Biosafety and Health 6 (2024) 61–69

Contents lists available at ScienceDirect

Biosafety and Health

journal homepage: www.elsevier.com/locate/bsheal

Original Research

Genomic surveillance of emerging SARS-CoV-2 Omicron variations in Tianjin Municipality, China 2022

Xin Gao [a,b], Ming Zou [a,b], Yue Lei [a,b], Zhaolin Tan [a,b], Zhichao Zhuang [a,b], Baolu Zheng [a,b], Aiping Yu [a,b], Yanzhen Han [a,b], Xiaohui Lu [a,b], Xiaochang Liu [a,b], Ying Wang [a,b], Yuan Wang [a,b], Liru Guo [a,b], Guangwen Liu [a,b], Wen Li [a,b], Yang Liu [a,b], Likun Lv [a,b], Peiyong Ning [a,b], Xiaoyan Li [a,b,*]

[a] Tianjin Centers for Disease Control and Prevention, Tianjin 300011, China
[b] Tianjin Key Laboratory of Pathogenic Microbiology of Infectious Disease, Tianjin Centers for Disease Control and Prevention, Tianjin 300011, China

ARTICLE INFO

Article history:
Received 1 December 2023
Revised 29 February 2024
Accepted 1 March 2024
Available online 2 March 2024

Keywords:
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
Coronavirus disease 2019 (COVID-19)
Omicron variant
Whole genome sequencing (WGS)

ABSTRACT

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has severely impacted public health. In 2022, the Omicron variant of SARS-CoV-2 rapidly became the dominant circulating variant in the local COVID-19 outbreaks in Tianjin Municipality, China. To gain a deeper understanding of the genetic variations of the Omicron variant in Tianjin, specimens from individuals who tested positive for SARS-CoV-2 between December 2021 and November 2022 were used for virus whole genome sequencing and phylogenetic analysis. A total of 1,674 high-quality Omicron sequences were obtained, consisting of 1,339 sequences from local cases belonging to 20 Phylogenetic Assignment of Named Global Outbreak (PANGO) lineages and 335 sequences from imported cases belonging to 70 lineages. Tianjin experienced five waves of local outbreaks, accompanied by multiple substitutions among subvariants, ranging from the initial BA.1.1 lineage to the subsequent BA.2, BF.7, and BA.5.2 lineages. The evolutionary rate of local strains, estimated to be 28.999 substitutions per year, and the evolutionary rate of imported strains, estimated to be 24.946 substitutions per year, were lower than that of the strains circulating globally. The additional substitutions and deletions of local strains have been used to identify and disrupt the virus transmission chains. The subvariants such as BA.5.2.48, BA.5.2.49, BF.7.14, and XBB.1 circulating in the fifth epidemic wave presented criterial immune escape mutations including S: R346T, S: L452R and S: F486V. It is essential to implement genomic surveillance strategies to investigate further the development of genomic mutation characteristics in the SARS-CoV-2 variant. This ongoing monitoring will contribute to a better understanding of the virus's genetic changes and aid in effective control measures.
© 2024 Chinese Medical Association Publishing House. Published by Elsevier BV. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is classified as a single positive-strand ribonucleic acid (RNA) virus belonging to the *Betacoronavirus* genus within the Coronaviridae family. Its genome comprises 14 functional open reading frames (ORFs), which encode 29 proteins, including four structural proteins: nucleocapsid (N) protein, membrane (M) protein, spike (S) protein and envelope (E) protein, and 16 nonstructural proteins (NSPs) encoded by the *Orf1a* and *Orf1ab* genes, which are cleaved by a specific protease, and nine accessory proteins: Orf3a, Orf3b, Orf6, Orf7a, Orf7b, Orf8, Orf9b, Orf9c, and Orf10 [1]. SARS-CoV-2, as an RNA virus, undergoes rapid evolution and accumulates genetic diversity over relatively short periods, especially during widespread human-to-human transmission [2]. The Omicron variant of SARS-CoV-2, the currently circulating variant of concern (VOC) defined by the World Health Organization (WHO), emerged in South Africa and Botswana in November 2021. It has since become the dominant variant globally, surpassing previously circulating VOCs such as Alpha, Beta, Gamma, and Delta variants [3]. The Omicron variant has evolved into numerous lineages according to the Phylogenetic Assignment of Named Global Outbreak (PANGO) Lineage nomenclature. WHO has identified five lineages, XBB.1.5, XBB.1.16, EG.5, BA.2.86, and JN.1, as variants of interest (VOIs). Additionally, five lineages, namely DV.7, XBB, XBB.1.9.1, XBB.1.9.2, and XBB.2.3, have been classified as variants under monitoring (VUMs). These lineages exhibit crucial amino acid mutations in the S protein.

Since January 2022, Tianjin Municipality in China has encountered multiple outbreaks of local coronavirus disease 2019 (COVID-19) epidemics caused by diverse subvariants of the Omicron variant. To con-

* Corresponding author: Tianjin Centers for Disease Control and Prevention, Tianjin 300011, China.
E-mail address: xiaoyanli1291@163.com (X. Li).

https://doi.org/10.1016/j.bsheal.2024.03.001
2590-0536/© 2024 Chinese Medical Association Publishing House. Published by Elsevier BV.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

HIGHLIGHTS

Scientific question

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has severely impacted public health. Tianjin Municipality was the first province in China to experience the COVID-19 outbreak caused by the Omicron variant and has experienced multiple outbreaks of local COVID-19 epidemics caused by diverse subvariants of Omicron. This paper comprehensively analyzes the genomic characteristics of the Omicron variant in the several rounds of COVID-19 outbreaks in Tianjin, China.

Evidence before this study

There are differences in epidemic scale, epidemic area, and source of infection among the COVID-19 outbreaks in Tianjin, and these differences may be related to SARS-CoV-2 mutation. The Omicron variant continues to mutate, and some critical mutations are closely related to the immune escape ability and virus transmissibility.

