKNOWLEDGMENTS

Results were partially presented at the 48th Virtual SIM (Italian Society of Microbiology) on November 18th, 2020.

Morris, D. E., David, W., and Cleary DW and Clarke, S. C. (2017). Secondary Bacterial Infections Associated With Influenza Pandemics. *Front. Microbiol.* 8, 1041. doi: 10.3389/fmicb.2017.01041

Ocampo-Sosa, A. A., Cabot, G., Rodríguez, C., Roman, E., Tubau, F., Macia, M. D., et al. (2012). Spanish Network for Research in Infectious Diseases (REIPI). Alterations of Oprd in Carbapenem-Intermediate and -Susceptible Strains of Pseudomonas Aeruginosa Isolated From Patients With Bacteremia in a Spanish Multicenter Study. *Antimicrob. Agents Chemother.* 56, 1703–1713. doi: 10.1128/AAC.05451-11

Pedersen, S. F., and Ho, Y. C. (2020). SARS-CoV-2: A Storm is Raging. *J. Clin. Invest.* 130, 2202–2205. doi: 10.1172/JCI137647

Rawson, T. M., Moore, L. S. P., Zhu, N., Ranganathan, N., Skolimowska, K., Gilchrist, M., et al. (2020). Bacterial and Fungal Coinfection in Individuals With Coronavirus: A Rapid Review to Support COVID-19 Antimicrobial Prescribing *Clin Infect Dis.* 71, 2459–2468. doi: 10.1093/cid/ciaa530

Rhee, C., Kadri, S. S., Dekker, J. P., Danner, R. L., Chen, H. C., Fram, D., et al. (2020). Prevalence of Antibiotic-Resistant Pathogens in Culture-Proven Sepsis and Outcomes Associated With Inadequate and Broad-Spectrum Empiric Antibiotic Use. *AMA Netw. Open* 3, e202899. doi: 10.1001/jamanetworkopen.2020.2899

Smith, A. M., and McCullers, J. A. (2014). Secondary Bacterial Infections in Influenza Virus Infection Pathogenesis. *Curr. Top. Microbiol. Immunol.* 385, 327–356. doi: 10.1038/nrmicro3231

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012). Isolation of a Novel Coronavirus From a Man With Pneumonia in Saudi Arabia. *N Engl. J. Med.* 367, 1814–1820. doi: 10.1056/NEJMoa1211721

Zhang, G., Hu, C., Luo, L., Fang, F., Chen, Y., Li, J., et al. (2020). Clinical Features and Short-Term Outcomes of 221 Patients With COVID-19 in Wuhan, China. *J. Clin. Virol.* 127, 104364. doi: 10.1016/j.jcv.2020.104364

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.

ORIGINAL RESEARCH
published: 18 August 2021
doi: 10.3389/fcimb.2021.670823

Research on Influencing Factors and Classification of Patients With Mild and Severe COVID-19 Symptoms

Xiaoping Chen[1†], *Lihui Zheng*[1†], *Shupei Ye*[2†], *Mengxin Xu*[3], *YanLing Li*[3], *KeXin Lv*[3], *Haipeng Zhu*[4], *Yusheng Jie*[5*] and *Yao-Qing Chen*[3*]

[1] *College of Mathematics and Statistics & FJKLMAA, Fujian Normal University, Fuzhou, China,* [2] *Pulmonary and Critical Care Medicine, The Third People's Hospital of Dongguan City, Dongguan, China,* [3] *School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, China,* [4] *Department of Infectious Diseases, The Ninth People's Hospital of Dongguan City, Dongguan, China,* [5] *Department of Infectious Diseases, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China*

Edited by:
Ian George Barr,
VIDRL, Australia

Reviewed by:
Shashank Ganatra,
Texas Biomedical Research Institute,
United States
Mohamed Gomaa Kamel,
Minia University, Egypt

***Correspondence:**
Yao-Qing Chen
chenyaoqing@mail.sysu.edu.cn
Yusheng Jie
jieyush@mail.sysu.edu.cn

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

Specialty section:
This article was submitted to Clinical Microbiology, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *22 February 2021*
Accepted: *25 June 2021*
Published: *18 August 2021*

Citation:
Chen X, Zheng L, Ye S, Xu M, Li Y, Lv K, Zhu H, Jie Y and Chen Y-Q (2021) Research on Influencing Factors and Classification of Patients With Mild and Severe COVID-19 Symptoms. Front. Cell. Infect. Microbiol. 11:670823. doi: 10.3389/fcimb.2021.670823

Objective: To analyze the epidemiological history, clinical symptoms, laboratory testing parameters of patients with mild and severe COVID-19 infection, and provide a reference for timely judgment of changes in the patients' conditions and the formulation of epidemic prevention and control strategies.

Methods: A retrospective study was conducted in this research, a total of 90 patients with COVID-19 infection who received treatment from January 21 to March 31, 2020 in the Ninth People's Hospital of Dongguan City were selected as study subject. We analyzed the clinical characteristics of laboratory-confirmed patients with COVID-19, used the oversampling method (SMOTE) to solve the imbalance of categories, and established Lasso-logistic regression and random forest models.

Results: Among the 90 confirmed COVID-19 cases, 79 were mild and 11 were severe. The average age of the patients was 36.1 years old, including 49 males and 41 females. The average age of severe patients is significantly older than that of mild patients (53.2 years old *vs* 33.7 years old). The average time from illness onset to hospital admission was 4.1 days and the average actual hospital stay was 18.7 days, both of these time actors were longer for severe patients than for mild patients. Forty-eight of the 90 patients (53.3%) had family cluster infections, which was similar among mild and severe patients. Comorbidities of underlying diseases were more common in severe patients, including hypertension, diabetes and other diseases. The most common symptom was cough [45 (50%)], followed by fever [43 (47.8%)], headache [7 (7.8%)], vomiting [3 (3.3%)], diarrhea [3 (3.3%)], and dyspnea [1 (1.1%)]. The laboratory findings of patients also included leukopenia [13(14.4%)] and lymphopenia (17.8%). Severe patients had a low level of creatine kinase (median 40.9) and a high level of D-dimer. The median NLR of severe patients was 2.82, which was higher than that of mild patients. Logistic regression showed that age, phosphocreatine kinase, procalcitonin, the lymphocyte count of the patient on admission, cough, fatigue, and pharynx dryness were independent predictors of COVID-19 severity. The classification of random forest was predicted and the importance of each variable was displayed. The variable importance of random forest indicates that age, D-dimer, NLR (neutrophil to lymphocyte ratio) and other top-ranked variables are risk factors.

Conclusion: The clinical symptoms of COVID-19 patients are non-specific and complicated. Age and the time from onset to admission are important factors that determine the severity of the patient's condition. Patients with mild illness should be closely monitored to identify those who may become severe. Variables such as age and creatine phosphate kinase selected by logistic regression can be used as important indicators to assess the disease severity of COVID-19 patients. The importance of variables in the random forest further complements the variable feature information.

Keywords: COVID-19, mild and severe, clinical characteristic, Lasso-Logistic regression, random forest

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Zhu et al., 2020). The virus was isolated and the whole genome sequenced in a short time, and was further found to have high homology with SARS coronavirus and bat coronavirus sequences through phylogenetic analysis (Lu et al., 2020; Wu et al., 2020; Zhou et al., 2020). The emergence of SARS-CoV-2 caused a profound change all over the world.

SARS-CoV-2 is an enveloped-RNA coronavirus. Symptoms of COVID-19 are highly variable, from none to severe illness. The symptoms are primarily fever, fatigue and dry cough, while gastrointestinal symptoms are not common. Most severe cases develop dyspnea after one week, and rapidly progress to acute respiratory distress syndrome, septic shock, coagulation dysfunction, and metabolic acidosis that are difficult to correct (Guan et al., 2020; Huang et al., 2020a).

Therefore, learning from the clinical characteristics and influencing factors of patients with mild and severe COVID-19, monitoring the changes in the patient's condition, and providing strategies for severe patients entering the intensive care unit (ICU) as soon as possible, which may contribute to reduce mortality and deliver proper care.

RESEARCH OBJECTS AND METHODS

Research Objects

This study retrospectively included 90 COVID-19-infected patients admitted to the Ninth People's Hospital of Dongguan City from January 21 to March 31, 2020.All patients were confirmed in accordance with the protocol of "Diagnosis and Treatment Plan for COVID-19 (Trial Version 7)" issued by the General Office of the National Health Commission and the Office of the State Administration of Traditional Chinese Medicine on February 19, 2020, and were detected by quantitative reverse transcription PCR (RT-QPCR), and both chest radiographs and nasopharyngeal swabs and were performed for all patients upon admission. This study was approved by the Ethics Committee of the Ninth People's Hospital of Dongguan, China.

Data Set Processing

The severity (mild and serious) of COVID-19 patients is defined according to the diagnosis and description of the disease course in the electronic medical record. The laboratory values of heart rate, the white blood cell count of the patient on admission (white blood cell count 1), the lymphocyte count of the patient on admission (lymphocyte count 1), the platelet count of the patient on admission (platelet count 1), and white blood cell count after admission (white blood cell count 2), lymphocyte count after admission (lymphocyte count 2), platelet count after admission (platelet count 2), NLR(Neutrophil to lymphocyte ratio), prothrombin time, phosphocreatine kinase, phosphocreatine kinase isoenzyme, procalcitonin, D-dimer were all measured. Some indexes lack data, so the random forest method is used for interpolation. The outcome of the patients were discharged from hospital.

The proportion of mild and severe patients is unbalanced (79 mild patients and 11 severe patients). The imbalance of categories will affect the modeling effect, the prediction effect of the classification model will be reduced, and the prediction result tends to favor the majority category. Therefore, the synthetic minority sample oversampling method (SMOTE) was used to balance the data. The SMOTE method is a data preprocessing technique applied to imbalance problems proposed by (Chawla et al., 2002). It uses the K-nearest neighbors and linear interpolation to add artificially synthesized minority class samples to the data, combined with the under-sampling method in the majority class, to balance the class distribution and reduce the possibility of overfitting. The SMOTE method operates in the feature space, so that the minority decision-making regions become more general. The SMOTE method can make the classifier build a larger decision-making area, including nearby minority class points, so it can improve the classification performance (Chawla et al., 2002). At present, SMOTE algorithm has been widely used in many fields, and SMOTE can effectively speed up classification, such as random forest, decision tree, Bayesian network, etc. Some studies have shown that the accuracy of these classification models using SMOTE training models is better than that of models not using SMOTE training (Alahmari, 2020).

The parameter percent.over in the minority sample was set to 1000, and percent.under for the majority sample was set to 100. This parameter setting means that the minority class will eventually generate 1+ percent.over/100 times the original, and the majority class will eventually be generated as (percent.under/ 100)*(percent.over/100) times the minority class. Finally got 110 mild cases and 121 severe cases in the model. Since the algorithm needs to extract samples from most types of samples, the data of mild and severe cases after balancing has increased.

Research Methods

In this paper, descriptive statistical methods are used to study the clinical characteristics of mild and severe patients. A T-test was used for the data subject to normal distribution, a Wilcox rank sum test was used for the continuous skewed data, and a chi-square test was used to compare the differences between the two groups of data for the classified data. After the data is balanced, based on the Lasso-Logistic regression model to screen the risk factors and symptoms related to severe illness, a P value of less than 0.05 is considered statistically significant. This article use the Shapiro-Wilk (S-W) method to judge the normality of continuous variables.

Lasso regression is a kind of machine learning regression, which is used to select the important factors influencing the results. Leucopenia was defined as a white blood cell count below 4×10^9/L; lymphocytopenia was defined as a lymphocyte count below 1×10^9/L. The software used in this article is R 3.6.2, where the balance data is the SMOTE function in the DMwR package in the R language. The Lasso regression model uses the gelnet function in the R language. This paper also uses random forest for classification modeling. The model is a classification tree-based algorithm proposed by Breiman and Cutler, which improves the prediction accuracy of the model by summarizing a large number of classification trees. Compared with general regression, it is not sensitive to multivariable collinearity, and it is more robust to missing data and unbalanced data. At the same time, it can give the importance of variables, evaluate the role of each variable in classification, and make good predictions (Yang et al., 2020b). The model is implemented using the Random Forest function in R language.

RESULTS

Basic Information, Epidemiological History and Clinical Manifestations

Among the 90 confirmed COVID-19 patients (49 males and 41 females), 79 patients were mild and 11 patients were severe. The average age of all patients was 36.1 years, the median was 36.5 years, and 50% of patients were between 21.3-47 years old. The age quartile of mild patients ranged from 19 to 45 years, while the severe patients ranged from 46 to 59.5 years. The mean age of severe patients was 53.1 years, which is significantly higher than the mean age of mild patients (33.7 years). With the increase of age, the proportion of serious patients with increased (P < 0.05). The average time from onset to admission was 4.1 days, with a median of 2.5 days. The average actual hospital stay was 18.7 days, with a median of 18 days. And the above two for severe patients were both longer compared to mild patients. Forty-eight of 90 patients (53.3%) had family cluster infections in severe patients, which was similar among in mild patients. Nearly half of the patients suffered from basic diseases [40 (44.4%)], mainly including hypertension [11 (12.2%)], chronic liver disease [7 (7.8%)] and other basic diseases [25 (27.8%)]. Basic diseases were more common in severe patients than mild ones, while hypertension, diabetes and other diseases were also more obvious. Eleven of the 90 patients were seriously ill, 8 of whom had underlying diseases, accounting for 72.7%. There were 79 patients with mild disease, among which 32 patients had underlying disease, accounting for 40.5%. On admission, most patients had fever [43 (47.8%)] or cough [45 (50%)]. Among them, 36(45.6%) of the mild patients had fever and 7(63.6%) of the mild patients had fever. There were 37 mild patients with cough (46.8%) and 8 severe patients with cough (72.7%). A total of 29 patients (32.2%) presented with both cough and fever. Other symptoms included dyspnea [1 (1.1%)], diarrhea [3 (3.3%)], emesis [3 (3.3%)], headache [7 (7.8%)]. (**Table 1**; **Figure 1**). This suggested that elderly patients with delayed diagnosis and admission for treatment were more likely to develop severe illnesses after being infected with COVID-19. (The continuous variables in the table have been tested for normality. Among them, age and actual hospital stay follow a normal distribution. Other variables use the median (quartile).

Table 1-2 showed the demographic and clinical characteristics of the patients after the SMOTE balance. In terms of the difference indicators of the patients with mild and severe diseases, the significant variables, except age and the time from onset to admission, were newly added with basic diseases, fever, cough, expectoration, fatigue and pharynx dryness.

Laboratory Examination Indexes

The statistical results of laboratory examination indexes of 90 patients were shown in **Table 2**. There was no significant difference in the heart rate among with mild to severe cases upon admission. On admission, the blood count of the patients showed leukopenia in 13 of them (14.4%), lymphocytosis in 16 of them (17.8%), and thrombocytopenia in 0. After a period of treatment, the blood count of the patients measured during the hospitalization showed that there were 6 patients with leukopenia (6.67%), 4 patients with lymphopenia (4.44%), and 2 patients with thrombocytopenia (2.22%). These indicators were not statistically significant between mild and severe patients (*P*>0.05). The median time of development of prothrombin was 12.5 in both mild and severe cases, and the level of procalcitonin was also similar. Levels of phosphocreatine kinase were low in critically ill patients (The median of creatine phosphokinase was 40.9). The D-dimer level was higher in severe patients than in mild patients, but there was a difference in severe patients (*P* < 0.05). The NLR level of severe patients is higher than that of mild patients. The median NLR of

TABLE 1 | Demographic and clinical characteristics of COVID-19 infected patients.

Basic characteristics	All patients (n=90)	Mild (n=79)	Severe (n=11)	*P* value
Gender [n]				
female	41	36	5	1
male	49	43	6	
age (mean (SD))(years)	36.07 (18.36)	33.70 (17.69)	53.09 (13.99)	<0.001
Time from onset to admission (d)	2.50 [1.00, 4.00]	2.00 [1.00, 4.00]	4.00 [2.50, 9.00]	0.016
Actual number of days hospitalized (mean (SD))(d)	18.67 (7.49)	18.53 (7.30)	19.73 (9.08)	0.682
Family gathering infection [n(%)]	48 (53.3)	41 (51.9)	7 (63.6)	0.683
Comorbidities [n(%)]				
Have disease	40 (44.4)	32 (40.5)	8 (72.7)	0.091
Hypertension	11 (12.2)	6 (7.6)	5 (45.5)	——
Diabetes	4 (4.4)	1 (1.3)	3 (27.3)	——
Cardiovascular diseases	3 (3.3)	2 (2.5)	1 (9.1)	——
Chronic liver disease	7 (7.8)	6 (7.6)	1 (9.1)	——
Respiratory disease	5 (5.6)	4 (5.1)	1 (9.1)	——
Nervous system disease	1 (1.1)	1 (1.3)	0	——
Metabolic diseases	4 (4.4)	4 (5.1)	0	——
Blood system diseases	0	0	0	——
Chronic kidney disease	2 (2.2)	2 (2.5)	0	——
Tumor	3 (3.3)	2 (2.5)	1 (9.1)	——
other	25 (27.8)	19 (24.1)	6 (54.5)	——
Signs and symptoms [n(%)]				
Fever	43 (47.8)	36 (45.6)	7 (63.6)	0.423
Cough	45 (50)	37 (46.8)	8 (72.7)	0.198
Expectoration	23 (25.6)	18 (22.8)	5 (45.5)	0.213
Fatigue	10 (11.1)	8 (10.1)	2 (18.2)	0.776
Dyspnea	1 (1.1)	0	1 (9.1)	0.246
Diarrhea	3 (3.3)	2 (2.5)	1 (9.1)	0.811
Poor appetite	12 (13.3)	10 (12.7)	2 (18.2)	0.975
Emesis	3 (3.3)	2 (2.5)	1 (9.1)	0.811
headache	7 (7.8)	6 (7.6)	1 (9.1)	1
Muscle ache	10 (11.1)	9 (11.4)	1 (9.1)	1
Pharynx dryness	13 (14.4)	9 (11.4)	4 (36.4)	0.080

severe patients was 2.82, which was higher than the median of 2.44 for mild patients. There is no significant difference between the two groups in the original data. After class-balanced data, NLR is significant(P=0.048). The treatment plan was based on antiviral and anti-infective treatment, and severe patients used oxygen therapy more than mild patients. **Table 2-2** showed the results of laboratory examinations and treatment plans for patients after the class balance. In terms of different indicators for mild and severe patients, in addition to D-dimer, the newly added (just admitted) lymphocytes count, procalcitonin, NLR, (after admission) platelet count, creatine phosphate kinase.

Lasso-Logistics Regression Model

The dependent variable was severity. Severe was 1 and mild was 0. There were a total of 30 independent variables, including population basic information (age, gender), hospital medical detection (heart rate), time from onset to admission, actual number of days hospitalized, blood routine examination on admission (white blood cell count 1, lymphocyte count 1, platelet count 1), hospitalization, blood routine examination (white blood cell count 2, lymphocyte count 2, platelet count 2) and blood coagulation (prothrombin time, D-dimer), NLR (Neutrophil to lymphocyte ratio), myocardial enzymes including creatine phosphate kinase (CK) and creatine phosphate kinase isoenzyme (CK-MB), myocardial markers (calcitonin former), underlying disease, family gathering infection, and symptoms (fever, cough, expectoration, fatigue, dyspnea, diarrhea, poor appetite, emesis, headache, muscle aches, pharynx dryness). Some variable names are abbreviated as follows: Time from onset to admission (onset time), actual number of days hospitalized (Actual days), Family gathering infection (FG infection), Have underlying disease (HU disease), White blood cell count 1 (WBC1), Lymphocyte count 1 (LYMBH1), Platelet count 1 (PLT1), White blood cell count 2 (WBC2), Lymphocyte count 2 (LYMBH2), Platelet count 2 (PLT2), creatine phosphate kinase (CK), creatine phosphate kinase isoenzyme (CK MB).

The significance of the variables under the univariate logistic regression was calculated first. The results were shown in **Table 3**. The significant variables were age, underlying disease, fever, cough, expectoration, fatigue, pharynx dryness, time from onset to admission, creatine phosphate kinase, procalcitonin, D-dimer, lymphocyte count 2 and lymphocyte count 1.

The above variables were substituted into the Lasso regression. The complexity of Lasso was controlled by lambda, and the model obtained by different lambda was different. **Figure 2A** showed that the best lambda value can be selected according to the AUC value when compressing different variables. In **Figure 2B**, each colored line represented one variable, and the horizontal axis represented the L1 norm, which was the sum of the absolute values of the regression coefficients. The vertical axis represented the coefficient. If a vertical dashed line was drawn in the figure, the vertical line

TABLE 1-2 | The demographic and clinical characteristics of patients infected with COVID-19 after SMOTE balance data.

Basic characteristics	All patients (n=231)	Mild (n=110)	Severe (n=121)	*P* value
Gender [n]				
Female	130	57	73	0.242
Male	101	53	48	
age(years)	45.00[35.00,56.14]	35.00[20.00,41.00]	53.43[47.52,59.23]	<0.001
Time from onset to admission (d)	3.00 [2.00, 6.32]	2.00 [1.00, 4.00]	4.00 [2.51, 8.83]	<0.001
Actual number of days hospitalized (d)	19.00[14.00,23.00]	18.00[13.00,23.00]	20.36[14.56,23.28]	0.697
Family gathering infection [n(%)]	125 (54.1)	57 (51.8)	68 (56.2)	0.593
Comorbidities [n(%)]				
Have disease	134 (58.0)	46 (41.8)	88 (72.7)	<0.001
hypertension	52 (22.5)	3 (2.7)	49 (40.5)	— —
diabetes	16 (6.9)	3 (2.7)	13 (10.7)	— —
Cardiovascular diseases	8 (3.5)	2 (1.8)	6 (5.0)	— —
Chronic liver disease	24 (10.4)	7 (6.4)	17 (14.0)	— —
Respiratory disease	12 (5.2)	6 (5.5)	6 (5.0)	— —
Nervous system disease	0	0	0	— —
Metabolic diseases	4 (1.7)	4 (3.6)	0 (0.0)	— —
Blood system diseases	0	0	0	— —
Chronic kidney disease	3 (1.3)	3 (2.7)	0 (0.0)	— —
Tumor	9 (3.9)	3 (2.7)	6 (5.0)	— —
other	100 (43.3)	29 (26.4)	71 (58.7)	— —
Signs and symptoms [n(%)]				
Fever	109 (47.2)	40 (36.4)	69 (57.0)	0.003
Cough	149 (64.5)	44 (40.0)	105 (86.8)	<0.001
Expectoration	93 (40.3)	18 (16.4)	75 (62.0)	<0.001
Fatigue	33 (14.3)	8 (7.3)	25 (20.7)	0.007
Dyspnea	6 (2.6)	0 (0.0)	6 (5.0)	0.051
Diarrhea	11 (4.8)	3 (2.7)	8 (6.6)	0.282
Poor appetite	39 (16.9)	16 (14.5)	23 (19.0)	0.466
Emesis	8 (3.5)	2 (1.8)	6 (5.0)	0.345
headache	22 (9.5)	6 (5.5)	16 (13.2)	0.074
Muscle ache	16 (6.9)	11 (10.0)	5 (4.1)	0.135
pharynx dryness	64 (27.7)	10 (9.1)	54 (44.6)	<0.001

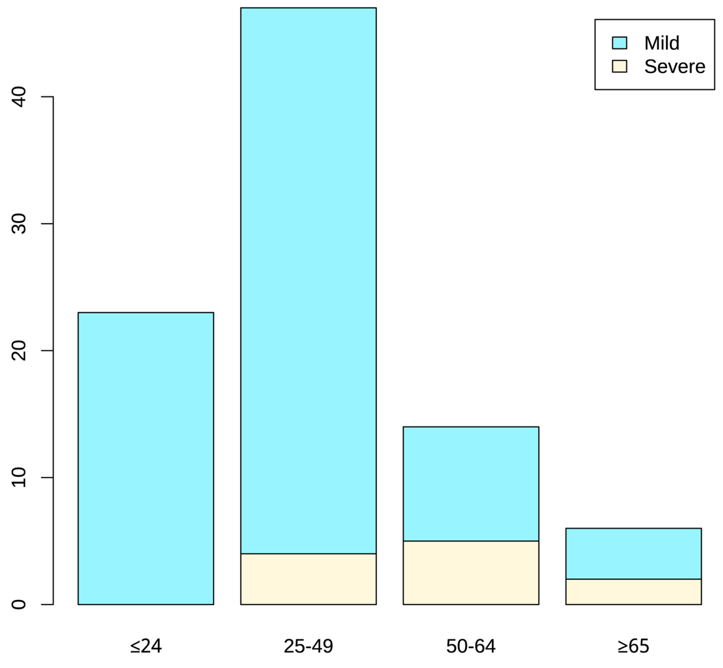


FIGURE 1 | Age distribution of patients.

TABLE 2 | Laboratory test results and treatment regimens for COVID-19 patients during hospitalization.

Laboratory index	All patients (n = 90)	Mild (n = 79)	Severe (n = 11)	*P* value
Heart rate (mean (SD))(Times/min)	90.68 (12.98)	90.73 (12.83)	90.27 (14.66)	0.923
white blood cell count 1 (×10^9/L)	5.26 [4.60, 6.38]	5.25 [4.61, 6.39]	5.75 [4.64, 6.21]	0.936
<4	13(14.4%)	11(13.9%)	2(18.2%)	— —
4-10	74(82.2%)	66(83.5%)	8(72.7%)	— —
>10	3(3.33%)	2(2.5%)	1(9.1%)	— —
Lymphocyte count 1(×10^9/L)	1.33 [1.05, 1.98]	1.39 [1.06, 2.00]	1.11 [0.89, 1.52]	0.128
<1.0	16(17.8%)	13(16.5%)	3(27.3%)	— —
≥1.0	74(82.2%)	66(83.5%)	8(72.7%)	— —
Platelet count 1(×10^9/L)	179.00 [164.00, 237.00]	179.00 [162.00, 237.00]	201.00 [171.00, 224.00]	0.54
<100	0	0	0	— —
≥100	90	79	11	— —
White blood cell count 2(×10^9/L)	5.49 [4.92, 6.78]	5.52 [4.96, 6.78]	5.04 [4.74, 6.75]	0.378
<4	6 (6.67%)	4(5.1%)	2(18.2%)	— —
4-10	83(92.2%)	74(93.7%)	9(81.8%)	— —
>10	1(1.11%)	1(1.3%)	0	— —
Lymphocyte count 2 (×10^9/L)	1.57 [1.37, 1.97]	1.58 [1.38, 1.98]	1.46 [1.23, 1.73]	0.232
<1.0	4(4.44%)	3(3.8%)	1(9.1%)	— —
≥1.0	86(95.6%)	76(96.2%)	10(90.9%)	— —
Platelet count 2 (×10^9/L)	221.34 [198.75, 271.25]	221.54 [196.00, 273.00]	217.20 [211.30, 233.64]	0.966
<100	2(2.22%)	2(2.5%)	0	— —
≥100	88(97.8%)	77(97.4%)	11(100%)	— —
Prothrombin time (s)	12.50 [12.30, 12.99]	12.50 [12.30, 13.11]	12.50 [12.31, 12.86]	0.777
Creatine phosphate kinase	43.86 [37.59, 59.90]	44.74 [37.98, 61.00]	40.91 [36.51, 51.56]	0.542
<50	54(60%)	46(58.2%)	8(72.7%)	— —
≥50 to <350	36(40%)	33(41.8%)	3(27.3%)	— —
Creatine phosphate kinase isoenzyme	8.58 [7.39, 11.63]	8.90 [7.35, 12.05]	8.35 [7.45, 8.54]	0.254
Procalcitonin (ng/ml)	0.11 [0.08, 0.15]	0.11 [0.08, 0.14]	0.12 [0.10, 0.18]	0.158
<0.1	34(37.8%)	32(40.5%)	2(18.2%)	— —
≥0.1 to <0.25	52(57.8%)	43(54.4%)	9(81.8%)	— —
≥0.25 to <0.5	4(4.44%)	4(5.1%)	0	— —
D - dimer (μg/L)	0.47 [0.28, 0.67]	0.44 [0.24, 0.65]	0.75 [0.57, 1.04]	0.002
NLR	2.47 [1.84, 3.38]	2.44 [1.82, 3.38]	2.82 [2.34, 3.58]	0.213
Treatment options				
Antiviral infection	86(95.6%)	75(94.9%)	11(100%)	— —
Anti-infective treatment	51(56.7%)	40(50.6%)	11(100%)	— —
Chinese medicine treatment	31(34.4%)	28(35.4%)	3(27.3%)	— —
oxygen therapy	38(42.2%)	28(35.4%)	10(90.9%)	— —

The data were median.

represented a penalty value, and the variable that intersects the color line was the selected variable.

According to the results of Lasso, the variables including age, CK, procalcitonin, D-dimer, LYMBH1, LYMBH2, underlying disease, fever, cough, expectoration, fatigue, pharynx dryness were selected as important variables. These variables were classified into the logistics model, and variables with insignificant P values were eliminated according to significance. The results were obtained (**Table 4**). Age, CK, procalcitonin, LYMBH1, cough, fatigue, and pharynx dryness were important variables.

It can be seen from the results that, except for the constant term and LYMBH1, the coefficients of the other variables were all positive. Among them, age, CK, LYMBH1 and procalcitonin were continuous variables. Cough, fatigue and pharynx dryness were categorical variables. The results were tested for collinearity. VIF values were all less than 5. And there was no collinearity among the variables.

The logistic regression model was obtained according to the balanced data, and prediction accuracy was 81.1% when the original data was substituted into the model. The ordinate of ROC curve (**Figure 3**) was the true positive rate (sensitivity), and the abscissa was the false positive rate (1-specificity). The ideal point on the ROC curve was (0,1), which means all positive classes were correctly classified and all negative classes were not mistakenly classified as positive classes. Therefore, the closer the ROC point was to the upper left corner, the better the performance. In this article, the abscissa of the ROC graph is specific, and the ideal point is (1,1). The closer the AUC value of the area under the ROC curve to 1, the better the diagnostic effect. The coordinates of the best critical point on the curve in **Figure 3** were (0.823, 0.727), and the threshold was 0.5. It has high sensitivity and specificity at this point. The AUC value in the figure was 0.775. It showed that the accuracy of the model is good

Random Forest

Random forest was used to extract a certain number of self-service samples from the original data using the bootstrap method with replacement and a decision tree was built for each sample. At each node, among all competing independent variables, several random competitive split variable were selected. Each tree in the random forest was left unpruned and

TABLE 2-2 | Laboratory test results and treatment plan of COVID-19 patients during hospitalization after SMOTE algorithm balancing data.

Laboratory index	All patients (n = 231)	Mild(n = 110)	Severe (n = 121)	*P* value
Heart rate (Times/min)	88.74 [82.00, 98.00]	90.00[82.00,103.75]	86.91[81.98,94.42]	0.075
white blood cell count 1 ($\times 10^9$/L)	5.11 [4.27, 6.01]	5.03 [4.23, 6.01]	5.21 [4.44, 5.99]	0.799
<4	41(17.7%)	21(19.1%)	20(16.5%)	
4-10	185(80.1%)	87(79.1%)	98(81.0%)	
>10	5(2.2%)	2(1.8%)	3(2.5%)	
Lymphocyte count 1 ($\times 10^9$/L)	1.28 [1.05, 1.74]	1.45 [1.05, 2.03]	1.23 [1.05, 1.60]	0.038
<1.0	48(20.8%)	21(19.1%)	27(22.3%)	
≥1.0	183(79.2%)	89(80.9%)	94(77.7%)	
Platelet count 1 ($\times 10^9$/L)	179.00[168.80,216.93]	174.50[170.00,237.00]	184.86 [168.64, 208.51]	0.689
<100	0	0	0	
≥100	231(100%)	110(100%)	121(100%)	
white blood cell count 2 ($\times 10^9$/L)	5.35 [4.78, 6.76]	5.39 [4.82, 6.76]	5.26 [4.74, 6.74]	0.315
<4	15(6.5%)	2(1.8%)	13(10.7%)	
4-10	215(93.1%)	107(97.3%)	108(89.3%)	
>10	1(0.4%)	1(0.9%)	0	
Lymphocyte count 2 ($\times 10^9$/L)	1.53 [1.36, 1.89]	1.54 [1.37, 1.94]	1.50 [1.35, 1.89]	0.101
<1.0	3(1.3%)	1(0.9%)	2(1.7%)	
≥1.0	228(98.7%)	109(99.1%)	119(98.3%)	
Platelet count 2 ($\times 10^9$/L)	229.20 [212.57, 283.81]	224.00 [198.75, 273.50]	233.61 [215.42, 288.51]	0.016
<100	3(1.3%)	3(2.7%)	0	
≥100	228(98.7%)	107(97.3%)	121(100%)	
Prothrombin time (s)	12.51 [12.31, 12.95]	12.48 [12.31, 12.89]	12.57 [12.31, 12.95]	0.703
Creatine phosphate kinase	44.71 [38.12, 68.08]	43.00 [37.51, 62.97]	45.19 [38.88, 76.26]	0.035
<50	131(56.7%)	65(59.1%)	66(54.5%)	
≥50 to <350	100(43.3%)	45(40.9%)	55(45.5%)	
Creatine phosphate kinase isoenzyme	8.47 [7.81, 10.07]	8.58 [7.43, 10.97]	8.44 [8.03, 9.38]	0.735
Procalcitonin (ng/ml)	0.12 [0.09, 0.17]	0.11 [0.08, 0.14]	0.14 [0.10, 0.18]	<0.001
<0.1	63(27.3%)	39(35.5%)	24(19.8%)	
≥0.1 to<0.25	165(71.4%)	68(61.8%)	97(80.2%)	
≥0.25 to<0.5	3(1.3%)	3(2.7%)	0	
D - dimer (μg/L)	0.62 [0.46, 0.95]	0.47 [0.29, 0.62]	0.90 [0.61, 1.10]	<0.001
NLR	2.55 [2.07, 3.39]	2.46 [1.84, 3.46]	2.62 [2.12, 3.36]	0.048
Treatment programs				
Antiviral infection	225 (97.4)	104 (94.5)	121 (100.0)	——
Anti-infective treatment	172 (74.5)	51 (46.4)	121 (100.0)	——
Chinese medicine treatment	66 (28.6)	37 (33.6)	29 (24.0)	——
oxygen therapy	135 (58.4)	33 (30.0)	102 (84.3)	——

The data were median.

allowed to grow fully. The final prediction result was a simple average of the results of all decision trees (Moghadas et al., 2020). In this paper, the random forest method was used to classify variables, give the importance of variables, and make prediction.

The data was extracted by 7:3 for the training set and the test set. For the node binary tree parameter mtry, the loop statement was used to calculate the mean value of the misjudgment rate of all models, and the optimal mtry was 6. It can be seen from **Figure 4** that when the number of decision trees contained in the random forest ntree was about 1300, the three lines tended to be stable, that is, the error within the model was basically stable, so ntree=1300. At this time, the OOB error is 5.45%. It can be seen

TABLE 3 | Significance of single-factor logistic regression variables.

Variable name	Gender (Female)	Age	Underlying disease	Heart rate	Fever	Cough
P value	0.193	5.54e-14***	3.11e-06***	0.0635	0.00182**	5.27e-12***
Variable name	Expectoration	Fatigue	Dyspnea	Diarrhea	Poor appetite	Emesis
P value	2.83e-11***	0.0053**	0.987	0.180	0.367	0.211
Variable name	headache	Muscle ache	FG_infection	Pharynx dry	onset_time	WBC2
P value	0.0513	0.0888	0.505	3.59e-08***	0.00657**	0.230
Variable name	LYMBH2	PLT2	Prothrombin time	CK	CK-MB	Procalcitonin
P value	0.0355*	0.0737	0.0784	0.00156**	0.824	2.94e-06***
Variable name	D-dimer	Actual_days	WBC1	LYMBH1	PLT1	NLR
P value	5.96e-07***	0.832	0.488	0.01005*	0.1224	0.804

**indicates that the P value of the variable is less than 0.05.*

***indicates that the P value is less than 0.01.*

****indicates that the P value is less than 0.001.*

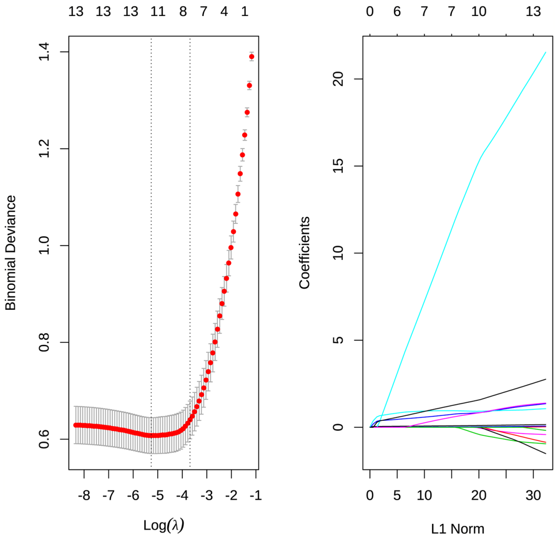


FIGURE 2 | **(A)** Select the best λ value for the AUC value when compressing different variables. **(B)** The degree of compression of each variable under different penalty parameters λ.

from **Table 5** that 74 patients with mild illness and 82 patients with severe illness were judged correctly, but still with a certain probability of misjudgment, and 6 patients with mild illness were judged as severe. In general, the correct rate of the model, that was, the proportion of samples that were correctly identified by all samples was (74 + 82)/165 = 94.54%; the accuracy mild rate indicates how many of the guessed positive samples are correct, and the mild disease was regarded as the positive sample. The precision rate was 74/(74 + 3)=96.1%; the recall rate indicates how many positive samples were recalled among all the positive samples, so the recall rate was 74/(74 + 6)=92.5%.

Table 6 showed the importance of variables, in which Mean Decrease Accuracy represents an average Decrease in Accuracy, and Mean Decrease Gini represents an average Decrease in node unpurity. A larger value indicates a stronger importance of variables. It can be seen from the table that age, D-dimer, time from onset to admission and other large values were important variables.

After sorting the Mean Decrease Accuracy and Mean Decrease Gini of the variables, **Figure 5** showed the visualization results of the variable importance ranking. According to the results of MeanAccuracy, we can see that the top 10 important variables were age, D dimer, onset time, Actual days, NLR, WBC1, procalcitonin, expectoration, CK, heart rate. According to the results of MeanDecreaseGini, we can see that the top 10 important variables were age, D dimer, onset time, expectoration, procalcitonin, cough, NLR, heart rate, PLT2, WBC1. Important variables in these two selection criteria include 8 variables: age, D dimer, onset time, procalcitonin, NLR, WBC1, expectoration, heart rate.

The prediction function was further used to verify the test set, and the model accuracy was 96.97%, the Sensitivity rate was 93.33%, and the recall rate was 93.3%. **Figure 6** was a random forest ROC curve and AUC evaluation graph. The optimal critical point threshold of the ROC curve was 1.5, and the coordinates are (0.933, 1), very close to the upper left. The AUC score under the ROC curve was 0.967 and the overall area of the curve was close to 1, indicating a strong recognition ability. All these indicate that the random forest method has a good effect.

TABLE 4 | Logistics model results.

	Estimate	Std. Error	P value
(Intercept)	-13.42683	2.34558	1.04e-08***
Age	0.14735	0.02545	7.08e-09***
CK	0.06749	0.01521	9.17e-06***
Procalcitonin	24.92632	6.64398	0.000176***
LYMBH1	-1.62463	0.76513	0.033727*
Cough	1.84101	0.54942	0.000806***
Fatigue	1.64915	0.73679	0.025201*
Pharynx dryness	2.84222	0.76075	0.000187***

**indicates that the P value of the variable is less than 0.05.*
****indicates that the P value is less than 0.001.*

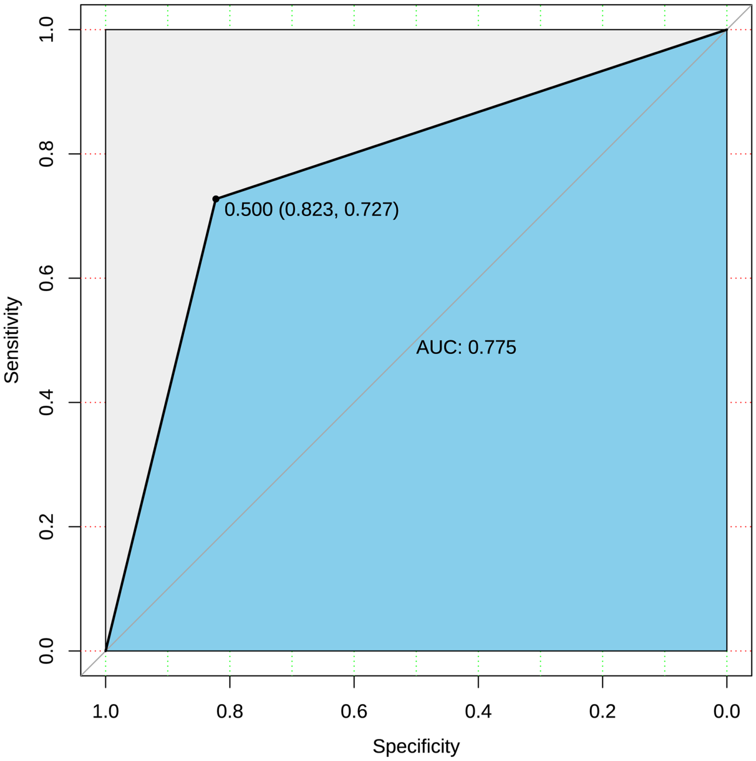


FIGURE 3 | ROC curve of prediction based on raw data.

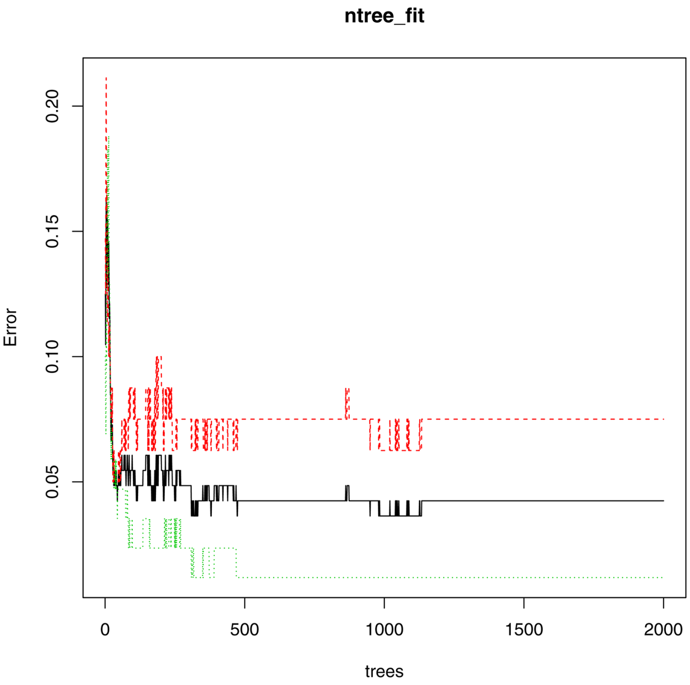


FIGURE 4 | The false positive rate of each tree in the random forest.

TABLE 5 | Confusion matrix.

	Mild	Severe	class.error
Mild	74	6	0.07500000
Severe	3	82	0.03529412

DISCUSSION

Looking back on the past 20 years, the world has experienced two severe beta coronavirus pandemics, SARS-CoV in 2002-2003 and MERS-CoV in 2012. They had a fatality rate of 10% and 35% respectively (Drosten et al., 2003; Chen B. et al., 2020), which dealt a huge blow to public health. In early 2020, SARS-CoV-2 causing COVID-19 spread rapidly worldwide. At the same time, we need to be wary of the superposition of influenza and COVID-19 in the winter.

In this study, the epidemiological history, clinical characteristics and laboratory test indicators of 90 SARS-COV-2 nucleic acid positive patients with severe and mild illness were analyzed. The clinical manifestations are non-specific, mainly fever and cough. A small number of patients appeared breathing difficulties. Gastrointestinal symptoms are not common, which has also been confirmed by similar studies (Chen N. et al., 2020; Li et al., 2020). The incubation period of SARS-CoV-2 infection is longer than that of SARS and MERS. The viral load of asymptomatic infected patients is similar to that of symptomatic infected patients, which may lead to the underestimation the potential transmissibility of SARS-CoV-2 among people (Zou et al., 2020). More studies have shown that the invisible transmission of the COVID-19 is caused by the incubation period and asymptomatic infection (Moghadas et al., 2020; Wang et al., 2020). This is also one of the reasons why it is difficult to effectively control asymptomatic infection in the COVID-19 epidemic. Our data suggested that clinical monitoring should not just focus on fever symptoms. Otherwise, it will increase the missed diagnosis rate. It is necessary to strictly implement nucleic acid testing and be assisted by comprehensive judgment on clinical chest radiographs (Ai et al., 2020). Tracing and early isolation of asymptomatic infected persons are two of the key points for pandemic control (Grasselli et al., 2020). In addition, there is a serious family cluster infection in this study, which also suggests that we should be on guard against human-to-human transmission of the virus in families and hospitals, even inter-city and overseas transmission. It also reflects the importance of timely isolation of the source of infection and isolation at home in the early stage of the epidemic.

In this study, logistic regression presents seven important factors that influence the risk of severe COVID-19, which may be of great value in screening patients with mild or severe disease. Clinicians can quickly assess what kind of risk a patient is at. The seven variables are age, creatine kinase, procalcitonin, lymphocyte count 1, cough, fatigue and pharynx dryness.

TABLE 6 | Importance of variables.

Sequence	Variable name	Mild	Severe	Mean Decrease Accuracy	Mean Decrease Gini
1	Gender	0.000842276	0.000304033	0.000571913	0.148250306
2	Age	0.114189754	0.147716116	0.130207653	18.01285314
3	HU_disease	0.002974213	0.003116931	0.003079772	0.367923962
4	Heart rate	0.012114962	0.019990241	0.016000088	2.846081568
5	Fever	0.003780202	0.000847247	0.002318694	0.389517379
6	Cough	0.017572645	0.015699049	0.016538291	3.237108144
7	Expectoration	0.03154458	0.015864688	0.023332413	3.984819339
8	Fatigue	0.001708698	0.001729405	0.001704126	0.338710957
9	Dyspnea	0.000337112	0.000361101	0.000348378	0.198804768
10	Diarrhea	0.00045062	0.000512224	0.000479591	0.133668522
11	Poor appetite	0.0017915	0.001509113	0.001652048	0.273898309
12	Emesis	0.000103448	-2.91E-05	4.04E-05	0.02585151
13	headache	0.000741436	0.000554534	0.000661425	0.280146418
14	Muscle ache	0.000738661	0.00234859	0.001551771	0.328014275
15	FG_infection	0.00134232	0.000961904	0.001115436	0.196406068
16	pharynx dryness	0.013076183	0.012061961	0.012526584	2.197295907
17	Onset_time	0.027652778	0.05316072	0.040377103	6.231243065
18	WBC2	0.006230574	0.010898642	0.008580465	2.048265
19	LYMBH2	0.009000937	0.016263749	0.012845822	2.371037233
20	PLT2	0.013663865	0.017568322	0.015517789	2.663186165
21	Prothrombin_time	0.009546072	0.019385667	0.014444887	2.49562609
22	cK	0.012179427	0.02243054	0.017384287	2.454430363
23	cK-MB	0.008558614	0.019866388	0.014431234	2.067444666
24	Procalcitonin	0.016416281	0.033016345	0.024915001	3.630346879
25	D-dimer	0.063398879	0.089809009	0.076390526	12.2980474
26	Actual days	0.011318214	0.020510316	0.015964144	2.433158984
27	WBC1	0.013069936	0.019856579	0.016628493	2.610376896
28	LYMBH1	0.008186132	0.015641668	0.011974092	1.797733889
29	PLT1	0.012741624	0.021607278	0.017247375	2.606923867
30	NLR	0.012406788	0.025559522	0.019054742	3.198992108

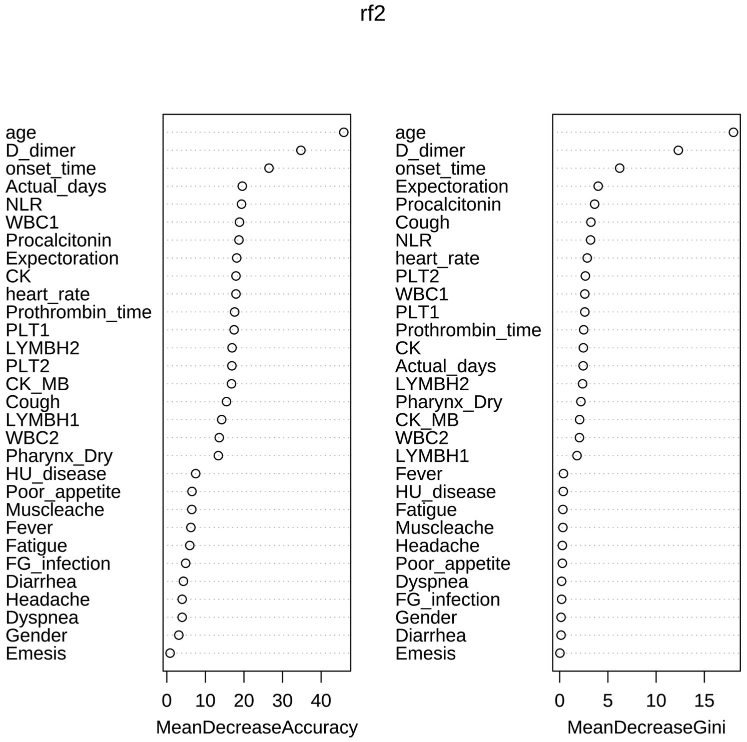


FIGURE 5 | Ranking chart of importance of variables.

In the COVID-19 epidemic, the high-risk groups affected are mainly the elderly and people with underlying diseases. Taking into account the decline in the rehabilitation ability of the elderly, the older patients are more likely to develop into severe patients (Hendren et al., 2020). The myocardial enzymes of patients with severe COVID-19 increased significantly, and the abnormal rate of myocardial enzymes of severe COVID-19 was higher than other groups (Babapoor-Farrokhran et al., 2020). Creatine phosphokinase is a kind of myocardial enzymes. Myocardial enzymes refer to a class of enzymes that catalyze the metabolism of myocardial cells and regulate the electrophysiological activities of myocardial cells. Once the cardiomyocytes are ruptured and necrotic, these enzymes are released into the blood and their values increase. Therefore, the changes in the myocardial enzyme indicators can measure the degree of damage to the cardiomyocytes. Especially in the early stage of myocardial infarction, when the myocardium has not yet undergone extensive necrosis, creatine kinase can be detected in human blood. Therefore, if this index is elevated, the degree of myocardial damage can be diagnosed in time (Babapoor-Farrokhran et al., 2020). Procalcitonin (PCT) is a biological marker that can be used to assess the possibility of bacterial infection, reflecting the severity of bacterial infection (Dymicka-Piekarska and Wasiluk, 2015). Different diseases have different concentrations of procalcitonin. The higher the concentration of procalcitonin, the higher the possibility of organ failure and the higher the risk of death. Severe disease in patients with COVID-19 is also closely related to lymphocytes. Studies have shown that lymphocytes in patients with COVID-19 show a significant decreasing trend, among which CD4+ and CD8+T lymphocytes are the most obvious. Therefore, CD4+ and CD8+T lymphocytes can be considered as early warning signals and prognostic indicators to judge the severity of COVID-19 patients (Huang et al., 2020b; Zhang et al., 2020). Three clinical symptoms, including cough, fatigue and pharyngeal dryness can be used as important clinical symptoms to judge the degree of disease development. Combined with the patient's laboratory indicators, they have a certain reference.

The random forest classification method solves the problem of variable collinearity, and has a better prediction effect. At the same time, it also provides important predictors in the method, such as the age, D-dimer, time from onset to admission, procalcitonin, NLR, heart rate, white blood cell count. Among these variables, in addition to the age and procalcitonin variables explained above, D-dimer and NLR also have an important position. In this epidemic, some severe patients died because of an inflammatory storm, which is an overreaction of the human immune system. D-dimer is one of the markers for detecting thrombus, and the rise of D-dimer can also be seen when the human body undergoes inflammation (Shah et al., 2020). NLR is the ratio of neutrophils to lymphocytes. Neutrophils are the main component of the white blood cell population. After microbial pathogens invade the body, they will quickly reach the inflammation site and exert phagocytosis. In addition, the

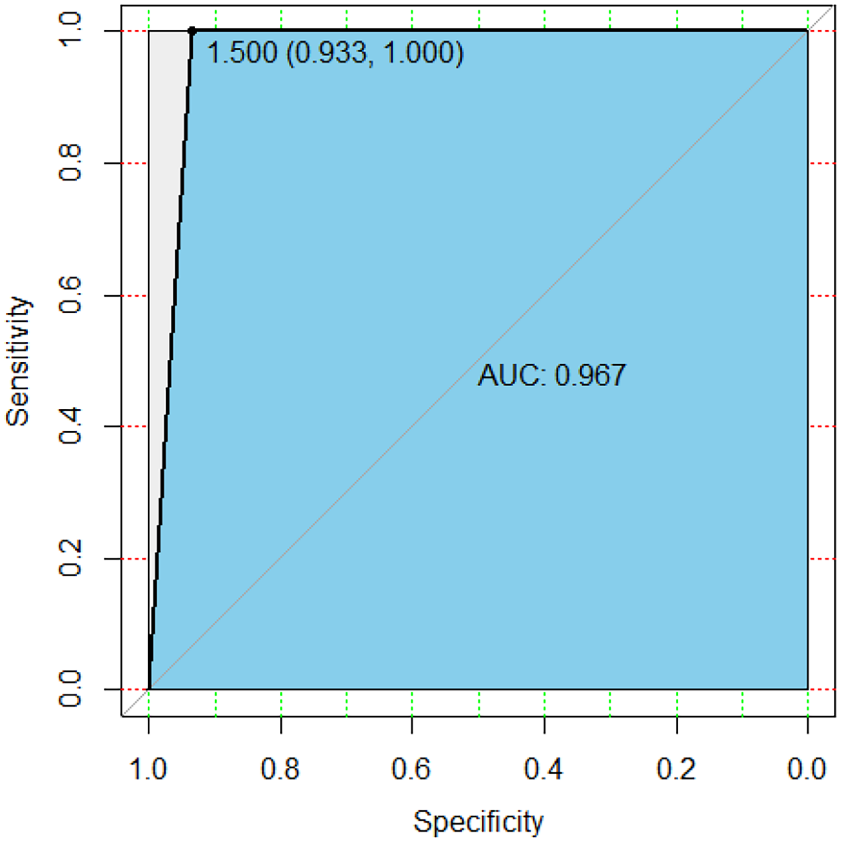


FIGURE 6 | ROC curve and AUC area of random forest.

human immune response triggered by viral infections mainly relies on lymphocytes. In the previous logistic regression, it was also analyzed that the lymphocytes decreased, the higher the possibility of the patient being severely ill (Yang et al., 2020a; Kerboua, 2021). Therefore, elevated D-dimer and NLR indicate that the patient's condition is getting worse. In this article, the D-dimer and NLR of severe patients are higher than those of mild patients. Random forests can still maintain accuracy for data with unbalanced classification and have a strong anti-overfitting ability. Therefore, the importance of variables given by random forests is further added to the variable information in addition to the logistics model. In the future, it can be considered as a direction in which to continue to look for key variables to distinguish the mild and severe diseases.

In this study, the clinical characteristics and modeling significant variables of patients with mild and severe diseases are of great value, and can help to identify whether or not the patients are at risk of critical illness in the early stage. However, there are some limitations in this study. Firstly, the clinical symptoms or signs and laboratory test results of patients extracted from electronic medical records lack some data, and some cases are incomplete. Secondly, because the research subjects are all hospitalized patients, the actual research lacks those with asymptomatic infections and mild patients who are not treated in hospital, causing the research results may be more inclined to serious outcomes. Thirdly, the sample size is limited, so the results may not be robust enough. The SMOTE algorithm performs class balancing and has certain errors. For example, it may repeatedly use outliers or wrong values in the data, or may incorrectly strengthen the local chance, thereby increasing the risk of overfitting. In addition, during the data collection period of this article, the Ninth People's Hospital of Dongguan City had no moderate cases, so patients were divided into two types.

To sum up, the COVID-19 outbreak in 2020 is a huge public health crisis to the whole world. In the context of the pandemic, early detection, early isolation and early intervention have always been the basic strategies for epidemic prevention and control. The data shows that we need to closely monitor patients with mild COVID-19, reduce the transformation from mild to severe, and rationally allocate medical resources in a scientific and efficient way to improve the cure rate of COVID-19.

DATA AVAILABILITY STATEMENT

The datasets and R language code presented in this study can be found at GitHub (https://github.com/johnsnowgo/covid-90data).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of the Ninth People's Hospital of Dongguan, China.

AUTHOR CONTRIBUTIONS

XC: Conceptualization, Methodology, Writing-review & editing, Project administration. LZ: Software, Formal analysis, Writing - original draft, Writing - review & editing. SY and HZ: Resources, Conceptualization, Methodology, Writing - review & editing. YJ: Conceptualization, Methodology, Writing-review & editing, Project administration. Y-QC: Conceptualization, Methodology, Writing-review & editing, Project administration. MX: Writing-original draft, Formal analysis. KL: Writing-original draft, Formal analysis. YL: Writing-original draft, Formal analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (11601083), the program for Probability and Statistics: Theory and Application (IRTL1704) and Innovative Research Team in Science and Technology in Fujian Province University (IRTSTFJ). This study was also supported by National Natural Science Foundation of China (31970881) and Shenzhen Science and Technology Program under Grant (SGG20200225152008136) to Y-QC. The special topic of prevention and control technology research and promotion of emergency tackling on Covid-19 in Dongguan city (2020717150119128).

REFERENCES

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., et al. (2020). Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (Covid-19) in China: A Report of 1014 Cases. *Radiology* 296, E32–E40. doi: 10.1148/radiol.2020200642

Alahmari, F. (2020). A Comparison of Resampling Techniques for Medical Data Using Machine Learning. *J. Inf Knowl. Manag.* 19, 1–13. doi: 10.1142/S021964922040016X

Babapoor-Farrokhran, S., Gill, D., Walker, J., Rasekhi, R. T., Bozorgnia, B., and Amanullah, A. (2020). Myocardial Injury and COVID-19: Possible Mechanisms. *Life Sci.* 253, 117723. doi: 10.1016/j.lfs.2020.117723

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953

Chen, B., Tian, E. K., He, B., Tian, L., Han, R., Wang, S., et al. (2020). Overview of Lethal Human Coronaviruses. *Signal Transduct. Target Ther.* 5, 89. doi: 10.1038/s41392-020-0190-2

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study. *Lancet* 395, 507–513. doi: 10.1016/S0140-6736(20)30211-7

Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., et al. (2003). Identification of a Novel Coronavirus in Patients With Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* 348, 1967–1976. doi: 10.1056/NEJMoa030747

Dymicka-Piekarska, V., and Wasiluk, A. (2015). Procalcitonin (PCT), Contemporary Indicator of Infection and Inflammation. *Postepy Hig. Med. Dosw.* 69, 723–728. doi: 10.5604/17322693.1158796

Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A., et al. (2020). Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-Cov-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA* 323, 1574–1581. doi: 10.1001/jama.2020.5394

Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., et al. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. doi: 10.1056/NEJMoa2002032

Hendren, N. S., Drazner, M. H., Bozkurt, B., and Cooper, L. T.Jr. (2020). Description and Proposed Management of the Acute Covid-19 Cardiovascular Syndrome. *Circulation* 141, 1903–1914. doi: 10.1161/CIRCULATIONAHA.120.047349

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical Features of Patients Infected With 2019 Novel Coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi: 10.1016/S0140-6736(20)30183-5

Huang, W., Berube, J., McNamara, M., Saksena, S., Hartman, M., Arshad, T., et al. (2020). Lymphocyte Subset Counts in COVID-19 Patients: A Meta-Analysis. *Cytometry A* 97, 772–776. doi: 10.1002/cyto.a.24172

Kerboua, K. E. (2021). Nlr: A Cost-Effective Nomogram to Guide Therapeutic Interventions in COVID-19. *Immunol. Invest.* 50, 92–100. doi: 10.1080/08820139.2020.1773850

Li, Q., Guan, X. H., Wu, P., Wang, X. Y., Zhou, L., Tong, Y. Q., et al. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N. Engl. J. Med.* 382, 1199–1207. doi: 10.1056/NEJMoa2001316

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8

Moghadas, S. M., Fitzpatrick, M. C., Sah, P., Pandey, A., Shoukat, A., Singer, B. H., et al. (2020). The Implications of Silent Transmission for the Control of COVID-19 Outbreaks. *Proc. Natl. Acad. Sci. U. S. A.* 117, 17513–17515. doi: 10.1073/pnas.2008373117

Shah, S., Shah, K., Patel, S. B., Patel, F. S., Osman, M., Velagapudi, P., et al. (2020). Elevated D-Dimer Levels Are Associated With Increased Risk of Mortality in Coronavirus Disease 2019: A Systematic Review and Meta-Analysis. *Cardiol. Rev.* 28, 295–302. doi: 10.1097/CRD.0000000000000330

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., et al. (2020). Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* 323, 1061–1069. doi: 10.1001/jama.2020.1585

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A New Coronavirus Associated With Human Respiratory Disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3

Yang, A. P., Liu, J. P., Tao, W. Q., and Li, H. M. (2020a). The Diagnostic and Predictive Role of NLR, d-NLR and PLR in COVID-19 Patients. *Int. Immunopharmacol.* 84, 106504. doi: 10.1016/j.intimp.2020.106504

Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., et al. (2020b). Study of Cardiovascular Disease Prediction Model Based on Random Forest in Eastern China. *Sci. Rep.* 10, 5245. doi: 10.1038/s41598-020-62133-5

Zhang, X., Tan, Y., Ling, Y., Lu, G., Liu, F., Yi, Z., et al. (2020). Viral and Host Factors Related to the Clinical Outcome of COVID-19. *Nature* 583, 437–440. doi: 10.1038/s41586-020-2355-0

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A Pneumonia Outbreak Associated With a New Coronavirus of Probable Bat Origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A Novel Coronavirus From Patients With Pneumonia in Chin. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017

Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., et al. (2020). SARS-Cov-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N. Engl. J. Med.* 382, 1177–1179. doi: 10.1056/NEJMc2001737

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.

REVIEW
published: 31 August 2021
doi: 10.3389/fmicb.2021.723850

The Burden of Respiratory Viruses and Their Prevalence in Different Geographical Regions of India: 1970–2020

*Rushabh Waghmode, Sushama Jadhav and Vijay Nema**

Division of Molecular Biology, ICMR-National AIDS Research Institute, Pune, India

As per the 2019 report of the National Health Portal of India, 41,996,260 cases and 3,740 deaths from respiratory infections were recorded across India in 2018. India contributes to 18% of the global population, with severe acute respiratory infection (SARI) as one of the prominent causes of mortality in children >5 years of age. Measures in terms of the diagnosis and surveillance of respiratory infections are taken up globally to discover their circulating types, detect outbreaks, and estimate the disease burden. Similarly, the purpose of this review was to determine the prevalence of respiratory infections in various regions of India through published reports. Understanding the pattern and prevalence of various viral entities responsible for infections and outbreaks can help in designing better strategies to combat the problem. The associated pathogens comprise respiratory syncytial virus (RSV), rhinovirus, influenza virus, parainfluenza virus, adenovirus, etc. Identification of these respiratory viruses was not given high priority until now, but the pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has sensitized our system to be alert about the burden of existing infections and to have proper checks for emerging ones. Most of the studies reported to date have worked on the influenza virus as a priority. However, the data describing the prevalence of other respiratory viruses with their seasonal pattern have significant epidemiological value. A comprehensive literature search was done to gather data from all geographical regions of India comprising all states of India from 1970 to 2020. The same has been compared with the global scenario and is being presented here.

Keywords: respiratory viruses, India, prevalence, outbreaks, multiplex detection

Edited by:
Hassan Zaraket,
American University of Beirut,
Lebanon

Reviewed by:
Martha Ann Delaney,
University of Illinois
at Urbana-Champaign, United States
Federico Manuel Giorgi,
University of Bologna, Italy

***Correspondence:**
Vijay Nema
vnema@nariindia.org;
dr.vijaynema@gmail.com

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

Received: *11 June 2021*
Accepted: *31 July 2021*
Published: *31 August 2021*

Citation:
Waghmode R, Jadhav S and Nema V (2021) The Burden of Respiratory Viruses and Their Prevalence in Different Geographical Regions of India: 1970–2020.
Front. Microbiol. 12:723850.
doi: 10.3389/fmicb.2021.723850

INTRODUCTION

The National Health Portal of India reported in 2019 that there were 41,996,260 cases and 3,740 deaths from respiratory infections in India in 2018. Acute respiratory infections (ARI) accounted for 69% of the total cases of communicable diseases, and this scenario is before the era of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). After the coronavirus disease 2019

Abbreviations: SARI, severe acute respiratory infection; RSV, respiratory syncytial virus; HRSV, human respiratory syncytial virus; HPIV, human parainfluenza virus; HRV, human rhinovirus; INF, influenza; HMPV, human metapneumovirus; HAdV, human adenovirus; DALY, disability-adjusted life year; WHO, World Health Origination; ALRI, acute lower respiratory infection; VRDLs, virus research and diagnostic laboratories.

(COVID-19) pandemic, the total number of infected numbers rose to millions and cases are still getting added while this review was being written. The spread of the virus was so fast and wide that the World Health Organization (WHO) had to declare this infection as a global pandemic on March 11, 2020. The United States, India, Brazil, France, Russia, Spain, Argentina, the United Kingdom, Colombia, and Mexico were the countries that were impacted severely. The United States has been reporting the largest number of cases with 35,283,729 and over 626,668 deaths, followed by India with 31,341,507 and over 420,196 deaths as per data until July 24, 2021 (Worldometer, 2021).

India hosts a population of 1.3 billion people distributed in states having diverse geographical and cultural makeup. In India, the eight states with the highest total cases of SARS-CoV-2 infections are in Maharashtra, with over 3.84 million confirmed cases, followed by Karnataka, Andhra Pradesh, Tamil Nadu, Kerala, Delhi, Uttar Pradesh, and West Bengal. As expected, with the huge population of the country, these states also have a huge population. Variations in the environmental conditions, food habits, and social practices all contribute to the diversity in the distribution of respiratory infections in different ways and with different intensities (Salvi et al., 2018). The global figures indicate that acute lower respiratory infection (ALRI) is the single largest infectious cause of death among children. In India, pneumonia itself causes 17% of all deaths in children >5 years. This leads to the initiation of multiple health measures and guidelines for infants and younger children. However, being such a vast and diverse country, there remain so many gaps in the treatment, vaccination, empirical use of available antimicrobials, etc. (Krishnan et al., 2019).

Leaving aside the influence of COVID-19, the world has always been struggling with the prevalence and the diversity of respiratory infections. These are highly dependent on seasonality and the change in climatic conditions. Hence, they visit again and again in the form of epidemics in sporadic regions in the world, sometimes crossing boundaries and becoming pandemics as well. The world faces more than 2–5 million cases every year, with deaths ranging from 290,000 to 650,000 (World Health Organization, 2021). The exact cause of such a large number has not been fully understood, but 99% of cases are reported in children below 5 years and are occurring in developing countries. Hence, it has something to do with nourishment and the availability of medical facilities along with sophistication.

Diagnostic Scenario and Prevalence of Respiratory Viruses

Diagnosing a particular respiratory viral infection based on the symptoms is difficult as all respiratory viral infections have overlapping symptoms. Testing for the presence of viruses in human samples using specific nucleic acid sequence detection is the only way to confirm diagnosis. Polymerase chain reaction (PCR) and its variants are the mainstream diagnostic modalities available, and the same is done with a single set of primers or multiplex formats after the design of specific probes and primers and optimization in the laboratory. Immunofluorescence assays are also in use, but suffer issues like sensitivity. Confirmation by culture has other issues, such as the availability of sophisticated cell culture laboratories and the containment levels available at different resource settings (Malhotra et al., 2016). Surveillance for influenza has been taken up by many countries wherein the use of multiplex real-time PCR assays has helped in detecting circulating viruses and outbreaks and estimating the disease burden. Following the development of newer methods and their optimization, more and more viruses are being diagnosed routinely or as and when suspected in many countries now. The following viruses are being detected for respiratory infections using multiplex real-time PCR methods.

Human Respiratory Syncytial Virus

Human respiratory syncytial virus (HRSV)/RSV is considered to be the most common viral cause of ALRI-related death. It is estimated to have 33.1 million cases globally, with about 10% hospitalizations and death of 59,600 children below 5 years of age every year. This amounts to about 22% of all episodes of ALRI and is the cause of the highest childhood mortality in low- and middle-income countries among all respiratory viral infections (Broor et al., 2018). RSV infections follow a distinctive seasonal pattern, but differ according to the geographical location. The tropical or warm climate of the Southeast Asian region allows it to stay longer, with a few areas harboring the virus and its outbreaks all through the year. India is also witnessing an increase in RSV infections in infants and children below 5 years of age. Hospital-based reports indicate that RSV is detected in around 16% of children presenting acute lower respiratory tract infections (Kini et al., 2019).

Human Parainfluenza Virus

Human parainfluenza virus (HPIV) is a virus from the family Paramyxoviriae with two genera: *Respirovirus* (HPIV-1 and HPIV-3) and *Rubulavirus* (HPIV-2 and HPIV-4). Not much could be found in the literature about the global burden estimates for the parainfluenza virus. Literature from Asia reported a variable prevalence of HPIV ranging from 1 to 66%, with 6.95% as the Indian average prevalence (Rafeek et al., 2020).

Human Metapneumovirus

Respiratory virus infections are highly prevalent in Southeast Asian countries, but the discovery of metapneumovirus occurred in 2001 in patients from the Netherlands (van den Hoogen et al., 2001). This virus was detected from all continents and in all age groups. Reports from Hong Kong, Japan, Korea, Thailand, etc., also described detections of human metapneumovirus (hMPV) after 2001. Rao et al. (2004) have presented a preliminary report on the first detection of hMPV in India from a pediatric patient in July to August 2003 by reverse transcription PCR.

Influenza Virus

Infection with the influenza virus is prevalent worldwide and is known to cause around 39 million episodes annually. India, with a tropical climate and variable hygiene practices, experiences a significant number of cases every year. Two strains of this virus are mainly detected: influenza A (INF-A) and influenza

B (INF-B). Multiple strains of influenza have emerged in due course, and subtypes H3N2 and H1N1, belonging to INF-A, are the significant ones. Although influenza virus infections are reported throughout the year, a higher number of cases are reported during the rainy season in India. Hence, the higher relative humidity has a role to play in its incubation and spread (Narayan et al., 2020). Influenza B, which is diverse from INF-A virus, comprises two antigenically distinct lineages ("B/Victoria/2/87-like" and "B/Yamagata/16/88-like," termed Victoria and Yamagata, respectively). This virus has also caused a few major pandemics in known history and is supposed to have appeared before the period of reporting.

Human Rhinovirus

Rhinovirus, although belonging to the Enterovirus family, causes respiratory infections along with gastroenteritis and has been a causal factor for community outbreaks and pediatric respiratory infections in many countries across the globe. Molecular analysis of the collected strains has revealed their presence for the last 250 years all around the world (Briese et al., 2008).

Human Adenovirus

Adenoviruses are not often reported to cause severe illnesses in normal or healthy individuals; however, they cause a wide range of illnesses in children and immunocompromised individuals. They start from respiratory symptoms such as the common cold, sore throat, bronchitis, and pneumonia to gastrointestinal disorders such as diarrhea, vomiting, nausea, and stomach pain. They are also known to cause conjunctivitis in children. Not many epidemiological reports are available from India. A report describing pediatric gastroenteritis caused by adenovirus was found in Kolkata, India. This study described the testing of 1,562 stool specimens in 2013–2014 and reported an 11.8% prevalence of enteric HAdV (Banerjee et al., 2017).

There are limited reports available regarding the global and regional burdens of all these respiratory virus infections. For instance, reports on the influenza virus and respiratory syncytial virus (RSV) are available for the global scenario, but not much could be found on the prevalence and coverage of the health system of the parainfluenza virus and metapneumovirus. Availability of reliable data is a primary requirement for the assessment of medical needs and also in the field of prediction for respiratory viral infections. Unfortunately, this is lacking in most countries, including India. We tried to assess this data gap and, hence, reviewed the literature from 1970 to 2020 mentioning cases of respiratory virus infections in India and from a global perspective. This review is all about the respiratory infections that occurred due to virus infections. It is high time that we should take some measures to manage these infections. Currently, every researcher is compelled by the circumstances to look at COVID-19 infections, although other respiratory infections should also be taken care of to avoid similar situations to other viral entities. With the COVID-19 pandemic, it has become clear that there exist deficiencies in the system and planning at a larger scale to address population-level needs. This broad evaluation of the burden of respiratory viruses in India from 1970 to 2020 points to the clear need for better and conclusive diagnostics for respiratory viruses and also for plans in controlling the spread and reemergence of infections. To our knowledge, this is the first systematic analysis of the prevalence of multiple respiratory viruses such as influenza virus, RSV, parainfluenza virus, metapneumovirus, adenovirus, rhinovirus, etc. covering all geographical regions and states of India. We also tried to cover the situation on a global scale in comparison with the Indian scenario.

MATERIALS AND METHODS

Search Strategy

We conducted data gathering of peer-reviewed papers indexed in PubMed and available search engine reports for estimations of the burden of chronic respiratory diseases; the prevalence of respiratory viruses; and viral infection throughout the country using the keywords "prevalence," "burden," "respiratory viruses," "chronic respiratory disease," "co-infection," "epidemiology," "India," "morbidity," "mortality," and "trends" from December 2020 to April 2021. We also tried to look for the diagnostic scenario of the virus by going through the number of publications and the types of virus reported. Our search remains restricted to the English language only. Studies with overlapping populations were sorted, and the ones with the largest and inclusive data were chosen to have conclusive outcomes.

Eligibility Criteria

Studies with cross-sectional, prospective, and cohort designs were all included. The objective of these studies was about detailing the prevalence of respiratory viruses and respiratory viral infections. Also, studies with diagnostic intent for viral infections describing PCR, real-time PCR, and cell culture-based detection were considered. Most of the studies with pediatric respiratory infections were given priority, but infections reported in adult populations were also considered. Studies reporting on the diagnosis of multiple viral species were also given priority as they have good coverage of all respiratory viruses. Reports describing the detection of individual respiratory viruses were considered for adding the counts to the overall prevalence. Studies with animal experiments and *ex vivo* and toxicological perspectives were excluded, along with non-peer-reviewed articles and reports.

RESULTS

Prevalence of Respiratory Viruses in India Based on Reported Cases

During our literature survey, we decided to gather data on reviews regarding the combined prevalence of respiratory viruses in India; however, no recent data could be found on Google Scholar, PubMed, and other search engines accessible to us. Hence, segregated data from different regions and states or cities in India were gathered and combined. A total of 35 research articles from 1970 to 2020 were considered significant, with around 22,000 cases examined (**Table 1**). The reports in terms

TABLE 1 | Summary of cases studied from various states/cities of India.

Year(s) of study	Area	RSV	hADV	PIV	Rhino	Influenza	INF-A	INF-B	hMPV	No. of samples	No. of cases	References
1964–1970	West Bengal	12	160	55	2	0	0	3	0	4,771	483	Kloene et al., 1970
2001	New Delhi	22	63	82	0	0	143	2	0	832	317	Jain et al., 2001
2008	West India	100	30	8	0	21	0	0	0	385	143	Yeolekar et al., 2008
2007–2008	Kolkata	95	0	0	0	0	121	59	0	1,091	275	Agrawal et al., 2009
2009–2011	North India	0	0	0	0	46	28	0	0	1,043	74	Gupta et al., 2013
2009–2011	North India	50	13	10	0	17	0	0	3	245	98	Broor et al., 2014
2009–2011	Pune	130	0	49	145	0	72	0	44	843	503	Choudhary et al., 2013
2010–2011	Kolkata	194	0	216	6	0	320	328	82	2,737	1616	Roy Mukherjee et al., 2013
2010–2011	West Bengal	88	0	0	0	0	30	135	8	880	232	Mazumdar et al., 2013
2010–2011	Kolkata	2	0	0	0	9	0	1	23	447	35	Roy Mukherjee et al., 2013
2011–2012	Mumbai	2	0	0	0	37	15	24	0	200	78	Chavan et al., 2015
2011–2012	Lucknow	40	10	0	0	0	14	3	2	188	86	Singh et al., 2014
2012	North India	0	82	0	0	0	0	0	0	508	82	Saha et al., 2015
2011–2013	East India	62	31	0	52	67	0	0	0	358	248	Mishra et al., 2016
2013–2014	Chennai	1	0	4	0	0	2	3	5	200	83	Ramya et al., 2017
2012–2014	North India	139	0	88	0	42	0	0	86	875	335	Krishnan et al., 2019
2012–2014	Rajasthan	6	12	13	15	27	2	4	35	155	136	Malhotra et al., 2016
2014–2015	Chandigarh	0	15	0	0	0	0	0	0	23	15	Singh et al., 2018
2012–2015	Pondicherry	32	0	20	0	179	0	0	24	648	287	Palani and Sistla, 2020
2012–2016	Rajasthan	279	83	24	106	30	11	20	44	997	747	Swamy et al., 2018
2016–2017	Odessa	14	0	7	82	0	0	0	8	332	108	Panda et al., 2017
2017	J&K	5	2	0	11	18	15	3	2	233	46	Koul et al., 2017
2017–2018	South India	15	4	2	7	0	2	5	0	69	33	Abinaya et al., 2020
2017–2018	Delhi	11	4	1 4	29	0	11	0	6	118	81	Meena et al., 2019
2016–2018	Chennai	60	0	15	13	39	0	0	12	350	135	Hindupur et al., 2020
2018	Delhi	57	7	21	80	0	29	30	0	306	224	Sachdev and Gupta, 2018
2017–2019	Telangana	8	5	23	53	56	87	33	48	513	261	Anand and Nimmala, 2020
2019	Coimbatore	10	0	0	12	25	10	8	0	185	84	Akila, 2019
2019	West India	29	18	11	21	5	0	0	7	100	82	Sonawane et al., 2019
2019	South India	94	0	6	0	0	15	10	0	383	130	Kini et al., 2019
2020	Pune	6	4	13	15	27	2	4	35	362	84	Potdar et al., 2020

The number of cases detected (Green to Orange Shading indicate increasing numbers) in the given study for a particular virus is listed under the virus name. RSV, respiratory syncytial virus; hADV, human adenovirus; PIV, parainfluenza virus; Rhino, rhinovirus; INF-A, influenza A; INF-B, influenza B; hMPV, human metapneumovirus.

of reliable publications did not emerge in a sustained fashion. Rather, there have been gaps in reporting. This was due to the unavailability of a program that could have monitored the prevalence of these viruses on a yearly basis. As apparent from **Table 1**, not many reliable reports could be found describing viral prevalence during 1970–2000. A similar trend was found from 2001 to 2007. However, with the availability and affordability of PCR and other detection modalities, regular reporting was observed after 2007. The two peaks in numbers, as observed in 2009 and 2012, were due to outbreaks, and hence there was enhanced diagnosis during those periods. This change in number was found to be an outcome of a change in the diagnostic rate. Earlier diagnoses or surveys did not have a comprehensive panel of viruses, and therefore the trend shows no detection of a few viral entities such as rhinovirus and hMPV in initial reports. Influenza, being an annual affair and better known to diagnostic systems, was always included in the diagnosis and was always detected in higher numbers.

Table 1 and **Figure 1** summarize the above findings with details of the type and number of viruses, their area of reporting, and the time durations from the overall findings of India. In total, RSV was found to be highly prevalent at all times with a

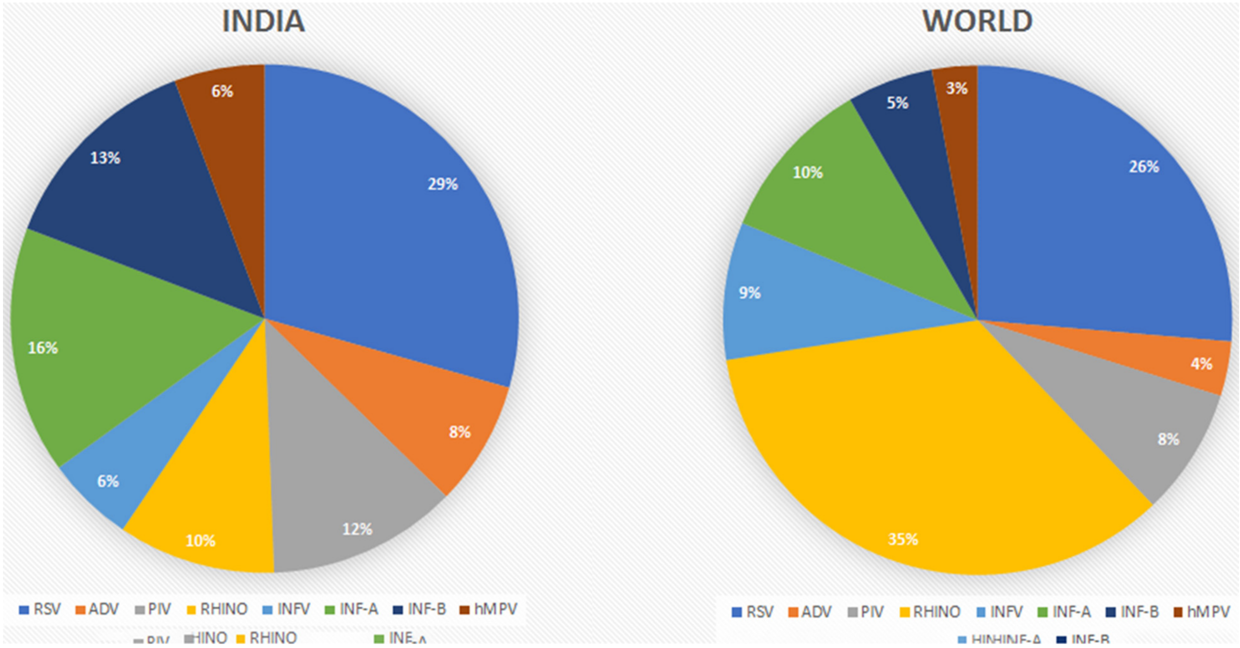


FIGURE 1 | Comparison of the prevalence rates of viral infections between India and the rest of the world.

percent contribution of 29%. Influenza A was the second most prevalent virus with a 16% share in all the detected viruses. The other viruses are also shown with their percent prevalence.

The reports were also classified based on the geographical regions as divided into four major zones, i.e., East, West, North, and South India. **Table 2** and **Figure 2** give an overview of the type and number of viruses and their prevalence based on the studied reports. The regions mentioned do not mean every part of the region, but that the area of the report falls under that geographical region. In brief, East India has reports from West Bengal and Odisha, and West India included reports from Maharashtra with independent reports from Mumbai and Pune along with Rajasthan. Reports from Ballabgarh, rural Ballabgarh, Chandigarh, Lucknow, J&K, Delhi, New Delhi, and Haryana were included in North India, while South India included reports from Karnataka with independent reports from Bengaluru, Puducherry, Telangana, and Tamil Nadu including Chennai and Coimbatore.

The reports were further classified based on the prevalence of respiratory viruses in those regions considering the differences in climatic conditions and the chance of finding a higher number of infections with a particular virus (**Table 2** and **Figure 2**).

North India

The North India region is a large area with cold temperatures and chilling conditions in the winter. This could be the reason behind influenza and parainfluenza having detection rates of 27 and 14%, respectively. The 2,085 reported cases had 747 positive detections for upper respiratory tract infection (URTI; Jain et al., 2001; Broor et al., 2014; Singh et al., 2014; Saha et al., 2015; Sachdev and Gupta, 2018; Krishnan et al., 2019; Meena et al., 2019). Parainfluenza and INF-B cases were reported to be higher here.

East India

It was found that publications are mostly available for West Bengal, only Kolkata and Odessa. In total, these studies have examined 12,875 patients for respiratory tract infection, of which 4,548 were detected positive for viral infections. East India is the only region from which relatively lower reported cases of RSV (20%) were found. INF-A and INF-B show the highest prevalence rates in India, i.e., 22 and 25%, respectively (Kloene et al., 1970; Agrawal et al., 2009; Mazumdar et al., 2013; Roy Mukherjee et al., 2013; Mishra et al., 2016; Panda et al., 2017). Data analysis revealed that the number of reported cases and the prevalence of influenza are higher compared to those of other regions in India.

West India

Reports from this region were mostly from Pune and Rajasthan, while other regions barely reported the viral diagnosis. As per these reports, 8,448 cases were studied and 1,315 samples were found positive for respiratory virus infections. Rhinovirus has a higher infection rate (18%) than in other regions and is detected regularly (**Figure 3**; Yeolekar et al., 2008; Choudhary et al., 2013; Chavan et al., 2015; Malhotra et al., 2016; Swamy et al., 2018; Sonawane et al., 2019). This region has reported the highest prevalence of rhinovirus: the cause of the common cold. RSV was also found to be higher in this region.

South India

The South India region with 2,298 cases reported 1,013 positive detections for respiratory virus, and around 50% of the cases

TABLE 2 | Prevalence of virus infections based on the geographical regions in India (1970–2020).

Region	Virus type									
	RSV	ADV	PIV	RHINO	INFV	INF-A	INF-B	hMPV	Total no. of cases	No. of positive cases
North India	232	83	102	59	195	40	12	11	2,085	747
East India	468	193	320	144	34	532	584	111	1,2875	4,548
West India	469	118	138	230	37	139	44	115	2,448	1,351
South India	324	10	54	76	18	89	44	56	2,298	1,013
Total cases	1493	404	614	509	284	800	684	293	1,9706	7,659

RSV, respiratory syncytial virus; ADV, adenovirus; PIV, parainfluenza virus; Rhino, rhinovirus; INFV, influenza virus; INF-A, influenza A; INF-B, influenza B; hMPV, human metapneumovirus.

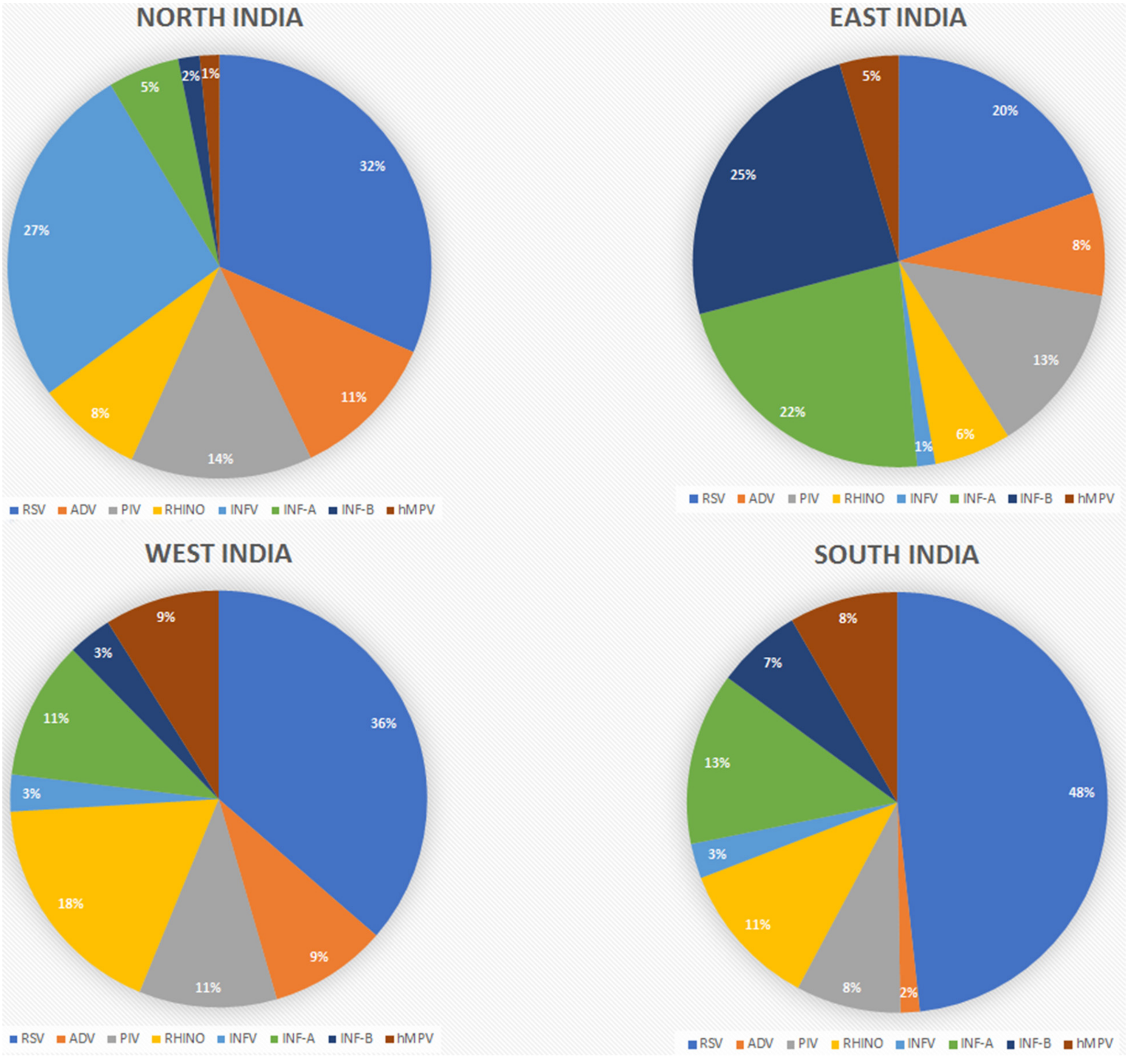


FIGURE 2 | Prevalence of viral infections based on the geographical regions in India.

detected were RSV, which is higher than that in any other region in the country (Akila, 2019; Kini et al., 2019; Abinaya et al., 2020; Anand and Nimmala, 2020; Hindupur et al., 2020; Palani and Sistla, 2020).

The very first report that we considered has summarized data from four villages in West Bengal, and the sampling duration was from 1964 to 1966. More than 4,000 samples were collected and 625 viruses were isolated. These included all types of respiratory

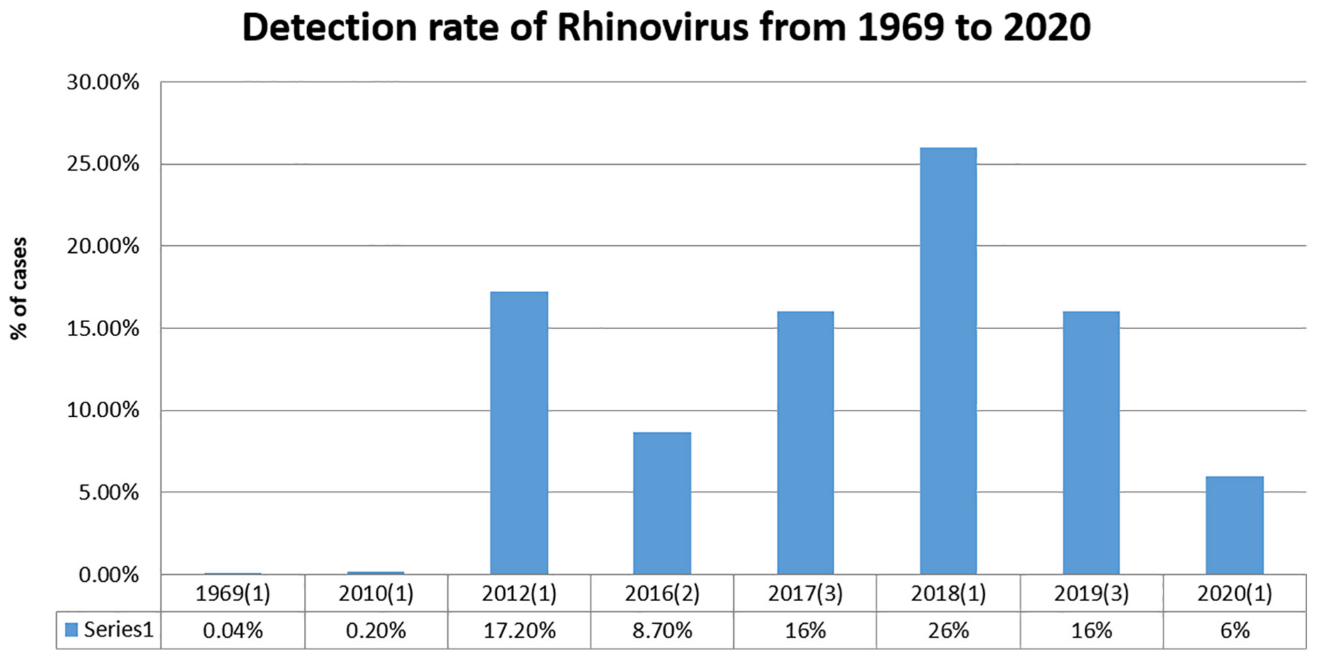


FIGURE 3 | Detection rate of rhinovirus from 1964 to 2020. The number of reports included in the study during this period is indicated in brackets.

viruses, with parainfluenza, RSV, adenovirus, rhinovirus, and influenza virus types. It was discussed that there were waves of rising infections during this period, while the persistence of these viruses was also noted. The results of other studies, as described in **Table 1** and **Figure 4**, can be discussed based on individual viruses, and their trend of detection in a given period is as described below.

Respiratory Syncytial Virus

From 1969 to 2001, reported cases of RSV were comparatively less than those of other viruses, but after 2007, the detection rate increased gradually, taking the total count of RSV-infected patients to 1,625 as calculated with the data included in this study. RSV has a higher prevalence rate in India than any other respiratory viruses, as shown in **Figure 2**. As compared regionally, South India showed a bit higher infection rate. This is the trend not only in India but also in other parts of the world (**Figure 1**). RSV has always been a part of the diagnostic machinery and was initially detected in moderate numbers routinely, except in 2009–2010 and 2012–2014 when its detection rate reached as high as 25% (Choudhary et al., 2013) and 41% (Krishnan et al., 2019), respectively, of the total positive cases. The prevalence took an upward trend afterward, and very high number of cases were reported until the recent time, except in 2020 when COVID-19 shadowed all detections and outnumbered all.

Human Adenovirus

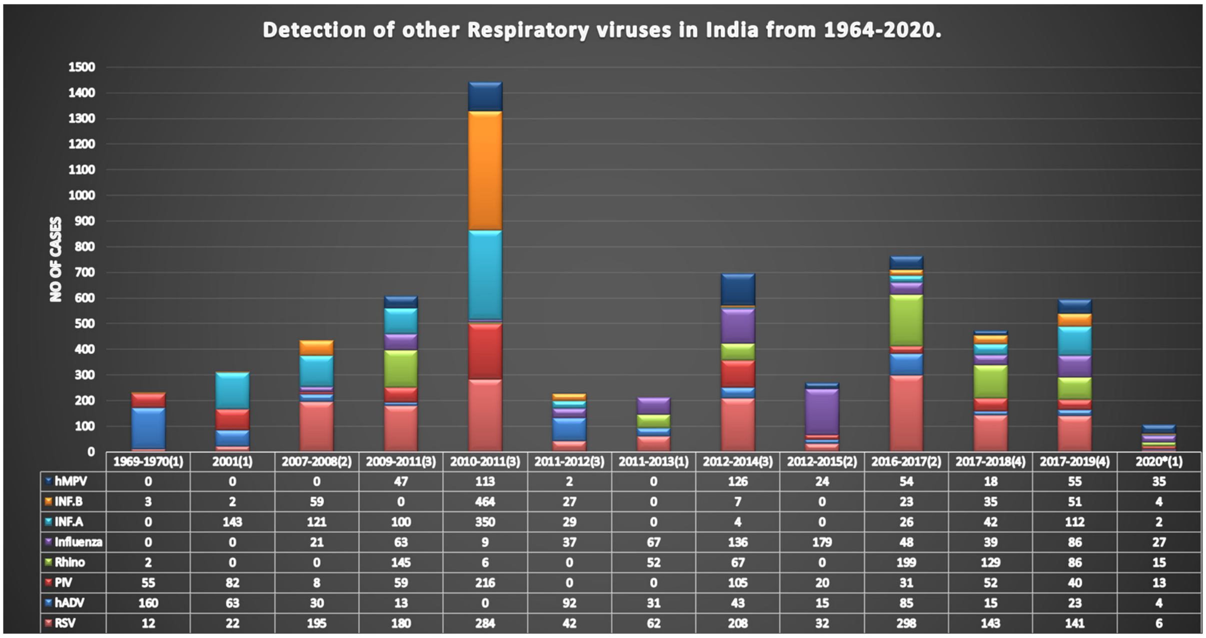
Detection of adenovirus was available since the earlier period of the reporting in this study, and its prevalence was reported in higher numbers initially. In later publications, this was not given a high priority due to the non-severe outcomes of the infection and was not included routinely while diagnosing other respiratory viruses in many studies. After 2011, the availability of multiplex testing again allowed including adenovirus in the diagnostic panel; since then, it has been reported routinely, with moderate counts (**Table 1** and **Figure 4**). Regional comparison showed its lower prevalence in South India, while the other three regions have a higher prevalence of this virus (**Figure 2**).

Human Parainfluenza Virus

Human parainfluenza viruses, a member of the Paramyxoviridae family, is classified as HPIV-1 to HPIV-4 based on the serotypes. In most cases of serious infections, HPIV-3 is diagnosed as a causal entity followed by HPIV-1 and HPIV-2, while HPIV-4 is associated with milder symptoms and is not included in the diagnosis for routine assays. The incidence rates of HPIV have always been higher whenever detected. The literature reported a higher incidence in colder temperatures; hence, more cases were recorded in North India and during the winter season compared to the other parts of the country and in other seasons (**Table 1** and **Figure 2**).

Influenza Virus

In some of the papers, influenza virus was diagnosed and reported as influenza only and was not further classified into its subtypes; hence, it has been depicted as a separate row in **Table 1**, along with INF-A and INF-B. It has been reported to have moderate to high counts, except in 2011–2012 when there was an influenza outbreak and a higher number of samples were collected. As per the published literature, influenza was found to be one of the highly prevalent infections after 2001.



	1969-1970(1)	2001(1)	2007-2008(2)	2009-2011(3)	2010-2011(3)	2011-2012(3)	2011-2013(1)	2012-2014(3)	2012-2015(2)	2016-2017(2)	2017-2018(4)	2017-2019(4)	2020*(1)
hMPV	0	0	0	47	113	2	0	126	24	54	18	55	35
INF.B	3	2	59	0	464	27	0	7	0	23	35	51	4
INF.A	0	143	121	100	350	29	0	4	0	26	42	112	2
Influenza	0	0	21	63	9	37	67	136	179	48	39	86	27
Rhino	2	0	0	145	6	0	52	67	0	199	129	86	15
PIV	55	82	8	59	216	0	0	105	20	31	52	40	13
hADV	160	63	30	13	0	92	31	43	15	85	15	23	4
RSV	12	22	195	180	284	42	62	208	32	298	143	141	6

FIGURE 4 | Prevalence of respiratory viruses from 1969 to 2020 in India. This figure shows the total prevalence of respiratory viruses in India from 1969 to 2020. The number of cases is reported on the *Y*-axis and the year of reporting is presented on the *X*-axis, while the number in brackets indicates the number of papers published in that period, with papers published in the same year being merged. The tabular format below the graphical presentation includes the number of viruses studied along with the year-wise reported positive cases.

Influenza A

After RSV, influenza A shows a higher detection rate in India, and in the current review, a total of 929 confirmed cases are included (**Figure 2**). INF-A is highly detected in East and South India, maybe because of the pandemic. The counts of influenza viruses remain very high, as reported in various studies so far, despite the availability of the vaccine in 2015. However, there was a gradual decline observed after this period.

Influenza B

The detection rate of INF-B is relatively low as compared to other viruses, although a peak was detected in 2010–2011 due to the enhanced detection during the outbreak of H1N1. At that time, a total of 4,064 patients were tested under routine diagnosis, whereas 675 cases were detected positive for INF-B.

Human Metapneumovirus

In India, hMPV has been taken into consideration in routine diagnostic tests since 2009. Before that, no cases of hMPV have been detected, as depicted in **Figure 4**. The prevalence of hMPV in India remained low compared to the other viruses until 2012; thereafter, the number of positive cases rose slightly each year. The detection rate is higher in warmer regions, i.e., South and West India, compared to colder regions, i.e., North and East India (**Figure 2**).

Rhinovirus: Is the Prevalence Increasing?

During our analysis, it was very apparent that rhinovirus has always been present in regional outbreaks or epidemics. Increased numbers of rhinovirus cases are being reported with improved diagnostics and better medical coverage. Although timelines could not be drawn with the limited data, we could observe a pattern that indicates increasing number of cases of rhinovirus infection in India, which is in line with global reports. **Figure 3** shows the pattern of cases during the period from 1969 to 2020, leaving those years in which there were no reports published.

The Scenario in Other Parts of the World

We also tried to compare our findings with reports from other parts of the world. We went through the literature available from other parts of the world and focused on studies reporting larger number of samples. Reports from five continents—Asia, Australia, Africa, Europe, and the United States—were found to be useful. At the same time, we included individual reports from the countries such as Bangladesh, China, Kenya, Thailand, Egypt, Italy, the Netherlands, Hong Kong, and Sweden, with 1971–2015 as the period of the research (**Table 3**). Overall, these case studies reported 36,559 cases. Cases from the different continents showed a diversity of regions, weather conditions, gender, race, age, religion, etc. (Henrickson et al., 2004; Farha and Thomson, 2005; Esposito et al., 2013; Zheng et al., 2018; Milucky et al., 2020). The reports that were picked up for this study gave cumulative data for a certain period and a larger coverage in terms of spread. Sporadic reports are not considered here as we aimed to get a generalized picture of the global scenario of all the respiratory viruses. WHO has conducted global surveillance of the important respiratory pathogens such as influenza and

TABLE 3 | Representative studies from the rest of the world to compare the scenario of viral disease prevalence in India.