New findings

In 2022, outbreaks in Tianjin were caused by different subvariants of Omicron. In the fifth outbreak, the BA.5.2 and BF.7 lineages became the dominant circulating variants, and these variants had S: R346T, S: L452R, and S: F486V variants associated with enhanced viral immune evasion. The mutation rate of circulating variants in Tianjin was lower than that of international circulating variants during the same period. The CH.1.1 variant has been detected in imported COVID-19 patients for the first time in China.

Significance of the study

This article reports the genomic characteristics of the Omicron variant that caused various outbreaks in Tianjin in 2022, illustrates the critical role of whole-genome sequencing technology in epidemic control, and emphasizes the necessity of genome surveillance.

trol these local epidemics effectively, a combination of whole genome sequencing (WGS) technology and epidemiological investigations have been employed to trace the origins of infection. This approach enables the government to gain real-time insights into the genetic mutations of the virus, monitor the prevalence of the Omicron variant within the local population, and evaluate the risk of imported cases from overseas contributing to the COVID-19 epidemic.

This study employed high-throughput WGS technology to sequence SARS-CoV-2 positive specimens obtained from local and imported COVID-19 cases in Tianjin between December 2021 and November 2022. The study's primary objectives were to analyze the genome characteristics of the virus, enhance the SARS-CoV-2 genome database, and provide valuable insights to aid in developing effective control strategies for COVID-19 outbreaks.

2. Materials and methods

2.1. Collection of specimens and SARS-CoV-2 detection

Nasopharyngeal swab specimens were collected from regular SARS-CoV-2 surveillance in Tianjin Municipality of China collected from overseas imported personnel and local individuals between December 2021 and November 2022. Total RNAs were extracted by using QIAamp Viral RNA Mini Kit (Qiagen, Germany. Cat. No. 52904) and viral deoxyribo nucleic acid (DNA) / RNA extraction kits with automatic nucleic acid extraction instrument (Xi'an Tianlong Science and Technology Co., China. Cat. No. NP968-C). SARS-CoV-2 detection was conducted *via* real-time reverse transcriptase polymerase chain reaction (RT-PCR) using nucleic acid detection kits (Shanghai BioGerm Medical Technology Co., China) and ABI 7500 real-time PCR instrument (Thermo Co., USA) according to the manufacturers' instructions. The SARS-CoV-2 positive RNA samples with cycle threshold (Ct) values less than 32 were prepared for SARS-CoV-2 WGS.

2.2. SARS-CoV-2 WGS

ULSEN SARS-CoV-2 whole-genome capture kits (MicroFuture Co., Beijing, China) were used for cDNA synthesis and genome enrichment as described by the manufacturer, and the amplified products were purified by QIAquic PCR purification kit (Qiagen, Germany. Cat. No. 28106). The purified amplicons were conducted for sequencing with Oxford nanopore platforms (Oxford Nanopore Technologies PLC., U. K.) and Illumina platforms (Illumina, Inc., USA). For the next-generation sequencing (NGS) with Illumina MiniSeq and NextSeq 2000 platforms (Illumina, San Diego, CA, USA), the purified amplicons were subjected to library preparation by using Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) according to manufacturer's protocol followed by 0.6 × volumes of AMPure XP beads purification. The libraries were sequenced using 2 × 150 reads with MiniSeq reagents (300 cycles) and NextSeq 2000 P1 reagents (300 cycles). For the third-generation sequencing with Oxford nanopore GridON platform (Oxford Nanopore Technologies, Oxford, U. K.), the purified amplicons were treated with NEBNext End repair/dA-tailing Module (New England Biolabs, USA) and were barcoded with native barcodes and sequencing adapters (EXP-NBD104 and EXP-NBD114 kits) (Oxford Nanopore Technologies, Oxford, U.K.). GridON was performed for genome sequencing following the ligation sequencing protocol with SQK-LSK109 (Oxford Nanopore Technologies, Oxford, UK) in the flow cell after adding the final libraries.

2.3. Consensus sequences generation

For analysis of SARS-CoV-2 sequencing, fastQ raw data was exported from Illumina and GridON sequencing platforms, and the workflow created by Qiagen CLC genomics workbench version 21, which consisted the functions of reads quality control, reads trim, reads mapping to reference, consensus extraction, and variant calling, was used to generate consensus sequences and call single-nucleotide variants relative to the reference sequence of isolate Wuhan-Hu-1 (GenBank accession: NC_045512.2) [4]. Whole genome sequences of strains were submitted to the Global Initiative of Sharing All Influenza Data (GISAID) database [5].

2.4. Phylogenetic analysis

The obtained sequences of Omicron variant were exported into fasta files for identification of PANGO Lineage [6] via PANGO Lineage web application (https://pangolin.cog-uk.io/) and information of mutations by using Nextclade CLI v2.10.0 [7] (https://docs.nextstrain.org/projects/nextclade/en/stable/user/nextclade-cli.html) under platforms of miniconda v4.12 and python v3.9.7. Genomic sequences were filtered for amino acid mutation analysis and phylogenetic analysis by checking the quality of sequences in Nexclade CLI analysis results. The multiple sequences alignments of Omicron variants sequences were generated via MAFFT v7.42 with FFI-NS-2 methods [8]. Model Finder was used to obtain the fittest substitution model for the genome sequences based on the Bayesian information criterion (BIC) results. Maximum likelihood (ML) trees were generated *via* IQ-tree v2.1.2 [9]. The time-scaled phylogenetic trees were constructed, and the evolutionary rates were estimated using the SARS-CoV-2 geno-

mic surveillance workflow implemented in Nextstrain CLI v7.1.0 [7] and TreeTime [10]. Phylogenetic trees were visualized with R v4.2.2 and ggtree package v2.4.132 [11].

3. Results

3.1. Distribution of SARS-CoV-2 evolutionary lineage in Tianjin from December 2021 to November 2022

According to the daily reports on the COVID-19 epidemic situation released by the official website of the National Health Commission of the People's Republic of China, between December 1, 2021, and November 12, 2022, Tianjin experienced a total of five waves of local COVID-19 epidemics (January 8 to February 6, February 24 to May 13, May 15 to June 3, June 26 to August 8, August 27 to November 12). The occurrence of these waves was determined based on the daily distribution of new infections, including confirmed cases and asymptomatic infections (Fig. 1). Nasopharyngeal swab specimens collected from laboratory-confirmed COVID-19 cases with a Ct value of less than 32 were sequenced during this period. A total of 1,896 sequences were generated from these specimens. All sequences belonged to VOCs: Omicron (1,512 sequences from local cases and 371 sequences from imported cases) and Delta (13 sequences from imported cases). Among the obtained sequences of the Omicron variant, 88.90 % (1,674/1,883) sequences, which included 1,339 sequences from local cases and 335 sequences from imported cases, met the quality criteria (genome of length more than 27,000; number of gaps less than 3,000) for further analysis.

The Omicron variant sequences derived from local infections in Tianjin were classified into 20 PANGO lineages. Time-resolved phylogenetic analysis revealed multiple rounds of substitutions from different clades. (Fig. 2). Notably, specimens collected from January to February 2022 were mostly represented by sequences belonging to the BA.1.1 lineage, with four sequences belonging to the BA.1.15 lineage. From February to June 2022, all specimens exclusively belonged to the BA.2 lineage, with thirteen sequences falling into the BA.2.2.1 lineage. During the period of July to August 2022, specimens were in majority represented by the BA.5.2.1 and BA.2.76 lineages. Subsequently, from August to October 2022, the predominant lineage observed in the specimens was the BF.21 lineage, accompanied by one sequence from the BA.5.1.1 lineage. Finally, from October to November 2022, the majority of specimens were primarily represented by the BA.5.2.48 and BA.5.2.49 lineages, along with the BF.7.14 lineage. Notable, there was one sequence from the XBB.1 lineage and one sequence from the BN.1.3 lineage during this timeframe (Fig. 3A).

The first two imported cases of the SARS-CoV-2 Omicron variant in Tianjin on December 13, 2021, were previously reported [12]. From these cases, two Omicron sequences were obtained, and one sequence belonging to the BA.1 lineage, with a high genetic coverage of 99.62 %, was included in this study. The high-quality Omicron variant sequences obtained from imported COVID-19 cases were divided into 70 PANGO lineages (Fig. 3B). By analyzing the distribution of sequences based on the sampling dates, it was observed that specimens collected between December 2021 and June 2022 were predominantly represented by BA.1, BA.2, and their sub-lineages. The specimens collected after June 2022 were mainly associated with BA.5 and its sub-lineages, with a smaller proportion of specimens represented by BA.4 and BA.2.75 lineages. The count of sequences belonging to the BA.2.75, BQ.1, and XBB lineages gradually increased after August 2022.

3.2. SARS-CoV-2 mutation rate during five epidemics

We further analyzed the evolutionary rate of the Omicron subvariants in Tianjin. For the whole genome sequencs obtained from local epidemics, the evolutionary rate was estimated to be 28.999 substitutions per year. On the other hand, for sequences derived from imported cases, the evolutionary rate was calculated to be 24.946 substitutions per year. Comparing these rates to the estimate provided by the Nextstrain website, which reported an annual nucleotide evolution rate of 29.663 substitutions per year based on current statistics during the same period [7], we found that the evolutionary rates observed in Tianjin, both for local and imported cases, were lower (Fig. 4).

3.3. Mutation analysis and phylogenetic analysis of SARS-CoV-2 among five local epidemics

Through a detailed mutation analysis, we identified critical immune escape amino acid mutations in the S protein of local emerging lineages in this study. The BF.21, BA.5.2.48, BA.5.2.49 and BF.7.14 lineages from local cases in the fifth epidemic wave displayed both S: L452R and S: F486V mutations, with a mutation frequency of 100 % (Fig. 5A). The S: L452R mutation was also identified in the genome of CH.1.1 lineage along with the S: F486S mutation from an imported case, and the CH.1.1 lineage was identified in China for the first time. Additionally, the BF.7.14 lineage from local cases also presented the S: R346T mutation with a 100 % mutation frequency (Fig. 5A). The S: R346T mutation was also found in the sequences of BQ.1.1 and XBB.1 lineages from imported cases.

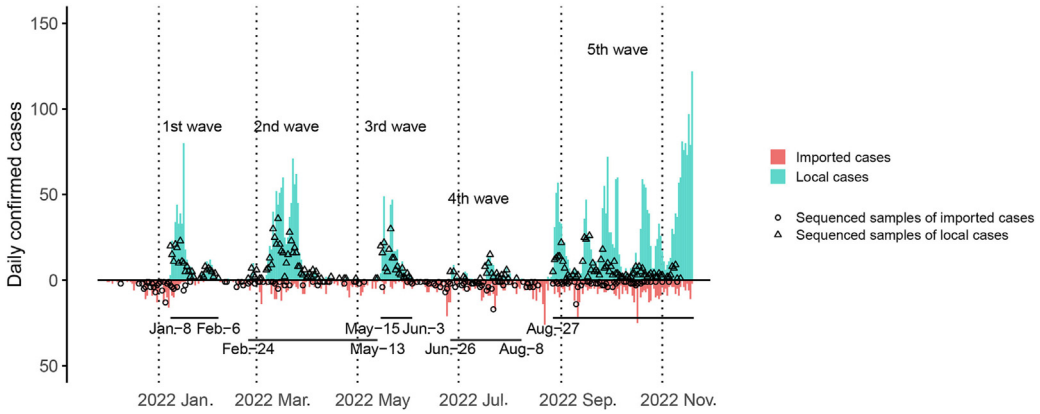

Fig. 1. Temporal dynamics of local and imported COVID-19 cases in Tianjin between December 2021 and November 2022. The histogram showed the number of daily reported laboratory-confirmed local and imported infections based on COVID-19 epidemic reports released by official website of the National Health Commission of the People's Republic of China. Daily number of obtained SARS-CoV-2 sequences from local cases and imported cases corresponding to the sampling date was marked by triangles and circles, respectively. The five local epidemic waves were annotated above the histogram. The start and end time of each epidemic wave were marked under the histogram. Abbreviations: COVID-19, coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