Area of reporting	hADV hADV	RSV	Influenza	FLU.A	FLU.B	Rhino	PIV	hMPV	Year(s) of research	Total no. of samples	No. of positive cases
Wisconsin, United States	27	778	–	241	45	30	226	–	1996–1998	3,325	1,347
Finland, United States, Australia, Switzerland, the Netherlands, Hong Kong, Sweden	95	572	–	285	270	74	211	63	1986–2003	9,768	1,059
Europe, Asia, America, Oceania, Africa	126	561	246	–	–	2,330	150	90	1971–2014	20,486	3,503
Milan, Italy	11	188	57	6	8	144	11	49	2007–2014	592	435
Bangladesh, China, Guatemala, Kenya, Thailand, Egypt	27	72	417	327	127	272	70	36	2013–2015	2,388	1,815
Total no. of cases	286	2,171	720	859	450	2,850	668	238	1971–2015	36,559	8,159

hADV, human adenovirus; RSV, respiratory syncytial virus; FLU.A, influenza A; FLU.B, influenza B; Rhino, rhinovirus; PIV, parainfluenza virus; hMPV, human metapneumovirus.

RSV. The Global Influenza Surveillance and Response System (GISRS) is one of the examples of the global surveillance of influenza, which has been done for more than 60 years and is still going on. Similarly, an RSV surveillance system is proposed using the GISRS platform. The National Respiratory and Enteric Virus Surveillance System (NREVSS) in the United States also conducts surveys of most of the respiratory viruses to understand the temporal and geographic circulation patterns by focusing on each virus individually. The patterns obtained from the published literature are shown in **Table 4** and **Figure 1**. After analysis, it was observed that India and other regions of the world follow a similar trend of prevalence of viral infection. Rhinovirus is the most prevalent virus with 35% of cases worldwide, whereas in India rhinovirus has a detection rate of 11%, but the trend line shows that its prevalence is increasing (**Figure 3**). The prevalence of RSV was found to be very similar in India compared to other parts of the world. In India, RSV is the most prevalent virus, whereas it is the second most prevalent virus in other parts of the world. In India, influenza and its INF-A subtype show almost similar detection rates (influenza: 11% in India and 9% in other parts of the world; INF-A: 15% in India and 10% in other parts of the world). INF-B and hMPV have higher prevalence rates in India of 12 and 8%, respectively, whereas the worldwide incidence rate is only 3% for both INF-B and hMPV. The prevalence rates of ADV and PIV were found to be almost similar in India to other parts of the world. However, their detection rates are also low, as per the reviewed literature (**Table 4**).

DISCUSSION

Although previous attempts have been made to estimate the prevalence rates of respiratory viruses or specific viruses in some states of the country, there have been no report summarizing the burden of respiratory viruses or infectious diseases, deaths, across all states of India for a long time. In total, the scenario of the whole country can be discussed as depicted in **Figure 2** and **Table 1**. RSV remained the most prevalent respiratory virus in India for the whole duration of the study period. If we add up all the influenza types together, we may find it to be the most prevalent viral cause of respiratory infections. Parainfluenza has also been a significant one, with rhinovirus as an emerging viral pathogen during this period, which may be due to improved diagnostics. At the same time, the prevalence rates of adenovirus and hMPV remained on the lower side.

RSV was the dominant one in all regions, except for East India that has variable percent contributions of all the other viruses, while influenza was highest in East India, followed by North India, compared to the other geographical regions. Reports from East India show higher numbers of influenza B, followed by influenza A, and a few reports indicating influenza as one group without details about its types. Also, this region has a higher prevalence of parainfluenza virus compared to all other regions. The higher prevalence rates of influenza and parainfluenza viruses in North and East India could be due to the extreme cold weather during the winter season and also to the drastic temperature variability in different seasons. Most of the studies considered all types of influenza in one group for reporting purposes; however, their subtypes may vary in occurrence and also may not be classified further due to resource constraints.

Human metapneumovirus had the lowest prevalence in all the regions, with the least prevalence in North India. West and

TABLE 4 | Comparison of the detection rates of viruses in India with those of other parts of the world.

Prevalence in India (%)	Virus	Prevalence in the rest of the world (%)
26	RSV	25
7	ADV	4
10	PIV	8
11	Rhino	35
11	Influenza	9
15	INF-A	10
12	INF-B	3
8	hMPV	3

RSV, respiratory syncytial virus; ADV, human adenovirus; PIV, parainfluenza virus; Rhino, rhinovirus; INF-A, influenza A; INF-B, influenza B; hMPV, human metapneumovirus.

South India have relatively higher prevalence rates of this virus. The higher percentages of hMPV in West and South India may be due to the relatively hot and humid climate in the majority of the areas, being in the coastal region. As reported earlier, infections with hMPV peak during the rainy season, and hence its higher prevalence in humid regions can be correlated (Evelyn et al., 2019). The western region has a relatively lower prevalence of influenza and other viruses compared to the other regions, while it has the highest number of reported cases of rhinovirus. The southern region was found to report the highest number of RSV cases, which also contributes to the overall high number of this virus in India.

The distribution of adenovirus was found to be uneven, with high prevalence in North India while very low in South India. A reason for this may be its moderate symptoms and low severity. Another possibility is that the infected individuals are normally not subjected to diagnosis, and hence many cases remain undiagnosed. In most cases, it was detected during surveys or after an attempt to detect other viral infections.

Rhinovirus has been the most prevalent virus in most parts of the world and is now also being reported frequently in India. The reason for the lower prevalence of rhinovirus in India may be its underdiagnosis in the earlier phase of infection (**Figure 3**). With improved and wide-ranging diagnostic procedures, it now seems that the pattern of its prevalence is similar to that of the rest of the world. The observation in this study also points toward an increase in prevalence from 2012 to 2019, after which COVID-19 has taken over and not many reports could be found. This also points toward the long persistence of this virus and a need to work in this area to avoid any major epidemic due to its variants or larger spread.

CONCLUSION

The outcomes of this study have projected an excessive load of disability-adjusted life years (DALYs; 32% of the global count) due to respiratory infections compared to the total population of the country, which is 18% of the global population. We provided a broad evaluation of the burden of respiratory viruses and diseases in different states of India, as classified under four geographical regions from 1970 to 2020, based on all accessible data. An important finding of this study is that most states in India have higher rates of infections of RSV, influenza, and parainfluenza. These are the top three infectious viruses as per the studied reports. Also, it was found that from 2016 to 2017, infection of rhinovirus has increased at a higher rate. The exercise was done during the collection of data, and its compilation also indicated the unavailability of a systematic centralized database of respiratory viruses and their regional prevalence and cumulative yearly incidences. The inferences drawn out from the prevalence data can help in deciding the diagnostic priorities in these regions and devising local protocols for routine diagnosis. In conclusion, the review points toward the need for improved detection of multiple viruses and informed management of their infections. Another key finding during our data analysis was that most of the reports were originated from areas where research facilities are established. Hence, either the coverage of these research facilities may be expanded or a few more diagnostic research laboratories may be established in those areas that are remote and not accessed by the public so far for the diagnosis of these viruses. Also, routine surveillance of such viruses may be added to their mandate. A new ray of hope has been seen with the decision for the establishment of VRDLs in 2016 by the Department of Health Research (DHR)/Indian Council of Medical Research (ICMR). This is a network of laboratories to enhance the country's capacity for the early diagnosis of all viral infections with public health importance in the Indian context. The target is to establish 160 such laboratories all over the country; 106 of such laboratories are already established and are either functioning or being made functional.

AUTHOR CONTRIBUTIONS

RW: manuscript draft preparation, review, and editing. SJ: review and approval. VN: concept, manuscript writing, editing, approval, and supervision. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

RW is thankful to the Director, ICMR-NARI, for allowing him to train under VN at the institute.

REFERENCES

Abinaya, S., Gaspar, B., and Benjamin, A. (2020). A study on aetiology and outcomes of viral lower respiratory tract infections in hospitalized children from South India. *Sri Lanka J. Child Health* 49, 218–222. doi: 10.4038/sljch.v49i3.9137

Agrawal, A. S., Sarkar, M., Chakrabarti, S., Rajendran, K., Kaur, H., Mishra, A. C., et al. (2009). Comparative evaluation of real-time PCR and conventional RT-PCR during a 2 year surveillance for influenza and respiratory syncytial virus among children with acute respiratory infections in Kolkata, India, reveals a distinct seasonality of infection. *J. Med. Microbiol.* 58, 1616–1622. doi: 10.1099/jmm.0.011304-0

Akila, K. (2019). *Viral Etiology of Acute Respiratory Tract Infections in Adults in Tertiary Care Hospital*. Doctoral dissertation. Coimbatore: PSG Institute of Medical Sciences and Research.

Anand, M., and Nimmala, P. (2020). Seasonal incidence of respiratory viral infections in Telangana, India: utility of a multiplex PCR assay to bridge the knowledge gap. *Trop. Med. Int. Health* 25, 1503–1509. doi: 10.1111/tmi.13501

Banerjee, A., De, P., Manna, B., and Chawla-Sarkar, M. (2017). Molecular characterization of enteric adenovirus genotypes 40 and 41 identified in children with acute gastroenteritis in Kolkata, India during 2013-2014. *J. Med. Virol.* 89, 606–614. doi: 10.1002/jmv.24672

Briese, T., Renwick, N., Venter, M., Jarman, R. G., Ghosh, D., Köndgen, S., et al. (2008). Global distribution of novel rhinovirus genotype. *Emerg. Infect. Dis.* 14, 944–947. doi: 10.3201/eid1406.080271

Broor, S., Dawood, F. S., Pandey, B. G., Saha, S., Gupta, V., Krishnan, A., et al. (2014). Rates of respiratory virus-associated hospitalization in children aged< 5 years in rural northern India. *J. Infect.* 68, 281–289. doi: 10.1016/j.jinf.2013.11.005

Broor, S., Parveen, S., and Maheshwari, M. (2018). Respiratory syncytial virus infections in India: epidemiology and need for vaccine. *Indian J. Med. Microbiol.* 36, 458–464. doi: 10.4103/ijmm.ijmm_19_5

Chavan, R. D., Kothari, S. T., Zunjarrao, K., and Chowdhary, A. S. (2015). Surveillance of acute respiratory infections in Mumbai during 2011-12. *Indian J. Med. Microbiol.* 33:43. doi: 10.4103/0255-0857.148376

Choudhary, M. L., Anand, S. P., Heydari, M., Rane, G., Potdar, V. A., Chadha, M. S., et al. (2013). Development of a multiplex one step RT-PCR that detects eighteen respiratory viruses in clinical specimens and comparison with real time RT-PCR. *J. Virol. Methods* 189, 15–19. doi: 10.1016/j.jviromet.2012.12.017

Esposito, S., Daleno, C., Prunotto, G., Scala, A., Tagliabue, C., Borzani, I., et al. (2013). Impact of viral infections in children with community-acquired pneumonia: results of a study of 17 respiratory viruses. *Influenza Other Respir. Viruses* 7, 18–26. doi: 10.1111/j.1750-2659.2012.00340.x

Evelyn, O., Jaime, F. S., David, M., Lorena, A., Jenifer, A., and Oscar, G. (2019). Prevalence, clinical outcomes and rainfall association of acute respiratory infection by human metapneumovirus in children in Bogotá, Colombia. *BMC Pediatr.* 19:345. doi: 10.1186/s12887-019-1734-x

Farha, T., and Thomson, A. H. (2005). The burden of pneumonia in children in the developed world. *Paediatr. Respir. Rev.* 6, 76–82. doi: 10.1016/j.prrv.2005.03.001

Gupta, V., Dawood, F. S., Rai, S. K., Broor, S., Wigh, R., Mishra, A. C., et al. (2013). Validity of clinical case definitions for influenza surveillance among hospitalized patients: results from a rural community in North India. *Influen. Other Respirat. Virus.* 7, 321–329.

Henrickson, K. J., Hoover, S., Kehl, K. S., and Hua, W. (2004). National disease burden of respiratory viruses detected in children by polymerase chain reaction. *Pediatr. Infect. Dis. J.* 23, S11–S18. doi: 10.1097/01.inf.0000108188.37237.48

Hindupur, A., Menon, T., and Dhandapani, P. (2020). Molecular surveillance of respiratory viruses in children with acute respiratory infections in Chennai, South India. *Int. J. Infect. Dis.* 101:515. doi: 10.1016/j.ijid.2020.09.1337

Jain, N., Lodha, R., and Kabra, S. (2001). Upper respiratory tract infections. *Indian J. Pediatr.* 68, 1135–1138.

Kini, S., Kalal, B. S., Chandy, S., Shamsundar, R., and Shet, A. (2019). Prevalence of respiratory syncytial virus infection among children hospitalized with acute lower respiratory tract infections in Southern India. *World J. Clin. Pediatr.* 8, 33–42. doi: 10.5409/wjcp.v8.i2.33

Kloene, W., Bang, F. B., Chakraborty, S. M., Cooper, M. R., Kulemann, H., Ota, M., et al. (1970). A two-year respiratory virus survey in four villages in West Bengal, India. *Am. J. Epidemiol.* 92, 307–320. doi: 10.1093/oxfordjournals.aje.a121212

Koul, P. A., Mir, H., Akram, S., Potdar, V., and Chadha, M. S. (2017). Respiratory viruses in acute exacerbations of chronic obstructive pulmonary disease. *Lung India* 34, 29–33.

Krishnan, A., Kumar, R., Broor, S., Gopal, G., Saha, S., Amarchand, R., et al. (2019). Epidemiology of viral acute lower respiratory infections in a community-based cohort of rural north Indian children. *J. Glob. Health* 9:010433.

Malhotra, B., Swamy, M. A., Janardhan Reddy, P. V., and Gupta, M. L. (2016). Viruses causing severe acute respiratory infections (SARI) in children =5 years of age at a tertiary care hospital in Rajasthan, India. *Indian J. Med. Res.* 144, 877–885. doi: 10.4103/ijmr.IJMR_22_15

Mazumdar, J., Chawla-Sarkar, M., Rajendran, K., Ganguly, A., Sarkar, U. K., Ghosh, S., et al. (2013). Burden of respiratory tract infections among paediatric in and out-patient units during 2010-11. *Eur. Rev. Med. Pharmacol. Sci.* 17, 802–808.

Meena, J. P., Brijwal, M., Seth, R., Gupta, A. K., Jethani, J., Kapil, A., et al. (2019). Prevalence and clinical outcome of respiratory viral infections among children with cancer and febrile neutropenia. *Pediatr. Hematol. Oncol.* 36, 330–343. doi: 10.1080/08880018.2019.1631920

Milucky, J., Pondo, T., Gregory, C. J., Iuliano, D., Chaves, S. S., McCracken, J., et al. (2020). The epidemiology and estimated etiology of pathogens detected from the upper respiratory tract of adults with severe acute respiratory infections in multiple countries, 2014–2015. *PLoS One* 15:e0240309. doi: 10.1371/journal.pone.0240309

Mishra, P., Nayak, L., Das, R. R., Dwibedi, B., and Singh, A. (2016). Viral agents causing acute respiratory infections in children under five: a study from Eastern India. *Int. J. Pediatr.* 2016:7235482.

Narayan, V. V., Iuliano, A. D., Roguski, K., Bhardwaj, R., Chadha, M., Saha, S., et al. (2020). Burden of influenza-associated respiratory and circulatory mortality in India, 2010-2013. *J. Glob. Health* 10:010402. doi: 10.7189/jogh.10.010402

Palani, N., and Sistla, S. (2020). Epidemiology and phylogenetic analysis of respiratory viruses from 2012 to 2015–A sentinel surveillance report from union territory of Puducherry, India. *Clin. Epidemiol. Glob. Health* 8, 1225–1235. doi: 10.1016/j.cegh.2020.04.019

Panda, S., Mohakud, N. K., Suar, M., and Kumar, S. (2017). Etiology, seasonality, and clinical characteristics of respiratory viruses in children with respiratory tract infections in Eastern India (Bhubaneswar, Odisha). *J. Med. Virol.* 89, 553–558. doi: 10.1002/jmv.24661

Potdar, V., Choudhary, M. L., Bhardwaj, S., Ghuge, R., Sugunan, A. P., Gurav, Y., et al. (2020). Respiratory virus detection among the overseas returnees during the early phase of COVID-19 pandemic in India. *Indian J. Med. Res.* 151, 486–489. doi: 10.4103/ijmr.ijmr_638_20

Rafeek, R. A. M., Divarathna, M. V. M., and Noordeen, F. (2020). A review on disease burden and epidemiology of childhood parainfluenza virus infections in Asian countries. *Rev. Med. Virol.* 31:e2164.

Ramya, R., Sagadevan, P., and JayaramJayaraj, K. (2017). Genetic characterization and the development of multiplex PCR for common respiratory viruses, in Chennai during September 2013 to January 2014. *Int. J. Chemtech Res.* 10, 389–401.

Rao, B. L., Gandhe, S. S., Pawar, S. D., Arankalle, V. A., Shah, S. C., and Kinikar, A. A. (2004). First detection of human metapneumovirus in children with acute respiratory infection in India: a preliminary report. *J. Clin. Microbiol.* 42, 5961–5962. doi: 10.1128/jcm.42.12.5961-5962.2004

Roy Mukherjee, T., Chanda, S., Mullick, S., De, P., Dey-Sarkar, M., and Chawla-Sarkar, M. (2013). Spectrum of respiratory viruses circulating in eastern India: prospective surveillance among patients with influenza-like illness during 2010–2011. *J. Med. Virol.* 85, 1459–1465. doi: 10.1002/jmv.23607

Sachdev, A., and Gupta, D. (2018). Incidence and severity profile of viral respiratory tract infections in children admitted to the tertiary level pediatric intensive care unit. *J. Pediatr. Crit. Care* 5:96. doi: 10.21304/2018.0501.00315

Saha, S., Pandey, B. G., Choudekar, A., Krishnan, A., Gerber, S. I., Rai, S. K., et al. (2015). Evaluation of case definitions for estimation of respiratory syncytial virus associated hospitalizations among children in a rural community of northern India. *J. Glob. Health* 5:010419.

Salvi, S., Kumar, G. A., Dhaliwal, R. S., Paulson, K., Agrawal, A., Koul, P. A., et al. (2018). The burden of chronic respiratory diseases and their heterogeneity across the states of India: the global burden of disease study 1990-2016. *Lancet Glob. Health* 6, E1363–E1374. doi: 10.1016/S2214-109X(18)30409-1

Singh, A. K., Jain, A., Jain, B., Singh, K. P., Dangi, T., Mohan, M., et al. (2014). Viral aetiology of acute lower respiratory tract illness in hospitalised paediatric patients of a tertiary hospital: one year prospective study. *Indian J. Med. Microbiol.* 32:13–28. doi: 10.4103/0255-0857.124288

Singh, M. P., Ram, J., Kumar, A., Rungta, T., Gupta, A., Khurana, J., et al. (2018). Molecular epidemiology of circulating human adenovirus types in acute conjunctivitis cases in Chandigarh, North India. *Indian J. Med. Microbiol.* 36, 113–115. doi: 10.4103/ijmm.ijmm_17_258

Sonawane, A. A., Shastri, J., and Bavdekar, S. B. (2019). Respiratory pathogens in infants diagnosed with acute lower respiratory tract infection in a Tertiary Care hospital of Western India using multiplex real time PCR. *Indian J. Pediatr.* 86, 433–438. doi: 10.1007/s12098-018-2840-8

Swamy, M. A., Malhotra, B., Reddy, P. J., and Tiwari, J. (2018). Profile of respiratory pathogens causing acute respiratory infections in hospitalised children at

Rajasthan a 4 year's study. *Indian J. Med. Microbiol.* 36, 163–171. doi: 10.4103/ijmm.ijmm_18_84

van den Hoogen, B. G., de Jong, J. C., Groen, J., Kuiken, T., de Groot, R., Fouchier, R. A., et al. (2001). A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat. Med.* 7, 719–724.

World Health Organization (2021). *World Health Organization fact sheet:Influenza (Seasonal)*. Geneva: World Health Organization.

Worldometer (2021). Availble online at: https://www.worldometers.info/coronavirus/ (accessed July 24, 2021).

Yeolekar, L. R., Damle, R. G., Kamat, A. N., Khude, M. R., Simha, V., and Pandit, A. N. (2008). Respiratory viruses in acute respiratory tract infections in Western India. *Indian J. Pediatr.* 75, 341–345.

Zheng, X. Y., Xu, Y. J., Guan, W. J., and Lin, L. F. (2018). Regional, age and respiratory-secretion-specific prevalence of respiratory viruses associated with asthma exacerbation: a literature review. *Arch. Virol.* 163, 845–853. doi: 10.1007/s00705-017-3700-y

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 Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fcimb.2021.739429

Clinical and Immunological Characteristics of Patients With Adenovirus Infection at Different Altitude Areas in Tibet, China

Bowen Wang[1†], **Mengjia Peng**[2†], **Li Yang**[1], **Guokai Li**[1], **Jie Yang**[3], **Ciren Yundan**[4], **Xiaohua Zeng**[5], **Qianqi Wei**[5,6], **Qi Han**[3], **Chang Liu**[1,7], **Ke Ding**[8], **Kaige Peng**[1*] and **Wen Kang**[9*]

[1] Department of Prevention and Control of Infectious Diseases, Center for Disease Control and Prevention (CDC) of Tibet Military Command, Lhasa, China, [2] Department of Emergency, General Hospital of Tibet Military Command, Lhasa, China, [3] Department of Radiology, General Hospital of Tibet Military Command, Lhasa, China, [4] Department of Thoracic Surgery, General Hospital of Tibet Military Command, Lhasa, China, [5] Department of Infectious Diseases, General Hospital of Tibet Military Command, Lhasa, China, [6] Department of Laboratory, 954 Hospital of Army, Lhoka, China, [7] Department of Laboratory, 956 Hospital of Army, Nyingchi, China, [8] Department of Radiology, Xuchang People's Hospital, Xuchang, China, [9] Department of Infectious Diseases, Tangdu Hospital, The Airforce Medical University, Xi'an, China

OPEN ACCESS

Edited by:
*Clyde Dapat,
Tohoku University Graduate School of Medicine, Japan*

Reviewed by:
*Fengjun Sun,
Army Medical University, China
Arryn Craney,
Weill Cornell Medical Center, United States*

***Correspondence:**
*Wen Kang
82399536@qq.com
Kaige Peng
pkgtmmu@163.com*

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

Specialty section:
This article was submitted to Clinical Microbiology, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *10 July 2021*
Accepted: *22 September 2021*
Published: *13 October 2021*

Citation:
Wang B, Peng M, Yang L, Li G, Yang J, Yundan C, Zeng X, Wei Q, Han Q, Liu C, Ding K, Peng K and Kang W (2021) Clinical and Immunological Characteristics of Patients With Adenovirus Infection at Different Altitude Areas in Tibet, China. Front. Cell. Infect. Microbiol. 11:739429. doi: 10.3389/fcimb.2021.739429

Background: The severities of human adenovirus (HAdV) infection are diverse in different areas of Tibet, China, where a large altitude span emerges. Serious consequences may be caused by medical staff if the clinical stages and immunological conditions of patients in high-altitude areas are misjudged. However, the clinical symptoms, immunological characteristics, and environmental factors of HAdV infection patients at different altitude areas have not been well described.

Methods: In this retrospective, multicenter cohort study, we analyzed the data of patients who were confirmed HAdV infection by PCR tests in the General Hospital of Tibet Military Command or CDC (the Center for Disease Control and Prevention) of Tibet Military Command from January 1, 2019, to December 31, 2020. Demographic, clinical, laboratory, radiological, and epidemiological data were collected from medical records system and compared among different altitude areas. The inflammatory cytokines as well as the subsets of monocytes and regulatory T cells of patients were also obtained and analyzed in this study.

Results: Six hundred eighty-six patients had been identified by laboratory-confirmed HAdV infection, including the low-altitude group ($n = 62$), medium-altitude group ($n = 206$), high-altitude group ($n = 230$), and ultra-high-altitude group ($n = 188$). Referring to the environmental factors regression analysis, altitude and relative humidity were tightly associated with the number of infected patients ($P < 0.01$). A higher incidence rate of general pneumonia (45.7%) or severe pneumonia (8.0%) occurred in the ultra-high-altitude group ($P < 0.05$). The incubation period, serial interval, course of the disease, and PCR-positive duration were prolonged to various extents compared with the low-altitude group ($P < 0.05$). Different from those in low-altitude areas, the levels of IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, G-CSF, GM-CSF, IFN-γ, IP-10, MCP-1, TNF-α, TNF-β, and VEGF in the plasma of the ultra-high-altitude group were increased ($P < 0.05$), while the proportion of non-classical monocytes and regulatory T cells was decreased ($P < 0.05$).

Conclusions: The findings of this research indicated that patients with HAdV infection in high-altitude areas had severe clinical symptoms and a prolonged course of disease. During clinical works, much more attention should be paid to observe the changes in their immunological conditions. Quarantine of patients in high-altitude areas should be appropriately extended to block virus shedding.

Keywords: altitude, adenovirus, clinic, immunology, epidemiology, character

INTRODUCTION

Since the discovery and isolation of adenovirus in the 1950s, more than 100 serotypes had been identified. There are 52 kinds of human adenoviruses (HAdVs), which could be divided into six subgroups: A, B, C, D, E, and F (Greber and Flatt, 2019). HAdVs are able to infect the human respiratory tract, gastrointestinal tract, urethra, bladder, eye, and liver. The typical symptoms of respiratory tract infection caused by HAdVs are cough, nasal obstruction, and pharyngitis, accompanied by fever, chills, headache, and muscle pain at the same time (Ison and Hirsch, 2019). Sporadic cases and outbreaks of HAdVs have been recorded in both USA and China, with cases totaling in hundreds, especially for children and military recruits. Fortunately, mortality has been estimated at a relatively low level (Gautret et al., 2014). It had been reported that HAdV pneumonia accounted for about 10% of childhood pneumonia, which was mostly caused by HAdV types 3 and 7 (Zou et al., 2021). In adolescence, the mortality of HAdV pneumonia is 8%–10% (Scott et al., 2016; Kujawski et al., 2020). Notably, due to the highly infectious and rapid spreading of HAdVs, military recruits are a special group at high risk during endemic outbreaks (Vento et al., 2011; Hang et al., 2020). In 2006, HAdV-14 caused abrupt epidemics among US military trainees in five states at least: California, Georgia, Illinois, Montana, and Texas (Tate et al., 2009). In 2011, 43 children, attending primary school beside an air force military center, were proved to be infected by HAdV-14 in the Tongwei County of Gansu, China (Huang et al., 2013). Therefore, military hospitals and CDC have always been paying particular attention to the control and prevention of HAdVs for fear of virus shedding.

Studies had reported that the chance of residents suffering from respiratory diseases in high-altitude areas was more likely to be increased. Vargas et al. found that the incidence rate of asthma was higher if the residents lived above 1,500 m (Vargas et al., 2018). Khan et al., through a long-term cohort study of 5,204 children in Pakistan, confirmed that the risk of pneumonia in children was closely related to the altitude they lived (Khan et al., 2009). In addition, during the global outbreak of new coronavirus pneumonia in 2019, higher altitude usually indicates more severe clinical symptoms and worse outcomes in patients (Zeng et al., 2020; Woolcott and Bergman, 2020). Tibet, located in the southwest of China, is impressed by the broad area, complex terrain and vast altitude span. For example, the altitude of Pali County of Shigatse is more than 5,000 m, while Motuo County of Nyingchi is not enough to even 500 m (Ni et al., 2021). However, there have not been reports on the clinical and immunological characteristics of HAdV infection in different areas of Tibet.

Although most HAdV infectious patients have the respiratory disease with mild or asymptomatic clinical manifestations, multisystem illnesses are also reported occasionally (Ison and Hirsch, 2019). Since lung physiology is heavily affected by the atmospheric pressure that decreases with altitudes, it is generally believed that respiratory diseases might be more serious in high-altitude areas (Truog et al., 2020). However, many medical staff still might not have a clear idea about the state of patients in high-altitude areas, without the preparations for the rapid changes of illness or the prolonged course of disease and incubation period. Especially for referral patients from high-altitude areas, medical staff in low-attitude areas are more likely to misjudge their clinical stage, miss the best opportunity for treatment, and even result in serious HAdV outbreaks in the local areas. This study aims to reveal the epidemiological, clinical, laboratory, radiological, and immunological characteristics of patients with HAdV infection as well as the environmental risk factors in different altitude areas, to provide scientific proof for the diagnosis and treatment of HAdV infections in high-altitude areas.

METHODS

Study Design and Participation

This retrospective, multicenter cohort study included four cohorts of patients from the General Hospital of Tibet Military Command in Lhasa, 954 Hospital of Army in Lhoka, 956 Hospital of Army in Nyingchi, and CDC of Tibet Military Command in Lhasa. All of the patients with confirmed HAdV infection who were admitted to the General Hospital, 954 Hospital, and 956 Hospital, as well as who were reported to CDC, were retrospectively enrolled from January 1, 2019, to December 31, 2020. This study was approved by the Research Ethics Commission of the General Hospital and the requirement for informed consent was waived by the Ethics Commission.

Data Collection

Epidemiological data were collected through interviews and field investigation. Environmental data were obtained in CDC of Tibet Military Command. Demographic, clinical, laboratory, and radiologic data were extracted from electronic medical records by a standard data collection form that was a modification of the World Health Organization (WHO)/International Severe Acute Respiratory record form (Zhou et al., 2020). All data above were verified by two physicians and a third researcher adjudicated any differences in interpretations between them.

Epidemiological Data Collection

The number of acute respiratory tract infections in military units of Tibet was reported to the CDC every month since January 1, 2019. Once clusters of cases emerged or suspected individuals were identified, the epidemiology team with members from the CDC of Tibet Military Command would be informed to initiate detailed field investigations and collect specimens including swabs, blood, stools, and urines at the acute phase for further tests.

Information about the date of illness onset, clinical symptoms, laboratory results, and the date of cure were collected through detailed interviews with infected persons, close contacts, and medical staff. Epidemiological data were acquired through interviews and field investigations. If necessary, the investigators interviewed each infected patient to verify the contact history within 2 weeks before the onset of the illness. Information about contact with other people with similar symptoms was also included. All of the data collected during the field survey, including contact history, the timeline of events, and identification of close contacts, had been cross-checked from multiple sources. The data were entered into the central database of the CDC and verified by EpiData Association.

Environmental Data Collection

In a previous study, high altitude was defined as a geographic elevation of ≥1,500 m (Woolcott et al., 2015). Here, altitude was divided into four categories: 0–1,500, 1,500–3,000, 3,000–4,500, and ≥4,500 m. The average altitude of each region in Tibet was obtained from the CDC of Tibet Military Command and verified by Google Earth. All of the study groups were divided based on the altitude of the residence of patients. Besides, the average temperature, relative humidity, and oxygen content of each month in different areas of Tibet were acquired from the local CDC.

Clinical Data Collection

The recent exposure history, clinical symptoms or signs, and results of laboratory tests on admission of inpatients and outpatients were obtained from the hospital electronic medical records. These data of the non-referral patients, if they had, were collected from the local medical systems. Radiologic evaluation including chest X-ray or computed tomography (CT), as well as laboratory testing, was performed according to the clinical needs of patients. Referring to the documents or descriptions in the medical record system, we determined whether the patient had radiologic abnormalities; if image scanning can be carried out, the attending physician of the respiratory department would review and extract the data. Disagreements between the two reviewers were resolved through negotiation with a third reviewer. Laboratory evaluation consisted of whole blood count, blood chemistry analysis, coagulation test, liver and kidney function evaluation, and measurement of electrolyte, C-reactive protein, procalcitonin, lactate dehydrogenase, and creatine kinase. Treatment data consisted of antiviral therapy, antibiotic therapy, use of corticosteroid, nasal cannula, non-invasive ventilation, and invasive mechanical ventilation. Prognosis of the infection could be divided into hospitalization, recovery, and discharge from hospital at the end of observation. All of the data mentioned above were entered into a computerized database and cross-checked in the CDC of Tibet Military Command. If the core data were lost, a request would be sent to the coordinator, who subsequently contacted the attending physician for further clarifications.

Laboratory Procedures

The nasopharyngeal swabs, blood, and fecal specimens during the acute phase of HAdV infection were subsequently sent to the central laboratory of the General Hospital or CDC for etiological detection. Blood samples were centrifuged at 500×*g* and the upper plasma was collected for following nucleic acid extraction experiments. Each 1 g fecal sample was added to 10 ml of 0.9% normal saline. Then, the samples were centrifuged at 500×*g* and the supernatant was collected for the following nucleic acid extracting tests. Nucleic acids of pathogens were extracted from different specimens using a QIAamp Viral RNA Mini Kit (Qiagen) according to the recommended protocol of the manufacturer. A diagnostic kit for HAdVs (Beijing Kinghawk) was applied for HAdV nucleic acid detection referring to the recommended protocol of the manufacturer. Briefly, the presence of HAdV nucleic acids in samples was detected by real-time PCR. The primer and probe sequences of the target were listed as follows: forward primer 5′-GCCACGGTGGGGTT TCTAAACTT-3′, reverse primer 5′-GCCCCAGTGGTCTTAC ATGCACATG-3′, and the probe 5′-FAM-TGCACCAGAC CCGGGCTCAGGTACTCCGA-3′-BHQ. Conditions for amplification were 95°C for 5 min, followed by 40 cycles of 95°C for 10 s and 58°C for 40 s. A cycle threshold value (Ct-value) of each specimen less than 37 was defined as a positive test, while a Ct-value of 40 or more was defined as a negative test. A medium viral load with a Ct-value of 37 to 40 required confirmation by retesting.

Case Definitions

A confirmed case of HAdV infection was defined as that whose respiratory, blood, or fecal specimens were positive for HAdVs by real-time PCR. A new identified case was in contact with or epidemiological association with a confirmed patient within 14 days before the onset of the disease. The definition of pneumonia cases with HAdV infection was based on the definition of community-acquired pneumonia cases recommended by the WHO (Metlay et al., 2019): fever, with or without temperature records; radiographic evidence of pneumonia; low or normal white blood cell count or low lymphocyte count; no remission of symptoms after 3 days of antimicrobial treatment; and etiological evidence. Diagnosis for severe pneumonia with HAdV infection also followed the WHO criteria (Jain et al., 2015): one primary criterion or more than three secondary criteria were met. Primary criteria were mechanical ventilation with tracheal intubation or vasoactive drugs required for septic shock after active fluid resuscitation. Secondary criteria were respiratory rate ≥30 times/min, oxygenation index ≤250 mmHg, multiple lobar infiltrations, disturbance of consciousness or disorientation, blood urea nitrogen ≥7.14 mmol/L, systolic blood pressure <90 mmHg, and requiring active fluid resuscitation.

Cytokine and Chemokine Measurement

The plasma cytokines and chemokines, interleukin (IL)-1β, IL-2, IL-4, IL-6, IL-8 (also known as CXCL8), IL-10, IL-12, IL-13, IL-15, granulocyte colony-stimulating factor (G-CSF, also known as CSF3), granulocyte–macrophage colony-stimulating factor (GM-CSF, also known as CSF2), interferon γ (IFN-γ), induced protein 10 (IP-10, also known as CXCL10), MCP-1 (also known as CCL2), monocyte chemoattractant protein 1 (MIP-1, also known as CCL3), tumor necrosis factor α (TNF-α), tumor necrosis factor β (TNF-β), and vascular endothelial growth factor (VEGF), were measured in the acute phase of illness. According to the instructions of the manufacturer, the inflammatory factors and chemokines of all patients were detected using ELISA kit (R&D System). The optical density (OD) value of the experimental results was read on a microplate reader (Bio-Rad).

Flow Cytometry

The blood samples during the acute phase of HAdV infection were collected and sent to the central laboratory of the General Hospital or CDC for flow cytometry. Total peripheral blood mononuclear cells (PBMCs) were separated from blood samples in the acute phase of illness. Then, PBMCs were stained with conjugated antibodies against the following proteins: CD45-PerCp (2D1), CD4-PE/Cy7 (OKT4), CD14-FITC (63D3), CD16-APC (3G8), CCR2-APC/Cy7 (K036C2), CX3CR1-PE/Cy7 (2A9-1), and CD86-PE (BU63) (all from Biolegend). While the intracellular detection of Foxp3 with Foxp3-PE (295D) (Biolegend) was performed on fixed and permeabilized cells using FOXP3 Fixation/Permeabilization Buffer (Biolegend) according to the instructions of the manufacturer. The lymphocyte and monocyte gates were both based on CD45. Regulatory T cells (Tregs) were defined as $CD4^+$ and $Foxp3^+$ among lymphocytes. The subsets of monocytes were recognized by both CD14 and CD16 (Cros et al., 2010). In addition, the expression of CCR2, CX3CR1, and CD86 was determined in different subsets of monocytes, respectively.

Statistics

The epidemic curve of HAdV and acute respiratory tract infections in Tibet was drawn according to the number of incident cases and month of diagnosis. Sensitivity analysis was also carried out to include the important outbreak events in the epidemic curve. In the infectious population, multilevel mixed-effect Poisson regression analysis was applied to estimate the correlation of altitude, temperature, relative humidity, and oxygen content with the number of patients. The incubation period distribution (the time delay from infection to illness onset), course of disease, and PCR-positive duration were estimated by fitting the log-normal distribution to the data of exposure history and onset date. The serial interval (the delay between illness onset dates in successive cases in a transmission chain) is estimated by fitting the gamma distribution of cluster survey data. Continuous variables and categorical variables were expressed as median (IQR) and *n* (%), respectively. We applied the Kruskal–Wallis test (followed by *post-hoc* analysis with Dunnett's *t*-test with Bonferroni adjustment when appropriate) to compare the continuous variables among different altitudes, while the χ^2 test or Fisher exact test was employed for the comparison of categorical variables. The Kruskal–Wallis test was also used to compare the changes of leukocyte count, lymphocyte count, platelet count, D-dimer, urea nitrogen, and creatinine at different days after disease onset. Dunnett's *t*-test was used to compare plasma inflammatory factors, Tregs, and monocyte subsets between the middle-altitude group, high-altitude group, ultra-high-altitude group, and low-altitude group. Statistical analyses were performed using SPSS Version 22.0. Two-tailed *P*-values less than 0.05 were considered statistically significant unless there are special explanations.

RESULTS

Epidemic Curve and Environmental Factors of HAdV Patients in Tibet

There were 70.63% of acute respiratory tract infections in Tibet that concentrated from September to March of next year (**Supplementary Figure 1**). The epidemic curve of HAdV infection suggested that the number of cases reached a peak in December and January (**Supplementary Figure 1**). However, epidemic curves of HAdV infection and respiratory diseases did not coincide completely due to a small fluctuation of HAdV cases in April and May. Despite gathering in spring and winter, some sporadic HAdV cases also emerged all around the year. The majority of the HAdV outbreaks occurred in Lhasa, accounting for 50% of the total incidents. Field investigations demonstrated that HAdV infections might spread across the areas due to the mobility of patients.

Generally speaking, environmental factors have vital roles in the onset, transmission, and maintenance of HAdV infections (Gautret et al., 2014; Yao et al., 2021). Here, we took the four environmental factors involving the altitude of residence of the patients, temperature, relative humidity, and oxygen content into account (**Figure 1**). Different from the traditional concepts, the temperature and oxygen content were not related to the number of HAdV patients ($P > 0.05$). However, altitude and relative humidity were the environmental factors affecting the number of patients ($P < 0.01$). Furthermore, altitude was positively correlated with the number of patients (**Figure 1A**), while a negative correlation was recognized between relative humidity and the number of patients (**Figure 1C**).

Demographic, Clinical, and Epidemiological Characteristics of HAdV Patients at Different Altitude Areas

Six hundred eighty-six patients had been identified by laboratory-confirmed HAdV infection, consisting of the low-

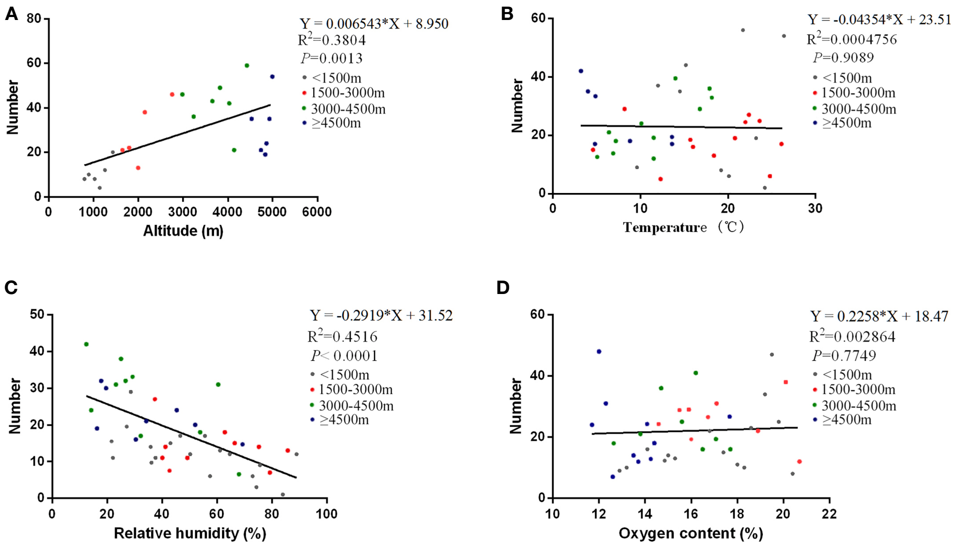


FIGURE 1 | Correlations of the number of patients with various environmental factors. The correlations of the number of HAdV infections with various environmental factors are shown in altitude **(A)**, temperature **(B)**, relative humidity **(C)**, and oxygen content **(D)**, respectively. The curves in the figure are derived from the Poisson regression model. The equation, correlation coefficient, and *P*-value of the curves are marked at the top right of every image.

altitude group (*n* = 62), medium-altitude group (*n* = 206), high-altitude group (*n* = 230), and ultra-high-altitude group (*n* = 188). The demographic and clinical characteristics of the patients are shown in **Table 1**. The median age of all patients was 26 years (interquartile range, 19–32), and 93.4% of the patients were male. There was no significant difference in age and gender among the groups (*P* > 0.05). A total of 2.6% of the patients were medical staff, and 38.9% of them were in the ultra-high-altitude group (**Table 1**).

Current smoking accounted for 66.0% of all patients with HAdV infection (**Table 1**). Compared with the low-altitude group (51.6%), the proportion of current smoking in the high-altitude group (70.4%) and the ultra-high-altitude group (76.6%) was significantly higher (*P* < 0.01). Among all the clinical symptoms, the most common were cough (77.4%), fever (45.5%), fatigue (38.5%), and sputum (30.5%), and these symptoms seemed to be more obvious in the high-altitude group and ultra-high-altitude group (*P* < 0.05). Other relatively rare symptoms, such as shortness of breath (20.0%) and sore throat (18.7%), also increased in both high-altitude areas and ultra-high-altitude areas (*P* < 0.05). In addition, there were significant differences in respiratory rate, heart rate, and mean arterial pressure with the comparison of the high-altitude group and ultra-high-altitude group to the low-altitude group (*P* < 0.05). Notably, the incidence of general pneumonia (45.7%) and severe pneumonia (8.0%) in the ultra-high-altitude group was significantly higher than that in the low-altitude group (*P* < 0.05). Similar to the severity of disease, patients in ultra-high-altitude areas had a higher percentage in receiving antiviral therapy (90.4%), a corticosteroid to restrict inflammation (48.9%), and invasive mechanical ventilation (6.9%) (*P* < 0.05). Although the majority of patients recovered after treatment, patients at ultra-high-altitude areas showed a higher hospitalization rate (5.3%) while a lower recovery rate (79.3%) at the end of observation (*P* < 0.05).

Referring to field investigation of clusters of cases, we found that the epidemiological characteristics of HAdV patients at different altitude areas were far from the same (**Figure 2**). By statistical estimation, the median incubation period of the low-altitude group (3.5 days, 95% CI 2.7–4.3), middle-altitude group (4.4 days, 95% CI 4.2–4.6), high-altitude group (5.2 days, 95% CI 4.9–5.5), and ultra-high-altitude group (6.2 days, 95% CI 5.5–6.8) were described. In addition, 95% of the incubation period distribution in the low-altitude group was 11.3 days (95% CI 8.9–16.1), while that in the ultra-high-altitude group was 14.9 days (95% CI 10.3–21.5) (**Figure 2A**). Similarly, serial interval (**Figure 2B**), course of the disease (**Figure 2C**), and PCR-positive duration (**Figure 2D**) of HAdV patients also increased with altitude. For the duration of disease, the difference between the groups was particularly huge (**Figure 2C**). The median duration of disease in the low-altitude group was 6.1 days (95% CI 5.8–6.3), while that in the ultra-high-altitude group was significantly prolonged (18.3 days, 95% CI 17.2–19.4).

Cytokines and Immunocytes of HAdV Patients at Different Altitude Areas

Initial plasma levels of IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12, IL-13, IL-15, G-CSF, GM-CSF, IFN-γ, IP-10, MCP-1, MIP-1, TNF-α, TNF-β, and VEGF were detected to compare the cytokines and chemokines of HAdV patients at different altitude areas (**Figure 3**). Plasma levels of IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, G-CSF, GM-CSF, IFN-γ, IP-10, MCP-1, TNF-α, TNF-β, and VEGF of the ultra-high-altitude group were significantly higher than those of low-altitude group (*P* < 0.05).

TABLE 1 | Demographics and clinical characteristics of study patients.

	Number (%)					*P*-value		
	Total (*n* = 686)	**Low-altitude group (<1,500 m) (*n* = 62)**	**Medium-altitude group (1,500–3,000 m) (*n* = 206)**	**High-altitude group (3,000–4,500 m) (*n* = 230)**	**Ultra-high-altitude group (≥4,500 m) (*n* = 188)**	P_1	P_2	P_3
Age (IQR)	26 (19–32)	26 (21–30)	25 (19–30)	25 (19–31)	27 (20–33)	0.521	0.534	0.462
Sex								
Male	641 (93.4)	58 (93.5)	192 (93.2)	214 (93.0)	177 (94.1)	>0.99	>0.99	0.68
Female	45 (6.6)	4 (6.5)	14 (6.8)	16 (7.0)	11 (5.9)			
Infected								
Patients	668 (97.4)	61 (98.4)	201 (97.6)	225 (97.8)	181 (96.3)	0.58	0.63	0.37
Medical staff	18 (2.6)	1 (1.6)	5 (2.4)	5 (2.2)	7 (3.7)			
Current smoking	453 (66.0)	32 (51.6)	115 (55.8)	162 (70.4)	145 (76.6)	0.56	0.007	<0.001
Comorbidity								
Hypertension	37 (5.4)	3 (4.8)	8 (3.9)	13 (5.7)	13 (6.9)	0.72	>0.99	0.77
Cardiovascular disease	12 (1.7)	1 (1.6)	3 (1.5)	3 (1.3)	5 (2.7)	>0.99	>0.99	>0.99
Diabetes	18 (2.6)	2 (3.2)	5 (2.4)	6 (2.6)	5 (2.7)	0.66	0.68	>0.99
Cerebrovascular diseases	9 (1.3)	0	2 (1.0)	4 (1.7)	3 (1.6)	>0.99	0.58	>0.99
COPD	32 (4.7)	2 (3.2)	6 (2.9)	9 (3.9)	15 (8.0)	>0.99	>0.99	0.26
Chronic kidney disease	8 (1.2)	0	1 (0.5)	2 (0.9)	5 (2.7)	>0.99	>0.99	>0.99
Chronic liver disease	48 (7.0)	2 (3.2)	9 (4.4)	13 (5.7)	24 (12.8)	>0.99	0.75	0.03
Other	32 (4.7)	5 (8.1)	6 (2.9)	9 (3.9)	12 (6.4)	0.13	0.19	0.77
Signs and symptoms								
Fever	312 (45.5)	22 (38.7)	65 (31.6)	116 (50.4)	109 (58.0)	0.64	0.045	0.003
Fatigue	264 (38.5)	8 (12.9)	34 (16.5)	107 (46.5)	115 (61.2)	0.56	<0.001	<0.001
Cough	531 (77.4)	39 (62.9)	152 (73.8)	178 (77.4)	162 (86.2)	0.11	0.032	<0.001
Sputum production	209 (30.5)	10 (16.1)	28 (13.6)	74 (32.2)	97 (51.6)	0.68	0.037	<0.001
Shortness of breath	137 (20.0)	5 (8.1)	19 (9.2)	54 (23.5)	59 (31.4)	>0.99	0.007	<0.001
Sore throat	128 (18.7)	5 (8.1)	14 (6.8)	48 (20.9)	61 (32.4)	0.78	0.025	<0.001
Abdominal pain	42 (6.1)	3 (4.8)	10 (4.9)	18 (7.8)	11 (5.9)	>0.99	0.083	>0.99
Diarrhea	35 (5.1)	3 (4.8)	8 (3.9)	14 (6.1)	10 (5.3)	0.73	0.277	>0.99
Nausea	38 (5.6)	4 (6.5)	8 (3.9)	15 (6.5)	11 (5.9)	0.48	>0.99	>0.99
Vomiting	26 (3.8)	2 (3.2)	6 (2.9)	10 (4.3)	8 (4.3)	>0.99	>0.99	>0.99
Respiratory rate (IQR)	23 (20–27)	20 (19–22)	22 (20–24)	24 (21–26)	25 (22–29)	0.75	0.043	0.014
Heart rate (IQR), bpm	86 (77–97)	80 (73–88)	85 (79–92)	89 (80–101)	92 (82–103)	0.83	0.004	<0.001
Mean arterial pressure (IQR), mmHg	88 (81–96)	85 (78–91)	86 (81–92)	91 (83–100)	95 (88–104)	>0.99	0.036	<0.001
Pneumonia								
General	214 (31.2)	10 (16.1)	47 (22.8)	71 (30.9)	86 (45.7)	0.29	0.025	<0.001
Severe	29 (4.2)	0	2 (1.0)	12 (5.2)	15 (8.0)	>0.99	0.077	0.026
Treatment								
Antiviral therapy	628 (91.5)	48 (77.4)	192 (93.2)	218 (94.8)	170 (90.4)	0.001	<0.001	0.014
Antibiotic therapy	665 (96.9)	58 (93.5)	202 (98.1)	221 (96.1)	184 (97.9)	0.086	0.49	0.11
Use of corticosteroid	223 (32.5)	8 (12.9)	45 (21.8)	78 (33.9)	92 (48.9)	0.15	<0.001	<0.001
Nasal cannula	378 (55.1)	30 (48.4)	117 (56.8)	120 (52.2)	111 (59.0)	0.25	0.67	0.18
Non-invasive ventilation	172 (25.1)	12 (19.4)	41 (19.9)	62 (27.0)	57 (30.3)	>0.99	0.25	0.10
Invasive mechanical ventilation	26 (3.8)	0	2 (1.0)	11 (4.8)	13 (6.9)	>0.99	0.13	0.042

(Continued)

TABLE 1 | Continued

	Number (%)					P-value		
	Total (*n* = 686)	**Low-altitude group (<1,500 m) (*n* = 62)**	**Medium-altitude group (1,500–3,000 m) (*n* = 206)**	**High-altitude group (3,000–4,500 m) (*n* = 230)**	**Ultra-high-altitude group (≥4,500 m) (*n* = 188)**	P_1	P_2	P_3
Prognosis								
Hospitalization	19 (2.8)	0	1 (0.5)	8 (3.5)	10 (5.3)	0.84	0.16	0.026
Recovery	593 (86.4)	58 (93.5)	190 (92.2)	196 (85.2)	149 (79.3)	–	–	–
Discharge from hospital	74 (10.8)	4 (6.5)	15 (7.3)	26 (11.3)	29 (15.4)	–	–	–

Data are median (IQR), n (%), or n/N (%), where N is the total number of patients with available data. P-values comparing among different altitude areas are from χ^2 test, Fisher's exact test, or Kruskal–Wallis test (followed by post-hoc analysis with Dunnett's t-test with Bonferroni adjustment). P_1 refers to the comparisons between the low-altitude group and medium-altitude group. P_2 refers to the comparisons between the low-altitude group and high-altitude group. P_3 refers to the comparisons between the low-altitude group and ultra-high-altitude group. IQR, interquartile range; COPD, chronic obstructive pulmonary disease.

Meanwhile, a slight rise of IL-2, IL-6, IL-8, IL-10, G-CSF, IP-10, MCP-1, TNF-α, TNF-β, and VEGF was observed in the high-altitude group compared with the low-altitude group ($P < 0.05$). However, there were no differences between the middle-altitude group and the low-altitude group except for IL-6, G-CSF, IP-10, TNF-α, TNF-β, and VEGF ($P < 0.05$).

Monocytes could be divided into classical monocytes ($CD14^{++}$ $CD16^{-}$), intermediate monocytes ($CD14^{++}$ $CD16^{+}$), and non-classical monocytes ($CD14^{+}$ $CD16^{++}$) according to the expression of CD14 and CD16 (**Figure 4A**) (Cros et al., 2010; Shi and Pamer, 2011). Our results indicated that with the elevation of altitude, the proportion of non-classical monocytes and intermediate monocytes gradually decreased ($P < 0.05$), while classical monocytes increased ($P < 0.05$) (**Figure 4B**). Especially, the median proportion of non-classical monocytes in the ultra-high-altitude group was merely 5.5% (interquartile range, 2.8–8.9). In addition, the expression of CCR2, CX3CR1, and CD86 on the surface of non-classical monocytes in the ultra-high-altitude group was significantly lower than that in the low-altitude group ($P < 0.05$) (**Table 2**). Similarly, we calculated the proportion of Tregs ($CD4^{+}$ $Foxp3^{+}$) of patients at different altitude areas (**Figure 5A**). Compared with the low-altitude group, the proportion of Tregs in the high-altitude group and ultra-high-altitude group was reduced significantly ($P < 0.05$) (**Figure 5B**).

Radiologic and Laboratory Findings of HAdV Patients at Different Altitude Areas

Radiology examinations, including chest radiograph and CT scan, were carried out for hospitalized patients and severe patients. Among all of the chest radiographs, 17.2% of HAdV patients had ground-glass opacity and 14.9% had patch shadowing. Compared with the low-altitude group (8.7%), the proportion of patchy shadowing in the high-altitude group (18.3%) and ultra-high-altitude group (20.2%) was significantly higher ($P < 0.05$) (**Supplementary Table 1**). Chest CT was much more sensitive in detecting abnormalities of HAdV infections. The typical manifestation of HAdV infection in chest CT was bilateral multiple lobar and subsegmental areas of consolidation. After a period of treatment, the consolidation would be gradually

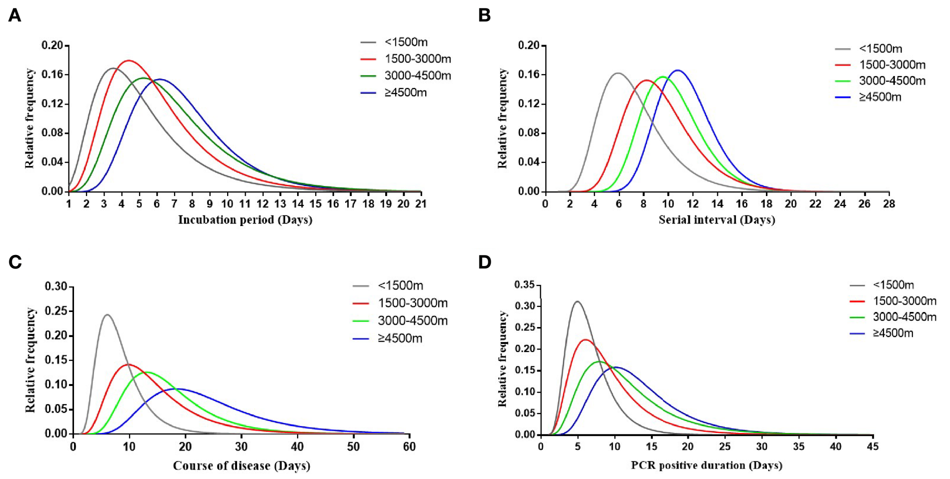


FIGURE 2 | Key time-to-event distributions. The estimated incubation period distribution (i.e., the time from infection to illness onset) is shown in **(A)**. The estimated serial interval distribution (i.e., the time from illness onset in successive cases in a transmission chain) is shown in **(B)**. The estimated course of disease (i.e., the time from illness onset to the disappearance of symptoms) is shown in **(C)**. The estimated PCR-positive duration (i.e., the time from the first PCR positive to PCR negative) is shown in **(D)**.

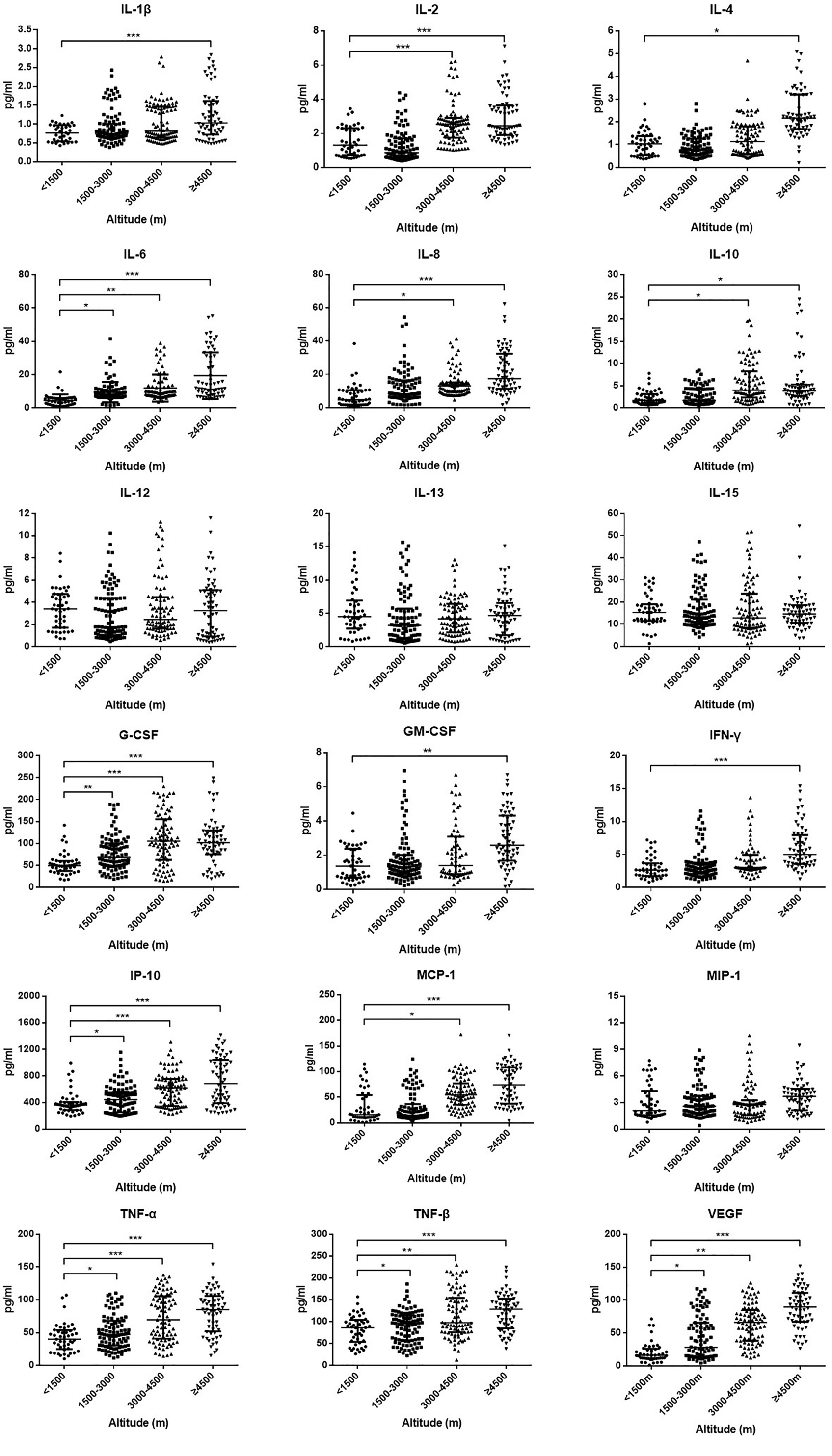


FIGURE 3 | Plasma level of cytokines and chemokines among HAdV infectious patients at different altitudes. Comparison of cytokine and chemokine levels among HAdV infectious patients at different altitudes. The vertical axes differ for each of the mediators to accommodate the extreme variation in the normal distributed range of cytokine concentrations. Data represent median and interquartile range. The Kruskal–Wallis test (followed by *post-hoc* analysis with Dunnett's *t*-test with Bonferroni adjustment) is applied to compare the differences between the low-altitude group (<1,500 m) and the other groups. IL, interleukin; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte–macrophage colony-stimulating factor; IFN-γ, interferon γ; IP-10, induced protein 10; MCP-1, monocyte chemoattractant protein 1; MIP-1, macrophage inflammatory protein 1; TNF-α, tumor necrosis factor α; TNF-β, tumor necrosis factor β; VEGF, vascular endothelial growth factor. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

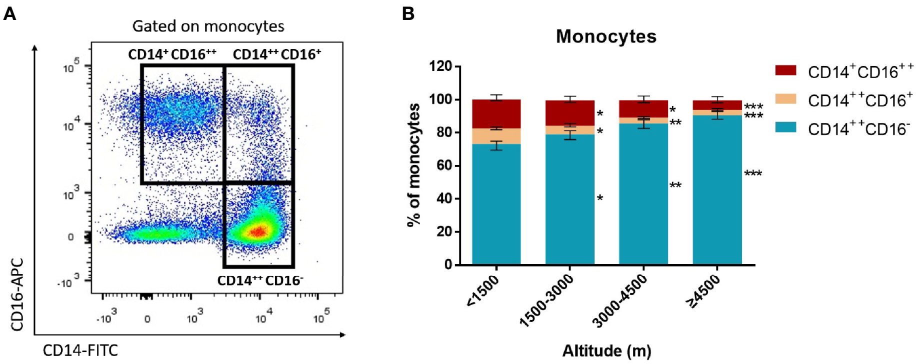


FIGURE 4 | Proportion of monocyte subsets among HAdV infectious patients at different altitudes. Comparison of monocyte subsets among HAdV infectious patients at different altitudes. **(A)** Representative FACS plots of the distributions of monocyte subsets distinguished by CD14 and CD16. **(B)** Comparison of monocyte subset distributions among HAdV infectious patients at different altitudes. The Kruskal–Wallis test (followed by *post-hoc* analysis with Dunnett's *t*-test with Bonferroni adjustment) is applied to compare the differences between the low-altitude group (<1,500 m) and the other groups. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

absorbed and ground-glass opacity would be more apparent (**Supplementary Figure 2**). Different from the results of chest radiographs, the most common abnormality in chest CT was patch shadowing (31.0%), followed by ground-glass opacity (26.1%). Ground-glass opacity (35.6%), patch shadowing (38.8%), vascular enlargement (34.6%), interlobular septal thinning (31.9%), and air bronchogram sign (28.7%) in the ultra-high-altitude group were significantly higher than those in the low-altitude group ($P < 0.05$) (**Supplementary Table 1**).

Despite within the normal range, the white blood cell count of patients was still at a relatively low level (4.9 × 10^9/L, interquartile range, 3.8–5.8) (**Supplementary Table 2**). Compared with the low-altitude group (5.5 × 10^9/L, interquartile range, 4.8–6.0), the white blood cell count level of the high-altitude group (4.5 × 10^9/L, interquartile range, 3.7–5.5) and ultra-high-altitude group (3.9 × 10^9/L, interquartile range, 2.8–5.3) was significantly lower ($P < 0.05$) (**Supplementary Table 2**). Equally, this feature was also suitable for lymphocyte count and platelet count. Furthermore, the lymphocyte counts of the high-altitude group (1.0 × 10^9/L, intermediate range, 0.6–1.3) and the ultra-high-altitude group (0.9 × 10^9/L, intermediate range, 0.5–1.2) were lower beyond the normal range. However, D-dimer, creatine kinase-MB (CK-MB), lactate

TABLE 2 | The receptor expression on the surface of monocyte subsets.

Receptor	Monocytes subsets	Percentage of monocyte subsets (IQR)				*P*-value		
		Low-altitude group (<1,500 m) (*n* = 35)	Medium-altitude group (1,500–3,000 m) (*n* = 109)	High-altitude group (3,000–4,500 m) (*n* = 94)	Ultra-high-altitude group (≥4,500 m) (*n* = 60)	P_1	P_2	P_3
CCR2	$CD14^{++}$ $CD16^{-}$	84.8 (79.2–91.1)	82.6 (76.3–89.4)	81.3 (74.2–88.9)	78.9 (71.5–86.3)	0.78	0.15	0.073
	$CD14^{++}$ $CD16^{+}$	74.1 (69.2–79.8)	72.5 (67.6–78.2)	70.4 (64.3–77.9)	68.9 (63.1–74.8)	>0.99	0.34	0.13
	$CD14^{+}$ $CD16^{++}$	97.4 (95.2–99.7)	94.3 (91.7–97.9)	92.3 (89.4–96.3)	89.2 (86.9–93.4)	0.48	0.11	0.046
CX3CR1	$CD14^{++}$ $CD16^{-}$	63.7 (59.8–68.1)	61.2 (57.3–67.8)	59.6 (54.3–65.9)	56.5 (52.3–61.8)	0.52	0.15	0.084
	$CD14^{++}$ $CD16^{+}$	96.7 (95.2–98.1)	94.4 (89.7–98.2)	91.3 (86.8–95.4)	87.5 (83.2–92.7)	0.35	0.065	0.013
	$CD14^{+}$ $CD16^{++}$	97.5 (96.3–98.8)	92.4 (88.1–96.7)	89.9 (86.5–95.1)	86.1 (82.8–91.2)	0.46	0.17	0.034
CD86	$CD14^{++}$ $CD16^{-}$	17.4 (14.8–21.6)	16.3 (14.2–20.9)	14.7 (11.5–17.9)	11.2 (8.9–16.4)	0.73	0.54	0.18
	$CD14^{++}$ $CD16^{+}$	97.2 (95.7–98.4)	95.8 (93.3–97.5)	93.9 (91.8–96.2)	90.1 (86.5–94.8)	0.87	0.25	0.11
	$CD14^{+}$ $CD16^{++}$	81.5 (77.9–86.3)	78.2 (73.7–84.5)	76.8 (70.2–82.9)	73.3 (68.2–78.7)	0.29	0.045	<0.001

Data are median (IQR), n (%), or n/N (%), where N is the total number of patients with available data. P-values comparing among different altitudes are from χ^2 *test, Fisher's exact test, or Kruskal–Wallis test (followed by post-hoc analysis with Dunnett's t-test with Bonferroni adjustment).* P_1 *refers to the comparisons between the low-altitude group and medium-altitude group.* P_2 *refers to the comparisons between the low-altitude group and high-altitude group.* P_3 *refers to the comparisons between the low-altitude group and ultra-high-altitude group. IQR, interquartile range; CCR2, chemokine C-C receptor 2; CX3CR1, chemokine C-X3-C receptor 1.*

-, negative; +, positive; ++, strong positive. The different monocyte subsets are shown in ***Figure 4A****.*

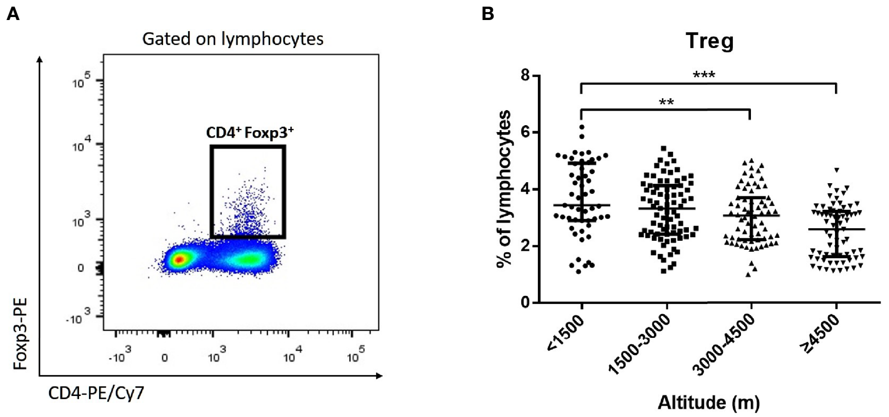


FIGURE 5 | Proportion of Tregs among HAdV infectious patients at different altitudes. Comparison of Tregs among HAdV infectious patients at different altitudes. **(A)** Representative FACS plots of the Tregs distinguished by CD4 and Foxp3. **(B)** Comparison of Tregs among HAdV infectious patients at different altitudes. The Kruskal–Wallis test (followed by *post-hoc* analysis with Dunnett's *t*-test with Bonferroni adjustment) is applied to compare the differences between the low-altitude group (<1,500 m) and the other groups. $^{**}P < 0.01$, $^{***}P < 0.001$.

dehydrogenase (LDH), alanine aminotransferase (ALT), aspartate aminotransferase (AST), and blood urea nitrogen (BUN) were significantly increased in the high-altitude group and the ultra-high-altitude group ($P < 0.05$) (**Supplementary Table 2**).

Subsequently, major laboratory markers were tracked from illness onset (**Supplementary Figure 3**). On the whole, the levels of white blood cell count, lymphocyte count, and platelet count were significantly decreased with the elevation of altitude, while D-dimer, BUN, and creatinine were increased. Although the white blood cell count and lymphocyte count in the low-altitude group reached the bottom at 6 days after disease onset, the decline was more likely to be extended in other groups. Especially, the decline in the ultra-high-altitude group lasted until 18 days and followed by slow recovery. BUN and creatinine levels in the low-altitude group almost did not change during the observation. However, visible variations were obtained in both the middle-altitude group and the high-altitude group. Notably, a rapid rise of BUN and creatinine in the ultra-high-altitude group was continued as long as 18 days after disease onset.

DISCUSSION

HAdVs, a common cause of acute respiratory diseases, is a double-stranded linear DNA virus with a diameter of 70–90 nm. Life-threatening HAdV pneumonia has been recorded in military employees, AIDS patients, and transplant recipients (Chen et al., 2020). Due to the innate characteristics of DNA, HAdVs are strongly resistant to general environmental changes and are difficult to inactivate. In addition, HAdVs are highly infectious transmitting directly from person to person and indirectly from the surrounding environment to a person (Bremner et al., 2009). A large number of epidemic episodes appeared in the community, demonstrating that proximity was the key factor for outbreaks (Bremner et al., 2009; Greber and Flatt, 2019). Our findings showed that HAdV epidemics might be not as optimistic as we thought in Tibet. During the 2 years of observation, six relatively large-scale HAdV outbreaks were recorded. However, we must emphasize that the HAdV epidemics are able to be effectively controlled and prevented if proper measures are carried out. After outbreaks, under the guidance of experts from the CDC and the General Hospital, the administrative departments quickly took vigorous and multifaceted measures, such as traffic restrictions, cancellation of gatherings, centralized quarantine, strengthening disinfection, and universal symptom survey. Confirmed cases, suspected cases, and close contacts were identified for quarantine or medical observation.

Environmental factors play an important role during the transmission of respiratory viruses. Our results indicated a positive correlation between the number of HAdV infections and altitude. HAdV susceptibility of residents in Tibet was described previously even though similar mutation patterns and closer genetic evolutionary distance were confirmed in the sequences of HAdVs isolated from Lhasa (3,650 m) and Chengdu (520 m) (Wang et al., 2017), which was partly attributed to the dysfunctions of the immune system at high altitude (Rijssenbeek-Nouwens et al., 2012; Boonpiyathad et al., 2020). Humidity was a vital environmental factor related to seasonal variation, especially in plateau climate. Our study found that the number of HAdV infections was negatively correlated with relative humidity, which was consistent with previous studies (Xu et al., 2021). Interestingly, the number of patients had nothing to do with temperature or oxygen content. It is traditionally believed that people were more likely to be infected in winter and spring when the ambient temperature and oxygen content were extremely low (Cao et al., 2014). However, our findings seemed to be discordant with the concepts. First of all, temperature of the plateau climate that consists of the rainy season and the dry season is relatively low all around the year. Therefore, the influence of temperature may be partly comprised of climatic characteristics. Secondly, although oxygen content is lower with the elevation of altitude, plants seem to provide additional oxygen to partly improve the anoxic situation. Thus, oxygen content and altitude have opposite effects on the number of patients. Finally, due to the small number of patients enrolled, statistical errors cannot be neglected.

Because the relationship of altitude and relative humidity with the number of patients has been revealed in our study, reducing the altitude and increasing the humidity may provide effective measures to relieve the symptoms and promote recovery in the process of clinical treatment. Especially for severe patients in high-altitude areas, transferring to a lower altitude may accelerate the cure rate of patients. If patients are not able to be transferred immediately, increasing the relative humidity appropriately in the dormitory and ward, such as the use of a humidifier, seems to help treatment and avoid virus spreading (Zhou et al., 2020). However, the effect of these measures on patients still needs further clinical trials to confirm in the future.

Field investigation found that some HAdV patients, despite only a few, still caused virus spread after discharge or quarantine. Due to the disorder and abnormality of the immune system at high altitude, the course of the disease might be longer than that at low altitude, which was also consistent with previous studies (Murray, 2014; Brakema et al., 2019). On the other hand, a prolonged course of infection might be assumed to result in the changes of the incubation period, serial interval, and PCR-positive duration. Therefore, especially in high-altitude areas, only those who have been confirmed both clinical remission and negative PCR tests are able to be released from the quarantine. However, the real reasons for this phenomenon have not been clearly explained. Further research should be carried out to ensure whether the alteration for natural characteristics of HAdVs or the dysfunctions of the immune system matters in the process of the epidemiological changes along with the altitude.

Similarities of clinical characteristics after HAdV infection at different altitude areas have been noted. In this cohort, most patients presented with fever, dryness, sputum, and fatigue. The proportion of current smoking in the high-altitude group and the ultra-high-altitude group was significantly higher. This situation may indicate that smoking is more likely to cause HAdV infections in high-altitude areas. Previous studies showed that a greater proportion of smoking in high-altitude areas may be partly attributed to social aspects of low education attainment, poverty, and lack of health insurance coverage (Mercado et al., 2021; Reitsma et al., 2021). In the chest CT scans, the main manifestations were patchy shadow and bilateral ground-glass shadow. Laboratory results showed that white blood cell count and lymphocyte of patients at high-altitude areas decreased more significantly. Medical staff should be alert enough to the possibility of co-infection with other microbes for HAdV infections. Overall, the higher the altitude is, the worse the clinical symptoms and laboratory results are. However, we also found that most of the patients (54%) were asymptomatic, which might contribute greatly to the virus epidemics around the camp and hospital. It is strongly recommended that precautions against airborne transmission should be taken, such as fit-tested N95 masks and other personal protective equipment. To prevent the further spread of the disease, the fever and respiratory symptoms of the infected patients should be closely monitored. Once the diagnosis is suspected, the respiratory tract specimens should be tested immediately. Both pathogen detection and exposure history should be taken into consideration for the identification of asymptomatic infection.

The analysis of cytokines and immunocytes in patients with acute respiratory infectious diseases has always been the research hot spot. Early studies had shown that the increase of serum proinflammatory cytokines (such as IL-1β, IL-6, IL-12, IFN-γ, IP-10, and MCP-1) was associated with extensive lung injury in SARS (Wong et al., 2004). SARS-CoV-2 infectious patients also had the promotion of IL-1β, IFN-γ, IP-10, and MCP-1 (Fajgenbaum and June, 2020). Our study found that a variety of cytokines including IL-1β, IL-2, G-CSF, GM-CSF, IFN-γ, IP-10, MCP-1, and TNF-α in the serum of patients at the highland were significantly increased, which might lead to the activated T helper cell-1 (Th1) response. Therefore, cytokine storms were more likely to occur at higher altitude areas. At the same time, the proportion of non-classical monocytes and Tregs, as well as the chemokine receptors expressed on the monocytes, was appropriately decreased in HAdV patients at high altitude. Previous studies had demonstrated that non-classical monocytes and Tregs played a critical role in antivirus and the maintenance of immune balance (Munoz-Rojas and Mathis, 2021; Robinson et al., 2021). Serious conditions and a prolonged course in the high-altitude area might be partly attributed to the decline of these two subsets. Non-classical monocytes play a great role in both antiviral effects. A previous study demonstrated that non-classical monocytes promoted neutrophil adhesion at the endothelial interface *via* the secretion of TNF-α and strengthened lymphocytes to produce a specific antibody for the virus (Chimen et al., 2017). In spite of only a few studies toward intermediate monocytes, this subset has the highest capacity to present antigen to T cells and induce the highest IL-10 and IFN-γ production when the body is infected by pathogens (Zawada et al., 2011; Hussen et al., 2020). Here, the decrease of non-classical monocytes and intermediate monocytes in high-altitude areas may result in a susceptibility to HAdVs. Besides, patients in high-altitude areas showing a higher possibility to suffer from severe pneumonia may be partly attributed to the decrease of these two subsets. However, further studies are needed to describe the pathogenesis and mechanisms of immune disorders. A biopsy study will be the key to understand this phenomenon.

Our study still has some limitations. First of all, HAdVs were not typed when PCR tests were performed on the respiratory specimens. Therefore, we had no idea about the type of HAdVs in our cohort. Considering the various clinical symptoms caused by different types of HAdVs, non-typing might be one of the most important defects in this study. Moreover, considering the limitations of the diagnostic kit, the exact viral load of HAdVs may be unavailable in the current study. Secondly, most patients in our study were primary men age 20–30. We must admit the selection bias that existed in our study. Thirdly, with the limited number of cases, it was inevitable to avoid certain statistical errors for the small number of groups. Larger cohort data would provide more meaningful results to further determine the clinical symptoms, environmental factors, and epidemiological characteristics of patients at different altitude areas. At the same time, the outcome of the statistical test and the *P*-value should be carefully explained. Non-significant *P*-value did not necessarily exclude the differences among groups. Fourth, the epidemiological resources and laboratory test records of some cases were not able to be extracted completely. Some cases were diagnosed in the local outpatient settings, where the information was just briefly documented and incomplete laboratory tests were carried out. Compared with the integrated documents, some of these local outpatient data are missing although great efforts have been made to contact the attending physicians at that time. Last but not least, no doubt, we had missed some asymptomatic or mild patients. The majority of samples were nasopharyngeal swabs, despite few blood and fecal specimens. Blood and fecal specimens were only applied when the oropharyngeal region was injured or the patient cannot tolerate the swab collections. However, we must admit that blood and fecal specimens may result in the missing of low viral load patients because of the relatively low detection rates of these two specimens (Lam et al., 2021). Therefore, our cohort is more likely to represent typical HAdV cases.

CONCLUSIONS

Here, we provide an initial assessment of clinical and epidemiological characteristics of HAdV patients at different altitude areas. Unexpectedly, the number of infected patients was correlated with altitudes and relative humidity rather than temperature and oxygen content. Thus, patients are more likely to benefit from the implementation of transference to a lower altitude or more humid areas. Secondly, the epidemiological characteristics of patients in high-altitude areas are varied from those in low altitude. Patients at high-altitude areas may need longer treatment and quarantine to guarantee a cure and avoid virus shedding. Thirdly, there are significant differences in clinical symptoms and laboratory and imaging findings among HAdV patients in Tibet, where a vast altitude span emerges. Diagnosis of this disease seemed to be complicated by the diversity in symptoms, radiological results, and severity of diseases. During the clinical works, much more attention should be paid to observe changes in the conditions of patients to help them recover, especially for those in high-altitude areas.

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 the Research Ethics Commission of the General Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

BW and WK were responsible for the study concept, designed the study, and took responsibility for the integrity of the data and the accuracy of the data analysis. BW, MP, and WK drafted the paper. BW, LY, GL, CY, XZ, QW, QH, and JY did the analysis and gave final approval for the version to be published. BW, LY, GL, CY, CL, KD, KP, and JY collected the data. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grants from the Major Program of National Natural Science Foundation of China (2017ZX10202101-004-005) and the Major Program of National Natural Science Foundation of China (2017ZX10202101-003-007).

ACKNOWLEDGMENTS

We are grateful to the many frontline medical staff for their dedication in the face of this infectious disease, despite the potential threat to their own lives and the lives of their families.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.739429/full#supplementary-material

REFERENCES

Boonpiyathad, T., Capova, G., Duchna, H. W., Croxford, A. L., Farine, H., Dreher, A., et al. (2020). Impact of High-Altitude Therapy on Type-2 Immune Responses in Asthma Patients. *Allergy* 75 (1), 84–94. doi: 10.1111/all.13967

Brakema, E. A., Tabyshova, A., Kasteleyn, M. J., Molendijk, E., van der Kleij, R., van Boven, J., et al. (2019). High COPD Prevalence at High Altitude: Does Household Air Pollution Play a Role? *Eur. Respir. J.* 53 (2). doi: 10.1183/13993003.01193-2018

Bremner, K. H., Scherer, J., Yi, J., Vershinin, M., Gross, S. P., and Vallee, R. B. (2009). Adenovirus Transport *via* Direct Interaction of Cytoplasmic Dynein With the Viral Capsid Hexon Subunit. *Cell Host Microbe* 6 (6), 523–535. doi: 10.1016/j.chom.2009.11.006

Cao, B., Huang, G. H., Pu, Z. H., Qu, J. X., Yu, X. M., Zhu, Z., et al. (2014). Emergence of Community-Acquired Adenovirus Type 55 as a Cause of Community-Onset Pneumonia. *Chest* 145 (1), 79–86. doi: 10.1378/chest.13-1186

Chen, Q., Liu, J., Liang, W., Chen, Y., Dou, M., Liu, Z., et al. (2020). Clinical Features, Replication Competence, and Innate Immune Responses of

Adenovirus Type 7 Infection in Humans. *J. Infect. Dis.* 223 (8), 1390–1399. doi: 10.1093/infdis/jiaa524

Chimen, M., Yates, C. M., Mcgettrick, H. M., Ward, L. S., Harrison, M. J., Apta, B., et al. (2017). Monocyte Subsets Coregulate Inflammatory Responses by Integrated Signaling Through TNF and IL-6 at the Endothelial Cell Interface. *J. Immunol.* 198 (7), 2834–2843. doi: 10.4049/jimmunol.1601281

Cros, J., Cagnard, N., Woollard, K., Patey, N., Zhang, S. Y., Senechal, B., et al. (2010). Human CD14dim Monocytes Patrol and Sense Nucleic Acids and Viruses *via* TLR7 and TLR8 Receptors. *Immunity* 33 (3), 375–386. doi: 10.1016 /j.immuni.2010.08.012

Fajgenbaum, D. C., and June, C. H. (2020). Cytokine Storm. *N Engl. J. Med.* 383 (23), 2255–2273. doi: 10.1056/NEJMra2026131

Gautret, P., Gray, G. C., Charrel, R. N., Odezulu, N. G., Al-Tawfiq, J. A., Zumla, A., et al. (2014). Emerging Viral Respiratory Tract Infections–Environmental Risk Factors and Transmission. *Lancet Infect. Dis.* 14 (11), 1113–1122. doi: 10.1016/ S1473-3099(14)70831-X

Greber, U. F., and Flatt, J. W. (2019). Adenovirus Entry: From Infection to Immunity. *Annu. Rev. Virol.* 6 (1), 177–197. doi: 10.1146/annurev-virology-092818-015550

Hang, J., Kajon, A. E., Graf, P., Berry, I. M., Yang, Y., Sanborn, M. A., et al. (2020). Human Adenovirus Type 55 Distribution, Regional Persistence, and Genetic Variability. *Emerg. Infect. Dis.* 26 (7), 1497–1505. doi: 10.3201/eid2607.191707

Huang, G., Yu, D., Zhu, Z., Zhao, H., Wang, P., Gray, G. C., et al. (2013). Outbreak of Febrile Respiratory Illness Associated With Human Adenovirus Type 14p1 in Gansu Province, China. *Influenza Other Respir. Viruses* 7 (6), 1048–1054. doi: 10.1111/irv.12118

Hussen, J., Shawaf, T., Al-Mubarak, A., Al, H. N., Almathen, F., and Schuberth, H. J. (2020). Dromedary Camel CD14(high) MHCII(high) Monocytes Display Inflammatory Properties and are Reduced in Newborn Camel Calves. *BMC Vet. Res.* 16 (1), 62. doi: 10.1186/s12917-020-02285-8

Ison, M. G., and Hirsch, H. H. (2019). Community-Acquired Respiratory Viruses in Transplant Patients: Diversity, Impact, Unmet Clinical Needs. *Clin. Microbiol. Rev.* 32 (4), 1–33. doi: 10.1128/CMR.00042-19

Jain, S., Self, W. H., Wunderink, R. G., Fakhran, S., Balk, R., Bramley, A. M., et al. (2015). Community-Acquired Pneumonia Requiring Hospitalization Among U. S. Adults. *N. Engl. J. Med.* 373 (5), 415–427. doi: 10.1056/NEJMoa1500245

Khan, A. J., Hussain, H., Omer, S. B., Chaudry, S., Ali, S., Khan, A., et al. (2009). High Incidence of Childhood Pneumonia at High Altitudes in Pakistan: A Longitudinal Cohort Study. *Bull. World Health Organ* 87 (3), 193–199. doi: 10.2471/BLT.07.048264

Kujawski, S. A., Lu, X., Schneider, E., Blythe, D., Boktor, S., Farrehi, J., et al. (2020). Outbreaks of Adenovirus-Associated Respiratory Illness on Five College Campuses in the United States. *Clin. Infect. Dis.* 72 (11), 1992–1999. doi: 10.1093/cid/ciaa465

Lam, C., Gray, K., Gall, M., Sadsad, R., Arnott, A., Johnson-Mackinnon, J., et al. (2021). Sars-CoV-2 Genome Sequencing Methods Differ In Their Ability To Detect Variants From Low Viral Load Samples. *J. Clin. Microbiol.* 0, 1021–1046. doi: 10.1128/JCM.01046-21

Mercado, C. I., Mckeever, B. K., Gregg, E. W., Ali, M. K., Saydah, S. H., and Imperatore, G. (2021). Differences in U.S. Rural-Urban Trends in Diabetes ABCS, 1999-2018. *Diabetes Care* 44 (8), 1766–1773. doi: 10.2337/dc20-0097

Metlay, J. P., Waterer, G. W., Long, A. C., Anzueto, A., Brozek, J., Crothers, K., et al. (2019). Diagnosis and Treatment of Adults With Community-Acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am. J. Respir. Crit. Care Med.* 200 (7), e45–e67. doi: 10.1164/rccm.201908-1581ST

Munoz-Rojas, A. R., and Mathis, D. (2021). Tissue Regulatory T Cells: Regulatory Chameleons. *Nat. Rev. Immunol.* 21 (9), 597–611. doi: 10.1038/s41577-021-00519-w

Murray, J. F. (2014). Tuberculosis and High Altitude. Worth a Try in Extensively Drug-Resistant Tuberculosis? *Am. J. Respir. Crit. Care Med.* 189 (4), 390–393. doi: 10.1164/rccm.201311-2043OE

Ni, J., Wu, T., Zhu, X., Wu, X., Pang, Q., Zou, D., et al. (2021). Risk Assessment of Potential Thaw Settlement Hazard in the Permafrost Regions of Qinghai-Tibet Plateau. *Sci. Total Environ.* 776, 145855. doi: 10.1016/j.scitotenv.2021.145855

Reitsma, M. B., Flor, L. S., Mullany, E. C., Gupta, V., Hay, S. I., and Gakidou, E. (2021). Spatial, Temporal, and Demographic Patterns in Prevalence of Smoking Tobacco Use and Initiation Among Young People in 204 Countries and Territories, 1990-2019. *Lancet Public Health* 6 (7), e472–e481. doi: 10.1016/S2468-2667(21)00102-X

Rijssenbeek-Nouwens, L. H., Fieten, K. B., Bron, A. O., Hashimoto, S., Bel, E. H., and Weersink, E. J. (2012). High-Altitude Treatment in Atopic and Nonatopic Patients With Severe Asthma. *Eur. Respir. J.* 40 (6), 1374–1380. doi: 10.1183/ 09031936.00195211

Robinson, A., Han, C. Z., Glass, C. K., and Pollard, J. W. (2021). Monocyte Regulation in Homeostasis and Malignancy. *Trends Immunol.* 42 (2), 104–119. doi: 10.1016/j.it.2020.12.001

Scott, M. K., Chommanard, C., Lu, X., Appelgate, D., Grenz, L., Schneider, E., et al. (2016). Human Adenovirus Associated With Severe Respiratory Infection, Oregon, USA, 2013-2014. *Emerg. Infect. Dis.* 22 (6), 1044–1051. doi: 10.3201/ eid2206.151898

Shi, C., and Pamer, E. G. (2011). Monocyte Recruitment During Infection and Inflammation. *Nat. Rev. Immunol.* 11 (11), 762–774. doi: 10.1038/nri3070

Tate, J. E., Bunning, M. L., Lott, L., Lu, X., Su, J., Metzgar, D., et al. (2009). Outbreak of Severe Respiratory Disease Associated With Emergent Human Adenovirus Serotype 14 at a US Air Force Training Facility in 2007. *J. Infect. Dis.* 199 (10), 1419–1426. doi: 10.1086/598520

Truog, R. D., Mitchell, C., and Daley, G. Q. (2020). The Toughest Triage - Allocating Ventilators in a Pandemic. *N. Engl. J. Med.* 382 (21), 1973–1975. doi: 10.1056/NEJMp2005689

Vargas, M. H., Becerril-Angeles, M., Medina-Reyes, I. S., and Rascon-Pacheco, R. A. (2018). Altitude Above 1500m Is a Major Determinant of Asthma Incidence. An Ecological Study. *Respir. Med.* 135, 1–7. doi: 10.1016/j.rmed. 2017.12.010

Vento, T. J., Prakash, V., Murray, C. K., Brosch, L. C., Tchandja, J. B., Cogburn, C., et al. (2011). Pneumonia in Military Trainees: A Comparison Study Based on Adenovirus Serotype 14 Infection. *J. Infect. Dis.* 203 (10), 1388–1395. doi: 10.1093/infdis/jir040

Wang, W., Liu, Y., Zhou, Y., Gu, L., Zhang, L., Zhang, X., et al. (2017). Whole-Genome Analyses of Human Adenovirus Type 55 Emerged in Tibet, Sichuan and Yunnan in China, in 2016. *PloS One* 12 (12), e189625. doi: 10.1371/ journal.pone.0189625

Wong, C. K., Lam, C. W., Wu, A. K., Ip, W. K., Lee, N. L., Chan, I. H., et al. (2004). Plasma Inflammatory Cytokines and Chemokines in Severe Acute Respiratory Syndrome. *Clin. Exp. Immunol.* 136 (1), 95–103. doi: 10.1111/j.1365-2249.2004.02415.x

Woolcott, O. O., Ader, M., and Bergman, R. N. (2015). Glucose Homeostasis During Short-Term and Prolonged Exposure to High Altitudes. *Endocr. Rev.* 36 (2), 149–173. doi: 10.1210/er.2014-1063

Woolcott, O. O., and Bergman, R. N. (2020). Mortality Attributed to COVID-19 in High-Altitude Populations. *High Alt. Med. Biol.* 21 (4), 409–416. doi: 10.1089/ ham.2020.0098

Xu, B., Wang, J., Li, Z., Xu, C., Liao, Y., Hu, M., et al. (2021). Seasonal Association Between Viral Causes of Hospitalised Acute Lower Respiratory Infections and Meteorological Factors in China: A Retrospective Study. *Lancet Planet Health* 5 (3), e154–e163. doi: 10.1016/S2542-5196(20)30297-7

Yao, L., Zhu, W., Shi, J., Xu, T., Qu, G., Zhou, W., et al. (2021). Detection of Coronavirus in Environmental Surveillance and Risk Monitoring for Pandemic Control. *Chem. Soc. Rev* 50 (6), 3656–3676. doi: 10.1039/d0cs00595a

Zawada, A. M., Rogacev, K. S., Rotter, B., Winter, P., Marell, R. R., Fliser, D., et al. (2011). SuperSAGE Evidence for CD14++CD16+ Monocytes as a Third Monocyte Subset. *Blood* 118 (12), e50–e61. doi: 10.1182/blood-2011-01-326827

Zeng, J., Peng, S., Lei, Y., Huang, J., Guo, Y., Zhang, X., et al. (2020). Clinical and Imaging Features of COVID-19 Patients: Analysis of Data From High-Altitude Areas. *J. Infect.* 80 (6), e34–e36. doi: 10.1016/j.jinf.2020.03.026

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical Course and Risk Factors for Mortality of Adult Inpatients With COVID-19 in Wuhan, China: A Retrospective Cohort Study. *Lancet* 395 (10229), 1054–1062. doi: 10.1016/S0140-6736(20)30566-3

Zou, L., Yi, L., Yu, J., Song, Y., Liang, L., Guo, Q., et al. (2021). Adenovirus Infection in Children Hospitalized With Pneumonia in Guangzhou, China. *Influenza Other Respir. Viruses* 15 (1), 27–33. doi: 10.1111/irv.12782

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: 28 October 2021
doi: 10.3389/fmicb.2021.749149

Genomic Epidemiology Reveals Multiple Introductions of Severe Acute Respiratory Syndrome Coronavirus 2 in Niigata City, Japan, Between February and May 2020

Keita Wagatsuma[1]*†, Ryosuke Sato[2†], Satoru Yamazaki[2], Masako Iwaya[2], Yoshiki Takahashi[2], Akiko Nojima[2], Mitsuru Oseki[3], Takashi Abe[4], Wint Wint Phyu[1], Tsutomu Tamura[5], Tsuyoshi Sekizuka[6], Makoto Kuroda[6], Haruki H. Matsumoto[7] and Reiko Saito[1]

[1] Division of International Health (Public Health), Graduate School of Medical and Dental Sciences, Niigata University, Niigata, Japan, [2] Niigata City Public Health and Sanitation Center, Niigata, Japan, [3] Division of Health Science, Niigata City Institute of Public Health and Environment, Niigata, Japan, [4] Division of Bioinformatics, Graduate School of Science and Technology, Niigata University, Niigata, Japan, [5] Virology Section, Niigata Prefectural Institute of Public Health and Environmental Science, Niigata, Japan, [6] Pathogen Genomics Center, National Institute of Infectious Diseases, Tokyo, Japan, [7] Division of Health and Welfare, Niigata Prefectural Government Office, Niigata, Japan

The coronavirus disease 2019 (COVID-19) has caused a serious disease burden and poses a tremendous public health challenge worldwide. Here, we report a comprehensive epidemiological and genomic analysis of SARS-CoV-2 from 63 patients in Niigata City, a medium-sized Japanese city, during the early phase of the pandemic, between February and May 2020. Among the 63 patients, 32 (51%) were female, with a mean (±standard deviation) age of 47.9 ± 22.3 years. Fever (65%, 41/63), malaise (51%, 32/63), and cough (35%, 22/63) were the most common clinical symptoms. The median C_t value after the onset of symptoms lowered within 9 days at 20.9 cycles (interquartile range, 17–26 cycles), but after 10 days, the median C_t value exceeded 30 cycles ($p < 0.001$). Of the 63 cases, 27 were distributed in the first epidemic wave and 33 in the second, and between the two waves, three cases from abroad were identified. The first wave was epidemiologically characterized by a single cluster related to indoor sports activity spread in closed settings, which included mixing indoors with families, relatives, and colleagues. The second wave showed more epidemiologically diversified events, with most index cases not related to each other. Almost all secondary cases were infected by droplets or aerosols from closed indoor settings, but at least two cases in the first wave were suspected to be contact infections. Results of the genomic analysis identified two possible clusters in Niigata City, the first of which was attributed to clade S (19B by Nextstrain clade) with a monophyletic group derived from the Wuhan prototype strain but that of the second wave was polyphyletic suggesting multiple introductions, and the clade was changed to GR (20B), which mainly spread in Europe in early 2020. These findings depict characteristics of SARS-CoV-2 transmission

OPEN ACCESS

Edited by:
Hassan Zaraket,
American University of Beirut,
Lebanon

Reviewed by:
Zigui Chen,
The Chinese University of Hong Kong,
China
Francesco Cerutti,
Laboratori di Microbiologia e Virologia,
Ospedale Amedeo di Savoia, Italy

****Correspondence:***
Keita Wagatsuma
waga@med.niigata-u.ac.jp

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

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

Received: *29 July 2021*
Accepted: *04 October 2021*
Published: *28 October 2021*

Citation:
Wagatsuma K, Sato R, Yamazaki S, Iwaya M, Takahashi Y, Nojima A, Oseki M, Abe T, Phyu WW, Tamura T, Sekizuka T, Kuroda M, Matsumoto HH and Saito R (2021) Genomic Epidemiology Reveals Multiple Introductions of Severe Acute Respiratory Syndrome Coronavirus 2 in Niigata City, Japan, Between February and May 2020. Front. Microbiol. 12:749149. doi: 10.3389/fmicb.2021.749149

in the early stages in local community settings during February to May 2020 in Japan, and this integrated approach of epidemiological and genomic analysis may provide valuable information for public health policy decision-making for successful containment of chains of infection.

Keywords: COVID-19, SARS-CoV-2, Japan, genomic epidemiology, active epidemiological surveillance, droplet, airborne, contact transmission

INTRODUCTION

Since the first report of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Wuhan, China at the end of 2019, the coronavirus disease 2019 (COVID-19) pandemic has become an unprecedented threat to public health on a global scale (Heymann and Shindo, 2020; Wang et al., 2020). As of early June 2021, more than 170 million cases have been reported worldwide, with over 3.6 million fatalities, resulting in a significant disease burden (World Health Organization, 2021).

In Japan, the COVID-19 epidemic began on January 16, 2020, with the first confirmed case being a returnee from Wuhan, China (Furuse et al., 2020a; Muto et al., 2020). Subsequently, the number of cases increased during January and April, reaching the first peak of 720 new cases in a day on April 11. The Japanese government declared a nationwide state of emergency between April 16 and May 25, 2020, in Japan.

In the early stages of Japan's response to the COVID-19 outbreak, the Japanese government assigned COVID-19 as a designated infectious disease within the framework of the Infectious Diseases Control Law, which requires physicians to immediately report diagnosed COVID-19 cases to the public health center in their jurisdiction, on February 01, 2020 (Ministry of Health Labour and Welfare, 2020a). The recognition of the disease as a designated infectious disease enabled the respective administrations to conduct active epidemiological surveillance and make recommendations and measures for mandatory hospitalization for quarantine purposes. Subsequently, based on the framework of the Infectious Disease Control Law, the Japanese government began implementing a cluster-based approach based on active epidemiological surveillance as part of the response to control the spread of the virus (Oshitani, 2020; Imamura et al., 2021). The aim of this strategy is to conduct intensive tracing of super-spreading events (i.e., clusters) of COVID-19 cases to investigate the activities of multiple infected individuals and to identify common sources of infection. In particular, it aims to identify cluster-borne occasions and to control transmission by restricting these occasions and encouraging behavioral changes in the general public. Preliminary epidemiological investigations by the National Task Force for COVID-19 outbreak in Japan identified hospitals and other medical facilities, as well as nursing homes and other social care facilities, as the main sources of clusters; at the individual level, many of the clusters were associated with karaoke parties, clubs, bars, and gyms, which were identified as being associated with high-risk behaviors in close proximity (Furuse et al., 2020a,b). By the end of April 2020, the Task Force reported three situations that could increase the risk of COVID-19 and started to advise the public to avoid the "three Cs": closed spaces with poor ventilation, crowded places, and close-contact settings, which led to the initial successful containment of many chains of infection in Japan (Oshitani, 2020).

On the other hand, owing to the virus spreading from Wuhan to various countries across the world in a relatively short period of a few months, there was a need to develop a genomic analysis based on whole-genome sequencing of SARS-CoV-2 in Japan to track the transmission routes of the virus (Sekizuka et al., 2020a,b,c). This approach to genomic epidemiology has been reported in many countries, including the United States of America (United States) (Deng et al., 2020; Laiton-Donato et al., 2020), the United Kingdom (United Kingdom) (da Silva Filipe et al., 2021), China (Lu et al., 2020; Geidelberg et al., 2021), New Zealand (Geoghegan et al., 2020), Italy (Alteri et al., 2021), Scotland (da Silva Filipe et al., 2021), Austria (Popa et al., 2020), Russia (Komissarov et al., 2021), Morocco (Badaoui et al., 2021), and Brazil (Faria et al., 2021), and has proven to be a useful tool for investigating outbreaks and tracking the evolution and spread of the virus. In particular, understanding the evolution and transmission patterns of viruses after they enter new populations is important for developing effective strategies for disease prevention and control. However, to date, only a limited number of studies have evaluated the viral genome information of SARS-CoV-2 combined with epidemiological studies at a local level (prefectural-level or city-level) in Japan as a whole (Seto et al., 2021).

In this study, we report the clinical characteristics of 63 COVID-19 cases, including ribonucleic acid (RNA) viral load over the course of the disease, as well as an integrated approach of epidemiological and virological genomic data to investigate the transmission dynamics and patterns of SARS-CoV-2 in a local city, Niigata City (Niigata Prefecture, Japan), between February and May 2020. Specifically, we performed phylogenetic and phylogeographic analysis and examined the temporal changes of genotypes to identify multiple introductions of SARS-CoV-2 over time in Niigata City. These results provide insights into the use of the genomic epidemiology approach in addition to the active epidemiological surveillance for the monitoring of COVID-19 in local clusters and may provide valuable information for public health policy decision-making.

MATERIALS AND METHODS

Study Location

Niigata City is located in the northern area of the main island on the coast of the Japan Sea and is the capital of Niigata Prefecture, Japan (**Supplementary Figure 1**) (Ogushi, 2021). In 2020, it had a population of approximately 800,000. Niigata City is well connected to the capital of Japan, Tokyo, with a bullet train that takes passengers to the capital in approximately 2 h.