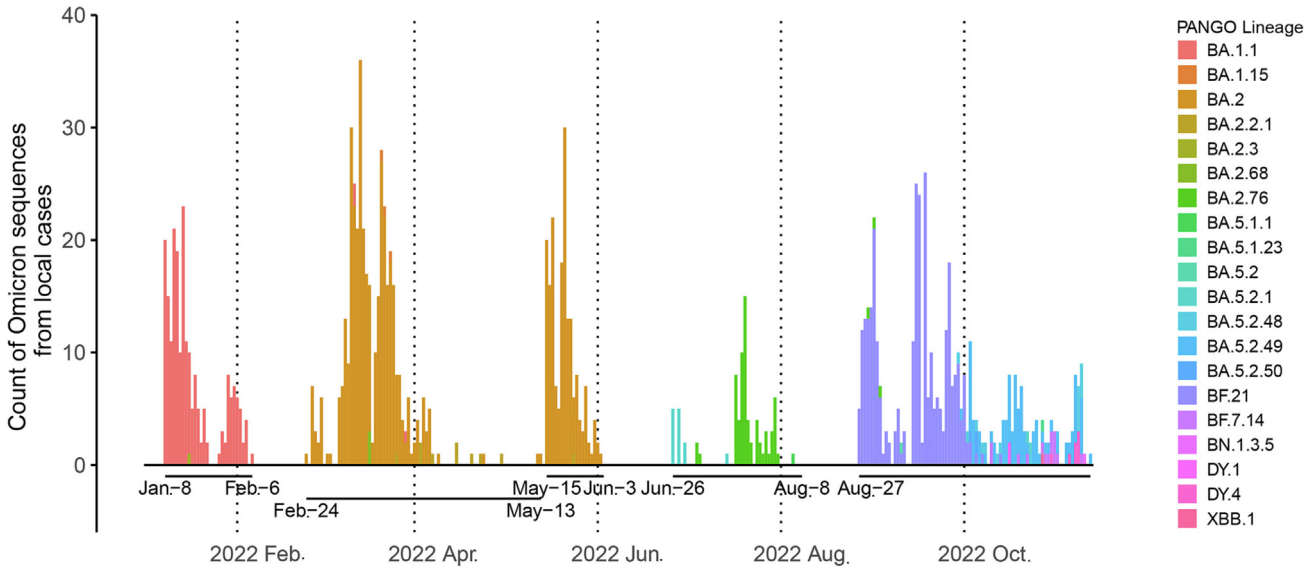

Fig. 2. Temporal dynamics in lineages of strains from local COVID-19 cases in Tianjin from January to November 2022. The histogram showed the number of obtained sequences of the Omicron variant corresponding to the sampling date of local case specimens from January to November 2022 in Tianjin. The lineages of strains were marked by different colors. The start and end times of each local epidemic wave were marked under the *X*-axis. Abbreviation: COVID-19, coronavirus disease 2019.

We performed phylogenetic analysis and mutation analysis of the dominant circulating variants in each local outbreak in more detail. The results showed distinct virus transmission chains originating from different sources of infection during each local epidemic wave (Fig. 5-B-F). Furthermore, certain strains with new mutations were observed to propagate further. The additional deletion mutations were observed during the virus transmission processes, which also aided in identifying the transmission chains. For instance, in the first epidemic wave, an independent transmission chain was formed by strains carrying 27,792–27,794 deletions in the *Orf7b* gene (Fig. 5B). During the second epidemic wave, two separate transmission chains were identified, one carrying the T2752C, C7471T and C27476T substitutions, and the other carrying the G5617T substitution (Fig. 5C). In the third epidemic wave, most virus strains evolved from a common original strain, which including strains carrying the new mutation G1509A and strains with other new mutations, such as C19875T, G19900A, and C28758T, forming two independent transmission chains (Fig. 5D). For the fourth epidemic wave, the strains carrying the A8658G mutation and other strains with C245T mutation formed two transmission chains in the outbreak caused by the BA.2.76 lineage, which carried notable different mutations compared with the mutations of the BA.5.2.1 lineage with the C22879T and G24740A mutations (Fig. 5E). In the fifth epidemic wave (October to November 2022), the virus strains could be categorized into two main lineages: BA.5.2.48 and BA.5.2.49. These lineages exhibited notable differences in mutation sites and were associated with different main epidemic areas in China (Fig. 5F). Hence, it was feasible to determine the virus's probable origin by identifying the strains' genotypes. Deletion mutations in the 509–524 region of the *Orf1ab* gene were observed in four local virus transmission processes caused by the BA.1.1, BA.2, and BF.21 lineages (Fig. 6). Identifying the newly added characteristic substitution mutations and deletion mutations provided valuable insights into determining the transmission chains of the outbreaks. This information assisted in epidemiological investigations and the development of effective epidemic control measures in Tianjin.

4. Discussion

This study comprehensively analyzes the spread of the SARS-CoV-2 Omicron variant in Tianjin during 2022. Starting from December 2021, the proportion of Omicron variant sequences gradually increased among all sequences from imported cases. After February 2022, all sequenced specimens from imported cases were identified as the Omicron variant, replacing the previously dominant Delta variant. In January 2022, Tianjin became the first province in China to experience a local outbreak caused by the Omicron variant, which was followed by a total of five local epidemic waves. Phylodynamic analysis of the obtained sequences revealed the prevalence and substitutions of multiple sub-lineages of the Omicron variant in Tianjin, ranging from the initial BA.1.1 lineage to the subsequent BF.7 and BA.5.2 lineages. The lineage distribution of dominant circulating strains in the first four epidemic waves was relatively single, with the first wave mainly attributed to the BA.1.1 lineage, the second and third waves caused by the BA.2 lineage, and the fourth wave caused by BA.5.2.1 and BA.2.76 lineages. These waves were successfully blocked before August 2022 through rapid quarantine, clinical treatment of infections, and effective surveillance and control measures for imported cases from other provinces and countries. Lineages such as BA.1.15, BA.2.2.1, and BA.2.68, which differed from the dominant strains previously circulating during the first four epidemic waves, did not cause large-scale outbreaks according to the genomic surveillance data. The fifth epidemic wave can be divided into two stages: an early stage from August 27, 2022, mainly caused by the BF.21 lineage, which was subsequently controlled in September 2022. However, the later stage of the fifth wave, starting from August 29, 2022, witnessed a rapid increase in the number of infections primarily attributed to the BF.7.14, BA.5.2.48, and BA.5.2.49 lineages, along with sporadic emergence of the XBB.1, BQ.1, and other lineages. The lineage distribution in the fifth wave's later stage was more complicated compared to the first four waves, with more infections and a higher risk of epidemic importation. The BF.7 and BA.5.2 lineages have demonstrated increased fitness compared to the prototype, with approximately 24 and 20 times higher fitness, respectively [13]. The evolutionary rates of subvariants observed in local epidemic waves and imported cases were lower than those of the internationally circulating variants during the same period, which could be attributed to the effective and rapid control measures implemented to curb the spread of the virus.

Identifying the source of infection and identifying and tracing the transmission chains are essential measures for halting the spread of the outbreaks [14]. WGS technology is integral in providing direct molecular evidence for determining transmission chains. By analyzing the sequences of index cases in each outbreak, specific mutation sites

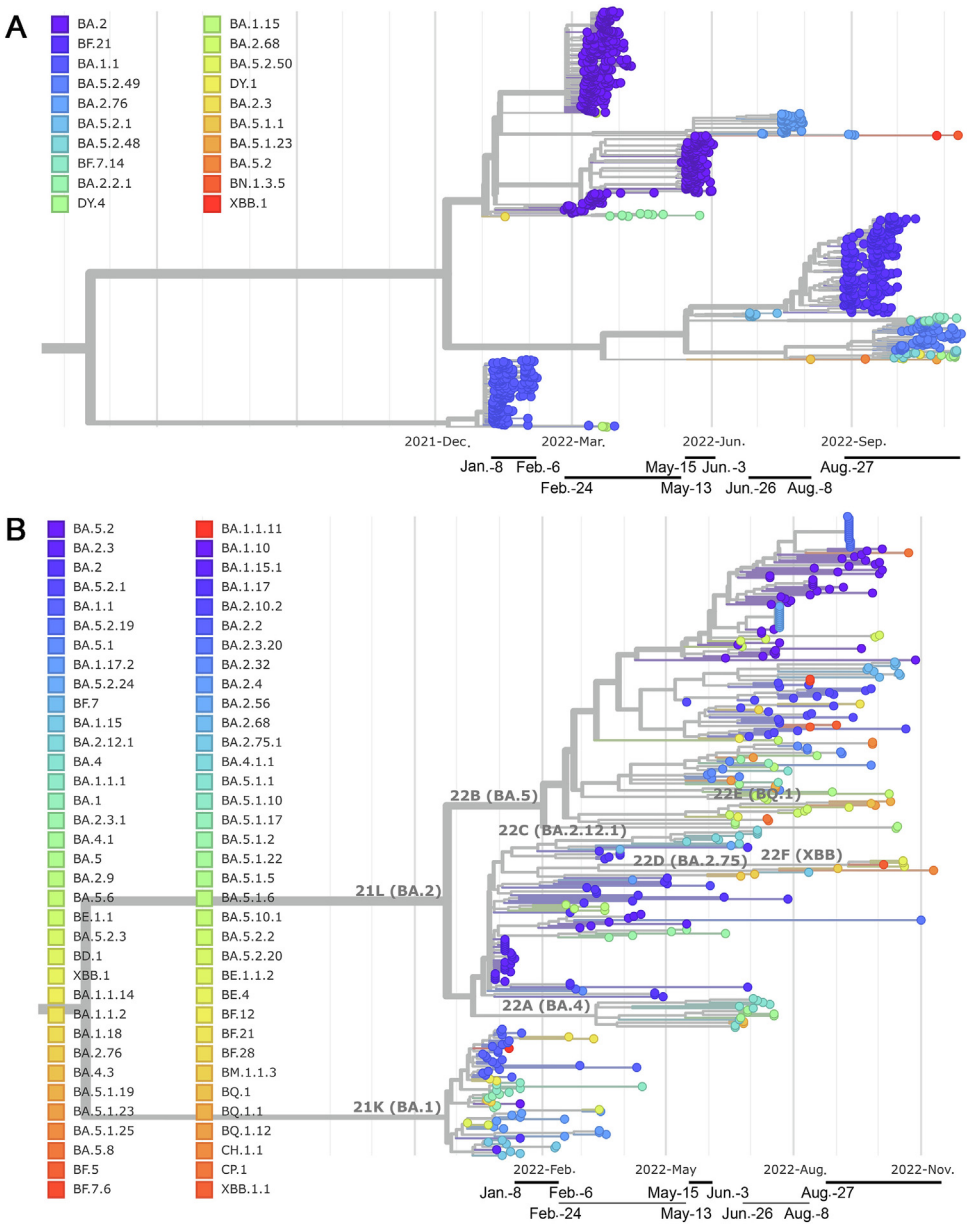

Fig. 3. Phylogenetic trees of Omicron subvariants circulating in Tianjin between December 2021 and November 2022. The time-scaled phylogenetic trees of Omicron subvariants from local cases (A) and imported cases (B). The sequences were labelled by PANGO lineages in points with different colors. The start and end times of each local epidemic wave were marked under the *X*-axis. Abbreviation: PANGO, Phylogenetic Assignment of Named Global Outbreak.

can be identified, and additional mutations that occur during subsequent virus transmission can be observed. These detailed mutation profiles enable the differentiation of infection chains within the same epidemic wave. For example, during the second local epidemic wave, the BA.2 lineages were the dominant subvariants circulating. However, further analysis revealed that these lineages could be divided into four sub-lineages carrying diverse mutations. This finding suggests that they originated from different sources and formed distinct transmission chains. To control the local epidemics caused by imported cases, comparative analysis of sequences with national and global SARS-CoV-2 genome databases can help trace the source of infection, identify at-risk groups and areas, and inform control measures. By comparing the sequences obtained from local cases with existing databases, similarities and differences can be observed, providing valuable insights into the origin and spread of the virus.