National Active Epidemiological Surveillance

In Japan, the national active epidemiological surveillance of COVID-19 was created based on the Infectious Diseases Control Law following the declaration of the COVID-19 as a "designated infectious disease" (National Institute of Infectious Diseases, 2020). Health authorities in 47 prefectures and 20 major cities throughout Japan take the responsibility for identifying cases, as well as putting in place various control measures to stop viral transmission within the community, including: (a) isolate patients in hospitals or hotels until qualitative reverse transcription polymerase chain reaction (RT-qPCR) test results became negative (this policy was valid from February 03 to May 29, 2020), (b) find the secondary cases and clusters by implementing active epidemiological investigation and RT-qPCR testing, and (c) identify infection sources to help prevent the spread of the virus and form clusters. The target of this active epidemiological surveillance was laboratory-confirmed cases and their close contacts.

Definitions of Case and Contacts in Epidemiological Investigation

In March 12, 2020, a "suspected case" of COVID-19 was defined by the Ministry of Health, Welfare, and Labor in Japan as a person with common cold-like upper respiratory symptoms and/or fever ≥37.5°C, and general fatigue or breathing difficulties lasting 4 days or longer (National Institute of Infectious Diseases, 2020). For individuals in high-risk groups, such as the elderly and patients with diabetes mellitus, cardiac failure, respiratory illness, hemodialysis, immunosuppressants, or undergoing chemotherapy for cancer, the definition was adapted to those presenting symptoms and signs for 2 days or longer. Contact history to a known or suspected COVID-19 case or travel history to endemic areas before 14 days was also counted when assessing the possibility of COVID-19 infections. Laboratory-confirmed cases were defined as those with a positive test result by RT-qPCR for SARS-CoV-2 with or without the above symptoms. Close contact was defined as a person who had contact with the confirmed cases in the following conditions: (a) persons who are living or staying closely with the confirmed cases including cars or air crafts, (b) persons who consulted or took care of the confirmed cases without appropriate personal protection, (c) persons who had direct contact with contaminants, such as respiratory secretions or the body fluid of the confirmed case, and (d) persons who touched or had a conversation within the confirmed case with 2 m. Initially, tracing of close contacts was advised to start from the day of symptom onset in the confirmed case by the Ministry of Health, Welfare and Labor in Japan (National Institute of Infectious Diseases, 2020). However, in Niigata city, contact tracing was started 2 days before the onset of symptoms in confirmed cases from early March 2020 to facilitate the detection of cases of pre-symptomatic transmission. Epidemiological links were categorized into four categories: (i) epidemiological links, (ii) possible links, (iii) no epidemiological links, and (iv) imported. An "epidemiological link" was defined as a case with a close contact history, as mentioned above, with a confirmed case and for whom the infection route was identified or estimated. A "possible link" was defined as one with no close or direct contact between the patients proved, but in which a weak link existed, such as using the same facility or living in the same community. "No epidemiological link" was defined by a lack of known contact or no possible link with a confirmed case. Imported cases were defined as those with a history of travel abroad up to 1 week before the onset of illness.

Patient Clinical Data Collection and Discharge Criteria

The Niigata City Public Health Center conducted active epidemiological surveillance based on the Infectious Diseases Control Law and collected clinical data of 63 symptomatic patients within the framework of the Nationwide Active Epidemiological Surveillance. Specifically, age, sex (male or female), presence of clinical symptoms (fever, malaise, cough, sore throat, taste and smell disturbances, runny nose, chills, loss of appetite, and diarrhea), days of hospitalization, and behavioral history (e.g., close contact or travel history) were recorded. All 63 cases identified in the present study were hospitalized for quarantine purposes under the regulations of the Infectious Diseases Control Law. The criteria for discharge of symptomatic patients in Japan from February 03 to May 29, 2020, stipulated that symptoms after admission improved, they were afebrile <37.5°C for more than 24 h, and their respiratory symptoms subsided, in addition to two negative RT-qPCR tests performed at least 24 h apart (Ministry of Health Labour and Welfare, 2020b). Therefore, multiple RT-qPCR tests were performed on a single patient at intervals of 24 h or more until two negative RT-qPCR results were confirmed.

Collection of Clinical Specimens and Reverse Transcription Polymerase Chain Reaction Testing

A total of 247 nasopharyngeal specimens were collected from 63 symptomatic patients in Niigata City between February and May 2020, and RT-qPCR testing targeting the nucleocapsid protein of the viral genome of SARS-CoV-2 was performed at the Niigata City Institute of Public Health and Environment (Niigata City Institute of Public Health and Environment, 2020), based on laboratory protocols by the National Institute of Infectious Diseases in Tokyo, Japan (Ministry of Health Labour and Welfare, 2020b). The positive RNA samples were subjected to whole-genome sequencing. The cycle threshold (C_t) value of an RT-qPCR run was inversely correlated with the copy number of the viral RNA. A C_t value higher than 40 (i.e., the limit of detection) was considered negative for SARS-CoV-2. Days from onset to sampling collection, days from onset to two negative RT-qPCR results, and the number of RT-qPCR tests per case were analyzed.

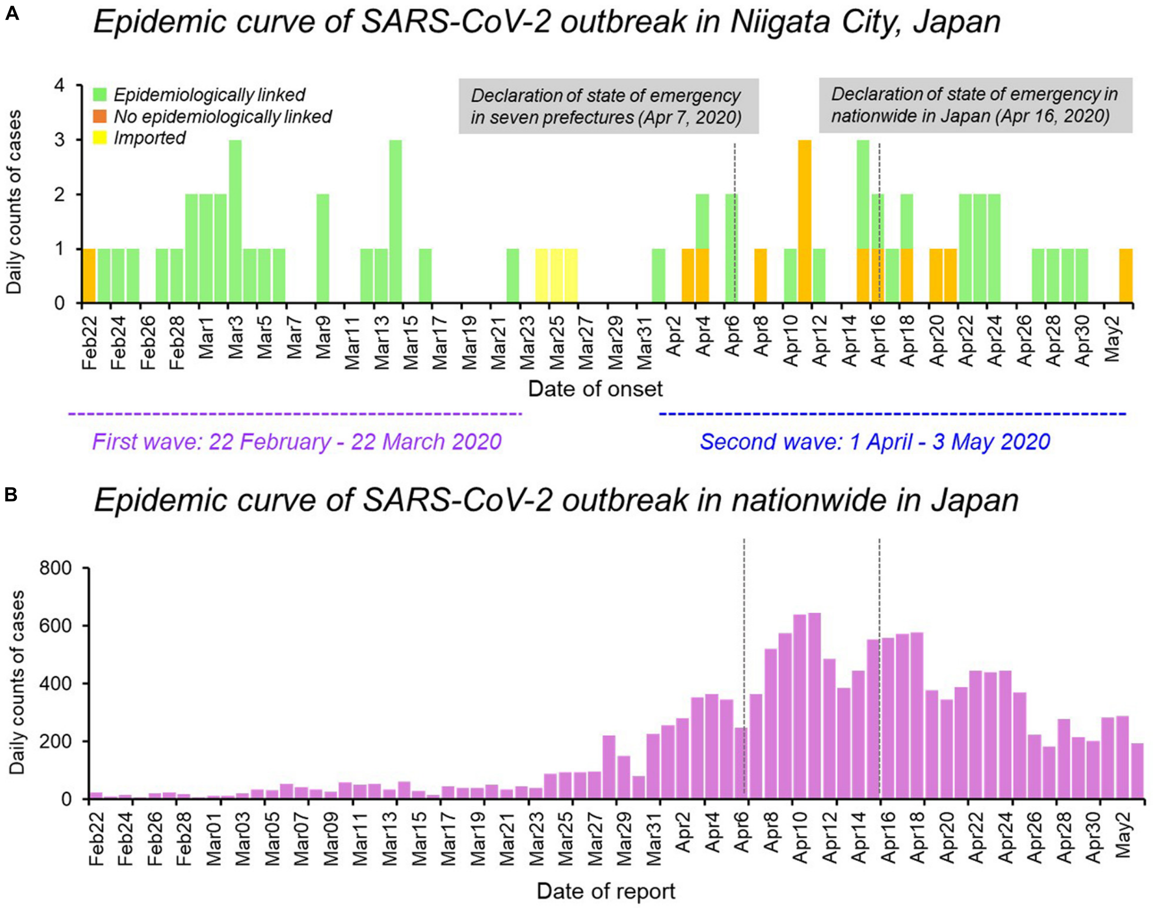


FIGURE 1 | Epidemiological dynamics of SARS-CoV-2 Niigata City and nationwide in Japan. **(A)** Epidemic curve of daily cases of laboratory-confirmed SARS-CoV-2 infection in Niigata City, Japan, by symptom onset date and colored as per epidemiological linkage, between 22 February and May 03, 2020 (*n* = 63). **(B)** Epidemic curve of daily cases of laboratory-confirmed SARS-CoV-2 infection in nationwide in Japan, by reporting date and colored as purple, between 22 February and May 03, 2020 (*n* = 14,915). The purple dotted line represents the first wave by the date of onset (between 29 February and March 22, 2020) and the light blue dotted line represents the second wave by the date of onset (between 01 April and May 03, 2020). Data on the daily number of new confirmed cases in nationwide in Japan were retrieved from the website of the Ministry of Health, Labour and Welfare, Japan (https://www.mhlw.go.jp/stf/covid-19/open-data.html).

Whole Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2

The whole genome sequences of SARS-CoV-2 were obtained using the PrimalSeq protocol to enrich complementary deoxyribonucleic acid (cDNA) of the SARS-CoV-2 genome by multiplex RT-qPCR amplicons using a multiplexed PCR primer set that was proposed by the Wellcome Trust ARTIC Network (2021). Two amplicons with low or zero coverage owing to the dimerization of the primers were identified. Therefore, we modified the protocol for SARS-CoV-2 genome sequencing produced by the ARTIC Network and exchanged some of the primers for multiplex PCR (Itokawa et al., 2020). Detailed information of multiplexed RT-qPCR primer sets was described in **Supplementary Table 1** (Itokawa et al., 2020). The PCR products from the same clinical sample were pooled, purified, and subjected to Illumina library construction using the QIAseq FX DNA Library Kit (QIAGEN, Hilden, Germany). The NextSeq 500 or iSeq100 platform (Illumina, San Diego, CA, United States) was used to sequence the indexed libraries. Next-generation sequencing (NGS) reads were mapped to the SARS-CoV-2 Wuhan/WIV04/2019 reference genome sequence [29.9-kb single standard RNA (ss-RNA)] (GenBank ID MN908947), resulting in the specimen-specific SARS-CoV-2 genome sequence by full mapping to the reference genome. The mapped reads of the SARS-CoV-2 sequences were assembled using A5-miseq version 20140604 or SKESA version 2.3.0 to determine the full genome sequence (Coil et al., 2015). Single nucleotide variation (SNV) sites and marked heterogeneity were extracted by read mapping at a depth of $\geq$10× and from the region spanning nucleotides (nt) 99 to 29,796 of the Wuhan/WIV04/2019 genome sequence. The lineage or clade classification of SARS-CoV-2 was performed using those of the Global Initiative on Sharing All Influenza Data (GISAID) database[1], PANGOLIN (version

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

TABLE 1 | Epidemiological characteristics of patients with COVID-19 in Niigata City, Japan between February and May 2020 (*n* = 63).

Characteristics	Mean ± *SD* [Median, IQR]	*n*	%
Age, years	47.9 ± 22.3 [45.0, 30.0–70.2]		
Age categories			
0–9 years		3	5
10–19 years		0	0
20–29 years		9	14
30–39 years		17	27
40–49 years		5	8
50–59 years		7	11
60–69 years		7	11
70–79 years		9	14
80–89 years		5	8
90–100 years		1	2
Sex			
Male		31	49
Female		32	51
Clinical symptom			
Fever (≥37.5°C at the onset symptoms)		41	65
Malaise		32	51
Cough		22	35
Sore throat		17	27
Loss of taste or smell		14	22
Runny nose		14	22
Chills		11	17
Loss of appetite		9	9
Diarrhea		2	2
Days from onset to sample collection	7.4 ± 6.5 [6.0, 6.0–9.0]		
Epidemiologic linkage			
Epidemiologically linked		47	75
Not epidemiologically linked		13	20
Imported		3	5

SD, standard deviation; IQR, interquartile range. Note that P1-E is not included in this table because the patient lived in a different city, and precise clinical and virological information is not available.

2.0[2] (Rambaut et al., 2020) and Nextstrain team version 8.0[3] (Hadfield et al., 2018) (**Supplementary Table 2**).

Phylogenetic and Phylogeographic Analysis and Haplotype Network

The full genome sequences of SARS-CoV-2 were downloaded from the Global Initiative on Sharing All Influenza Data (GISAID) database (see Text Footnote 1) (Shu and McCauley, 2017). For phylogenetic analysis, we used a total of 1,374 genomic sequences, including 781 global strains, 546 Japanese strains, and 47 strains from Niigata City obtained in this study. Global strains in the present study were collected from December 2019 to May 2020, which is used in the global phylogenetic tree, constructed by Nextstrain team version 8.0 (see Text Footnote 3) (Hadfield et al., 2018), and the Japanese strains from January to May 2020 excluding clade O (other) and airport quarantine isolates in Japan downloaded on December 04, 2020. As a reference sequence, Wuhan/WIV04/2019 (GISAID ID, EPI_ISL_402124), the official reference sequence from the GISAID database, was used. Multiple alignments for the full-length genome sequences of SARS-CoV-2 were performed using MAFFT version 7.475 (Katoh and Standley, 2013). The removal of spurious sequences or poorly aligned regions from a multiple sequence alignment was performed using trimAl version 1.4.rev22 (Capella-Gutiérrez et al., 2009). The maximum likelihood (ML) phylogenetic analysis was performed using IQ-TREE version 2.1.2 with Model Finder and ultrafast bootstrap test parameters (Nguyen et al., 2015; Kalyaanamoorthy et al., 2017; Hoang et al., 2018), and visualized by iTOL version 6.0 (Letunic and Bork, 2021).

Haplotype networks from genomic SNVs (HN-GSNV) for Niigata City isolates were conducted using median joining network analysis (Bandelt et al., 1999) by PopART version 1.7 (Leigh and Bryant, 2015) against the SNV created by aligning the sequences by MAFFT and excluded the gap containing sequences. SNV detection was performed using snp-sites version 2.5.1 (Page et al., 2016) after multiple alignment based on the reference sequence by MAFFT version 7.475. The annotation of the detected SNV was added by SnpEff version 5.0e (Cingolani et al., 2012).

Retrospectively trace the dispersal dynamics of SARS-CoV-2 transmission patterns in Niigata City and other global/local areas, phylogeographic analysis was performed by Nextstrain team version 8.0 (see Text Footnote 3) (Hadfield et al., 2018) against 292 global strains, 3,968 Japanese strains, and all 47 strains obtained in this present study, which had accurate collection date information. Global strains in the present study were collected from December 2019 to May 2020 and were used to construct the global phylogenetic tree using Nextstrain team version 8.0 (see Text Footnote 3) (Hadfield et al., 2018) and all Japanese strains from January to May 2020 downloaded on August 31, 2021.

Statistical Analysis

Data were described as the mean ± standard deviation (SD) and/or median [interquartile range (IQR)] for continuous variables and frequency (%) for categorical variables. Age (years) was divided into ten categorical variables (0–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, and 90–100 years) and time from onset to sampling collection (days) into four categorical variables (≤9, 10–20, 20–30, and ≥30 days) for analysis. The Kruskal–Wallis rank-sum test was used to compare the median values of C_t values of RT-qPCR (cycles) by the four categorical variables of time from onset to sample collection. Bonferroni correction was applied as a *post hoc* test for each pair. Spearman's rank-order correlation coefficient (ρ) was used to investigate the association between the time from onset to sample collection and the C_t values of RT-qPCR. Statistical significance was set at $p < 0.05$, using a two-tailed test. All analyses were

[2] https://github.com/cov-lineages/pangolin
[3] https://nextstrain.org/ncov/global

performed using EZR version 1.27 (Kanda, 2013) and STATA version 15.1 (Stata Corp., College Station, TX, United States).

Ethical Approval and Consent to Participate

This study was conducted in accordance with the Nationwide Active Epidemiological Surveillance under the Infectious Diseases Control Law in Japan. Therefore, the approval of the ethical review board was waived, as well as the need for written consent from the patients and close contacts to collect the epidemiological and clinical data, viral sample collection, and analysis. The personal data of patients used in this study were anonymized.

RESULTS

Epidemiological Dynamics

The epidemic curve of the initial 63 cases of COVID-19 in Niigata City by symptom onset date was characterized by two distinct peaks, corresponding to the first wave (between February 22 and March 22, 2020) and the second wave (between April 01 and May 03, 2020) (**Figure 1**). Overall, 47 (75%) of 63 patients had an epidemiological link, 13 (20%) had no apparent epidemiological links, and the remaining three (5%) were imported cases (**Figure 1A** and **Table 1**). Specifically, in the first wave, almost all cases (96%, 26/27) had epidemiological links, whereas in the second wave, less than half of the cases (36%, 12/33) had apparent epidemiological links. Between the two epidemic waves, three imported cases from Brazil, Czechia, and United States, were identified from March 24 to March 26, 2020 (**Figure 1A** and **Table 1**). The second wave of Niigata city corresponded to an increase of patients nationwide from the end of March 2020, and the first declaration of the state of emergency for seven prefectures (e.g., Tokyo and Osaka) announced on April 07, 2020 and subsequently, it was expanded to all 47 prefectures on April 16, 2020 mandating soft lockdown including Niigata (**Figure 1B**).

Demographic and Clinical Characteristics

Under Japan's Infectious Disease Control Law, all the patients were quarantined and hospitalized. Of the 63 patients who tested positive for SARS-CoV-2 by RT-qPCR between February and May 2020, 32 (51%) were female, with a mean age of 47.9 years (SD, 22.3 years), ranging from 0–9 to 90–100 years old (**Table 1**). All 63 patients were symptomatic, and no asymptomatic cases were found. Fever was the most common clinical symptom (41/63, 65%), followed by malaise (32 cases, 51%) and cough (22 cases, 35%). Loss of taste or smell was reported in 14 cases (22%) (**Table 1**). The time from onset to sample collection was 7.4 days (SD, 6.5 days). One 60-year-old patient (P5-2) developed severe pneumonia under ventilator control, however, he recovered and was discharged almost 4 months after his onset (**Supplementary Table 3**). No deaths occurred during the study period.

Time-Series of Hospitalization and Negative Reverse Transcription Polymerase Chain Reaction Tests

Among the 63 cases, the mean time from onset to two negative RT-qPCR results was 19.0 days (SD, 7.98 days) (**Supplementary Table 3**). One case with subsequent RT-qPCR negative results after confirmation was excluded from this calculation (P5-2). In all 63 patients, the mean number of tests was 6.8 times (SD, 4.2 times) and the mean length of hospitalization was 24.2 days (SD, 15.0 days).

Relationship Between Days Post-symptom Onset and C_t Value

The relationship between days post-symptom onset and C_t values is shown for 247 samples obtained from the 63 cases by RT-qPCR (**Figure 2**). The median C_t values were 20.9 cycles (IQR, 17–26 cycles; $n = 52$) for ≤ 9 days post-symptom onset, 33.0 cycles (IQR, 30.4–34.5 cycles; $n = 84$) for 10–20 days, 34.1 cycles (IQR, 32.9–35.7 cycles; $n = 87$) for 20–30 days, and 34.4 cycles (IQR, 32.3–35.6 cycles; $n = 24$) for ≥ 30 days, respectively. A significant difference in the median C_t values was observed between the groups ≤ 9 days and >10 days after the onset of symptoms (Kruskal–Wallis rank-sum test, $p < 0.001$) (**Figure 2**). The days post-symptom onset and C_t values showed a significant positive correlation (Spearman's rank-order correlation coefficient $\rho = 0.56$, $p < 0.001$), suggesting an attenuation of the viral load over time. However, detailed clinical information on the relationship between viral load and severity of illness in the patients by present official surveillance framework were not available, thus, we were not able to fully assess these associations.

Transmission Chain

We plotted all 63 symptomatic patients in the first and second waves in Niigata City in 2020, and presented a transmission chain showing the relationship between the cases and the exposure histories (**Figure 3** and **Supplementary Table 4**). Niigata City reported the first confirmed case of SARS-CoV-2 on February 29, 2020 (P1-1), almost one and a half months after the first case was reported in Japan. P1-1 was a domestically imported case, with a symptom onset of malaise and cough on 22 February, and fever on 23 February, 2020. The case had a history of travel to an endemic area outside Niigata prefecture (Tokyo metropolitan area) within 1 week before the onset to join a social gathering. It was not clear whether P1-1 had a contact with the known COVID-19 patient at the gathering. Subsequently, a cluster of four cases (P1-2, P1-3, P1-6, and P1-E) occurred indoor sports at Facility A in February 2020 (**Figure 3** and **Supplementary Table 4**), P1-1 played the indoor sports together with the other four on February 20, 2 days before symptom onset. Thus, it was speculated as pre-symptomatic infections. From this initial cluster of cases, subsequent infections occurred in multiple secondary clusters, resulting in the first wave in Niigata City (22 February to 22 March, 2020) (**Figure 1**). One of the secondary infections spread from the initial patient group, namely from patient P2-1, who played a different indoor sport in the same

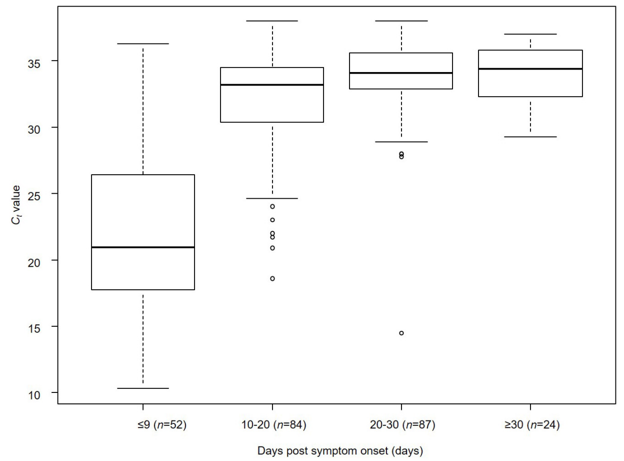


FIGURE 2 | Relationship between days post-symptom onset and C_t value of SARS-CoV-2 in Niigata City, Japan, between February and May 2020 (n = 247). Kruskal–Wallis rank-sum test was used to compare the median values of C_t values of RT-qPCR (cycles) by the four categorical variables of time from onset to sample collection ($\leq$9, 10–20, 20–30, and $\geq$30 days) (days). Bonferroni correction was applied as a *post hoc* test between each pair. A significant difference in median C_t values was observed between the groups $\leq$9 and >10 days after the onset of symptoms (Kruskal–Wallis rank-sum test, $p < 0.001$): $\leq$9 days vs. 10–20 days (Bonferroni correction, $p < 0.001$), $\leq$9 days vs. 20–30 days (Bonferroni correction, $p < 0.001$), and $\leq$9 days vs. $\geq$30 days (Bonferroni correction, $p < 0.001$). The lower and upper limits of boxes indicate the 25th and 75th percentiles, respectively; lines within boxes indicate the median values. The lines extending from the boxes indicate the range of non-outlying values. The dots represent individual points that fall outside this range (dots).

facility (facility A) where P1-1 participated in an indoor sport activity the previous day. Although whether they shared the same locker room, athletic room, or toilet is not certain, contact infection is highly likely because they were infected despite not meeting each other and playing different sports. From P2-1, the infection spread to friends and a colleague by direct contact (P2-2, P2-3, and P2-4). P1-3 infected two family members (P1-4 and P1-5) in their household. P1-E is assumed to have infected P3-1 after both used a sports facility at the same time (facility E); however, the two patients did not see each other and played different sports, suggesting that the route of transmission may be contact infection. P1-E belongs to the first sport cluster, despite the patient living in a different city; therefore, precise clinical and virological information, including genome sequence data, was not included in this study. P3-1 spread the infection within an elderly daycare center, where P3-1 worked, to a colleague and an elderly individual (P3-3 and P3-6), then to their family members (P3-2, P3-4, and P3-5). For P1-1, the infection was spread to P4-1 and P5-1 while they participated in indoor sports activities on different days in different facilities (Facilities B and D). Case P4-1 infected two colleagues (P4-3 and P4-2) and a child (P4-4) at a children's daycare center (Facility G), where P4-1 worked. Then, a tertiary infection occurred from P4-2 to a friend (P4-5). P5-1 caused subsequent infections in the workplace (company H) to colleagues (P5-2 and P5-3). Although direct contact with patients was not confirmed, P5-6 developed COVID-19 infection after the visit to Company H for business.

In the second wave in Niigata City, tracking the epidemiological links between patients was more difficult than in the first wave (**Figure 3** and **Supplementary Table 4**). While several small clusters, comprised of family members and colleagues, were observed, as well as sporadic cases, the links among the clusters were difficult to confirm, unlike in the first wave. The index cases of the small clusters (P11-1, P13-1, P16-1, and P20-1) were not epidemiologically linked to each other. In addition, there was a lack of self-reported contact with persons in an endemic area (e.g., Tokyo Metropolitan area). One cluster (P9-1, P9-2, and P9-3) occurred by an extended family gathering, infected from a family member returned from another prefecture. Three cases (P10, P14, and P18) had no apparent link to the known cases or each other, while one case (P12) had a history of contact with a confirmed case in another prefecture. In addition, small clusters and sporadic cases occurred in a local area in one of the districts of Niigata City from mid-April to May. Individuals in this locality were seemingly infected through contacts that occurred during social activities, including drinking and singing karaoke with friends, neighbors, and family members (from P15-1 to -3, from P17-1 to -7, P21-1, and P21-2). However, the direct epidemiological links among these clusters in the community remain unclear. Two cases (P19 and P22) resided in the same district but had no apparent epidemiological links to others. Between the two waves, three imported cases resulting from international travels were identified (P6, P7, and P8).

Genomic Analysis

The whole viral genome was available in 47 (75%) of 63 cases. Specifically, ML phylogenetic analysis showed that the 47 isolates from Niigata City were different in the first and second wave clusters (**Figure 4** and **Supplementary Table 5**). The first wave cluster was found to be generated by clade S, attributed to the Wuhan/WIV04/2019 haplotype, the ancestor of the global pandemic. In contrast, all isolates in the second wave cluster belonged to clade GR, the main lineage circulated in Europe, strongly suggesting that the introduction of SARS-CoV-2 in Niigata City was associated with two different SARS-CoV-2 lineages. The three imported cases belonged to GH (P6), and GR (P7 and P8), respectively.

Haplotype network analysis of the whole genome showed that P1-1, P1-2, and P1-6, who were exposed to infection at Facility A during the indoor sports, had identical genome sequences, while P1-3 in the same cluster had an SNV (**Figure 5** and **Supplementary Table 5**). In the first wave, patients with secondary, tertiary, or quaternary infections (i.e., P2-2, P2-4, P3-6, P5-2, P5-3, P5-4, P3-6, and P5-6) were closely associated with two or three SNVs around the P1 index patient group (**Figures 3**, **5** and **Supplementary Table 5**). Many of the major clusters identified in the first wave had lower SNV rates than the isolates collected in the second wave, and four common nucleotide changes, T4402C, G5062T, C8782T, and T28144C, were seen from Wuhan/WIV04/2019 (**Figure 5** and **Supplementary Table 5**), suggesting that they were closely related to the prototype SARS-CoV-2 strain. In contrast, the genome

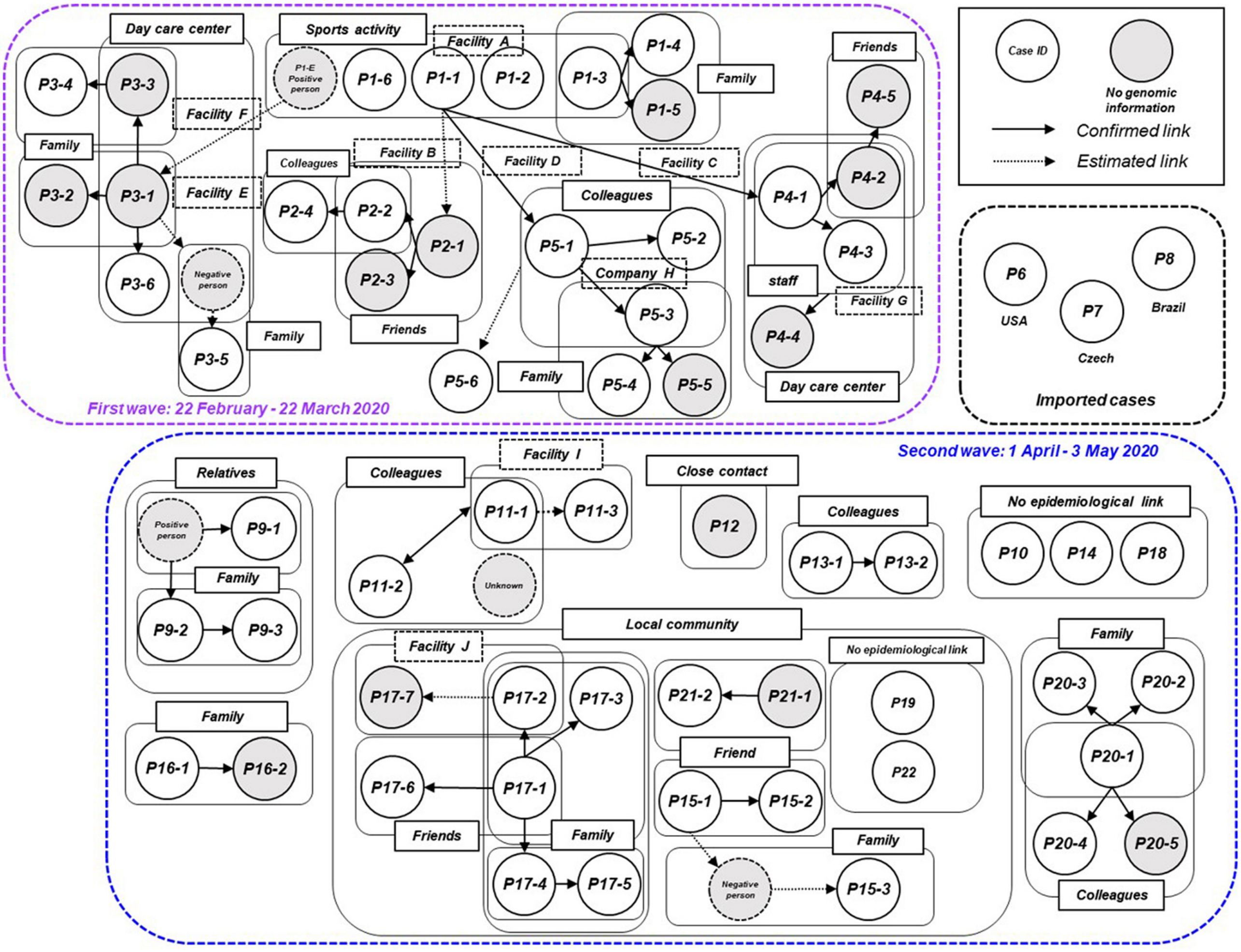


FIGURE 3 | Transmission chain of the COVID-19 outbreak, Niigata City, Japan, between February and May 2020 (*n* = 63). Diagram showing the contact relationships of patients, including the 63 symptomatic cases in this study. The purple dotted box represents the first wave by the date of onset (between 22 February and March 22, 2020) and the light blue dotted box represents the second wave by the date of onset (between 01 April and May 03, 2020). The circles indicate each patient and are assigned a case ID. Gray circles indicate cases for which no genomic information was available, dotted circles indicate cases with negative qualitative reverse transcription polymerase chain reaction (RT-qPCR) results or unidentified cases. A large pale blue frame indicates common indoor environment, facility of transmission patients (e.g., family, colleagues, friends, and faculties), or community transmission. Arrows indicate established link and dotted arrows represent the most plausible link, both based on contact tracing. The case ID of each patient corresponds to the haplotype network in **Figure 5** and detailed epidemiological information of **Supplementary Table 4**.

sequences of SARS-CoV-2 from the second wave possessed eight common nucleotide changes, C241T, C313T, C3037T, A14408T, A23403T, G28881A, G28882A, and G28883C from the reference, and more than 20 individual SNVs with diversified haplotype patterns (**Figure 5** and **Supplementary Table 5**). Of note, the strains in the second wave were characterized by A23403T, Asp614Gly in spike protein that was reported to increase infectivity compared to prototype SARS-CoV-2 strain (Plante et al., 2021).

To further explore the dispersal dynamics of SARS-CoV-2 transmission patterns in Niigata City and other global/local areas, we also employed a phylogeographic inference analytic approach, using Nextstrain team version 8.0 (see Text Footnote 3) (Hadfield et al., 2018), on the whole genomes of all 47 isolates from Niigata that had accurate collection date information (**Supplementary Figure 2** and **Supplementary Table 5**). The results indicated that the initial COVID-19 outbreak in Niigata City belonged to 19B (clade S) by Nextstrain and formed a monophyletic group that supported the epidemiological information suggesting a single introduction to the city (**Supplementary Figure 2**). The strains were closely related to strains in Tokyo during the same period (data not shown). In contrast, the strains in the second wave belonged to 20B (clade GR), showing polyphyletic distributions in the time-aware phylogeny by Nextstrain, and showed strong relationship with the strains mainly in Tokyo (**Supplementary Figure 2**). These results are indicative of the existence of multiple untraced introduction pathways in the second wave in Niigata City.

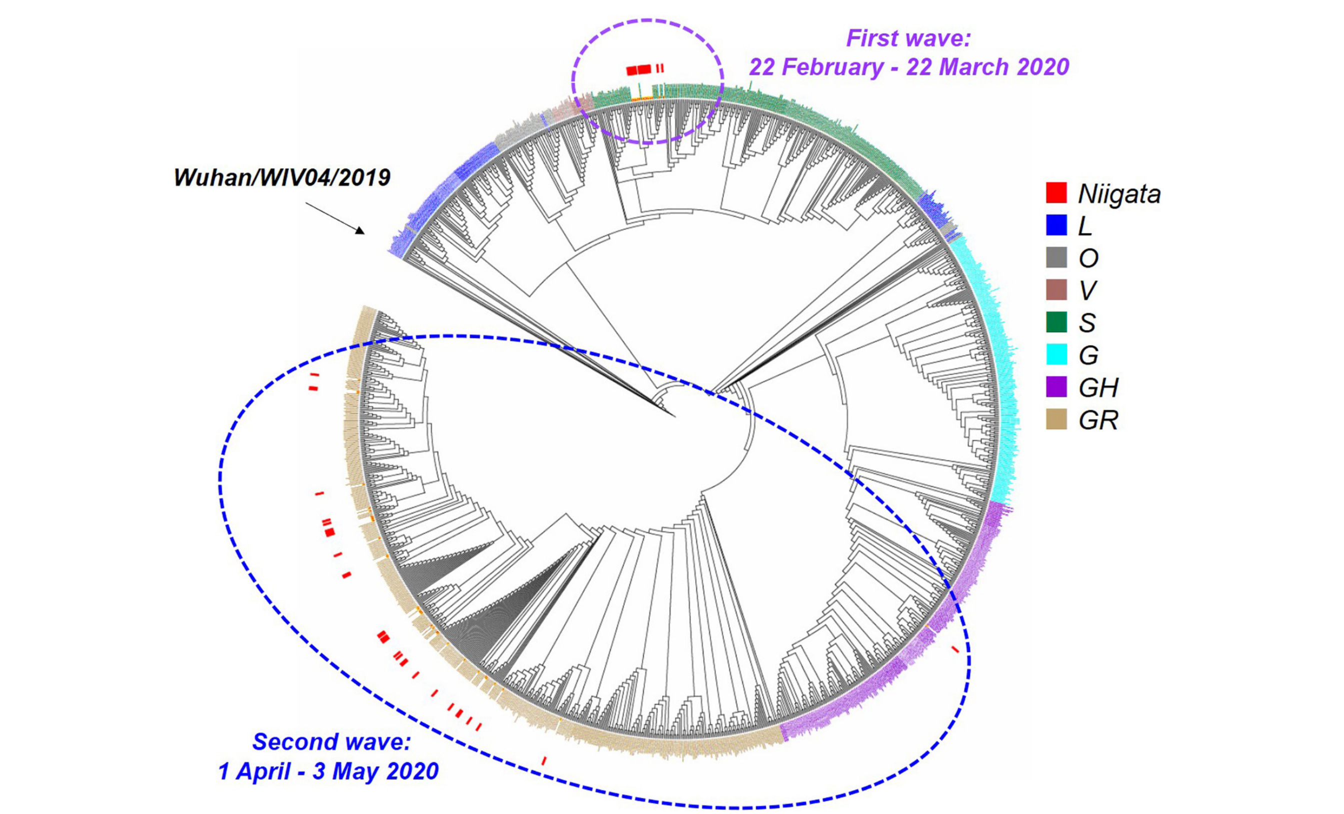


FIGURE 4 | Maximum likelihood phylogenetic analysis of SARS-CoV-2 in Niigata City, Japan, between February and May 2020 (*n* = 47). This phylogenetic tree was based on the GISAID sequences of 781 global strains, 47 Niigata City strains, and 546 strains from other cities of Japan. The purple dotted circle represents the first wave by the date of onset (between 22 February and March 22, 2020) and the light blue dotted circle represents the second wave by the date of onset (between 01 April and May 03, 2020). Red squares indicate the strains detected in Niigata City, blue squares the L clade, mouse-colored squares the O clade, brown squares the V clade, dark green squares the S clade, light blue squares the G clade, dark purple squares the GH clade, and light brown squares the GR clade.

Genome analysis provides useful information with which to assess where the SARS-CoV-2 strains originated from and can be used as **Supplementary Information** to identify and confirm epidemiological links.

DISCUSSION

In the present study, both epidemiological and genome-wide analyses were used to elucidate the transmission dynamics of SARS-CoV-2 in a local area of Japan during the early stages of the COVID-19 pandemic in 2020. Two waves of infection were identified in Niigata City, Japan between February and May 2020, with different patterns of infection in each wave. In the first wave, all of the cases identified were infected with the Clade S (19B) virus from China, which was prevalent in Japan in February and March (Wagatsuma et al., 2021), with the majority of these patients spreading the infection from a single index group to their family members or colleagues. Interestingly, the first wave was characterized by a local epidemic that occurred in a closed environment, namely a sports facility, which resulted in a chain of cluster-mediated infections, as described above. On the other hand, the second wave was characterized by the clade GR (20B) virus from Europe, which was prevalent in Japan since mid-March (Sekizuka et al., 2020a), with an increased number of cases with unknown routes of infection compared to the first wave. There was no commonality in age, sex, or behavior of the patients in the second wave, indicative of multiple routes of entry. Similar to our study, this type of genomic epidemiological analysis has been widely used in various local communities in other countries (Deng et al., 2020; Geoghegan et al., 2020; Laiton-Donato et al., 2020; Lu et al., 2020; Popa et al., 2020; Sekizuka et al., 2020a; Alteri et al., 2021; Badaoui et al., 2021; da Silva Filipe et al., 2021; Faria et al., 2021; Geidelberg et al., 2021; Komissarov et al., 2021). In support of the literature, our results highlight the benefits of genome surveillance to complement epidemiological surveillance and further support a comprehensive country-wide study of SARS-CoV-2 epidemiology and transmission patterns in Japan.

The most common clinical symptoms reported in the patient population included in the present study were almost identical to those of previous studies. In this study, the most common symptoms were fever (65%), followed by malaise (51%), cough (35%), and sore throat (27%). The largest Japanese registry study describing the clinical epidemiological characteristics of 2,638 hospitalized COVID-19 patients enrolled from 227 healthcare facilities in Japan as of July 07, 2020, showed similar results,

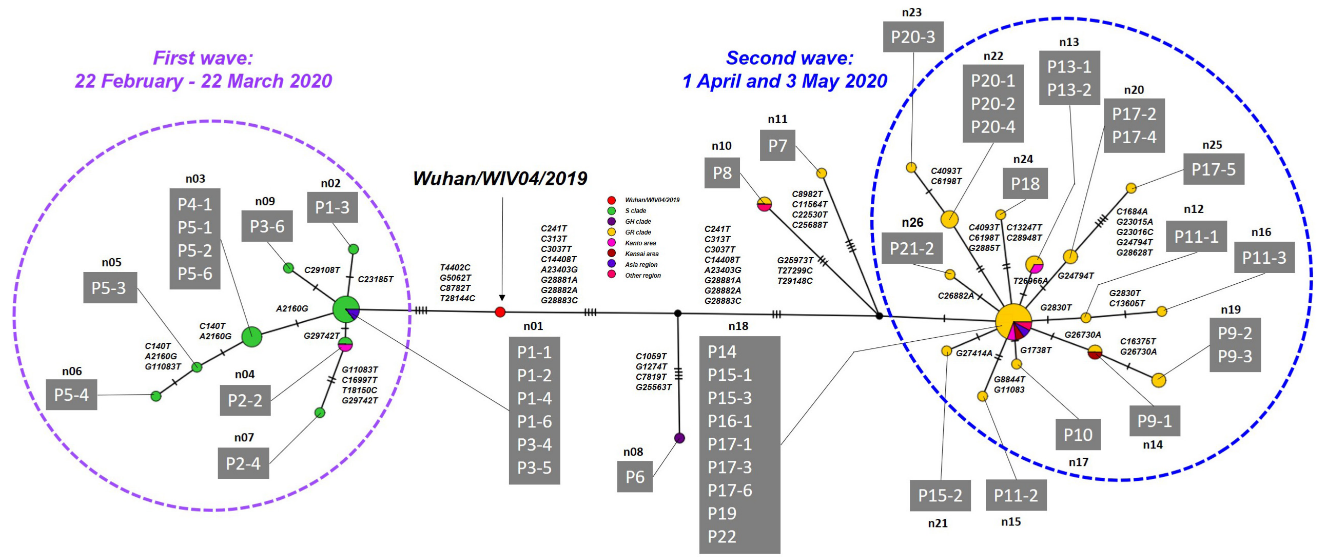


FIGURE 5 | Haplotype network using genome-wide SNVs (HN-GSNVs) in Niigata City, Japan, between February and May 2020 (n = 47). The purple dotted circle represents the first wave by the date of onset (between 22 February and March 22, 2020) and the light blue dotted circle represents the second wave by the date of onset (between 01 April and May 03, 2020). Red circles indicate the reference sequence (Wuhan/WIV04/2019), green circles the S clade, dark purple circles the GH clade, and yellow circles the GR clade. Pink circles indicate the clusters of Kanto area prefectures isolates (Tokyo, Chiba, Kanagawa, Saitama, Gunma, and Tochigi), dark red circles the clusters of Kansai area prefectures isolate (Osaka, Kyoto, Hyogo, Shiga, Nara, and Wakayama), purple circles the clusters of Asia region isolate, and dark pink circles the clusters of other regions isolate. Haplotype clusters including each patient are represented as n01 to n26, respectively. Branches are labeled with SNVs. The case ID of each patient corresponds to the transmission chain in **Figure 3** and detailed epidemiological information of **Supplementary Table 4**. Detailed list of nucleotide substitutions found in each haplotype constructed in Niigata City isolates compared to reference strain Wuhan/WIV04/2019 is presented in **Supplementary Table 5**.

with fever, cough, dyspnea, fatigue, and olfactory and gustatory disturbances in a large proportion of patients (Matsunaga et al., 2020). On the other hand, the demographic characteristics of the present study population showed a broadly uniform age distribution and an almost equal ratio of males to females. It should be emphasized that this is contradictory to the Japanese registry study, which reported a median age of 56 years (IQR, 40–71 years) for COVID-19 inpatients, with more than half of the cases comprised of male patients (58.9%, 1542/2619). The reason for this difference compared to the previous literature may be that the epidemiology in the early stages of the initial epidemic in Niigata City may not have been captured owing to the small number of cases (i.e., 63 cases). Therefore, there is a need to increase the number of cases studied in the future, as well as clarify the patient characteristics because disease severity changes over time, owing to surveillance capacity and circulating genotypes, in the case of COVID-19.

The results of the analysis of the C_t values of the initial 63 symptomatic individuals detected in Niigata City showed that the highest viral load was recorded after the onset of symptoms, and that the viral load decreased steadily after the onset of symptoms, with C_t values $>$ 30 at 10–20 days (He et al., 2020; Singanayagam et al., 2020). Notably, our results showed a relatively large median difference of approximately 10 cycles between C_t values $\leq$ 9 days and $>$10 days after the onset of symptoms, which is consistent with the findings of several previous studies (Arons et al., 2020; Bullard et al., 2020; Perera et al., 2020; Singanayagam et al., 2020). On the other hand, many previous studies have reported that SARS-CoV-2 is likely to be transmitted from pre-symptomatic or asymptomatic patients. Although this pre-symptomatic infection was not well known at the beginning of epidemic in Japan, we checked the contact histories of patients from 2 days prior to the onset through the information of literatures and the thorough examinations of epidemiological information of the first cluster of patients in Niigata City. P1-1 seemed to have transmitted to others in the pre-symptomatic period. The previous studies in Singapore and Tianjin City in China, where active case-finding was carried out, an estimated proportion of pre-symptomatic transmission of 48% (95% credible interval [CrI], 32–67%) and 62% (95% CrI, 50–76%), were observed, respectively (Ganyani et al., 2020). The proportion of asymptomatic infections tended to be higher in areas where active case-finding was undertaken. However, in our analysis, asymptomatic patients were not detected. It is possible that several asymptomatic infections were missed owing to the limited capacity of RT-qPCR and case tracing at the beginning of the epidemic in 2020.

The results of our epidemiological analysis in Niigata City are consistent with the results of a large epidemiological study of COVID-19 clusters previously conducted using nationwide data in Japan in the early stages of the epidemic, from January to April 2020 (Furuse et al., 2020b). Furuse et al. (2020b) found that many of the COVID-19 clusters were associated with heavy breathing in closed environments, such as singing at karaoke parties, cheering in clubs, talking in bars, and exercising in gyms, with several studies characterizing these activities as the three Cs: closed spaces, crowded places, and close-contact settings (Furuse et al., 2020a,b; Muto et al., 2020). These conditions are characterized by infection via droplet or aerosol dispersion (Jayaweera et al., 2020; Morawska and Milton, 2020; Tang et al., 2021). In the present study, we found that the initial cluster of epidemics in Niigata City between February and May 2020 was also characterized by closed environments, including indoor sports activities, between relatives and colleagues, confirming the findings of previous studies. A thorough epidemiological investigation conducted by the Niigata City Public Health Center revealed that almost all secondary infections occurred via droplet or aerosol infections. Surprisingly, at least two cases were seemingly infected by contact infections since the patients of the index cases and secondary cases did not come into direct contact. Previous studies have reported that care should be taken to avoid contact infections through common areas, such as toilets and locker rooms (Huang et al., 2020; van Doremalen et al., 2020; Groves et al., 2021). On the one hand, the Centers for Disease Control and Prevention (CDC) have recently reported that the risk of contact transmission of COVID-19 is quite low (Centers for Disease Control and Prevention, 2021). However, we would like to emphasize again that the infection risk associated with this type of transmission is not zero, and that measures, such as surface disinfection of objects, are essential, especially in areas where clusters are likely to occur, such as hospitals and sports facilities.

The results of a genealogical network study using SARS-CoV-2 genomic data in Niigata City suggested that the impact of the influx of strains from abroad may have a direct influence on local epidemics in regional cities over a short period of time. Importantly, haplotype network analysis of SARS-CoV-2 suggested that the second wave of COVID-19 in early April had a distinctly different origin from the first wave of strains. This suggests that strains that subsequently spread domestically in Japan may have entered the city via multiple routes (the haplotype network includes strains from other prefectures). Additionally, a detailed assessment of potential inter-cluster variation in SNVs showed that higher nucleotide changes were observed in the second wave than in the first wave in Niigata City further highlighting the possibility of multiple sources of importation from other regions in the second wave. Furthermore, phylogeographic inference indicates the dispersion dynamics of SARS-CoV-2 transmission patterns suggested in the first wave were monophyletic from a single source, but that of the second wave were polyphyletic with multiple sources, and Tokyo was an important central transmission hub for Niigata City. In view of the increasing number of COVID-19 cases in Europe and the detection of imported cases in Japan, the Japanese government decided to suspend the entry of European travelers into Japan on March 27, 2020 (Furuse et al., 2020a; Wagatsuma et al., 2021). Local restriction measures were only weakly imposed in Japan, and more prolonged and rigid self-restraint measures could have been more effective in mitigating the increase in COVID-19 cases until mid-March. However, owing to the lack of strict measures, an increase in the number of cases occurred in Japan after April. A recent report on the impact of local and national travel restrictions on COVID-19 in Japan strongly suggests that restrictions on domestic travel by public transport are quite effective in preventing the spread of infection (Murano et al., 2021). Given the fact that large numbers of people move at the local level, local cities are likely to be immediately affected by outbreaks in densely populated areas (e.g., Tokyo, Osaka, and Aichi). However, the rapid application of travel restrictions (especially domestic) could help delay the widespread transmission of COVID-19.

This study should be interpreted in the context of several limitations. Primarily, the present surveillance and contact-tracking data only account for the early phase of the COVID-19 epidemic in Niigata City, presenting the epidemiology of only 63 initial cases in this city. Therefore, our results may not be representative of the period of sustained local transmission following the introduction of SARS-CoV-2 into Niigata City; in addition, they do not represent the whole picture of Niigata Prefecture, which has a larger population (more than double). In future studies, the sample size must be increased to analyze the association between imported events and national and/or global pandemics through social events and variations in the viral loads. A detailed analysis is critically needed to fully assess these associations. However, because COVID-19 transmission is likely to start in populated areas (Furuse et al., 2020a; Wagatsuma et al., 2021), we believe that our study is sufficient to depict the early pictures of COVID-19 in Niigata Prefecture. Additionally, literature describing the transmission dynamics of SARS-CoV-2 at the local level (prefectural-level or city-level) during the early stages of the COVID-19 pandemic by combining epidemiological analysis and genomic information are quite limited, and the authors believe that the present study has a decent value to elaborate active surveillance by public health centers in Japan. Second, within the framework of an active nationwide epidemiological surveillance, especially in early 2020, the identification of infection contacts relies on self-reported information by index cases, which may have led to the occurrence of unconfirmed cases or routes, and a degree of incompleteness, and, therefore, bias. This may have led to an underestimation of the initial epidemiology of COVID-19 infection in Niigata City. Third, in the present study, our analysis did not consider symptomatic or asymptomatic cases. In particular, it should be noted that some cases may have been missed in our study, especially in view of previous literature, which has shown that over half of all infections caused by COVID-19 can be traced back to asymptomatic individuals (Ganyani et al., 2020; He et al., 2020; Johansson et al., 2021). Fourth, during the study period, tests were only offered to patients who met the epidemiological and clinical criteria, which was initially strict owing to the low capacity of RT-qPCR. Fifth, in the present study, only the whole viral genomes of 47 (75%) out of the 63 cases were analyzed, such that not all eligible cases were analyzed. Sixth, it was difficult to conduct a detailed analysis of the initial viral load, severity of patients' illness, and the other potential epidemiological information associated with hospitalization and negative RT-qPCR tests. Of note, assessment of the association between these epidemiological indicators may provide useful information for better clinical impact on patient management and treatment in the early stage of emerging and reemerging infectious disease outbreaks including COVID-19 pandemic and detailed analysis is a key topic for future study. Furthermore, it is fruitful to understand the differences in patient outcomes by region and country attributable to these epidemiological factors. Finally, the relatively small sample size in the present study may have underestimated the circulation of viral strains in the surrounding areas. To explicitly characterize the dispersal dynamics of SARS-CoV-2 transmission patterns in Niigata City and other global/local areas, further phylogeographic studies are critically needed to increase the sample size. Lastly, owing to the protection of privacy, it was not possible to include detailed information, such as the date of onset and date of confirmation of individual cases or types of sports activities, owing to the need to avoid disclosing personal information.

Despite the number of limitations noted above, it should be stressed that the first 4 months of the COVID-19 outbreak in Niigata City were associated with multiple introductions associated with people returning or visiting from other prefectures in which COVID-19 was more prevalent, as well as imported cases from overseas, early community transmission, and clusters associated with large indoor events, families, and workplaces. At the same time, however, Niigata City was able to control the spread of the disease better than other prefectures in Japan in the early stages of the epidemic because of its rapid epidemiological investigations in both the first and second waves (cumulative number of COVID-19 cases in neighboring prefectures as of May 2020, approximately 200 in Toyama Prefecture and approximately 150 in Gunma Prefecture vs. approximately 70 in Niigata Prefecture) (Ministry of Health Labour and Welfare, 2020c). This led to the identification of close contacts, a 2-week stay-at-home order, and prompt testing of those with symptoms. Additionally, measures such as early restrictions on in-country movement from high-risk areas with high numbers of infected people, such as Tokyo, to regional cities would have better ensured that outbreaks did not spread in the early stages of the epidemic and that multiple clusters of ongoing community transmission were prevented. The COVID-19 epidemic in Niigata City was brought briefly under control when a state of emergency was declared in Niigata Prefecture in late April 2020. Given that Japan did not adopt a strict lockdown policy to weaken the spread of the disease, it is worth noting that restricting human movement may have been effective in preventing the spread of COVID-19. Particularly in the early stages of an emerging infectious disease outbreak, including SARS-CoV-2, protecting vulnerable local and regional health systems, and securing time for these areas to prepare for an epidemic can reduce the economic impact of a pandemic. As demonstrated in this preliminary study, combining epidemiological and genomic data is likely to provide useful insights for use in future public health intervention policies, with the possibility of applying these policies to other affected regions.

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

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

KW, RyS, SY, MI, YT, AN, and ReS designed the study. RyS, SY, MI, YT, AN, and TT collected the data. KW, RyS, TS, MK, TA, WWP, and ReS analyzed the data. KW drafted the manuscript. ReS supervised the study. RyS, SY, MI, YT, AN, MO, TS, MK, TA, TT, and HM made critical revisions to the manuscript for important intellectual content and gave final approval of the manuscript. All authors took responsibility for the integrity of the data and the accuracy of the data analysis. The opinions, results, and conclusions reported in this manuscript are those of the authors and are independent of the funding bodies.

FUNDING

KW was supported by the Community Medical Research Grant from the Niigata City Medical Association (Grant No. GC03220213) and the BIKEN Taniguchi Scholarship from the Research Institute for Microbial Diseases, Osaka University, Osaka, Japan (RIMD). ReS was supported by the Japan Initiative for Global Research Network on Infectious Diseases (J-GRID) from the Japan Agency for Medical Research and Development (AMED) (Grant Nos. 15fm0108009h0001–1wm0125005h0002), the Health and Labor Sciences Research Grants, Ministry of Health, Labor and Welfare, Japan (Grant Nos. H30-Shinkougyousei-Shitei-002 and H30-Shinkougyousei-Shitei-004), and the Niigata Prefecture Coronavirus Infectious Disease Control Research and Human Resources Development Support Fund. The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the paper. The authors had full access to all the data in the study and had the final responsibility to submit it for publication.

ACKNOWLEDGMENTS

We would like to thank the staff of all participating medical institutions and the staff of the Division of International Health

(Public Health), Graduate School of Medical and Dental Sciences, Niigata University, Niigata, Japan. We are deeply thankful to the staff of the local public health centers, especially the Niigata City Public Health Center and the Field Epidemiology Training Program of Japan (FETP-J) for their field epidemiological studies and for their role in preserving and providing the clinical specimens. We thank the Ministry of Health, Labor and Welfare, Japan, and local governments, especially the Niigata Prefectural Office, the Niigata City Office, the Niigata Prefectural Institute of Public Health and Environmental Science, and the Niigata City Institute of Public Health and Environmental Science for their assistance with administrative matters, field investigation, data collection, and laboratory testing. We acknowledge the authors from the laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via the GISAID database. We would like to thank Editage (www.editage.com) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.749149/full#supplementary-material

REFERENCES

Alteri, C., Cento, V., Piralla, A., Costabile, V., Tallarita, M., Colagrossi, L., et al. (2021). Genomic epidemiology of SARS-CoV-2 reveals multiple lineages and early spread of SARS-CoV-2 infections in Lombardy, Italy. *Nat. Commun.* 12:434. doi: 10.1038/s41467-020-20688-x

Arons, M. M., Hatfield, K. M., Reddy, S. C., Kimball, A., James, A., Jacobs, J. R., et al. (2020). Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N. Engl. J. Med.* 382, 2081–2090. doi: 10.1056/NEJMoa2008457

Badaoui, B., Sadki, K., Talbi, C., Salah, D., and Tazi, L. (2021). Genetic diversity and genomic epidemiology of SARS-CoV-2 in Morocco. *Biosaf. Health* 3, 124–127. doi: 10.1016/j.bsheal.2021.01.003

Bandelt, H. J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Bullard, J., Dust, K., Funk, D., Strong, J. E., Alexander, D., Garnett, L., et al. (2020). Predicting infectious severe acute respiratory syndrome coronavirus 2 from diagnostic samples. *Clin. Infect. Dis.* 71, 2663–2666. doi: 10.1093/cid/ciaa638

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Centers for Disease Control and Prevention (2021). *Science Brief: SARS-CoV-2 and Surface (Fomite) Transmission for Indoor Community Environments*. Available online at: https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/surface-transmission.html (accessed June 2, 2021).

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Coil, D., Jospin, G., and Darling, A. E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31, 587–589. doi: 10.1093/bioinformatics/btu661

da Silva Filipe, A., Shepherd, J. G., Williams, T., Hughes, J., Aranday-Cortes, E., Asamaphan, P., et al. (2021). Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland. *Nat. Microbiol.* 6, 112–122. doi: 10.1038/s41564-020-00838-z

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

Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., Candido, D. D. S., Mishra, S., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* 372, 815–821. doi: 10.1126/science.abh2644

Furuse, Y., Ko, Y. K., Saito, M., Shobugawa, Y., Jindai, K., Saito, T., et al. (2020a). Epidemiology of COVID-19 outbreak in Japan, from January-March 2020. *Jpn. J. Infect. Dis.* 73, 391–393. doi: 10.7883/yoken.JJID.2020.271

Furuse, Y., Sando, E., Tsuchiya, N., Miyahara, R., Yasuda, I., Ko, Y. K., et al. (2020b). Clusters of coronavirus disease in communities, Japan, January-April 2020. *Emerg. Infect. Dis.* 26, 2176–2179. doi: 10.3201/eid2609.202272

Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., et al. (2020). Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveill.* 25:2000257. doi: 10.2807/1560-7917.ES.2020.25.17.2000257

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

Geoghegan, J. L., Ren, X., Storey, M., Hadfield, J., Jelley, L., Jefferies, S., et al. (2020). Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. *Nat. Commun.* 11:6351. doi: 10.1038/s41467-020-20235-8

Groves, L. M., Usagawa, L., Elm, J., Low, E., Manuzak, A., Quint, J., et al. (2021). Community transmission of SARS-CoV-2 at three fitness facilities – Hawaii, June-July 2020. *MMWR Morb. Mortal. Wkly. Rep.* 70, 316–320. doi: 10.15585/mmwr.mm7009e1

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

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* 26, 672–675. doi: 10.1038/s41591-020-0869-5

Heymann, D. L., and Shindo, N. (2020). COVID-19: what is next for public health? *Lancet* 395, 542–545. doi: 10.1016/S0140-6736(20)30374-3

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Huang, L., Zhang, X., Zhang, X., Wei, Z., Zhang, L., Xu, J., et al. (2020). Rapid asymptomatic transmission of COVID-19 during the incubation period demonstrating strong infectivity in a cluster of youngsters aged 16–23 years outside Wuhan and characteristics of young patients with COVID-19: a prospective contact-tracing study. *J. Infect.* 80, e1–e13. doi: 10.1016/j.jinf.2020.03.006

Imamura, T., Saito, T., and Oshitani, H. (2021). Roles of public health centers and cluster-based approach for COVID-19 response in Japan. *Health Secur.* 19, 229–231. doi: 10.1089/hs.2020.0159

Itokawa, K., Sekizuka, T., Hashino, M., Tanaka, R., and Kuroda, M. (2020). *nCoV-2019 Sequencing Protocol for Illumina*. Available online at: https://dx.doi.org/10.17504/protocols.io.bnn7mdhn (accessed September 17, 2021).

Jayaweera, M., Perera, H., Gunawardana, B., and Manatunge, J. (2020). Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. *Environ. Res.* 188:109819. doi: 10.1016/j.envres.2020.109819

Johansson, M. A., Quandelacy, T. M., Kada, S., Prasad, P. V., Steele, M., Brooks, J. T., et al. (2021). SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Netw. Open* 4:e2035057. doi: 10.1001/jamanetworkopen.2020.35057

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

Kanda, Y. (2013). Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant.* 48, 452–458. doi: 10.1038/bmt.2012.244

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

Komissarov, A. B., Safina, K. R., Garushyants, S. K., Fadeev, A. V., Sergeeva, M. V., Ivanova, A. A., et al. (2021). Genomic epidemiology of the early stages of the SARS-CoV-2 outbreak in Russia. *Nat. Commun.* 12:649. doi: 10.1038/s41467-020-20880-z

Laiton-Donato, K., Villabona-Arenas, C. J., Usme-Ciro, J. A., Franco-Muñoz, C., Álvarez-Díaz, D. A., Villabona-Arenas, L. S., et al. (2020). Genomic epidemiology of severe acute respiratory syndrome coronavirus 2, Colombia. *Emerg. Infect. Dis.* 26, 2854–2862. doi: 10.3201/eid2612.202969

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

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

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

Matsunaga, N., Hayakawa, K., Terada, M., Ohtsu, H., Asai, Y., Tsuzuki, S., et al. (2020). Clinical epidemiology of hospitalized patients with COVID-19 in Japan: report of the COVID-19 registry Japan. *Clin. Infect. Dis.* 2020:ciaa1470. doi: 10.1093/cid/ciaa1470

Ministry of Health Labour and Welfare (2020a). *Cabinet Order Establishing COVID-19 as Designated Infectious Disease*. Available online at: https://www.mhlw.go.jp/content/10900000/000589748.pdf (accessed June 1, 2021).

Ministry of Health Labour and Welfare (2020b). *Treatment of Discharge and Work Restrictions for COVID-19 Patients Under the Law Concerning the Prevention of Infectious Diseases and Medical Care for COVID-19 Patients*. Available online at: https://www.mhlw.go.jp/stf/newpage_09346.html (accessed June 2, 2021).

Ministry of Health Labour and Welfare (2020c). *Press Releases on COVID-19 Infection in Japan (Outbreaks, Patient Outbreaks in Japan, Airport and Seaport Quarantine Cases, Overseas Situation, Mutant Strains)*. Available online at: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00086.html (accessed June 2, 2021).

Morawska, L., and Milton, D. K. (2020). It is time to address airborne transmission of coronavirus disease 2019 (COVID-19). *Clin. Infect. Dis.* 71, 2311–2313. doi: 10.1093/cid/ciaa939

Murano, Y., Ueno, R., Shi, S., Kawashima, T., Tanoue, Y., Tanaka, S., et al. (2021). Impact of domestic travel restrictions on transmission of COVID-19 infection using public transportation network approach. *Sci. Rep.* 11:3109. doi: 10.1038/s41598-021-81806-3

Muto, K., Yamamoto, I., Nagasu, M., Tanaka, M., and Wada, K. (2020). Japanese citizens' behavioral changes and preparedness against COVID-19: an online survey during the early phase of the pandemic. *PLoS One* 15:e0234292. doi: 10.1371/journal.pone.0234292

National Institute of Infectious Diseases (2020). *Provisional Guidelines for Active Epidemiological Surveillance of COVID-19 Patients in Japan*. Available online at: https://www.niid.go.jp/niid/ja/diseases/ka/corona-virus/2019-ncov/2484-idsc/9357-2019-ncov-02.html (accessed June 2, 2021).

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

Niigata City Institute of Public Health and Environment (2020). *Annual Reports of Niigata City Institute of Public Health and Environment*. Available online at: https://www.city.niigata.lg.jp/iryo/shoku/syokuei/shokueishisetsu/eisei_ken/eiken_chousa.html (accessed June 2, 2021).

Ogushi, Y. (2021). "Current status and issues for urban (regional area) formulation of the location normalization plan: the case of Niigata City," in *Frontiers of Real Estate Science in Japan. New Frontiers in Regional Science: Asian Perspectives*, Vol. 29, eds Y. Asami, Y. Higano, and H. Fukui (Singapore: Springer), 289–295. doi: 10.1007/978-981-15-8848-8_20

Oshitani, H. (2020). Cluster-based approach to coronavirus disease 2019 (COVID-19) response in Japan, from February to April 2020. *Jpn. J. Infect. Dis.* 73, 491–493. doi: 10.7883/yoken.JJID.2020.363

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., et al. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2:e000056. doi: 10.1099/mgen.0.000056

Perera, R., Tso, E., Tsang, O. T. Y., Tsang, D. N. C., Fung, K., Leung, Y. W. Y., et al. (2020). SARS-CoV-2 virus culture and subgenomic RNA for respiratory specimens from patients with mild coronavirus disease. *Emerg. Infect. Dis.* 26, 2701–2704. doi: 10.3201/eid2611.203219

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

Popa, A., Genger, J. W., Nicholson, M. D., Penz, T., Schmid, D., Aberle, S. W., et al. (2020). Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.* 12:eabe2555. doi: 10.1126/scitranslmed.abe2555

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

Sekizuka, T., Itokawa, K., Hashino, M., Kawano-Sugaya, T., Tanaka, R., Yatsu, K., et al. (2020a). A genome epidemiological study of SARS-CoV-2 introduction into Japan. *mSphere* 5:e00786-20. doi: 10.1128/mSphere.00786-20

Sekizuka, T., Itokawa, K., Kageyama, T., Saito, S., Takayama, I., Asanuma, H., et al. (2020b). Haplotype networks of SARS-CoV-2 infections in the Diamond Princess cruise ship outbreak. *Proc. Natl. Acad. Sci. U.S.A.* 117, 20198–20201. doi: 10.1073/pnas.2006824117

Sekizuka, T., Kuramoto, S., Nariai, E., Taira, M., Hachisu, Y., Tokaji, A., et al. (2020c). SARS-CoV-2 genome analysis of Japanese travelers in Nile River cruise. *Front. Microbiol.* 11:1316. doi: 10.3389/fmicb.2020.01316

Seto, J., Aoki, Y., Komabayashi, K., Ikeda, Y., Sampei, M., Ogawa, N., et al. (2021). Epidemiology of coronavirus disease 2019 in Yamagata Prefecture, Japan, January-May 2020: the importance of retrospective contact tracing. *Jpn. J. Infect. Dis.* (in press). doi: 10.7883/yoken.JJID.2020.1073

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

Singanayagam, A., Patel, M., Charlett, A., Lopez Bernal, J., Saliba, V., Ellis, J., et al. (2020). Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill.* 25:2001483. doi: 10.2807/1560-7917.ES.2020.25.32.2001483

Tang, J. W., Marr, L. C., Li, Y., and Dancer, S. J. (2021). Covid-19 has redefined airborne transmission. *BMJ* 373:n913. doi: 10.1136/bmj.n913 van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., et al. (2020). Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* 382, 1564–1567. doi: 10.1056/NEJMc2004973

Wagatsuma, K., Phyu, W. W., Osada, H., Tang, J. W., and Saito, R. (2021). Geographic correlation between the number of COVID-19 cases and the number of overseas travelers in Japan, Jan-Feb, 2020. *Jpn. J. Infect. Dis.* 74, 157–160. doi: 10.7883/yoken.JJID.2020.471

Wang, C., Horby, P. W., Hayden, F. G., and Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *Lancet* 395, 470–473. doi: 10.1016/S0140-6736(20)30185-9

Wellcome Trust ARTIC Network (2021). *ARTIC Network Protocol*. Available online at: https://artic.network/ncov-2019 (accessed June 1, 2021).

World Health Organization (2021). *Coronavirus Disease (COVID-19) Pandemic*. Available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed June 1, 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 | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 04 April 2022
doi: 10.3389/fmicb.2022.693196

Phylogenetic and Spatiotemporal Analyses of the Complete Genome Sequences of Avian Coronavirus Infectious Bronchitis Virus in China During 1985–2020: Revealing Coexistence of Multiple Transmission Chains and the Origin of LX4-Type Virus

Wensheng Fan[†], Jiming Chen[†], Yu Zhang[†], Qiaomu Deng, Lanping Wei, Changrun Zhao, Di Lv, Liting Lin, Bingsha Zhang, Tianchao Wei, Teng Huang, Ping Wei and Meilan Mo**

College of Animal Science and Technology, Guangxi University, Nanning, China

Infectious bronchitis (IB) virus (IBV) causes considerable economic losses to poultry production. The data on transmission dynamics of IBV in China are limited. The complete genome sequences of 212 IBV isolates in China during 1985–2020 were analyzed as well as the characteristics of the phylogenetic tree, recombination events, dN/dS ratios, temporal dynamics, and phylogeographic relationships. The LX4 type (GI-19) was found to have the highest dN/dS ratios and has been the most dominant genotype since 1999, and the Taiwan-I type (GI-7) and New type (GVI-1) showed an increasing trend. A total of 59 recombinants were identified, multiple recombination events between the field and vaccine strains were found in 24 isolates, and the 4/91-type (GI-13) isolates were found to be more prone to being involved in the recombination. Bayesian phylogeographic analyses indicated that the Chinese IBVs originated from Liaoning province in the early 1900s. The LX4-type viruses were traced back to Liaoning province in the late 1950s and had multiple transmission routes in China and two major transmission routes in the world. Viral phylogeography identified three spread regions for IBVs (including LX4 type) in China: Northeastern China (Heilongjiang, Liaoning, and Jilin), north and central China (Beijing, Hebei, Shanxi, Shandong, and Jiangsu), and Southern China (Guangxi and Guangdong). Shandong has been the epidemiological center of IBVs (including LX4 type) in China. Overall, our study highlighted the reasons why the LX4-type viruses had become the dominant genotype and its origin and transmission routes, providing more targeted strategies for the prevention and control of IB in China.

Keywords: infectious bronchitis virus, recombination, selective pressure, Bayesian phylogenetics, viral phylodynamics, spatial transmission

OPEN ACCESS

Edited by:
Ian George Barr,
WHO Collaborating Centre for Reference and Research on Influenza, VIDRL, Australia

Reviewed by:
Silvia Spoto,
Policlinico Universitario Campus Bio-Medico, Italy
Muhammad Zubair Shabbir,
University of Veterinary and Animal Sciences, Pakistan

Correspondence:
Ping Wei
pingwei8@126.com
Meilan Mo
momeilan@163.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: *10 April 2021*
Accepted: *09 February 2022*
Published: *04 April 2022*

Citation:
Fan W, Chen J, Zhang Y, Deng Q, Wei L, Zhao C, Lv D, Lin L, Zhang B, Wei T, Huang T, Wei P and Mo M (2022) Phylogenetic and Spatiotemporal Analyses of the Complete Genome Sequences of Avian Coronavirus Infectious Bronchitis Virus in China During 1985–2020: Revealing Coexistence of Multiple Transmission Chains and the Origin of LX4-Type Virus. Front. Microbiol. 13:693196. doi: 10.3389/fmicb.2022.693196

INTRODUCTION

Avian infectious bronchitis (IB) is an economically critical poultry disease worldwide. IB virus (IBV) is a member of the *Coronaviridae* family. Coronavirus infects humans and many other animals. Since the beginning of this century, three coronaviruses (SARS-CoV, MERS-CoV, and SARS-CoV-2) have infected humans and caused death (Drosten et al., 2003; Zaki et al., 2012; Lauer et al., 2020). In particular, the COVID-19 outbreak caused by SARS-CoV-2 in late 2019 has created a serious public health problem worldwide. Since people have close contact with animals, and coronavirus is prone to mutation and recombination, it is unpredictable whether recombination will occur between SARS-CoV-2 and animal coronavirus. Therefore, it is necessary to strengthen the prevention and control of animal coronavirus infections, including IB.

Infectious bronchitis virus has a single-stranded positive-sense RNA genome of approximately 27.6 kilobases (kb) in length encoding four structural proteins: spike (S) glycoprotein, envelope (E) protein, membrane (M) glycoprotein, and nucleocapsid (N) protein (Cavanagh, 2007; Fan et al., 2019). The S protein is cleaved into S1 and S2 subunits by furin-like host cell proteases (Cavanagh et al., 1992). The S1 protein mediates the binding of the virus to host cells and is the most variable region, which contains serotype-specific and virus-neutralizing epitopes (Ignjatovic and Sapats, 2005). Therefore, the sequence characteristics and evolution rules of S1 gene are often used for molecular epidemiological analysis of IBV.

Multiple IBV genotypes, serotypes, or variants have been identified globally. IBV strains worldwide are classified into seven genotypes comprising thirty-five distinct viral lineages based on the complete S1 gene sequences (Molenaar et al., 2020). In China, the IBV strains are divided into multiple distinct groups, and the LX4 type (QX or GI-19) appears to be the dominant genotype (Xu et al., 2018). The LX4-type strain was first isolated in China in 1996 (Wang et al., 1997) and has spread to several neighboring countries and even to Europe (Liu et al., 2006; Huang et al., 2019). At present, LX4-type viruses have already become the dominant epidemic strains in the world (Zhao et al., 2017; Huang et al., 2019). However, where did the LX4 type come from? Why did the LX4 type evolve into a major dominant epidemic strain in the world? These problems still remain unclear. Further investigation is needed.

As mentioned above, many molecular epidemiological studies based on S1 gene of IBV have been reported (Valastro et al., 2016; Bande et al., 2017; Huang et al., 2019; Molenaar et al., 2020). However, these studies could not fully explain the changes in the pathotypes and serotypes of IBV variants (Lin and Chen, 2017; Fan et al., 2019). Recently, there have been some studies based on the variation of other structural proteins of IBV (Finger et al., 2018; Yuan et al., 2018; Liang et al., 2019; Wu et al., 2019), indicating that these proteins also play important roles in the evolution of IBV. However, the abovementioned studies mainly focused on S1 gene, occasionally M and N genes, but rarely analyzed all the four structural protein genes and the full-length genome. In order to obtain comprehensive genetic information and molecular mechanism of variation of circulating isolates, it is necessary to analyze all the structural proteins of IBV simultaneously and even the whole genome.

Recombination is an important contributing factor in the emergence and evolution of IBV or even the emergence of new coronaviruses or novel diseases (Jackwood et al., 2010). Recombination events could occur between the field and vaccine viruses or between the field viruses (Fan et al., 2019). So far, the recombinants between the field and vaccine viruses were common. However, there are multiple live-attenuated vaccines used in the field in China, and little attention has been paid to which live vaccine is more likely to participate in the recombination. In addition, many previous recombination descriptions merely analyzed the recombination of the S1 or S genes, but the S1 or S genes were only a small part of the IBV genome and could not represent all the biological characteristics and recombination events of the whole genome, resulting in the failure to fully understand the evolution rules of IBV. Moreover, few studies on recombination hotspots are available. Hence, a thorough and comprehensive recombination analysis using the full genome sequences of as many strains as possible is needed.

China is the major producer of poultry worldwide with so many chicken breeds, differing ages, and a wide variety of rearing patterns (Huang et al., 2019). Yellow chickens are the major local breeds in southern China where about 5 billion birds were produced in 2019 (Deng et al., 2021). Compared to the white-feather chickens, the yellow chickens have a long raising time (about 120 days), and the free-range style is used with ineffective biosecurity measures due to the open fields and mountainsides (Wei, 2019; Deng et al., 2021). Moreover, some other inadequate biosecurity measures were applied in these flocks, such as lack of following the all-in, all-out basis, multi-age chicken farming, close distance between different-age flocks, different vaccination programs, and incomplete disinfection between each batch of the birds (Deng et al., 2021), which increase the odds of multiple infections of several IBV strains. In recent years, due to the frequent vaccination failure, many chicken farms tended to administer more frequent immunization and simultaneous administration of multiple-type live-attenuated vaccines, which promoted more frequent recombinations and mutations. Therefore, the prevention and control of IB are facing greater challenges and difficulties. In order to provide more targeted strategies to prevent and control IB, a whole-genome sequence analysis with longer time spans, that is more comprehensive, and with more isolates is needed.

Although there were many previous studies that focused on the genetic analysis of IBVs, most of the studies focused on the phylogenetic and recombination analyses of S1 gene (Yan et al., 2019; Ren et al., 2020; Zhang et al., 2020). The number of IBV strains and the spanning time were limited. There were few studies on the issues of defining the history of IBV and the spread trajectories in China. There was little information about its population dynamics over time and the major routes of its domestic spread at the national level. It is necessary to carry out long-term, in-depth, systematic epidemiological monitoring of IBVs in China. Hence, we conducted the analyses of phylogenetic, recombination, dN/dS ratios, temporal dynamics, phylogeographic analysis, and demographic history of 212 IBV strains isolated in China during the years 1985–2020 in the present study, aiming to obtain a comprehensive insight into the evolution and spread of IBV in China and provide more targeted strategies for the prevention and control of IB in China and even around the world, as well as a reference for the prevention and control of other animal coronavirus diseases and even the COVID-19 pandemic.

MATERIALS AND METHODS

Infectious Bronchitis Virus Isolates and Sequence Data

All complete genome sequences of Chinese IBV isolates were downloaded from the GenBank database on December 31, 2020[1]. Only the complete genome sequences with known sampling information, including dates (years) and locations (provinces) were selected from the GenBank database. On this basis, those sequences with a length greater than 27,000 nt were further selected for analysis to ensure robust statistical support. The complete genome sequences of 212 IBV isolates from 1985 to 2020 were available for analysis after the abovementioned screening (**Supplementary Tables 1**, **2**). According to the GenBank database, the whole-genome sequences of 212 Chinese IBV isolates were obtained by Sanger dideoxy sequencing technology. The earliest eligible IBV strain sequence available from the GenBank database was GX-C isolated in 1985, which was sequenced and submitted to the GenBank database by our group (Mo et al., 2013; Fan et al., 2019). The digital map (**Supplementary Figure 1**) was obtained from the Institute of Geographic Sciences and Natural Resources Research in China (resource and environment data cloud platform[2]). These 212 IBV isolates were isolated in 22 provinces or autonomous cities (zones) with a wide geographical range in China (**Supplementary Figure 1**). The available complete sequences of 33 reference IBV strains (**Supplementary Table 3**) included the commonly used live vaccine strains in China, and the dominant genotypes and serotypes around the world (Thor et al., 2011; Mo et al., 2013; Zhao et al., 2016, 2017; Fan et al., 2019; Huang et al., 2020). Among all complete genome sequences of IBV isolates from other countries retrieved from the GenBank database, 68 IBV isolates with known sampling dates and geographic locations were chosen for analysis together (**Supplementary Tables 4**, **5**). Different datasets, including the sequences of the S1, S2, E, M, and N structural genes, were extracted from the complete genome sequences and used for analyses.

Gene Alignment and Phylogenetic Analysis

The complete genome sequences and structural genes S1, S2, E, M, and N of 212 Chinese isolates and 33 reference strains were aligned using the Mafft version 7[3] (Katoh and Standley, 2013). The lowest Bayesian information criterion (BIC) was calculated using Jmodeltest[4] (Darriba et al., 2012). The best-fitting substitution model (GTR + G + I) was selected. Phylogenetic trees based on 212 Chinese isolates and 33 reference strains were constructed using MEGA version 6.06 according to the previous description (Tamura et al., 2013). Phylogenetic trees based on the complete genome sequences of 68 IBV isolates from other countries were also constructed as the above method.

Alignment of the deduced amino acid (aa) sequences of the S1 protein was performed for the 212 Chinese isolates by comparing with the vaccine strains M41, H120, 4/91, QXL87, and LDT3-A, which were commonly used in China, and a field strain LX4, representing the dominant IBV genotype circulating in China. The minimal receptor-binding domain (RBD) and three hypervariable regions (HVRs) were compared by the aa sequences of S1 protein.

Recombination Detection

Two different approaches were used to identify the potential recombination events that happened among the IBV complete genomes, S1, S2, E, M, and N genes of the 212 Chinese isolates. First, the complete genomes and the five genes were detected for potential recombinants in the sequence alignment using seven methods (RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, and 3SEQ) available in the Recombination Detection Program (RDP4, Version 4.95, Simmonics, University of Warwick, Coventry, United Kingdom) (Martin, 2009). To minimize false positives for the complete genomes, recombination events were only considered to be significant if they were supported by at least four of the seven methods with a significance value lower than 10^{-12} (p-value < 10^{-12}) (Thor et al., 2011). To minimize false positives for the S1, S2, E, M, and N genes, recombination events were only considered to be significant if they were supported by at least four of the seven methods with a significance value lower than 10^{-6} (p-value < 10^{-6}) (Franzo et al., 2017). Second, the complete genomes and the five genes further verified the potential recombination events and the breakpoints by similarity plots (SimPlots) analysis in SimPlot version 3.5.1 (Fan et al., 2019), with a window of 200 bp and step size of 20 bp. Both Simplot and RDP4 statistically identified each of the beginning and ending breakpoints (with no gaps) of the recombinant strains.

Estimation of dN/dS Ratios

The dN/dS ratios within the proteins S1, S2, E, M, and N from the 212 Chinese IBV isolates were analyzed by the SLAC, FEL, and IFEL methods of Datamonkey version[5] (Jackwood and Lee, 2017). The dN/dS ratios of the S1 protein of different genotypes of IBV were also analyzed as the abovementioned method. The maximum likelihood (ML) phylogenetic trees and the best-fit model (GTR + G + I) of nucleotide substitution were used for analyzing all the proteins. The recombinants were excluded in order to reduce false detection of the dN/dS ratios.

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

[2] http://www.resdc.cn/data.aspx?dataid=243

[3] http://mafft.cbrc.jp/alignment/software/

[4] https://github.com/ddarriba/jmodeltest2

[5] http://www.datamonkey.org/

Temporal Dynamics of Infectious Bronchitis Virus

To infer the evolutionary rate, the time to the most recent common ancestor (tMRCA), and population dynamics of the Chinese IBV, the Bayesian Markov chain Monte Carlo (MCMC) implemented in BEAST version 1.10.4 package[6] (Suchard et al., 2018) was used. The recombinants were excluded in order to reduce the effect of evolutionary parameter estimation. To test the temporal signal in the dataset, a regression of root-to-tip genetic distances against the year of sampling in TempEst v 1.5.1[7] (Rambaut et al., 2016) was performed using the unrooted ML tree generated with IQ-TREE v1.6.12 (Nguyen et al., 2015). The best-fitting root of the tree that optimizes the temporal signal was selected. The GTR + G + I substitution models were used for the complete genome sequences based on the result of jModelTest. BEAST version 1.10.4 package was set to both the strict molecular clock model and the uncorrected lognormal relaxed molecular clock model with two demographic models (constant size and Bayesian skyline). Four pairs of models were generated, and each pair was run for 5 million generations with every 50,000 cycles sampled to ensure the convergence of relevant parameters (i.e., effective sample sizes [ESSs] > 200) using Tracer v1.7 (Rambaut et al., 2018). The fitting clock rate and virus population evolutionary model were compared using ACIM, and a strict molecular clock and Bayesian skyline were selected (**Supplementary Table 6**). The MCMC was run for at least 500 million generations, allowing 10% burn-in and sub-sampling every 5 million steps. Convergence was evaluated by calculating the ESS of the parameters with ESS > 200 using Tracer v1.7 (Rambaut et al., 2018). The tree was summarized as maximum clade credibility (MCC) tree using Tree Annotator v1.10.4, and then the created MCC tree was visualized using FigTree v1.4.3. The change of effective population size over time was inferred by Bayesian skyline plots (Rambaut et al., 2018).

Phylogeographic Analysis and Demographic History of Infectious Bronchitis Virus

To gain insight into the circulation of IBV across the poultry producing regions in China, the spatial transmission patterns were reconstructed using a phylogeographic analysis (Baele et al., 2012) in BEAST 1.10.4, based on the dataset of complete genome sequences of Chinese IBV isolates. To infer the diffusion rates among geographic locations, the asymmetric substitution model with the Bayesian stochastic search variable selection (BSSVS) was selected in BEAST v1.10.4, also allowing Bayes factors (BF) calculations to test for significant diffusion rates in SPREAD3 v0.9.7 (Bielejec et al., 2016). Spread pathways were considered to be important when they yielded a BF greater than 3 and when the mean posterior value of the corresponding was greater than 0.50. The MCMC simulations for 500 million steps across three independent Markov chains were run, and samples of every 50,000 steps were collected. The ESS of each parameter was also required to be more than 200.

In order to further explore the evolutionary origin of the LX4 type of IBV, phylogeographic analysis of the LX4-type viruses was performed based on the complete genome sequences of IBV from China and other countries. The MCMC analysis was run for 300 million steps with a sampling of every 30,000 steps. Other parameters were the same as the above method.

RESULTS

Phylogenetic Analysis

Based on the phylogenetic tree analyses of the complete genome and S1, S2, E, M, and N genes, the 212 Chinese IBV isolates were segregated into 6, 7, 7, 6, 6, and 5 distinct groups, respectively (**Figures 1A–F**). There were 71.70% (152/212), 56.60% (120/212), 63.68% (135/212), 64.62% (137/212), 64.62% (137/212), and 66.04% (140/212) of the IBV isolates clustered in the LX4 type, respectively according to the complete genome and S1, S2, E, M, and N genes. The LX4 type has been the predominant genotype since 1999 (**Supplementary Figure 2**). Among the complete genome sequences of 68 IBV isolates from other countries, 15 isolates (22.06%) belonged to the LX4 type (**Supplementary Figure 3**). The phylogenetic tree based on the complete genome sequences showed that all the Chinese IBV isolates were segregated into the LX4 type, Mass type, 4/91 type, Peafowl/GD/KQ6/2003 type, Taiwan-I type, and CK/CH/LSC/99I type, except a unique isolate GX-C (**Figure 1A**). According to the phylogenetic tree characteristics of the complete genome sequence of 212 Chinese IBV isolates analyzed in this study, the LX4 type was the predominant genotype since 1999; the Mass type, 4/91 type, and Taiwan-I type were the second, third, and fourth most dominant genotypes, respectively, but the Mass type and 4/91 type nearly disappeared after 2015. The phylogenetic tree based on S1 gene sequences (**Figure 1B**) showed that all the Chinese isolates were segregated into the LX4 type (GI-19, 56.60%, 120/212), Mass type (GI-1, 18.87%, 40/212), 4/91 type (GI-13, 11.32%, 24/212), Taiwan-I type (GI-7, 5.19%, 11/212), New type (GVI-1, 3.77%, 8/212), CK/CH/LSC/99I type (GI-22, 2.36%, 5/212), and LDT3-A type (GI-28, 1.4%, 3/212), respectively, except for the unique isolate GX-C. The LX4 type was the predominant genotype since 1999. The Mass type and the 4/91 type were the second and third most dominant genotypes, and their change trends were similar to those based on the complete genome sequences. However, the Taiwan-I type (GI-7) has increased since 2011, and the New type (GVI-1) increased since 2015.

The alignment results of S1 gene sequences with those of the vaccine strain M41, H120, 4/91, QXL87, LDT3-A, and the reference field strain LX4 revealed the presence of multiple mutations (**Supplementary Figure 4**). There was a 5-aa-insertion (FFLII), located in the HVR I found in S1 genes of 11 isolates.

Analysis of Recombinants

The recombination analysis showed that recombination events were found in the complete genome of 59 isolates (27.83%,

[6] http://beast.community/index.html

[7] http://tree.bio.ed.ac.uk

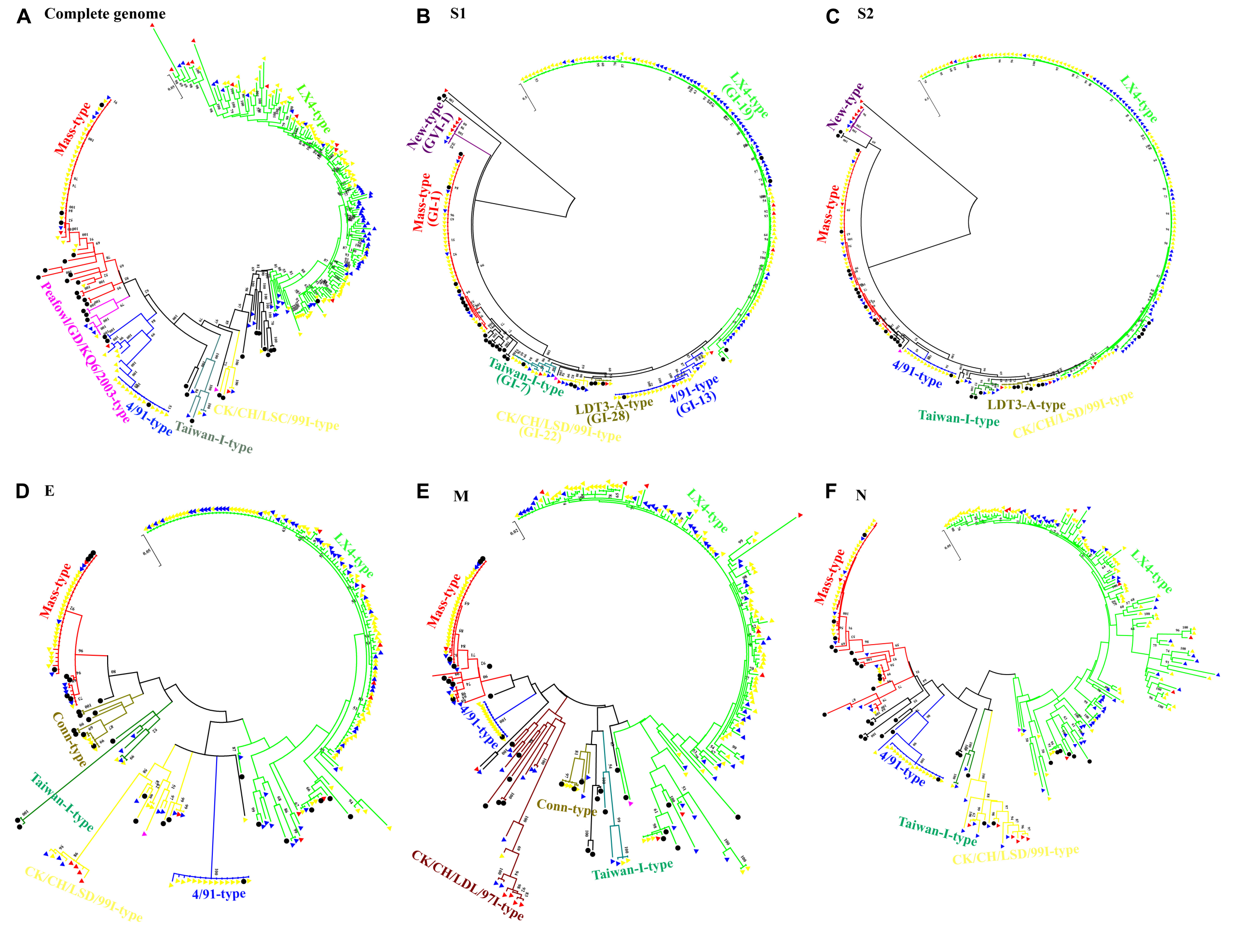


FIGURE 1 | Phylogenetic trees of the complete genome **(A)** and genes S1 **(B)**, S2 **(C)**, E **(D)**, M **(E)**, and N **(F)** of infectious bronchitis viruses (IBVs). Each type of IBV was highlighted in different colors. The 33 IBV reference strains were marked with filled ●. IBV isolates isolated during 1985–1998 were marked with filled ▲. The 69 IBV isolates isolated during 1999–2010 were marked with filled ▲. The 128 IBV isolates isolated during 2011–2014 were marked with filled ▲, and the 14 IBV isolates isolated during 2015–2020 were marked with filled ▲. Phylogenetic trees were constructed with the maximum likelihood method using MEGA version 6.06. The bootstrap clade support values were computed based on 1,000 bootstrap replicates. The number represents the percentage of times that the clade appeared in the bootstrap tree distribution. Bootstrap values greater than 50% are shown.

59/212), and 16 of them (7.55%, 16/212) were found in S1 gene (**Figure 2** and **Supplementary Tables 7**, **8**). The recombination regions were distributed throughout the entire genome. Based on the complete genome of 59 recombinant isolates, nsp3 and S genes were found to have the greatest number of recombination regions, with 28 and 26 recombination regions, respectively (**Figure 2A**). nsp16 (19 recombination regions) and N (17 recombination regions) genes were found to have the third and fourth greatest number of recombination regions. Based on S1 gene of 16 isolates, the midstream (600–700 bp) of S1 gene (9 recombination regions) was found to have the greatest number of recombination regions (**Figure 2B**). The downstream (1,500–1,611 bp) of S1 gene (8 recombination regions) was found to have the second greatest number of recombination regions. The detectable recombination breakpoint positions were located throughout the genome (**Supplementary Figure 5A**). A large number of breakpoints were observed in the 5′UTR, 1a, nsp3, nsp5, N, and 3′UTR. 5′UTR, nsp3, nsp16, S, and N genes were found to have the greatest number of transferred fragments (**Table 1**), which was consistent with the location and number of breakpoints in **Supplementary Figure 5A**. Moreover, the recombination breakpoint positions clustered in nsp3 and S genes. Thus, nsp3 and S genes were associated with the greatest number of transferred fragments. There was no breakpoint in the nsp8, nsp9, nsp10, and 3a regions. In addition, the detectable recombination breakpoint positions of S1 gene were shown in **Supplementary Figure 5B**. The recombination breakpoint positions were clustered within the midstream (600–700 bp) and downstream (1,500–1,611 bp) of S1 gene. In order to confirm the results of the RDP analysis, genomic sequence analyses of the complete genome of 59 isolates and S1 gene of 16 isolates were carried out by the Simplot software, and the results were consistent with those of the RDP analysis (**Supplementary Figures 6**, **7**).

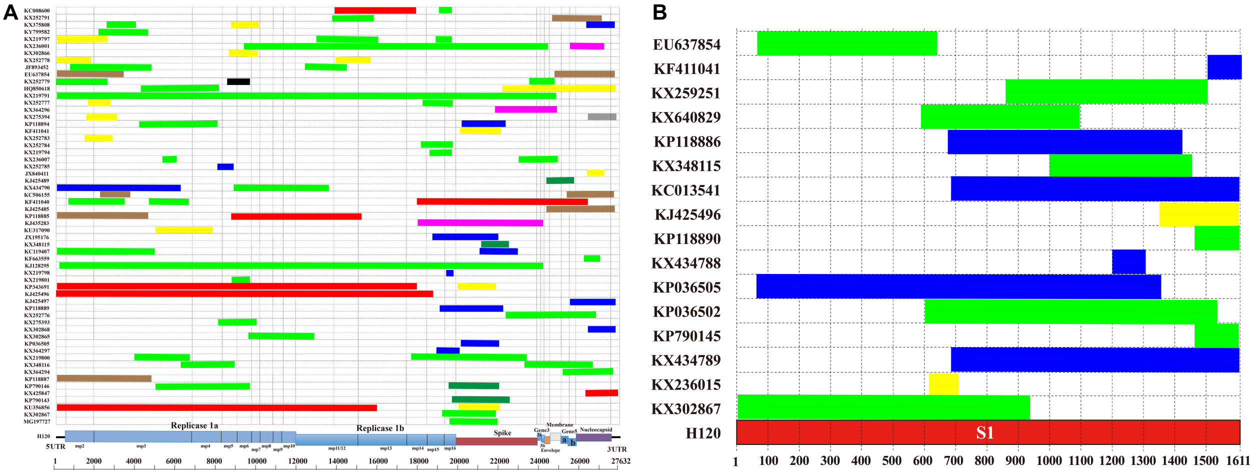


FIGURE 2 | The vaccine strain H120 was used as the reference strain. **(A)** The virus genome of the coronavirus avian infectious bronchitis showing the locations of 5′ and 3′ UTRs and the coding regions. **(B)** Recombination regions identified across S1 gene. Recombinant regions detected by RDP4 are highlighted in the line. The green, yellow, blue, red, dark brown, pink, dark green, black, and gray lines indicate the recombinant regions that belong to the genotype of LX4, CK/CH/LSC/99I, 4/91, Mass, Conn, Peafowl/GD/KQ6/2003, Taiwan-I, Lowa, and Georgia, respectively.

TABLE 1 | Number of transferred fragments associated with individual areas of the genome for all of the isolates in China.

Genomic region	Number of fragments	% of total
5′UTR	13	7.47
nsp2	6	3.45
nsp3	29	16.67
nsp4	5	2.87
nsp5	8	4.60
nsp6	5	2.87
nsp7	3	1.72
nsp8	0	0.00
nsp9	0	0.00
nsp10	0	0.00
nsp11/12	8	4.60
nsp13	5	2.87
nsp14	7	4.02
nsp15	8	4.60
nsp16	10	5.75
Spike	28	16.09
3a	0	0.00
3b	1	0.57
Envelope	3	1.72
Membrane	6	3.45
5a	0	0.00
5b	6	3.45
Nucleocapsid	13	7.47
3′UTR	10	5.75

UTR, untranslated region; nsp, non-structural protein.

In addition, based on the complete genome sequences, a total of 52 of 59 (88.14%), 14 of 59 (23.73%), and 11 of 59 (18.64%) recombinants were likely from the LX4-type, 4/91-type, and Mass-type isolates, respectively (**Supplementary Figure 5A** and **Supplementary Table 7**). There were 20 recombinants formed by two major parents and two minor parent strains. There were four recombinants formed by three major parents and three minor parent strains. So multiple recombination events between the field and vaccine strains might occur in a total of 24 of the recombinant isolates (40.68%, 24/59) (**Supplementary Table 7**). Based on S1 gene, a total of 11 of 16 (68.75%) and 8 of 16 (50%) recombinants were likely from the LX4-type and 4/91-type isolates, respectively (**Supplementary Figure 5B** and **Supplementary Table 8**).

Estimation of dN/dS Ratios on the S1, E, M, and N Proteins of Infectious Bronchitis Viruses

The results showed that the ratios of the mean substitution rates at the non-synonymous sites over those at the synonymous sites (dN/dS) for S1, S2, E, M, and N proteins of the 212 isolates were 0.301, 0.136, 0.225, 0.168, and 0.159, respectively (**Supplementary Table 9**). The dN/dS ratios for S1 protein of the LX4 type, New type, Taiwan-I type, Mass type, CK/CH/LSC99I type, 4/91 type, and LDT3-A type were 0.369, 0.358, 0.308, 0.307, 0.267, 0.239, and 0.174, respectively.

Evolutionary Rates and Population Dynamics Analysis

The temporal signal was assessed using TempEst software *via* root-to-tip regression analysis for the IBV isolates (correlation coefficient = 0.6576; R^2 = 0.4325). In this test, the dataset was considered to have a relatively strong temporal signal for the sampling year. The evolution rate of IBVs was estimated to be 3.15 × 10^{-3} nucleotide substitutions/site/year (95% highest posterior density [HPD] 3.01E–3 to 3.29E–3). The IBVs in the last 35 years were classified into 6 main clusters by plotting the MCC tree (**Figure 3**). The estimated tMRCA of all the IBVs in China was the year 1901.94 (95% HPD [1885.44 to 1910.84]) found in Liaoning province, and the times of origin (95% HPD) of the major branch of viruses (LX4 type) in China was 1955.98 (95% HPD [1948.06 to 1960.31]) in Liaoning province also (**Figure 3**). A Bayesian skyline plot coalescent model was used to assess the reconstruction of population history. **Figure 3** shows temporal changes in the effective population size of the viruses. Initially, the population diversity of the viruses showed a slow reduction before the 1980s. By 1980, the effective population size of the viruses had increased and undergone an upward trend after having fluctuated. The genetic diversity of the isolates increased rapidly and peaked until 2006 and a sudden and sharp decline in the effective population size was observed after 2006; thereafter, it persisted for 2 years. But since late 2008, the viral population has been rapidly rising again.

Phylogeography Analyses of Infectious Bronchitis Virus in China

The spatial dispersion patterns of IBVs in China during 1985–2020 were reconstructed. **Figure 4** and **Supplementary Table 10** show the major spread links of IBV. Generally, they spread between northern and southern China. The first transmission route was between the northern provinces Shandong, Hebei, and Beijing and the northeastern provinces Liaoning and Heilongjiang, with decisive support (BF > 100). The second transmission route was from Jiangsu to Shanxi with decisive support (BF > 100). The third transmission route was from Hebei to Henan and Inner Mongolia (BF > 30). The fourth transmission route was a long-distance movement event, notably from Heilongjiang to Guangdong and Guangxi (BF > 30). The fifth transmission route was from Hebei to Jilin, Hunan, Liaoning to Shandong, Shandong to Jilin, and Jiangsu (BF > 10). A long-distance movement event was from Shandong to Gansu (BF > 3). In accordance with the major spread links of IBV, Liaoning was the origin, which then propagated outwards in the early 1900s (**Figure 4A**). The directions of virus spread occurred predominantly from Liaoning to Heilongjiang in the late 1950s (**Figure 4B**), from the northeastern to northern in the early 1980s (**Figure 4C**), and from the northern to eastern and southern from 1980s to 2020 (**Figure 4D**). It is indicated that Shandong has been the epidemiological center of IBVs in China from the 1980s to 2020 (**Figure 4E** and **Supplementary Table 10**), based on the decisive spread routes from Shandong to Heilongjiang and Shandong to Hebei (BF > 1,000).

An additional temporal signal of the dataset (152 LX4-type isolates in China based on **Figure 1A** and 15 LX4-type isolates in other countries based on the **Supplementary Figure 3**) was assessed. The results of the MCC tree (**Figure 5A**) showed that the sequences of the LX4-type IBV isolates in China were closely related to the early isolates of Liaoning in the late 1950s (correlation coefficient = 0.5945; R^2 = 0.3534). The results of the BF value and the major spread links showed that the LX4-type isolates had multiple transmission routes in China (**Figure 6** and **Supplementary Table 11**). The first transmission route was between Shandong and Liaoning, and Jilin and Heilongjiang (BF > 100). The second transmission route was between Shandong, Hebei, Shanxi, and Jiangsu (BF > 100). The third transmission route was between Shandong and Beijing (BF > 30). The fourth transmission route was between Shandong, Hebei, and Anhui, Zhejiang (BF > 3). The fifth transmission route was a long-distance movement event, notably from Heilongjiang, Jilin to Guangxi and Guangdong (BF > 3), Liaoning to Xinjiang, Shandong to Gansu, and Hebei to Hunan (BF > 10). Besides, a short-distance movement event was also observed from Guangxi to Guangdong (BF > 3). In accordance with the major spread links of the LX4 genotype, Liaoning was the origin of LX4 genotype IBV, which then propagated outwards in the late 1950s (**Figures 3**, **5A**, **6**). The directions of virus spread occurred predominantly from Liaoning to Heilongjiang in the early 1960s (**Figure 6B**), from the northeastern to northern regions in the early 1980s (**Figure 6C**), and from northern to eastern and southern regions (**Figure 6D**). Similarly, it is indicated that Shandong has been the epidemiological center of the LX4 genotype (**Figure 6E** and **Supplementary Table 11**), based on the decisive spread routes from Shandong to Heilongjiang, Shandong to Hebei, and Shandong to Jiangsu (BF > 1,000).