Continuous genomic surveillance of SARS-CoV-2 variants has proven valuable in monitoring their mutation characteristics and assessing their impact on various aspects of the virus, such as transmissibility, the associated disease severity, the performance of vaccines, diagnostic tools, or other public health measures. The numerous mutations of the Omicron variant and its sub-lineages, including BA.1, BA.2, BA.4, and BA.5, exhibit a high number of mutations that contribute to increase viral fitness and the ability to escape the neutralization efficiency induced by prior vaccination or infection [15]. The S protein of SARS-CoV-2, particularly its receptor-binding domain (RBD), is a target of essentially all neutralizing antibodies. The RBD of the Omicron variant contains at least 15 mutations, some of which are of concern, such as the K417N and N501Y mutations, as they contribute to immune evasion and enhanced infectivity [16]. Previous studies have identified twelve important immune escape mutation sites in the S protein, including R346, K356, K444, V445, G446, N450, L452, N460, F486, F490, R493, and S494, which have been shown by deep mutational scanning (DMS) to facilitate evasion of neutralizing antibodies [17,18]. Antibodies elicited by BA.1 infection after vaccination predominantly are largely evaded by S mutants L452Q, L452R, and F486V. Additionally, new clones of BA.1-specific antibod-

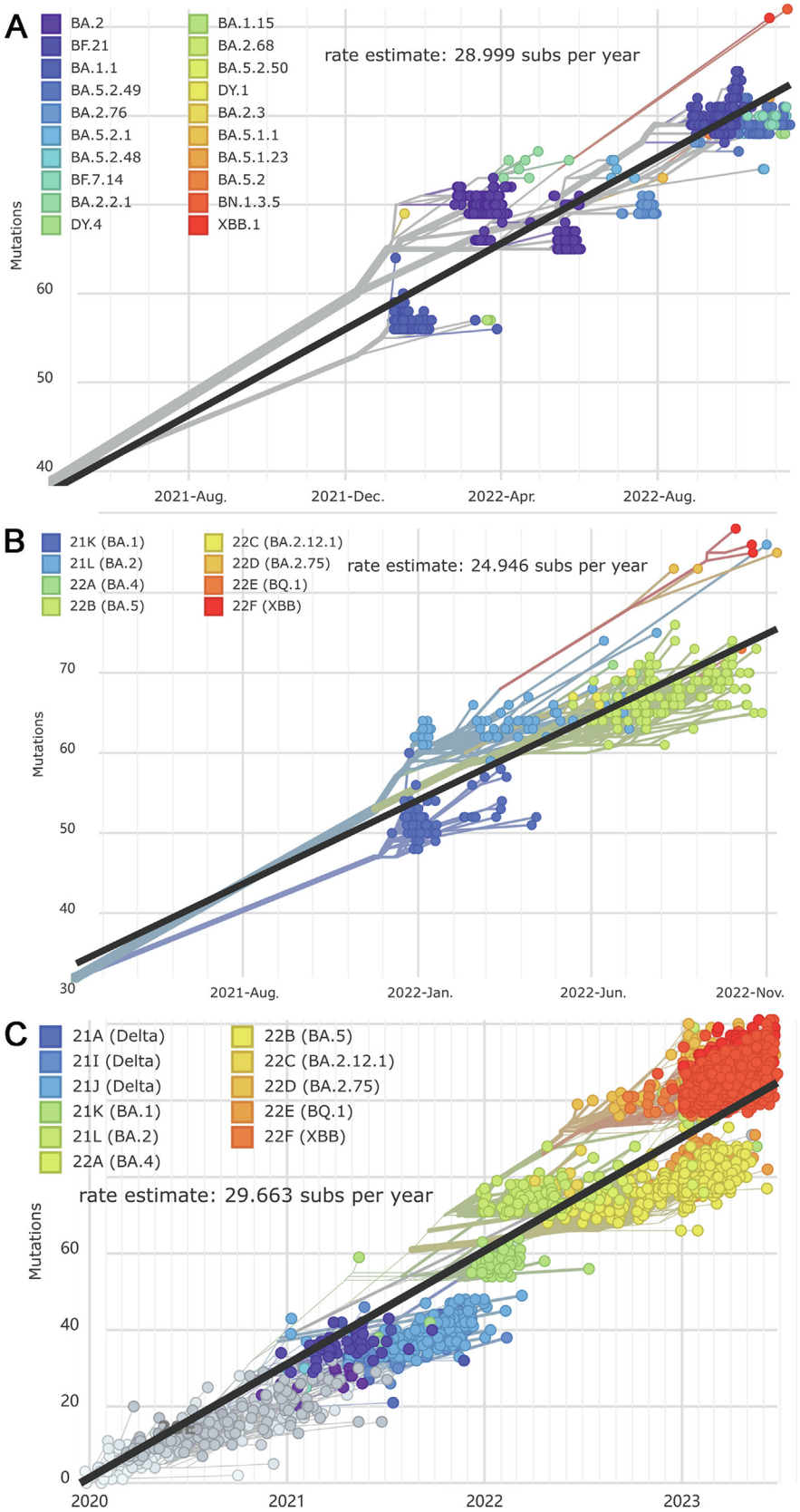

Fig. 4. Evolutionary rate of strains from local and imported cases, and strains circulating globally between December 2021 and November 2022. A) The evolutionary rate of strains from local epidemics in Tianjin between January 2022 and November 2022. B) The evolutionary rate of strains from imported cases in Tianjin between December 2021 and November 2022. C) The evolutionary rate of variants circulating globally obtained from the Nextstrain website (https://nextstrain.org/ncov/gisaid/global/6m) based on GISAID data between December 2021 and November 2022. Abbreviations: GISAID, the Global Initiative of Sharing All Influenza Data; subs, substitutions.