The results of the MCC tree (**Figure 5B**) showed that the sequences of Europe's LX4-type isolates were closely related to the early isolates of China in the 1980s, which were in group 2. South Korea's LX4-type isolates were closely related to the late isolates of China in the 1980s, which were in group 1 (correlation coefficient = 0.612; R^2 = 0.3746). The results of BF value and the major spread links (**Figure 7**) showed that the LX4-type IBVs had two major transmission routes in the world: from China to Europe (Poland, Sweden, and so on) (BF > 10) and from China to South Korea (BF > 100).

DISCUSSION

The co-circulation of multiple IBV genotypes and the ongoing emergence of IBV variants have the greatest epidemiological challenges in China. In our study, a total of 6, 7, 7, 6, 6, and 5 genotypes were identified based on the sequences of the full-length genome and structural genes S1, S2, E, M, and N, respectively. According to the phylogenetic trees of both the whole genome and S1 gene, the top four dominant genotypes were the LX4 type, Mass type, 4/91 type, and Taiwan-I type, respectively. Therefore, our results showed that China was still in a state of coexistence of multiple genotypes with the LX4 type as the dominant genotype, which agreed with many previous descriptions (Liu et al., 2009; Zou et al., 2010; Li et al., 2012; Huang et al., 2019). We found that the LX4-type isolates have been dominant since 1999 in our study, which was earlier than those previous descriptions (Zhao et al., 2017; Jiang et al., 2018; Ren et al., 2020). The possible explanations are as follows: first, the isolates in our study were collected from 1985, while those in the previous studies were collected later than 1985. Second, the previous studies focused on the S1 gene sequences rather than the whole genome, but it is impossible to get a comprehensive understanding of the origins and evolutionary processes of those emerging viruses only by analyzing this small portion of the genome. In addition, the LX4 genotype was found

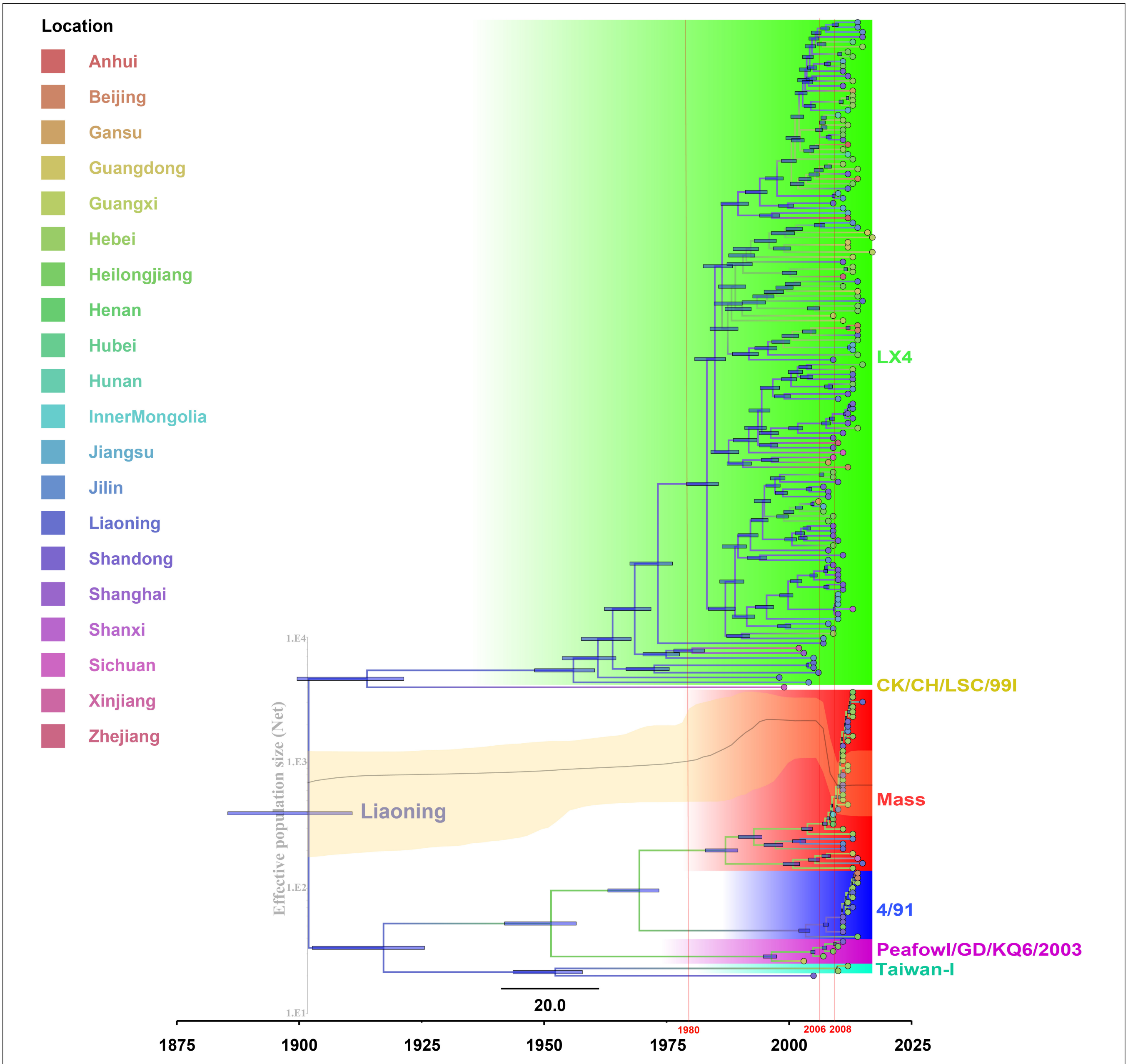


FIGURE 3 | The maximum clade credibility tree and Bayesian skyline plot (BSP) of full-length genome sequences of infectious bronchitis viruses (IBVs) (see section "Materials and Methods" for further details). The maximum clade credibility (MCC) tree was constructed with 10% burn-in by Tree Annotator version 1.10.4 implemented in the BEAST software package. The color of the branches corresponds to geographical areas. A change in branch color indicates a time and geographical area change during the genotype's evolutionary history. The scale bar represents the unit of time (year). Bayesian skyline plot (BSP) is superimposed to the tree in the same time-scale (*x*-axis). The *y*-axis represents the effective population size (log10 scale). The black line represents the effective population size through time, and the orange area around the line represents the 95% highest posterior density (HPD) of the estimates.

to have the highest dN/dS ratios compared with other genotypes. High dN/dS ratios may promote the viral fitness in the infected hosts (Zhao et al., 2016). Thus, this is an explanation for why LX4 was the dominant genotype in recent years. Besides, we found that the Taiwan-I-type and New-type isolates showed an increasing trend since 2011 and 2015, respectively, in the country, which agreed with other previous studies (Xu et al., 2018; Fan et al., 2019; Huang et al., 2020), indicating that the commonly used commercial vaccines could not provide complete protection against these two genotype isolates. The development of new vaccines targeting Taiwan-I-type and New-type isolates should be of practical significance currently.

Recombination among coronaviruses could decrease mutational load, create genetic variation, and lead to the

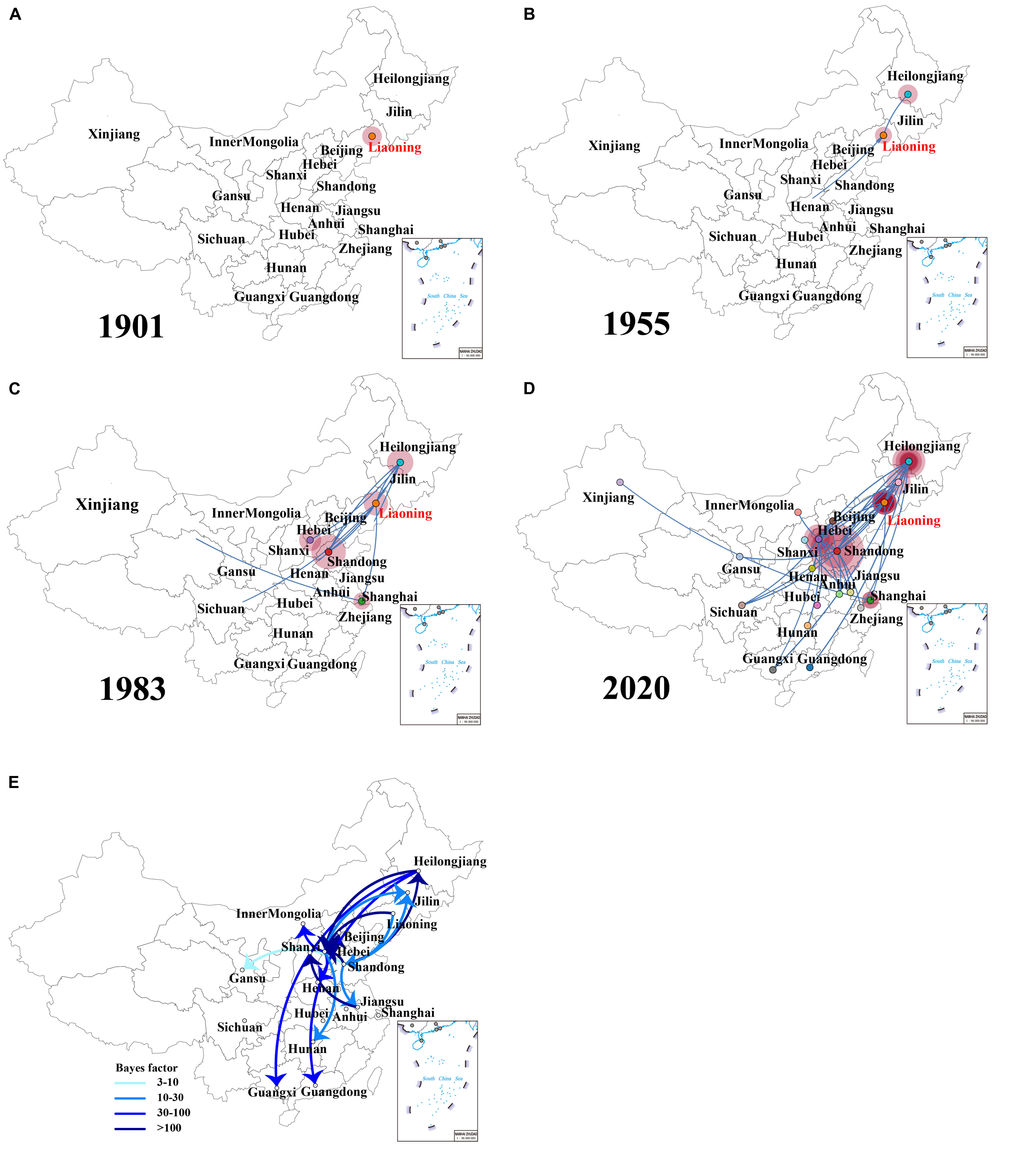


FIGURE 4 | The dynamics analysis of geographical regions of the full-length genome sequences of field isolates from China sampled during the years 1985–2020 **(A–D)**. Spatiotemporal dynamics of the full-length genome gene sequences of infectious bronchitis virus (IBV) among 20 localities. Bayes factor (BF) test for significant non-zero rates and regional transmission analysis **(E)**. Curves show the among-country virus lineage transitions statistically supported with Bayes factor > 3 for all the IBV isolates in China. Curve colors represent corresponding statistical support (Bayes factor value) for each transition rate (inset).

appearance of new isolates (Worobey and Holmes, 1999; Thor et al., 2011). Another study showed that some "hot spots" for recombination were in the regions of 1a (nsp2 and nsp3) and 1b (nsp16) of the genome and the upstream of the S gene based on the analysis of 25 America IBV isolates, and there were a high number of breakpoints in these three areas (Thor et al., 2011).

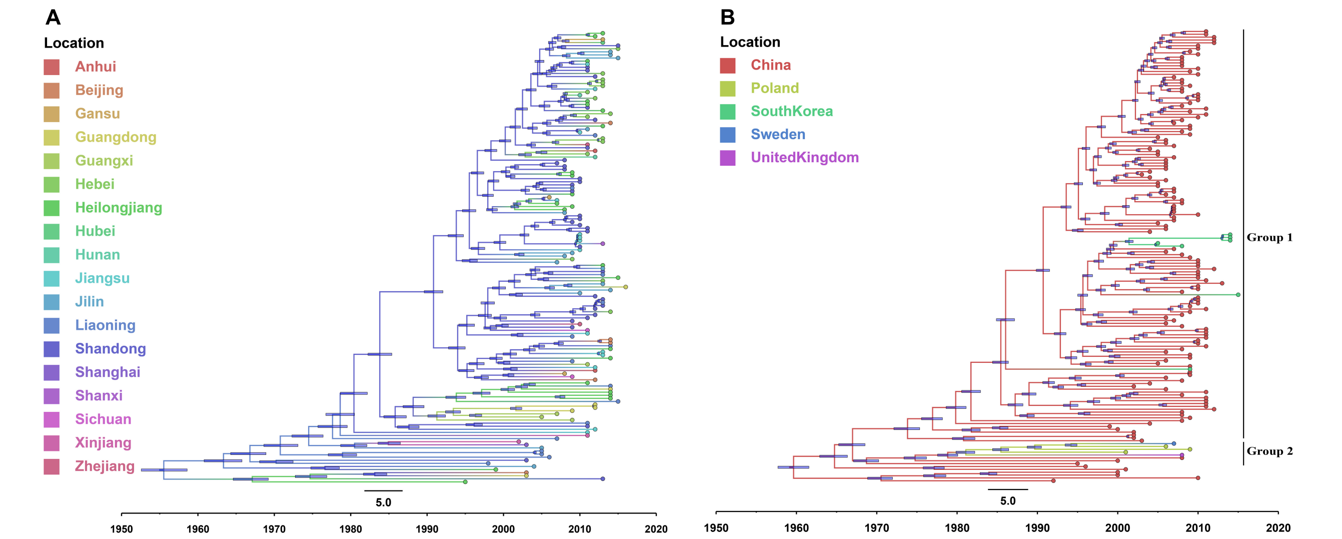


FIGURE 5 | The maximum clade credibility tree of LX4-type full-length genome sequences of infectious bronchitis virus (IBV) isolates in China **(A)** and the world **(B)**. The color of the branches corresponds to geographical areas.

However, our results showed that most of the breakpoints of recombination events may be located in nsp3 (1a region) and S gene (midstream and downstream) based on the 212 IBV isolates in China. Therefore, the breakpoints of recombination events of the IBV isolates in China were unique. Interestingly, no breakpoint in the regions of nsp8, nsp9, nsp10, and 3a was found. Areas that have the highest occurrence of recombination are located in regions of the genome that code for non-structural proteins 3 and the structural spike glycoprotein. A previous study showed that nsp3 encodes the protease PLP2, which is a type I interferon (IFN) antagonist and has a deubiquinating-like activity (Thor et al., 2011). Changes in the amino acid composition of the area in nsp3 could affect the ability of the virus to replicate in a variety of cell types (Thor et al., 2011). Moreover, the recombination in the area of nsp3 could alter the pathogenicity of the virus by modulating host cytokine expression (Eriksson et al., 2008; Thor et al., 2011). The spike glycoprotein exerts influence in attachment to host cell receptors, membrane fusion, and entry into the host cell (Thor et al., 2011). The spike glycoprotein has conformationally dependent epitopes that induce virus-neutralizing and serotype-specific antibodies (Cavanagh, 2007; Thor et al., 2011). Hence, the degree of recombination observed in the spike glycoprotein may be one of the principal mechanisms for generating genetic and antigenic diversity within IBV. The recombination regions within or near the S1 protein can lead to the emergence of new coronaviruses or new IBV serotypes (Thor et al., 2011). Our research showed that the recombination regions within S1 gene may be clustered in the midstream (600–700 bp) and downstream (1,500–1,611 bp) of the gene. The RBD was located in the midstream of S1 gene. The RBD was very important for the biological implications of genetic variation in circulating IBV genotypes (Bouwman et al., 2020). It was found that the amino acids 99–159 of RBD in the nephropathogenic IBV strain QX were required to establish kidney binding, and the reciprocal mutations in QX-RBD amino acids 110–112 could abolish kidney binding completely (Bouwman et al., 2020). Thus, more attention should be paid to these regions. In addition, the results of recombination analysis based on the whole genome and S1 gene would not be always consistent (Thor et al., 2011; Franzo et al., 2017). For example, based on the analysis result of recombination events in the complete genomes, there was a recombination region detected in the S1 region in EU637854 too. However, the p-value of S1 gene in EU637854 was lower than 10^{-6} but higher than 10^{-12} ($10^{-12} < p\text{-value} < 10^{-6}$). We did not show such a result in **Figure 2A**.

China is the major producer of chickens worldwide. In spite of the widespread use of Mass type (H120, H52, Ma5, 28/86, M41, and W93), 4/91 type (NNA), LDT3-A, QXL87, and other vaccines, IB has been a continual problem in vaccinated flocks in China (Fan et al., 2018; Yan et al., 2019; Ren et al., 2020; Zhang et al., 2020). To obtain a comprehensive understanding of the epidemiological situation, evolutionary trend, and spatiotemporal diffusion of IBVs in China, a 35-year epidemiological analysis of a total of 212 IBV isolates from the years 1985 to 2020 on a national scale was performed here.

So far, there are three sequencing technologies, namely, Sanger dideoxy sequencing, next-generation sequencing (NGS), and third-generation sequencing (single molecular sequencing). The Sanger sequencing is read in longer sequences (800–1,000 bp), with high accuracy and ease of use, and its disadvantages were the limited capacity and relatively high costs (Heather and Chain, 2016). The low-fidelity polymerase errors could affect the result of genome sequences during the PCR, and the errors can be reduced by using a high-fidelity polymerase. The NGS allowed the sequencing of multiple genes, exomes, or genomes in a single experiment, which is far quicker and cheaper than the Sanger sequencing, but its operation processes require more steps

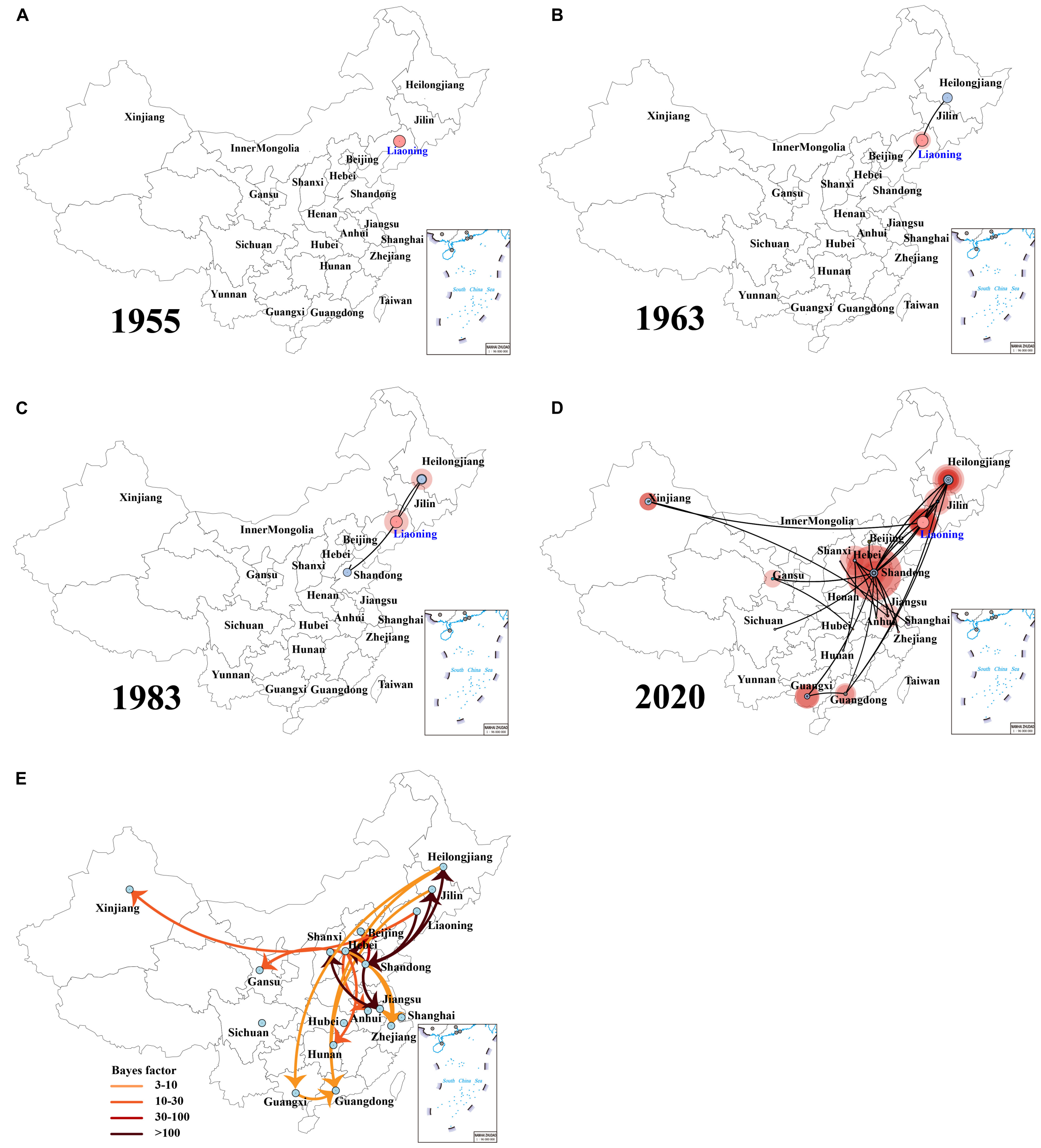


FIGURE 6 | The dynamics analysis of geographical regions of the LX4 genotype isolates from China **(A–D)**. Spatiotemporal dynamics of the full-length genome gene sequences of LX4 genotype infectious bronchitis virus (IBV) among 18 localities. Bayes factor (BF) test for significant non-zero rates and regional transmission analysis **(E)**. Curves show the among-province virus lineage transitions statistically supported with Bayes factor > 3 for all the LX4 genotype IBV isolates in China.

and read in shorter sequences (400–500 bp) (Diekstra et al., 2015). The third-generation sequencing technology is developed rapidly with lower cost and more reads in long sequences, but it has more proportion of the wrong bases (Heather and Chain, 2016). According to the GenBank database, all the whole-genome sequences of 212 Chinese IBV isolate for analysis were obtained by Sanger dideoxy sequencing technology. Therefore, the sequencing method used for the sequences analyzed in the present study has little effect on the results.

Recent studies revealed that the 4/91-type strains have become an important gene donor to variants in the genetic evolution of IBV in China (Zhao et al., 2016; Fan et al., 2019; Xu et al., 2019; Huang et al., 2020). Our studies showed that the percentage of 4/91-type recombinants (23.73 and 50% recombinants based on

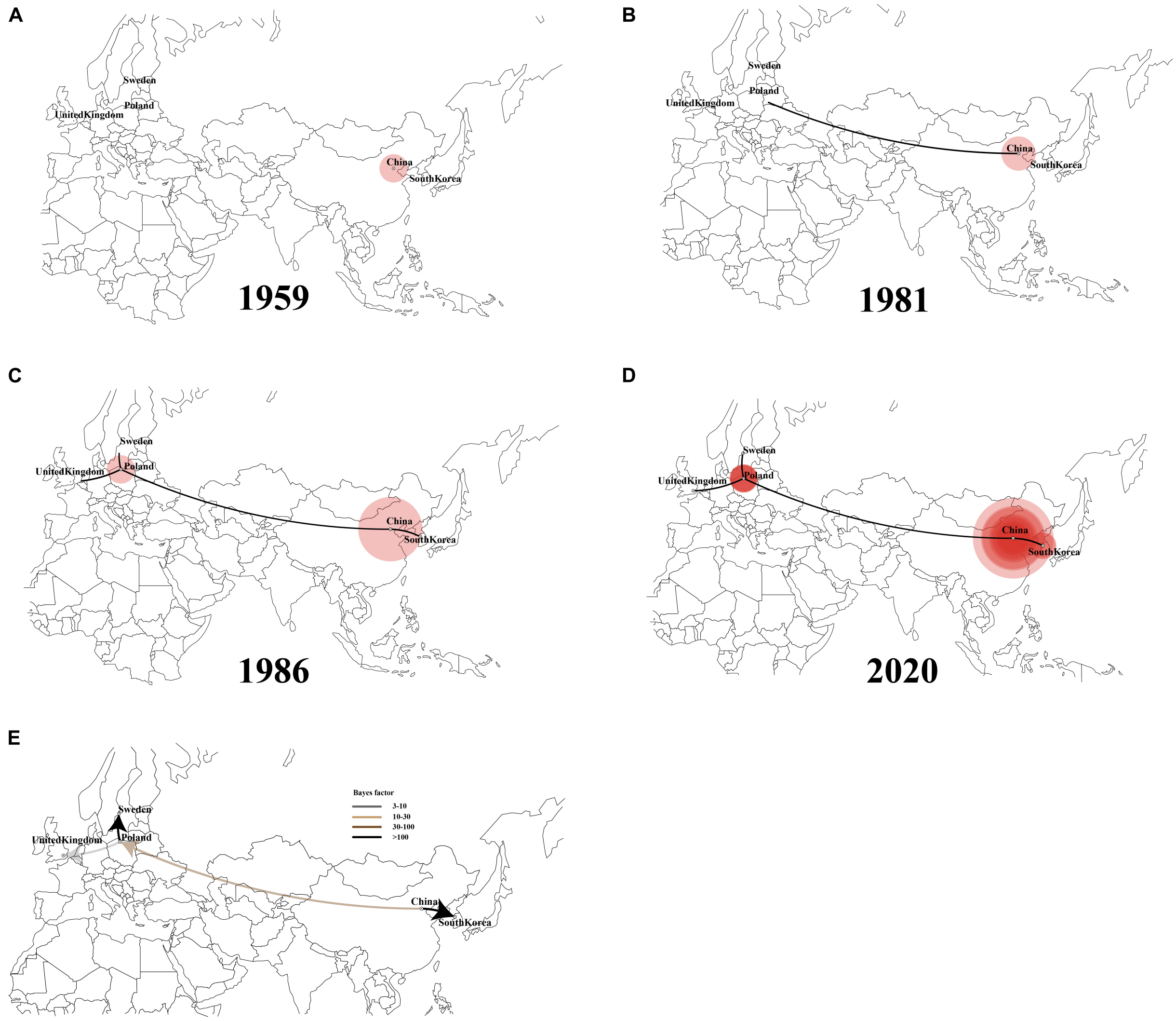


FIGURE 7 | Spatiotemporal dynamics of LX4-type infectious bronchitis virus (IBV) isolates among different localities **(A–D)**. Bayes factor (BF) test for significant non-zero rates and regional transmission analysis **(E)**. Curves show the among-country virus lineage transitions statistically supported with Bayes factor > 3 for all the LX4 genotype IBV isolates in the world.

the full-length genome and S1, respectively) may be higher than that of Mass-type recombinants (18.64 and 12.50% recombinants based on the full-length genome and S1, respectively) in the present study, which agreed with our previous reports (Mo et al., 2013; Fan et al., 2019). In the previous study, we analyzed several reasons why the 4/91-type strains were more prone to recombination than other vaccine-type strains (Fan et al., 2019). Given the infrequent prevalence of the 4/91-type isolates since 2015, it is recommended that the 4/91-type vaccine is to be used sparingly or with caution, but continued monitoring is needed. In addition, we found that multiple recombination events between field and vaccine strains may have occurred in the 24 recombinant isolates based on the complete genomes of the 212 Chinese isolates, suggesting the complexity of the recombination mechanism of IBV. Our results also suggest that genome-wide recombination analysis, not just S1 gene recombination analysis, should be performed in order to have a comprehensive understanding of recombination events of IBV isolates.

Bayesian phylodynamic models have not been widely used for disease surveillance in the poultry industry. However, they have been used widely to analyze the evolution and spatial spreading of SARS-CoV (Rest and Mindell, 2003), MERS-CoV (Breban et al., 2013), and SARS-CoV-2 (Li et al., 2020). In our study, the abovementioned models were used for analyzing the domestic history of IBVs in China to serve as a real-time early warning monitoring system and help the managers to make the correct decisions to reduce the risk of infection. According to our results, the Liaoning isolates were at the root of the tree, with the highest root state posterior probability in the 1900s. Thus, Liaoning was the origin of IBV in China in 1901. To our knowledge, this is the first report referring to the origin of the Chinese IBV on both the exact time and place by the Bayesian phylodynamic models. However, the inferred past geographical locations are heavily limited by the input data. All sequences were from provinces in China, and thus the inferred origin of the virus in the dataset would undoubtedly be one of the provinces of China. In China, Liaoning used to be one of the major provinces of poultry production and has a close relationship with live bird trading with other provinces. In addition, based on the MCC tree, the LX4-type isolates were first found in Liaoning in the late 1950s and then spread to other provinces. Franzo et al. (2017) found that the QX genotype (LX4 type, GI-19) originated from China in the late 1970s, and the GI-19 lineage circulated within the country until it migrated to Europe in the late 1990s, based on the analysis of S1 gene of total of 807 strains from 18 countries by plotting the MCC tree. The first LX4-type isolate (QX strain) was reported in Qingdao, Shandong province, in 1996 (Wang et al., 1997). The first case reported in Qingdao in 1996 did not mean that the virus first appeared in that time and area. Zhao et al. found that many LX4-type IBVs disappeared before 2005 because they were not “fit” for adaptation in chickens, but those “fit” LX4-type IBVs continued to evolve and have become widespread and predominant in commercial poultry after 2005 (Zhao et al., 2017). The LX4 (GI-19) lineage had circulated for a long time in China before its recognition as a distinct genotype in 1996 (Franzo et al., 2017). Hence, our result first demonstrated that Liaoning province was the origin of LX4-type IBV in the late 1950s. Besides, in order to get a full picture of the virus evolution, we try to divide the alignment into several recombination-free regions such as S and E genes and do a separate analysis on each of them instead. However, different genomic regions have different evolutionary histories. Therefore, in order to obtain comprehensive genetic information and molecular mechanism of variation of IBV isolates and mapped to the S gene (the greatest number of recombination regions and the very strong signal), we excluded the complete genome of 16 isolates, which were the potential recombinant sequences in S gene, and used the dataset to analyze the evolution and spatial spreading of viruses.

A previous report showed that the avian influenza virus’s transmission occurred along the poultry industry trade routes (Yang et al., 2020). Interestingly, a similar phenomenon appeared in our study based on the IBV trade routes. In our study, five trade routes for both IBV and LX4-type isolates in China were strongly indicated with high BF values. Moreover, the first transmission route (from Shandong to Heilongjiang and Jilin), the second one (from Jiangsu to Shanxi, from Shandong to Jiangsu and Hebei), the fifth one (from the northeastern province Heilongjiang and Jilin to the southern provinces Guangxi and Guangdong), and the long-distance movement events (from the northern province Shandong to the northwestern province Gansu and from Hebei to the southern province Hunan) of the LX4-type isolates were similar to those of the IBVs in China. The northern region including Shandong and Liaoning provinces is the biggest region of white chicken production in China with annually 1.8 and 0.8 billion birds, respectively, while Guangxi and Guangdong had the biggest production of the local breeds of chickens (1.9 billion birds) in 2019 (Wei, 2019; Deng et al., 2021). Hence, the five transmission routes were basically consistent with poultry industry trade routes.

In order to provide more targeted strategies for the prevention and control of the disease, we employed a Bayesian phylogeographic approach. Based on the IBV spread routes, all IBVs especially the LX4-type isolates in China exhibited evidence of a continuous process of geographic spread. Nowadays, northeastern, northern, and southern regions of China were the most frequent places of the viral output and input, indicating that these three areas have been the epicenter of the LX4 genotype in China. A previous report showed that the intensity of the poultry trade among locations and the migration of wild birds among locations in China were positively associated with avian influenza virus lineage spread, and the distances along the road network in China were strongly inversely associated with viral lineage spread (Yang et al., 2020). We found that the IBVs (including LX4 type) spread routes and that the live poultry trade network was positively associated with the viral lineage spread. Hence, the IBVs in the Chinese mainland were mainly spread through the transport and trade of live birds, while the long-distance spreads, which were from the northeastern Liaoning province to northwestern Xinjiang province and the northeastern provinces Heilongjiang, and Jilin to the southern provinces Guangxi and Guangdong, were more likely through migratory wild birds. These data on mode of geographic spread would play a critical role in the prevention and control of IB.

To further investigate the source of the LX4-type isolates in the world, we also conducted phylogeographic inference. The full-length genome sequences of the early LX4-type isolates during 2004–2018 in other countries were found to be closely related to the early isolates in China during 1995–2016, as they were at the lower level of the node in the tree, which means that the LX4-type isolates were likely to have originated from China. In addition, tMRCA of the earliest LX4-type isolates was estimated to be in the late 1950s according to the Bayesian analyses. The LX4-type isolates had spread from China to Europe in the early 1980s and spread from China to South Korea in the late 1980s based on the Bayesian analyses. Of course, the inferred transmissions of LX4-type IBV might not be direct, but the virus may have gone unobserved through multiple countries before entering Europe/South Korea. Other researchers reported that the LX4-type strain was from South Korea based on the analysis of S1 genotype, the time of isolating, and the geographical distance between Shandong province and South Korea (Liu et al., 2006; Huang et al., 2019). However, analysis of S1 gene alone had a limitation. The Bayesian phylodynamic models have been used widely to analyze the evolution and spatial spreading of SARS-CoV-2 (Li et al., 2020). Hence, a thorough and comprehensive Bayesian analysis using as many strains across the world with the entire genome sequences could contribute to a greater understanding of the spread of IBV. Of course, our research also had some data limitations. For example, the number of samples was inhomogeneous in different countries.

The generation, maintenance, and distribution of genetic variation of the virus can be affected by a sudden change in population size (Gao et al., 2019). Indeed, our demographic analyses reveal that IBV populations in China had been small before the 1980s but have undergone expansion since the 1980s, possibly associated with a capacity expansion of chicken in China after China adopted the reform and opening-up policy. However, a sudden and sharp decrease in genetic diversity was observed after 2006, probably due to an outbreak of the H5N1 virus in migratory waterfowl in Qinghai in 2005 and spread to other areas (Chen et al., 2005), which had brought substantial damage to the poultry industry. A large number of farms were closed due to being affected by the pandemic avian influenza (Wei, 2019). At that time, the inter-provincial trade of live poultry was banned, which was in accordance with our result that the IBV population declined in size from 2006. However, since late 2008, a sharp increase was observed in viral population size. There may be two main reasons to explain this phenomenon. First, new farms were increasing because of the successful control of the avian influenza outbreak. The inter-provinces' live poultry trade was allowed. Second, the LX4-type isolates had become the predominant IBVs since 2008 based on the analysis of S1 gene (Mo et al., 2013; Fan et al., 2019), and the implementation of the current vaccines could not provide good protection against the circulating strains. Other studies indicated that the change of viral population size had a strong association with the vaccine administration/withdrawal (Franzo et al., 2014, 2016). However, the protection provided by the new commercial LX4-type vaccine (QXL87) on the market against the LX4-type isolates has not been studied yet, which needs further investigation.

CONCLUSION

In conclusion, Chinese IBV was still in a state of coexistence of multiple genotypes, the LX4 type with the highest dN/dS ratios has been the most dominant genotype since 1999, and the Taiwan-I type and New type showed an increasing trend recently. A high frequency of recombination and multiple recombination events have occurred in the complete genome, and the 4/91-type strains were more prone to be involved in the recombination. The Chinese IBV originated from Liaoning in the early 1900s and has three main clusters: the northeastern cluster, the northern cluster, and the southern cluster. In particular, the LX4 type originated from Liaoning in the late 1950s and had multiple transmission routes in China plus two major transmission routes to other countries. Shandong has been the epidemiological center of IBVs (including LX4 type) in China. These data extend our insight into the evolution and spread of IBV in China and provide more targeted strategies for the prevention and control of IB in China.

DATA AVAILABILITY STATEMENT

The data analyzed during this study were available within the article and its supporting information.

AUTHOR CONTRIBUTIONS

WF analyzed the data and wrote the manuscript. JC and YZ contributed to the experiment. PW and MM provided the funding of research and reviewed and approved the final manuscript. All authors participated in this study, read the final manuscript, and approved it for submission.

FUNDING

This work was supported by National Natural Science Foundation of China (Grant No. 31860715), Guangxi Key Research and Development Plan (Grant No. AB21238003), Guangxi Special Funding on Science and Technology Research (Grant No. AA17204057), and Guangxi Natural Science Foundation Grants (Grant Nos. 2019GXNSFAA245009 and 2020GXNSFDA297004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.693196/full#supplementary-material

Supplementary Figure 1 | Distribution of 212 IBV isolates in China. Distribution is presented as the number of isolates per province or autonomous city (zone), from most cases (45, dark red) to no case (0, white).

Supplementary Figure 2 | The percentages of different genotypes based on the complete genome sequence **(A)**, S1 **(B)**, S2 **(C)**, E **(D)**, M **(E)**, and N **(F)** of IBV strains in different years. There was only one IBV strain (GX-C) isolated during 1985–1998, which showed considerable low homology with the above other genotypes and belonged to a separate group.

Supplementary Figure 3 | Phylogenetic trees of the complete genome of 68 IBVs from other countries and the 33 IBV reference strains. Phylogenetic trees were constructed with the maximum likelihood method using RaxML software. Green represent LX4 genotype. The green solid dots represent 15 LX4 genotype IBV isolates.

Supplementary Figure 4 | Alignment of the deduced amino acid sequences of the S1 subunit of the spike protein of 212 IBV isolates and the references M41, H120, 4/91, QXL87, LX4, and LDT3-A strains. Amino acid sequences encoding the three hypervariable regions are signed in red line. The minimal receptor-binding domains corresponding to the most N-terminal residues 19–272 of the spike protein of IBV M41 are boxed, and the critical amino acids 70–74 are highlighted in red. KM213963: the isolate of CK/CH/XDC-2/2013; KP118886: the isolate of ck/CH/LSD/111235; KC008600: the isolate of GX-C; HQ848267: the isolate of GX-YL5; HQ850618: the isolate of GX-YL9; KF411041: the isolate of CK/CH/LGX/091109; KP118894: the isolate of ck/CH/LGD/090907; KC013541: the isolate of Ck/CH/LGD/120723; KC119407: the isolate of Ck/CH/LGD/120724; JF893452: the isolate of YN; MK644086: the isolate of E160_YN.

Supplementary Figure 5 | The vaccine strain H120 was used as the reference strain. **(A)** The virus genome of the coronavirus avian infectious bronchitis showing the locations of 5′ and 3′ UTRs and the coding regions. Recombination breakpoints identified across the full-length genome. The putative recombinants were listed in the key, with the colored circles matching each corresponding breakpoint. The beginning breakpoint was highlighted in ●, ●, ● circles and the ending breakpoint was highlighted in ●, ●, ● circles from 5′ to 3′ UTRs in every recombinant isolate. **(B)** Recombination breakpoints identified across the S1 gene. The beginning breakpoint was highlighted in ● circle and the ending breakpoint was highlighted in ● circle in every recombinant isolate.

Supplementary Figure 6 | Similarity plots of the full-length genome of 59 IBV isolates created with SimPlot version 3.5.1. A 200-bp window with a 20-bp step was used.

Supplementary Figure 7 | Similarity plots of the S1 structural gene region of 16 IBV isolates created with SimPlot version 3.5.1. A 200-bp window with a 20-bp step was used.

REFERENCES

Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., and Lemey, P. (2012). Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol. Bio. Evol.* 30, 239–243. doi: 10.1093/molbev/mss243

Bande, F., Arshad, S. S., Omar, A. R., Hair-Bejo, M., Mahmuda, A., and Nair, V. (2017). Global distributions and strain diversity of avian infectious bronchitis virus: a review. *Anim Health Res. Rev.* 18, 70–83. doi: 10.1017/S1466252317000044

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

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, e1363–19. doi: 10.1128/JVI.01363-19

Breban, R., Riou, J., and Fontanet, A. (2013). Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *Lancet* 382, 694–699. doi: 10.1016/S0140-6736(13)61492-0

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

Cavanagh, D., Davis, P. J., Cook, J. K., Li, D., Kant, A., and Koch, G. (1992). Location of the amino acid differences in the S1 spike glycoprotein subunit of closely related serotypes of infectious bronchitis virus. *Avian Pathol.* 21, 33–43. doi: 10.1080/03079459208418816

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

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

Deng, Q., Shi, M., Li, Q., Wang, P., Li, M., Wang, W., et al. (2021). Analysis of the evolution and transmission dynamics of the field MDV in China during the years 1995-2020, indicating the emergence of a unique cluster with the molecular characteristics of vv+MDV that has become endemic in southern China. *Transboundary Emerg. Dis.* 68, 3574–3587. doi: 10.1111/tbed.13965

Diekstra, A., Bosgoed, E., Rikken, A., van Lier, B., Kamsteeg, E. J., Tychon, M., et al. (2015). Translating sanger-based routine DNA diagnostics into generic massive parallel ion semiconductor sequencing. *Clin. Chem.* 61, 154–162. doi: 10.1373/clinchem.2014.225250

Drosten, C., Gunther, S., Preiser, W., van der, W. S., and Brodt, H. R., Becker, S., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967–1976. doi: 10.1056/NEJMoa030747

Eriksson, K. K., Cervantes, B. L., Ludewig, B., and Thiel, V. (2008). Mouse hepatitis virus Liver pathology is dependent on ADP-ribose-1"-phosphatase, a viral function conserved in the alpha-Like supergroup. *J. Virol.* 82, 12325–12334. doi: 10.1128/JVI.02082-08

Fan, W., Li, H., He, Y., Tang, N., Zhang, L., Wang, H., et al. (2018). Immune protection conferred by three commonly used commercial live attenuated vaccines against the prevalent local strains of avian infectious bronchitis virus in southern China. *J. Vet. Med. Sci.* 80, 1438–1444. doi: 10.1292/jvms.18-0249

Fan, W., Tang, N., Dong, Z., Chen, J., Zhang, W., Zhao, C., et al. (2019). Genetic analysis of avian coronavirus infectious bronchitis virus in yellow chickens in southern China over the past decade: revealing the changes of genetic diversity, dominant genotypes, and selection pressure. *Viruses* 11:898. doi: 10.3390/v11100898

Finger, P. F., Pepe, M. S., Dummer, L. A., Magalhaes, C. G., de Castro, C. C., de Oliveira, H. S., et al. (2018). Combined use of ELISA and Western blot with recombinant N protein is a powerful tool for the immunodiagnosis of avian infectious bronchitis. *Virol. J.* 15:189. doi: 10.1186/s12985-018-1096-2

Franzo, G., Massi, P., Tucciarone, C. M., Barbieri, I., Tosi, G., Fiorentini, L., et al. (2017). Think globally, act locally: phylodynamic reconstruction of infectious bronchitis virus (IBV) QX genotype (GI-19 lineage) reveals different population dynamics and spreading patterns when evaluated on different epidemiological scales. *PLoS One* 12:e0184401. doi: 10.1371/journal.pone.0184401

Franzo, G., Naylor, C. J., Lupini, C., Drigo, M., Catelli, E., Listorti, V., et al. (2014). Continued use of IBV 793B vaccine needs reassessment after its withdrawal led to the genotype's disappearance. *Vaccine* 32, 6765–6767. doi: 10.1016/j.vaccine.2014.10.006

Franzo, G., Tucciarone, C. M., Blanco, A., Nofrarias, M., Biarnes, M., Cortey, M., et al. (2016). Effect of different vaccination strategies on IBV QX population dynamics and clinical outbreaks. *Vaccine* 34, 5670–5676. doi: 10.1016/j.vaccine.2016.09.014

Gao, F., Liu, X., Du, Z., Hou, H., Wang, X., Wang, F., et al. (2019). Bayesian phylodynamic analysis reveals the dispersal patterns of tobacco mosaic virus in China. *Virology* 528, 110–117. doi: 10.1016/j.virol.2018.12.001

Heather, J. M., and Chain, B. (2016). The sequence of sequencers: the history of sequencing DNA. *Genomics* 107, 1–8. doi: 10.1016/j.ygeno.2015.11.003

Huang, M., Zhang, Y., Xue, C., and Cao, Y. (2019). To meet the growing challenge: research of avian infectious bronchitis in China. *Microbiol. China.* 46, 1837–1849. doi: 10.13344/j.microbiol.china.180898

Huang, M., Zou, C., Liu, Y., Han, Z., Xue, C., and Cao, Y. (2020). A novel low virulent respiratory infectious bronchitis virus originating from the recombination of QX, TW and 4/91 genotype strains in China. *Vet. Microbiol.* 242:108579. doi: 10.1016/j.vetmic.2020.108579

Ignjatovic, J., and Sapats, S. (2005). Identification of previously unknown antigenic epitopes on the S and N proteins of avian infectious bronchitis virus. *Arch. Virol.* 150, 1813–1831. doi: 10.1007/s00705-005-0541-x

Jackwood, M. W., Boynton, T. O., Hilt, D. A., McKinley, E. T., Kissinger, J. C., Paterson, A. H., et al. (2010). Emergence of a group 3 coronavirus through recombination. *Virology* 398, 98–108. doi: 10.1016/j.virol.2009.11.044

Jackwood, M. W., and Lee, D. (2017). Different evolutionary trajectories of vaccine-controlled and non-controlled avian infectious bronchitis viruses in commercial poultry. *PLoS One* 12:e0176709. doi: 10.1371/journal.pone.0176709

Jiang, L., Han, Z., Chen, Y., Zhao, W., Sun, J., Zhao, Y., et al. (2018). Characterization of the complete genome, antigenicity, pathogenicity, tissue tropism, and shedding of a recombinant avian infectious bronchitis virus with a ck/CH/LJL/140901-like backbone and an S2 fragment from a 4/91-like virus. *Virus Res.* 244, 99–109. doi: 10.1016/j.virusres.2017.11.007

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

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., et al. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* 172, 577–582. doi: 10.7326/M20-0504

Li, M., Wang, X., Wei, P., Chen, Q., Wei, Z., and Mo, M. (2012). Serotype and genotype diversity of infectious bronchitis viruses isolated during 1985-2008 in Guangxi. *China. Arch. Virol.* 157, 467–474. doi: 10.1007/s00705-011-1206-6

Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B. T., et al. (2020). 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

Liang, J., Fang, S., Yuan, Q., Huang, M., Chen, R., Fung, T., et al. (2019). N-Linked glycosylation of the membrane protein ectodomain regulates infectious bronchitis virus-induced ER stress response, apoptosis and pathogenesis. *Virology* 531, 48–56. doi: 10.1016/j.virol.2019.02.017

Lin, S., and Chen, H. (2017). Infectious bronchitis virus variants: molecular analysis and pathogenicity investigation. *Int. J. Mol. Sci.* 18:2030. doi: 10.3390/ijms18102030

Liu, S., Zhang, Q., Chen, J., Han, Z., Liu, X., Feng, L., et al. (2006). Genetic diversity of avian infectious bronchitis coronavirus strains isolated in China between 1995 and 2004. *Arch. Virol.* 151, 1133–1148. doi: 10.1007/s00705-005-0695-6

Liu, S., Zhang, X., Wang, Y., Li, C., Han, Z., Shao, Y., et al. (2009). Molecular characterization and pathogenicity of infectious bronchitis coronaviruses: complicated evolution and epidemiology in China caused by cocirculation of multiple types of infectious bronchitis coronaviruses. *Intervirology* 52, 223–234. doi: 10.1159/000227134

Martin, D. P. (2009). Recombination detection and analysis using RDP3. *Methods Mol. Biol.* 537, 185–205. doi: 10.1007/978-1-59745-251-9_9

Mo, M., Li, M., Huang, B., Fan, W., Wei, P., Wei, T., et al. (2013). Molecular characterization of major structural protein genes of avian coronavirus infectious bronchitis virus isolates in southern china. *Viruses* 5, 3007–3020. doi: 10.3390/v5123007

Molenaar, R. J., Dijkman, R., and de Wit, J. J. (2020). Characterization of infectious bronchitis virus D181, a new serotype (GII-2). *Avian Pathol.* 49, 243–250. doi: 10.1080/03079457.2020.1713987

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

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

Ren, G., Liu, F., Huang, M., Li, L., Shang, H., Liang, M., et al. (2020). Pathogenicity of a QX-like avian infectious bronchitis virus isolated in China. *Poult. Sci.* 99, 111–118. doi: 10.3382/ps/pez568

Rest, J. S., and Mindell, D. P. (2003). SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. *Infect. Genet. Evol.* 3, 219–225. doi: 10.1016/j.meegid.2003.08.001

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

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Thor, S. W., Hilt, D. A., Kissinger, J. C., Paterson, A. H., and Jackwood, M. W. (2011). Recombination in avian gamma-coronavirus infectious bronchitis virus. *Viruses* 3, 1777–1799. doi: 10.3390/v3091777

Valastro, V., Holmes, E. C., Britton, P., Fusaro, A., Jackwood, M. W., Cattoli, G., et al. (2016). S1 gene-based phylogeny of infectious bronchitis virus: an attempt to harmonize virus classification. *Infect. Genet. Evol.* 39, 349–364. doi: 10.1016/j.meegid.2016.02.015

Wang, Y., Zhang, Z., Wang, Y., Fan, G., Jiang, Y., Liu, X., et al. (1997). A preliminary report on the study of glandular stomach type infectious bronchitis of chickens. *Chin. J. Anim. Quarant.* 14, 6–8.

Wei, P. (2019). Challenge and development opportunity of quality chicken industry in China: a review about 9 issues on the topic. *China Poult.* 41, 1–6. doi: 10.16372/j.issn.1004-6364.2019.11.001

Worobey, M., and Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* 80, 2535–2543. doi: 10.1099/0022-1317-80-10-2535

Wu, Q., Lin, Z., Qian, K., Shao, H., Ye, J., and Qin, A. (2019). Peptides with 16R in S2 protein showed broad reactions with sera against different types of infectious bronchitis viruses. *Vet. Microbiol.* 236:108391. doi: 10.1016/j.vetmic.2019.108391

Xu, L., Han, Z., Jiang, L., Sun, J., Zhao, Y., and Liu, S. (2018). Genetic diversity of avian infectious bronchitis virus in China in recent years. *Infect. Genet. Evol.* 66, 82–94. doi: 10.1016/j.meegid.2018.09.018

Xu, L., Ren, M., Sheng, J., Ma, T., Han, Z., Zhao, Y., et al. (2019). Genetic and biological characteristics of four novel recombinant avian infectious bronchitis viruses isolated in China. *Virus Res.* 263, 87–97. doi: 10.1016/j.virusres.2019.01.007

Yan, S., Sun, Y., Huang, X., Jia, W., Xie, D., and Zhang, G. (2019). Molecular characteristics and pathogenicity analysis of QX-like avian infectious bronchitis virus isolated in China in 2017 and 2018. *Poult. Sci.* 98, 5336–5341. doi: 10.3382/ps/pez351

Yang, Q., Zhao, X., Lemey, P., Suchard, M. A., Bi, Y., Shi, W., et al. (2020). Assessing the role of live poultry trade in community-structured transmission of avian influenza in China. *Proc. Natl. Acad. Sci. U. S. A.* 117, 5949–5954. doi: 10.1073/pnas.1906954117

Yuan, Y., Zhang, Z., He, Y., Fan, W., Dong, Z., Zhang, L., et al. (2018). Protection against virulent infectious bronchitis virus challenge conferred by a recombinant baculovirus co-expressing S1 and N proteins. *Viruses* 10:347. doi: 10.3390/v10070347

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820. doi: 10.1056/NEJMoa1211721

Zhang, X., Deng, T., Lu, J., Zhao, P., Chen, L., Qian, M., et al. (2020). Molecular characterization of variant infectious bronchitis virus in China, 2019: implications for control programmes. *Transbound. Emerg. Dis.* 67, 1349–1355. doi: 10.1111/tbed.13477

Zhao, W., Gao, M., Xu, Q., Xu, Y., Zhao, Y., Chen, Y., et al. (2017). Origin and evolution of LX4 genotype infectious bronchitis coronavirus in China. *Vet. Microbiol.* 198, 9–16. doi: 10.1016/j.vetmic.2016.11.014

Zhao, Y., Zhang, H., Zhao, J., Zhong, Q., Jin, J., and Zhang, G. (2016). Evolution of infectious bronchitis virus in China over the past two decades. *J. Gen. Virol.* 97, 1566–1574. doi: 10.1099/jgv.0.000464

Zou, N., Zhao, F., Wang, Y., Liu, P., Cao, S., Wen, X., et al. (2010). Genetic analysis revealed LX4 genotype strains of avian infectious bronchitis virus became predominant in recent years in Sichuan area, China. *Virus Genes.* 41, 202–209. doi: 10.1007/s11262-010-0500-9

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 | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 April 2022
doi: 10.3389/fmicb.2022.735363

Rapid and Accurate Detection of SARS Coronavirus 2 by Nanopore Amplicon Sequencing

Xiao-xiao Li[1]†, Chao Li[1]†, Peng-cheng Du[3]†, Shao-yun Li[1], Le Yu[3], Zhi-qiang Zhao[3], Ting-ting Liu[3], Cong-kai Zhang[3], Sen-chao Zhang[3], Yu Zhuang[1], Chao-ran Dong[2] and Qing-gang Ge[1]**

[1] Department of Pharmacy, Department of Intensive Care Unit, and Department of Medical Affairs, Peking University Third Hospital, Beijing, China, [2] Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [3] Beijing YuanShengKangTai (ProtoDNA) Genetech Co. Ltd., Beijing, China

OPEN ACCESS

Edited by:
Clyde Dapat,
Tohoku University Graduate School of Medicine, Japan

Reviewed by:
José Patricio Miranda,
Pontificia Universidad Católica de Chile, Chile
Mei-li Duan,
Capital Medical University, China

Correspondence:
Chao-ran Dong
dongcr@imm.ac.cn
Qing-gang Ge
qingganggelin@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: *02 July 2021*
Accepted: *12 January 2022*
Published: *06 April 2022*

Citation:
Li X-X, Li C, Du P-C, Li S-Y, Yu L, Zhao Z-Q, Liu T-T, Zhang C-K, Zhang S-C, Zhuang Y, Dong C-R and Ge Q-G (2022) Rapid and Accurate Detection of SARS Coronavirus 2 by Nanopore Amplicon Sequencing. Front. Microbiol. 13:735363. doi: 10.3389/fmicb.2022.735363

Objective: We aimed to evaluate the performance of nanopore amplicon sequencing detection for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in clinical samples.

Method: We carried out a single-center, prospective cohort study in a Wuhan hospital and collected a total of 86 clinical samples, including 54 pharyngeal swabs, 31 sputum samples, and 1 fecal sample, from 86 patients with coronavirus disease 2019 (COVID-19) from Feb 20 to May 15, 2020. We performed parallel detection with nanopore-based genome amplification and sequencing (NAS) on the Oxford Nanopore Technologies (ONT) minION platform and routine reverse transcription quantitative polymerase chain reaction (RT-qPCR). In addition, 27 negative control samples were detected using the two methods. The sensitivity and specificity of NAS were evaluated and compared with those of RT-qPCR.

Results: The viral read number and reference genome coverage were both significantly different between the two groups of samples, and the latter was a better indicator for SARS-CoV-2 detection. Based on the reference genome coverage, NAS revealed both high sensitivity (96.5%) and specificity (100%) compared with RT-qPCR (80.2 and 96.3%, respectively), although the samples had been stored for half a year before the detection. The total time cost was less than 15 h, which was acceptable compared with that of RT-qPCR (∼2.5 h). In addition, the reference genome coverage of the viral reads was in line with the cycle threshold value of RT-qPCR, indicating that this number could also be used as an indicator of the viral load in a sample. The viral load in sputum might be related to the severity of the infection, particularly in patients within 4 weeks after onset of clinical manifestations, which could be used to evaluate the infection.

Conclusion: Our results showed the high sensitivity and specificity of the NAS method for SARS-CoV-2 detection compared with RT-qPCR. The sequencing results were also used as an indicator of the viral load to display the viral dynamics during infection. This study proved the wide application prospect of nanopore sequencing detection for SARS-CoV-2 and may more knowledge about the clinical characteristics of COVID-19.

Keywords: severe acute respiratory syndrome coronavirus 2, nanopore amplicon sequencing, RT-qPCR, viral detection, genome coverage

HIGHLIGHTS

- Reference genome coverage was a prior indicator in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) detection applying nanopore amplicon sequencing (NAS) compared with viral read number.
- Based on the reference genome coverage, a higher sensitivity and specificity of NAS in detecting SARS-CoV-2 was revealed through its comparison with reverse transcription quantitative polymerase chain reaction (RT-qPCR).
- Some long-term stored samples with negative RT-qPCR results tested positive using the NAS approach.
- Sequencing results using the NAS approach revealed that dynamic variations in both pharyngeal swabs and sputum samples had potential to evaluate disease progression, especially for patients diagnosed as coronavirus disease 2019 within 4 weeks after the first clinical manifestations exhibited.

INTRODUCTION

The sudden emergence of the coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has rapidly spread globally (Pollard et al., 2020; Zhu et al., 2020). Strengthening the national and regional laboratory diagnostic capacity for COVID-19 was one of the key measures to counter this global public health crisis (Tang Y. W. et al., 2020). Specifically, a rapid and accurate nucleic acid test method for SARS-CoV-2 detection and COVID-19 diagnostics is critical for disease control and patient treatment.

The COVID-19 outbreak was followed by the characterization of the whole viral genome, which helped to develop several molecular diagnostic methods (Tang X. et al., 2020). On February 4, 2020, real-time reverse transcription polymerase chain reaction (RT-PCR) assay for COVID-19 was authorized by the Centers for Disease Control and Prevention (FDA, 2020). Till now, the RT-PCR is considered as the most conclusive molecular diagnostic approach. In addition, the cycle threshold (C_t) value of reverse transcription quantitative polymerase chain reaction (RT-qPCR) is also used as an indicator of the viral load in samples to estimate the viral dynamics in different periods or body sites during the infection. However, it suffers from high false-negative rates due to variations in the viral ribonucleic acid (RNA) sequences in the *ORF1ab* and *N* genes, which the primers target, raising the call for more sensitive methods (Wang Y. et al., 2020).

Next-generation sequencing (NGS) has been used for SARS-CoV-2 detection and research as early as its emergence. However, the higher costs, longer detection periods, and complicated operating steps have limited its application, even that several improved protocols based on library capture or viral genome amplification have been proposed (Chen et al., 2020; Nasir et al., 2020; Xiao et al., 2020). Oxford nanopore sequencing is a promising new generation of sequencing technology that features real-time data generation and long read length (Gu et al., 2021). Therefore, SARS-CoV-2 detection combining viral genome amplification and nanopore sequencing provides comprehensive genome coverage and lower time cost (Wang M. et al., 2020). Our team reported a patient finally confirmed as a COVID-19 case using nanopore-based genome amplification and sequencing (NAS), while the routine RT-qPCR was negative due to a viral mutation within the primer region (Mehta et al., 2020). Furthermore, this method allows a more convenient and efficient SAR-CoV-2 genomic surveillance, helping to identify mutations that have an impact on virulence and transmissibility (Bull et al., 2020; Lu et al., 2020). However, the performance of this approach has not been widely evaluated, particularly using clinical samples and in comparison with RT-qPCR.

We obtained NAS data from a cohort including hospitalized COVID-19 patients with mild (or moderate) or severe (or critical) disease and healthy controls to gain insight into the identification of SARS-CoV-2-specific genome compared with RT-qPCR, which may further serve as a better method for SARS-CoV-2 detection and an indicator of the viral load dynamics during infection.

MATERIALS AND METHODS

Specimen Collection

We carried out a single-center, prospective cohort study in a Wuhan pulmonary hospital. Eighty-six patients with a minimum age of 18 years and were confirmed as having COVID-19 were consecutively enrolled from Feb 20 to May 15, 2020. On the other hand, 27 patients who tested negative for COVID-19 were recruited into the control group. Patients were diagnosed with COVID-19 according to the *Diagnosis and Treatment Protocol of COVID-19* (the 6th–8th tentative version[1]) issued by the National Health Commission of the People's Republic of China; if they had any epidemiological history plus any two

Abbreviations: AUC, area under the ROC curve; COVID-19, coronavirus disease 2019; C_t, cycle threshold; MAPQ, mapping quality score; NAS, nanopore-based genome amplification and sequencing; NC, negative controls; NGS, next-generation sequencing; AFS, after first showing symptoms; RNA, ribonucleic acid; ROC, receiver operating characteristic; RT-qPCR, reverse transcription quantitative polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WHO, world health organization.

[1] www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df351d7da8aefc2.shtml

clinical manifestations (fever and/or respiratory symptoms; the aforementioned imaging characteristics of COVID-19; normal or decreased white blood cell count, and normal or decreased lymphocyte count in the early stage of onset) or all three clinical manifestations if there was no clear epidemiological history; and if the etiological evidence was confirmed as a *positive* result for new coronavirus nucleic acid detected using RT-qPCR (*positive* was defined as $C_t \leq 35$ according to the protocol offered by the clinical laboratory center). Patients were then classified into two groups: mild (or moderate) (fever, respiratory symptoms, etc., with or without abnormal lung imaging) and severe (or critical) (respiratory rate $\geq$ 30 per min, oxygen saturation $\leq$ 93% on room air at rest, arterial oxygen pressure/fraction of inspiration oxygen $\leq$ 300 mmHg, or progressive clinical symptoms and lung imaging showing >50% significant progression of the lesion within 24–48 h; even cause an aggravation of respiratory failure requiring mechanical ventilation, shock, or combined with other organ failure and requiring intensive care unit supervision). In addition to discarded pharyngeal swab, 5 ml discarded sputum sample or discarded fecal sample was collected from each patient during hospitalization and stored at −80°C after nucleic acid isolation. The demographic data were recorded. All of the clinical specimens and demographic data were obtained in accordance with the study protocol, which was approved by the Ethical Review Board of Peking University Third Hospital in March 2020, with informed consent from participants being exempted (IRB00006761-M2020083).

Ribonucleic Acid Extraction and Routine Reverse Transcription Quantitative Polymerase Chain Reaction Detection of Severe Acute Respiratory Syndrome Coronavirus 2

We performed parallel detections, one applying RT-qPCR and the other one applying SARS-CoV-2 genome amplification and sequencing using the Oxford Nanopore Technologies (ONT) platform based on the ARTIC protocol[2]. Firstly, total RNA was extracted from 200 μl specimen with an automatic nucleic acid extractor (EXM3000; Zybio Inc., Chongqing, China) according to the manufacturer's instructions. A 60-μl elution volume was obtained for each sample in total after three extractions in parallel. Five microlters total RNA was used for real-time RT-qPCR, which targeted the *ORF1ab* and *N* genes using the Novel Coronavirus (2019-nCoV) nucleic acid detection kit (ZC-HX-201-2; BioGerm, Shanghai, China). RNase P was used as the endogenous internal standard. Real-time RT-qPCR was performed in the following conditions: 50°C for 10 min and 95°C for 5 min, then 40 cycles of amplification at 95°C for 10 s and 55°C for 40 s with the 7,500 Real-Time PCR System (Applied Biosystems, Bedford, MA, United States). The criteria for evaluation of the results were as follows: C_t value $\leq$ 38 as positive and simultaneous positive determination of both *ORF1ab* and *NP* genes.

Severe Acute Respiratory Syndrome Coronavirus 2 Detection by Nanopore Amplicon Sequencing

During SARS-CoV-2 genome amplification and library construction, 16 μl total RNA was used as a template for reverse transcription with the LunaScript RT SuperMix kit [E3010L; New England Biolabs (NEB), Ipswich, MA, United States] according to the manufacturer's instructions. A 5-μl cDNA production was then used as a template for targeted amplification of the SARS-CoV-2 genome with the Q5 Hot-Start High-Fidelity 2x Master Mix (M0494; NEB, Ipswich, MA, United States) and primer pools A and B (PCR tiling of COVID-19 virus, version 4; ONT, Didcot, United Kingdom). PCR was performed in the following conditions: 98°C for 30 s, followed by 30 cycles of amplification at 98°C for 15 s and 65°C for 2 min with the Veriti DX 96-well Thermal Cycler (Applied Biosystems, Bedford, MA, United States). DNAse/RNAse-free water was assayed in each batch as a negative control. A synthetic double-strand SARS-CoV-2 sequence with a length of 820 bp (NC_045512.2: nt 13,005–13,824) was used as a positive control. PCR products were cleaned up with the same volume of Agencourt AMPure XP beads (A63881; Beckman, Brea, CA, United States). In the next stage, the Ultra II End-Repair/dA-Tailing Module (E7546; NEB, Ipswich, MA, United States) was used for end-prep reaction and the Next Ultra II Ligation Module (E7595; NEB, Ipswich, MA, United States) used for barcode ligation with Native Barcoding Expansion 96 (NBD-196; ONT, Didcot, United Kingdom). MinION library preparation was performed according to the manufacturer's instructions in the Ligation Sequencing Kit (SQK-LSK109; ONT, Didcot, United Kingdom). Different barcoded samples were pooled with equal masses. The concentration of each library was measured using Qubit 4.0 Flurometer (Invitrogen, Carlsbad, CA, United States), and the volume of each library was then calculated to make an equimolar pool of libraries. The pool of libraries was finally sequenced using the ONT MinION platform (MinKNOW 19.12.5).

Sequencing Data Analysis

After Nanopore sequencing, the raw data were filtered using NanoFilt software[3]. High-quality reads ($Q \geq 10$) were retained. The clean data were mapped to the reference genome of SARS-CoV-2 isolate Wuhan-Hu-1 (accession no. NC_045512.2) using Minimap2 software (Li, 2018). The reads that could be mapped to the reference with $a \geq 50$ mapping quality score (MAPQ) were then compared with the reference using BLAST software (Altschul et al., 1990). When the alignment of a read with reference by BLAST covered >80% of the whole read and the sequence identity was ≥90%, this read was recognized as a SARS-CoV-2 read. For each sample, the sequencing depth of each position and the number of times that position was read were calculated with SAMtools mpileup. The calculation was based on the mapping results of the confirmed SARS-CoV-2 reads (Li et al., 2009). Genome coverage was determined using the proportion of sequenced genome position. It was calculated based on the

[2] https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w

[3] https://github.com/wdecoster/nanofilt

number of sequenced reference position, which was divided by the reference genome size (29,903 bp).

Statistical Analysis

To evaluate the detection ability of NAS for SARS-CoV-2 in clinical samples, we calculated the sensitivity and specificity of both the sequencing and the RT-qPCR results. Samples collected from COVID-19 patients were considered positive, while the nasopharyngeal swabs from others were considered negative for the virus. The cutoff of the RT-qPCR outputs was followed by the instruction of 2019-nCoV nucleic acid detection kit used ($C_t \leq 38$). The cutoff values of the read number and genome coverage from NAS were estimated by calculating the receiver operating characteristic (ROC) curve and the area under the curve (AUC) generated using the pROC package in R (Robin et al., 2011). Using different parameters of sequencing results, the sensitivity and specificity were then calculated once the maximum AUCs were obtained. Comparison of the C_t values and the genome coverage between the two groups was performed with a non-parametric comparison using the wilcox.test() function in R package. A p-value < 0.05 was considered statistically significant.

RESULTS

Rapid Severe Acute Respiratory Syndrome Coronavirus 2 Whole-Genome Detection by the Nanopore Amplicon Sequencing Approach

We enrolled 86 hospitalized patients diagnosed with COVID-19, with a median age of 65 years [interquartile range (IQR) = 50–76 years], with 52.3% being men (45 patients). There were 86 clinical samples collected from these patients, including 54 pharyngeal swabs, 31 sputum samples, and 1 fecal sample. To test the SARS-CoV-2 detection performance, we carried out a parallel design based on the NAS and RT-qPCR approaches for 86 clinical samples from COVID-19 patients (COVID-19 group) and 27 negative controls (NC group). In total, the detection of all 113 samples was performed in three R9.4 chips, which was completed in 8 h (**Figure 1A**).

After filtering low-quality data and carrying out the two-step identification of reads from SARS-CoV-2, the median number of viral read was 585.5 (IQR = 384–2,755) in the 86 clinical samples. According to the ARTIC protocol, the overlapping amplicons covered nearly the whole viral genome, except for the 3′ and 5′ untranslated regions, and the median size of amplicons was 392 bp (IQR = 383–400). Among the high-quality clean data, the median length of nanopore reads was 495 bp (IQR = 346–532). Surprisingly, viral reads were also identified in all 27 samples of the NC group, with a median number of 20 (IQR = 8–64). Based on the mapping results of the viral reads to the reference genome of SARS-CoV-2 isolated from Wuhan-Hu-1, the median reference genome coverage was 30.5% (IQR = 21.4–40.0%) in the COVID-19 group and was 5.2% (IQR = 4.0–6.5%) in the NC group. Both the read number and the genome coverage of the virus for the COVID-19 group were significantly higher than those of the NC group (**Figures 1B,C**).

High Sensitivity and Specificity of Nanopore Amplicon Sequencing

The sensitivity and specificity of NAS were then estimated by drawing the ROC curve and calculating the AUC based on the viral read number and reference genome coverage in the COVID-19 and NC groups (**Figures 1D,F**). Two samples from COVID-19 patients were determined as negative, with a read number cutoff of 123. Three NC samples were then determined as positive. Subsequently, we obtained the highest sensitivity of 97.7% (84/86), along with a specificity of 88.9% (24/27) and an AUC of 0.983 (95% CI = 0.9626–1). With the genome coverage cutoff set as 9.005%, three samples from COVID-19 patients were determined as negative, while none of the NC samples were determined as positive. The highest AUC of 0.994 (95% CI = 0.9862–1) was then obtained, along with a sensitivity of 96.5% (83/86) and a specificity of 100% (27/27). These results revealed the high sensitivity and specificity of NAS for the detection of SARS-CoV-2 (**Table 1**).

We next performed parallel RT-qPCR detection of SARS-CoV-2 in these samples to compare the efficiency of both two methods. With the standard C_t value cutoff set as 38, 17 samples from COVID-19 patients were negative and one NC sample was positive. A sensitivity of 80.2% (69/86) and a specificity of 96.3% (26/27) were obtained (**Figure 1E**). Compared with the performance of RT-qPCR, NAS showed obviously higher sensitivity and specificity. In addition, the genome coverage generated by the NAS method was in line with the C_t value of RT-qPCR, with an R^2 of 0.382 with the C_t value of the *ORF1ab* gene amplification and an R^2 of 0.311 with the C_t value of the *N* gene amplification (**Figure 2**).

The Viral Load in Sputum Estimated by the Nanopore Amplicon Sequencing Approach Might Be Related With the Severity and Duration of Infection

The C_t value of RT-qPCR has been widely applied as a surrogate indicator for viral load. However, the sensitivity of RT-qPCR was markedly lower than that of NAS in the detection of the long-term stored samples in this study (80.2% *vs.* 96.5% as reference genome coverage). The reference genome coverage calculated from the NAS detection and the C_t value of RT-qPCR were highly consistent. We therefore estimated the viral load of the samples using both values (the C_t value of the *ORF1ab* gene was applied due to its higher consistency with the genome coverage generated by the NAS approach) and then compared their estimated efficiency, followed by analysis of the variations in the viral loads of COVID-19 patients during the period of infection.

Among the 54 pharyngeal swabs, 35 samples were from patients within less than 4 weeks after onset of clinical manifestations (hereinafter referred to as "≤4w AFS") and 18 samples were from patients within longer than 4 weeks after the first clinical manifestations exhibited (hereinafter referred to as ">4w AFS") (the confirmed time of a patient was missing;

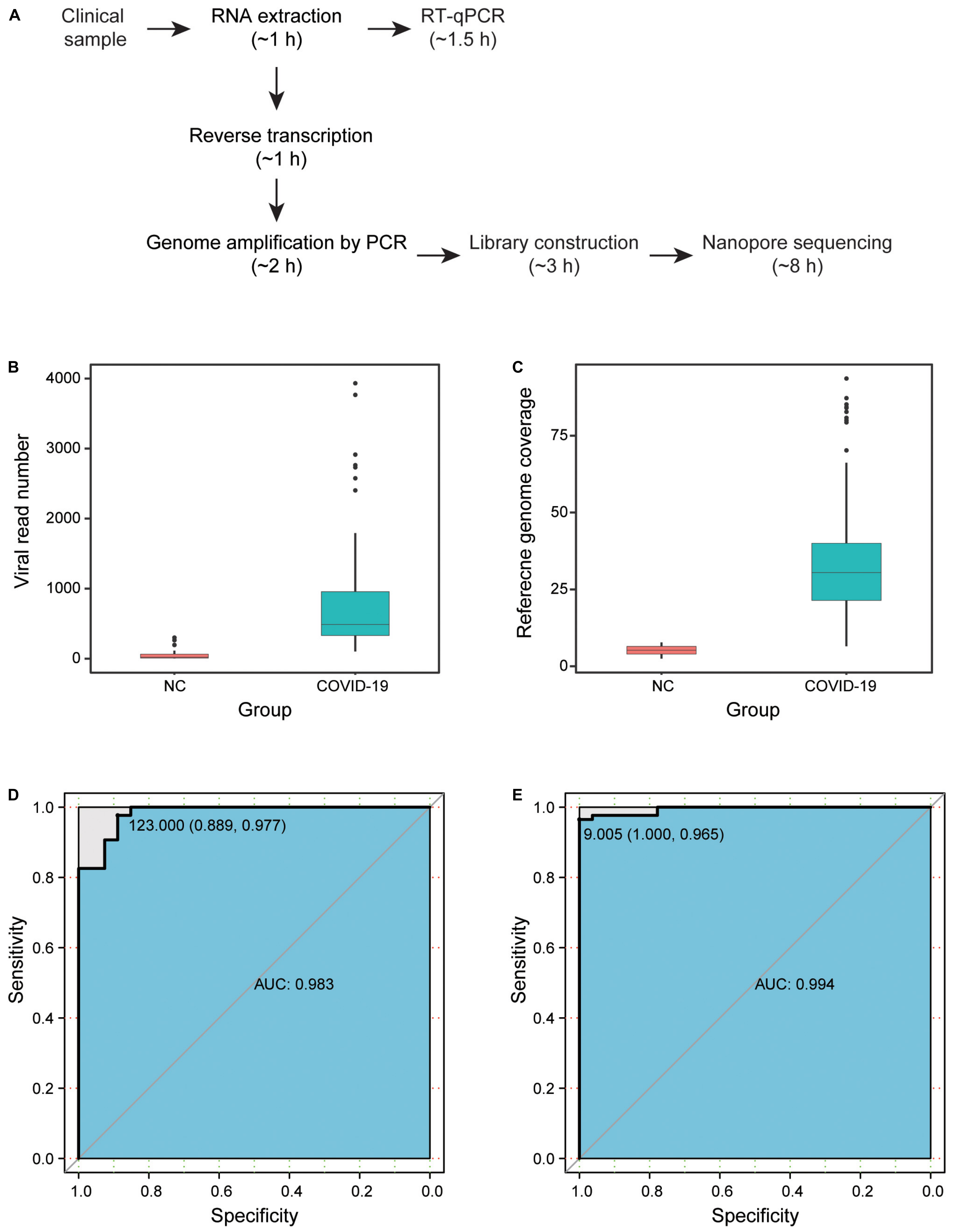


FIGURE 1 | Results of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) NAS. **(A)** Detection time of NAS and RT-qPCR. **(B)** Viral read number of samples in the COVID-19 and NC groups. **(C)** Reference genome coverage of samples in the COVID-19 and NC groups. **(D)** ROC curve based on the viral read number. **(E)** ROC curve based on the reference genome coverage. NAS, nanopore amplicon sequencing; RT-qPCR, reverse transcription quantitative polymerase chain reaction; ROC, receiver operating characteristic; NC, negative controls; COVID-19, coronavirus disease 2019.

TABLE 1 | Sensitivity and specificity of nanopore amplicon sequencing (NAS) and reverse transcription quantitative polymerase chain reaction (RT-qPCR).

	Group		Sensitivity (%)	Specificity (%)
	COVID-19	NC		
NAS viral read number				
≥123	84	3	97.7	88.9
<123	2	24		
NAS coverage				
≥9.005%	83	0	96.5	100
<9.005%	3	27		
RT-qPCR				
+	69	1	80.2	96.3
−	17	26		

NAS, nanopore amplicon sequencing; RT-qPCR, reverse transcription quantitative polymerase chain reaction; NC, negative controls; COVID-19, coronavirus disease 2019; +*, positive;* −*, negative.*

Figure 3A). For these two groups, the median C_t values were 29.95 and 28.75 and the median genome coverage values were 28.77% and 28.31%, respectively (**Figures 3B,C**). Neither the reference genome coverage nor the C_t value in pharyngeal swabs was significantly different between the two groups of patients (**Figures 3B,C**). Thirty-two pharyngeal swabs were from mild or moderate patients (mild group) and 22 samples were from severe or critical patients (severe group). In these two groups, the median C_t values were 28.8 and 29.4 and the median genome coverage were 30.2% and 23.8%, respectively (**Figures 3D,E**). There was also no significant difference in both C_t values and genome coverage in the pharyngeal swabs between the two groups. We then divided the 54 pharyngeal swabs into four groups based on both the onset of clinical manifestations and the severity of disease: ≤4w AFS and mild (19 samples), ≤4w AFS and severe (16), >4w AFS and mild (12), and >4w AFS and severe (6). For these four groups, the median C_t values were 30.6, 29.2, 28.7, and 29.9 (**Figure 3F**) and the median genome coverage values were 29.5%, 26.1%, 35.1%, and 23.7%, respectively (**Figure 3G**). No significant difference was observed between any two groups.

In contrast, of the 31 sputum samples, 8 samples were from patients within ≤4w AFS and 23 samples were from patients within >4w AFS (**Figure 4A**). For these two groups, the median C_t values were 25.2 and 28.6 and the median genome coverage values were 50.4% and 31.3%, respectively (**Figures 4B,C**). The C_t values in the sputum were not significantly different between the two groups (**Figure 4B**); however, the median genome coverage values in the sputum from patients within ≤4w AFS were markedly higher than those from patients >4w AFS, but the difference was not significant (**Figure 4C**). Twenty-three sputum samples were from the mild group and 8 samples were from the severe group. For these two groups, the median C_t values were 28.1 and 25.5 and the median genome coverage values were 31.5% and 30.3%, respectively (**Figures 4D,E**). There was also no significant difference in both C_t values and genome coverage in the sputum between these two groups (**Figures 4D,E**). We then divided the 31 sputum samples into four groups based on both the time diagnosis and the severity of the disease: ≤4w AFS and mild (5 samples), ≤4w AFS and severe (3), >4w AFS and mild (18), and >4w AFS and severe (5). For these four groups, the median C_t values were 26.7, 24.6, 28.6, and 27.7 and the median genome coverage values were 34.7%, 82.8%, 31.4%, and 28.6%, respectively (**Figures 4F,G**). The genome coverage in ≤4w AFS and severe group was dramatically higher than those of the other

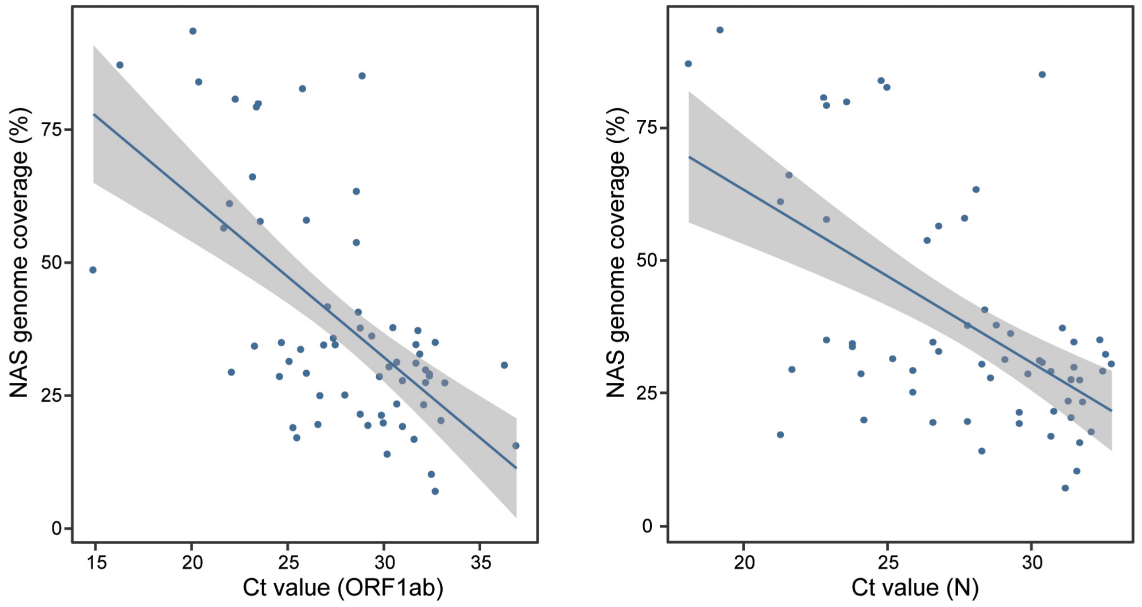


FIGURE 2 | Consistency of the genome coverage of NAS and the C_t value of RT-qPCR detection of the *ORF1ab* and *N* genes. *NAS*, nanopore amplicon sequencing; C_t, cycle threshold; *RT-qPCR*, reverse transcription quantitative polymerase chain reaction.

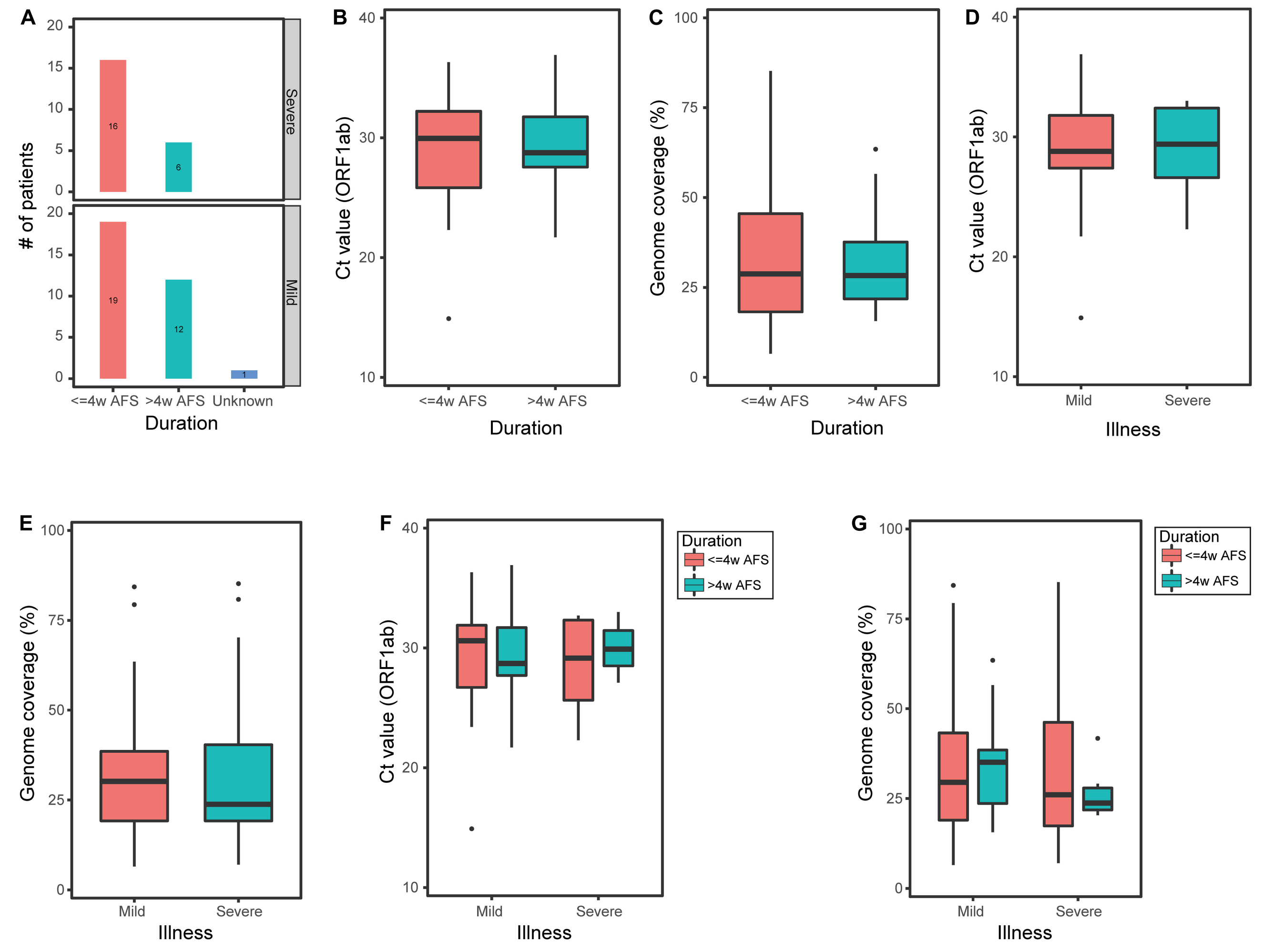


FIGURE 3 | C_t value of *ORF1ab* and reference genome coverage in pharyngeal swabs from COVID-19 patients. **(A)** Number of pharyngeal swabs from mild or moderate patients (mild group) and critical or severe patients (severe group) within less or longer than 4 weeks after onset of clinical manifestations (*≤4w AFS* or *>4w AFS*, respectively). **(B,C)** C_t value and reference genome coverage in pharyngeal swabs from patients with ≤4w AFS or >4w AFS. **(D,E)** C_t value and reference genome coverage in pharyngeal swabs from mild or severe groups. **(F,G)** C_t value and reference genome coverage in pharyngeal swabs in four groups based on both the time of diagnosis and the severity of patients. COVID-19, coronavirus disease 2019; AFS, after first showing clinical manifestations; C_t, cycle threshold.

groups, although the difference was not significant due to the limited number of samples, which was also in line with the feature of the C_t value (**Figures 4F,G**). Therefore, the viral load in sputum samples instead of that in pharyngeal swabs might be related to the severity of infection, particularly in patients with ≤4w AFS, which could be used as a secondary tool to evaluate the severity of SARS-CoV-2 infection.

DISCUSSION

In this study, we detected SARS-CoV-2 in 86 clinical samples from COVID-19 patients using the NAS method on the Nanopore minION sequencing platform. We calculated the viral read number and reference genome coverage in each sample to evaluate the detection results. We estimated the sensitivity and specificity of this method using these results and those from 27 negative control samples. Our analysis revealed both a higher sensitivity and a better specificity (96.5% and 100%, respectively) of NAS (based on genome coverage) compared with RT-qPCR (80.2% and 96.3%, respectively). The reference genome coverage of the viral reads was in line with the C_t value of RT-qPCR, which revealed that the reference genome coverage could also be used as an indicator of the viral load in samples. In addition, the total time cost was less than 15 h, which was also acceptable compared with that of RT-qPCR (∼2.5 h). These results showed that the NAS method was accurate and rapid for SARS-CoV-2 detection in clinical samples.

Our results revealed that the NAS detection approach used in this study would be more sensitive and accurate than that based on PCR alone. Firstly, dozens of positions along the viral genome were detected by overlapping amplification, whereas only one or two loci were detected by RT-qPCR. Secondly, the amplicons were sequenced and validated by subsequent

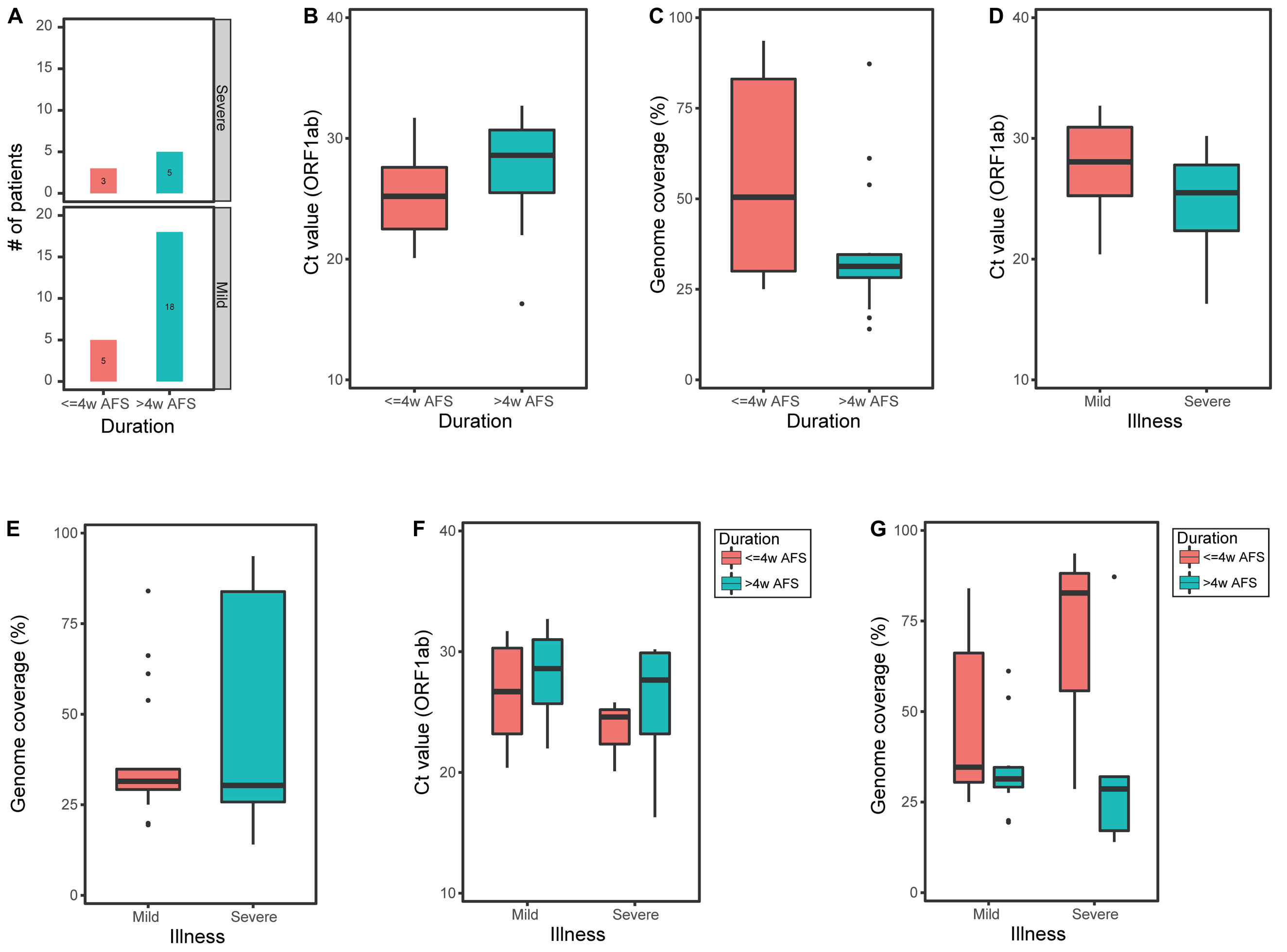


FIGURE 4 | C_t value of *ORF1ab* and reference genome coverage in sputum samples from COVID-19 patients. **(A)** Number of sputum samples from mild or moderate patients (mild group) and critical or severe patients (severe group) within less or longer than 4 weeks after onset of clinical manifestations (*≤4w AFS* or *>4w AFS*, respectively). **(B,C)** C_t value and reference genome coverage in sputum samples from patients with ≤4w AFS or >4w AFS. **(D,E)** C_t value and reference genome coverage in sputum samples from mild or severe groups. **(F,G)** C_t value and reference genome coverage in sputum samples in four groups based on both the time of diagnosis and the severity of patients. *COVID-19*, coronavirus disease 2019; *AFS*, after first showing clinical manifestations; C_t, cycle threshold.

sequence comparison. This whole viral genome detection led to a high sensitivity and a good specificity simultaneously. This approach was used to diagnose a COVID-19 patient, in which the routine RT-qPCR was negative due to a viral mutation on the primer site and only one read was obtained by NGS detection (Mehta et al., 2020).

The NAS approach might also be more valuable than those based on NGS for the detection of SARS-CoV-2. Numerous studies performed SARS-CoV-2 detection and research based on NGS and revealed great valuable knowledge about the genomic characteristics and transmission of the virus (Gudbjartsson et al., 2020; Lu et al., 2020; Rambaut et al., 2020). However, due to the fixed and prolonged run time of the NGS platform, the detection results are usually obtained in 2–3 days, which limited its application in clinical tests and rapid screening. In contrast, sequencing using the ONT minION platform is flexible due to its real-time nature. With the NAS method, we performed SARS-CoV-2 genome amplification firstly to enrich the viral nucleotide fragments. Therefore, the data required would be moderate in the subsequent sequencing, by which the detection time (15 h, or even shorter) would be reduced significantly compared to deep sequencing using the NGS platform. Meanwhile, a high accuracy was also achieved due to the subsequent data analysis based on the long sequencing read compared with NGS (400 bp for NAS *vs*. 75–250 bp for NGS).

Our results showed the low C_t value of RT-qPCR and high genome coverage in sputum samples than in pharyngeal swabs from both mild (or moderate) and severe (or critical) COVID-19 patients within ≤4w AFS. The results of the two methods revealed consistent trends of the viral load difference between the two types of samples and the two periods. It has been reported that the viral load in pharyngeal swabs was lower than that in sputum samples, and the C_t value of RT-qPCR in pharyngeal swabs decreased more rapidly than that in sputum samples (Huang et al., 2020). The significant difference revealed by the genome coverage of NAS indicated that this method and the genome coverage data might make a better tool to evaluate the viral load and perform quantitative monitoring of clinical samples. In addition, the detection and quantitative monitoring of sputum might be more sensitive than pharyngeal swabs, which have been reported to be helpful in evaluating disease progression (Yu et al., 2020).

There are several advantages of this study. Firstly, we performed whole-genome amplification of SARS-CoV-2 and sequencing with ONT real-time sequencing technology, which allowed highly sensitive detection in samples for which the RT-qPCR results were false negative. Secondly, as the C_t value of RT-qPCR could be a relatively accurate parameter for the viral load, which was highly correlated with the copy number of droplet digital PCR (Yu et al., 2020), we analyzed the correlation of the C_t value with the genome coverage from our data and found consistency in the two values. We evaluated the performance of NAS based on the viral read number and also found a high sensitivity and an acceptable specificity (97.7 and 88.9%, respectively). Thirdly, we analyzed the time-varying characteristics of the sputum samples and pharyngeal swabs. Although the viral load in a clinical sample taken from a specific site may not be representative of the overall viral burden in SARS-CoV-2-infected carriers, pharyngeal swabs collected from patients confirmed to have COVID-19 beyond 4 weeks should be interpreted with caution (Thi et al., 2020). This method would also be valuable in the detection of other critical clinical pathogens, in particular those that cannot be well tested using routine methods, such as virus and fungi.

This study has a few limitations. Firstly, we only collected clinical samples from one hospital in Wuhan during the COVID-19 epidemic. Moreover, the quality of sample collection was limited to discarded samples, and there was about 6 months between RNA extraction and the RT-qPCR or NAS test. These partly explain the several COVID-19 patients showing negative results, with RT-qPCR showing a much worse performance in terms of sensitivity. On one hand, the storage time of the samples could negatively affect the RNA integrity, then causing failure of primer binding or PCR extension. On the other hand, the kits for reverse transcription and amplification were also different during the RT-qPCR detection and generation of the Nanopore library. Therefore, the differences in the sensitivity could also be partially explained by the differences in kits used for reverse transcription and amplification. Furthermore, the numbers of the different types of samples were not balanced; however, it was consistent with clinical practice. In addition, we did not obtain complete genome from any sample. On one hand, the samples were stored for months from the collection to the RNA extraction; therefore the RNA might have degraded partly. On the other hand, the amplification efficiencies of the primers covering the whole viral genome were not equal; therefore, parts of the genomes would not have been amplified and sequenced well.

Here, we evaluated a detection solution based on Oxford nanopore sequencing that showed increased accuracy and sensitivity, as well as saving time, by performing viral genome amplification and nanopore sequencing on SARS-CoV-2 detection. Our results showed the high sensitivity and specificity of this method compared with RT-qPCR. However, the higher costs and longer detection time limited its application at present. It is more suitable as a supplementary method for RT-qPCR now, e.g., while facing the emergence of a new Omicron variant. The sequencing results were also used as an indicator of viral load to display the viral dynamics during infection. This study proved the wide application prospect of nanopore sequencing detection of SAR-CoV-2 and provided more knowledge about the clinical characteristics of COVID-19. This approach would also be suitable for the development of more rapid and sensitive detection methods for critical clinical pathogens to improve the infectious disease diagnostics and disease control measurements.

DATA AVAILABILITY STATEMENT

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) in National Genomics Data Center (Nucleic Acids Res 2021), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA004486) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Review Board of Peking University Third Hospital. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

X-XL: conceptualization, methodology, data curation, and writing – original draft. CL: conceptualization, data curation, and writing – editing. P-CD: investigation, formal analysis, visualization, and writing – original draft. Z-QZ, T-TL, and S-YL: investigation. LY: methodology, supervision, and writing – editing. C-KZ: writing – editing. S-CZ: formal analysis. YZ: writing – editing. C-RD: conceptualization, methodology, validation, and supervision. Q-GG: conceptualization, validation, and supervision. All authors took part in the final version for submission and accepted overall accountability for accuracy and integrity of the manuscript.

FUNDING

This study was supported by the China Cohort Consortium Foundation (Grant Number CCC2020007, receiver: X-XL, 2020), and Beijing Natural Science Foundation (Grant Number M22036, receiver: Q-GG, 2022).

ACKNOWLEDGMENTS

We thank all the physicians, nurses, pharmacists, and the anti-epidemic medical taskforces worldwide. You are the fighters on the frontline of COVID-19. We are very grateful to Professor Chen Jun (Department of Laboratory Medicine, Wuhan Pulmonary Hospital) for help with sample and data collection.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

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

Chen, C., Li, J., Di, L., Jing, Q., Du, P., Song, C., et al. (2020). MINERVA: a facile strategy for SARS-CoV-2 whole-genome deep sequencing of clinical samples. *Mol. Cell* 80, 1123–1134. doi: 10.1016/j.molcel.2020.11.030

FDA (2020). *Coronavirus Disease 2019 (COVID-19) Emergency Use Authorizations for Medical Devices*. Silver Spring, MD: FDA.

Gu, W., Deng, X., Lee, M., Sucu, Y. D., Arevalo, S., Stryke, D., et al. (2021). Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nat. Med.* 27, 115–124. doi: 10.1038/s41591-020-1105-z

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.

Huang, Y., Chen, S., Yang, Z., Guan, W., Liu, D., Lin, Z., et al. (2020). SARS-CoV-2 viral load in clinical samples from critically ill patients. *Am. J. Respir. Crit. Care Med.* 201, 1435–1438. doi: 10.1164/rccm.202003-0572LE

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

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. doi: 10.1016/j.cell.2020.04.023

Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., Manson, J. J., et al. (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* 395, 1033–1034. doi: 10.1016/S0140-6736(20)30628-0

Nasir, J. A., Kozak, R. A., Aftanas, P., Raphenya, A. R., Smith, K. M., Maguire, F., et al. (2020). A comparison of whole genome sequencing of SARS-CoV-2 using amplicon-based sequencing. Random hexamers, and bait capture. *Viruses* 12:895. doi: 10.3390/v12080895

Pollard, C. A., Morran, M. P., and Nestor-Kalinoski, A. L. (2020). The COVID-19 pandemic: a global health crisis. *Physiol. Genomics* 52, 549–557. doi: 10.1152/physiolgenomics.00089.2020

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

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). PROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. doi: 10.1186/1471-2105-12-77

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., et al. (2020). On the origin and continuing evolution of SARS-CoV-2. *Nat. Sci. Rev.* 7, 1012–1023. doi: 10.1093/nsr/nwaa036

Tang, Y. W., Schmitz, J. E., Persing, D. H., and Stratton, C. W. (2020). Laboratory diagnosis of COVID-19: current issues and challenges. *J. Clin. Microbiol.* 58, e00512–e00520. doi: 10.1128/JCM.00512-20

Thi, V. L. D., Herbst, K., Boerner, K., Meurer, M., Kremer, L. P., Kirrmaier, D., et al. (2020). A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Sci. Transl. Med.* 12:eabc7075. doi: 10.1126/scitranslmed.abc7075

Wang, M., Fu, A., Hu, B., Tong, Y., Liu, R., Liu, Z., et al. (2020). Nanopore targeted sequencing for the accurate and comprehensive detection of SARS-CoV-2 and other respiratory viruses. *Small* 16:2002169. doi: 10.1002/smll.202002169

Wang, Y., Kang, H., Liu, X., and Tong, Z. (2020). Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak. *J. Med. Virol.* 92, 538–539. doi: 10.1002/jmv.25721

Xiao, M., Liu, X., Ji, J., Li, M., Li, J., Yang, L., et al. (2020). Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med.* 12:57. doi: 10.1186/s13073-020-00751-4

Yu, F., Yan, L., Wang, N., Yang, S., Wang, L., Tang, Y., et al. (2020). Quantitative detection and viral load analysis of SARS-CoV-2 in infected patients. *Clin. Infect. Dis.* 71, 793–798. doi: 10.1093/cid/ciaa345

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: P-CD, LY, Z-QZ, T-TL, C-KZ, and S-CZ were employed by Beijing YuanShengKangTai (ProtoDNA) Genetech Co. Ltd., Beijing, China.

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

Publisher's Note: 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.

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