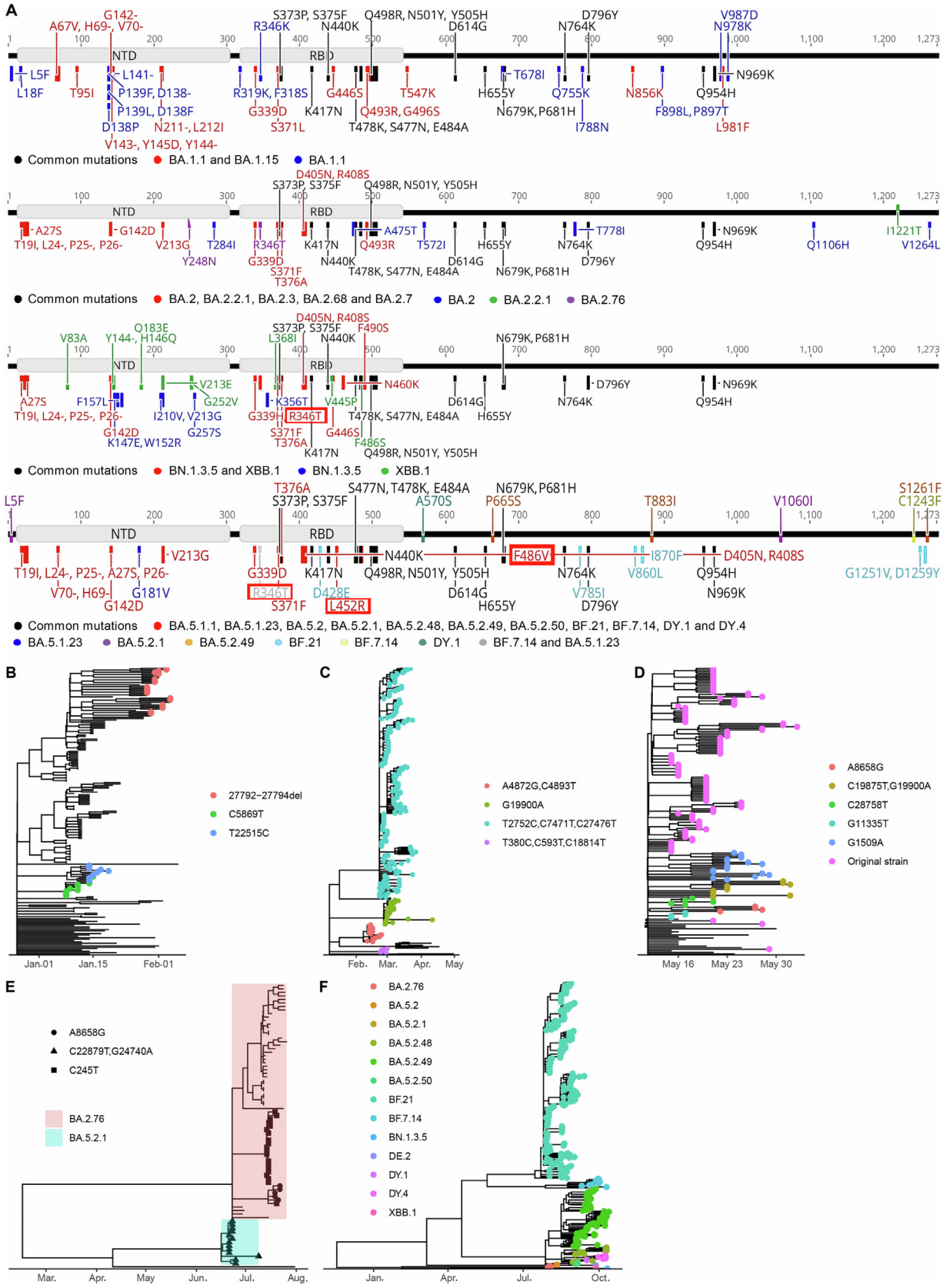

Fig. 5. Spike mutations observed in strains from local cases in Tianjin and phylogenetic trees of strains from each local epidemic wave. A) The schematic diagram showing the mutations in the Spike protein in strains from local cases in Tianjin. The common mutations obtained from all sequences of local epidemic waves were marked in black. The specific mutations obtained from sequences of different lineages were marked in different colors. The criterial immune escape amino acid mutations in the Spike protein were marked by red rectangles. B)-F) The time-scaled phylogenetic trees of Omicron subvariants circulating in the five local epidemic waves. Clustered strains with shared specific mutations were labelled by nucleic acid mutation sites in points with different colors (B-E). Different subvariants were labelled by PANGO lineages in different colors (E,F). Abbreviation: PANGO, Phylogenetic Assignment of Named Global Outbreak.

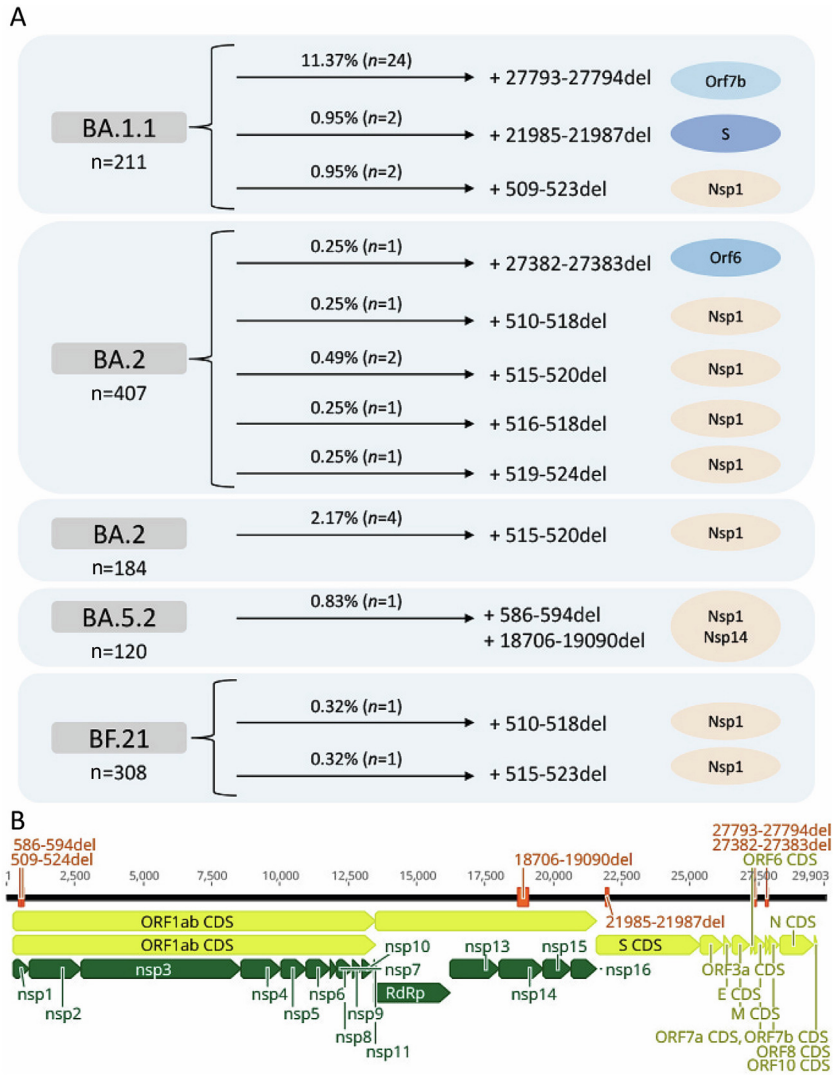

Fig. 6. The additional deletion mutations obtained in strains during local transmission. A) The additional deletion mutations formed in the virus transmission processes. B) The schematic diagram shows the additional deletion mutations in strains from local cases in Tianjin on SARS-CoV-2 genome. The deletion mutations were marked in red. Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; ORF, open reading frames; CDS, coding sequences; S, spike; N, nucleocapsid; E, envelope; M, membrane.

ies induced by BA.1 infection are largely evaded by BA.2 and BA.4/BA.5 due to D405N and F486V mutations [19]. The BA.2.3.20, BA.2.75.2, BQ.1.1, and XBB lineages have emerged and displayed a significant growth advantage over BA.5, with mutations observed on several hotspots of their RBD, indicating convergent evolution. Previous studies have demonstrated that BQ.1.1.10, BA.4.6.3, XBB, and CH.1.1 are among the most antibody-evasive strains tested [17]. In Tianjin, COVID-19 cases have been identified with BF.7 and XBB.1 lineages, previously defined as Omicron subvariants under monitoring, and other previously defined VUMs, including BQ.1 and BA.2.75. CH.1.1 lineage carrying the S: L452R mutation, BQ.1, BQ.1.1 with S: R346T, and XBB.1 lineage have been detected in imported cases that need constant monitoring. Moreover, the CH.1.1 lineage was identified in China for the first time. Crucially, immune escape mutations, including R346T identified in BF.7.14 lineage, and L452R and F486V mutations observed in BA.5.2.48, BA.5.2.49, and BF.7.14 lineages, may be associated with the high prevalence of these subvariants and the rapid increase in infections in Tianjin between October 2022 to early 2023. The L452R mutation is also present in the Delta, Kappa, and Epsilon variants, and a sequence belonging to the A lineage without the D614G mutation was previously identified in Tianjin in 2021 [20]. The L452Q mutation is found in the Lambda variant [21]. As of November 2022, XBB.1.5 lineage with S: F486P mutation was not identified.

The study also revealed the presence of deletion mutations during virus transmission, specifically the relatively common deletion mutations in the Nsp1-coding region (509–524 deletions). Another deletion mutation in a different region (686–694 deletions) of the Nsp1 was found in the transmission process of the BA.5.2 lineage, which had been previously reported in the early stages of the COVID-19 pandemic in 2020 [22]. Nsp1, known as the leader protein, plays a central role in hampering the anti-viral innate immune response, particularly interferon-alpha expression, and it has been considered a potential target for therapeutic interventions aimed at reducing viral pathogenicity [23].

This study had several limitations that need to be acknowledged. Firstly, the application of WGS technology and the availability of sequencing facilities limited the number of sequences with high genomic coverage that could be obtained. Resultantly, only a portion of specimens with high viral load could be sequenced, and complete virus genome sequencing was challenging for all positive specimens. Consequently, the number of obtained sequences was lower than the actual count of COVID-19 cases, which may have implications for accurately assessing the prevalence of genomic lineages. Furthermore, relying on

the obtained genome sequences to infer the actual situation of genomic lineage prevalence introduces an element of uncertainty. The sequences obtained may not fully represent the diversity and distribution of all circulating lineages in the population. These limitations highlight the need for further advancements in sequencing technology, increased sequencing capacity, and broader sampling to provide a more comprehensive understanding of the genomic landscape and lineage prevalence in COVID-19 cases.

5. Conclusion

This study emphasizes the importance of genomic surveillance in monitoring and controlling COVID-19 epidemics, identifying transmission chains, and developing effective control measures. Continuous and comprehensive genomic surveillance is essential for adapting strategies to mitigate the spread of emerging SARS-CoV-2 variants.

Ethics statement

The specimens involved in this study were all obtained from the remaining specimens of routine SARS-CoV-2 surveillance project by Tianjin Centers for Disease Control and Prevention. The researchers did not have access to any personal information of the participants, and the participants would not receive any risk from the study. Objectively, the researchers were unable to find the participants. Therefore, we submitted the request for exemption from informed consent to the Tianjin Centers for Disease Control and Prevention Ethics Committee for ethical review, which has been approved (No. TJCDC-R-2022-017).

Acknowledgements

The research was funded by the National Key Research and Development Program of China (No. 2023YFC3041500), a key project of the Tianjin Municipal Health Commission (No. MS20014), Tianjin Municipal Health Commission Project (No. TJWJ2022ZD010), Tianjin Municipal Administration for Technology Project (No. 20JCZDJC00130) and Tianjin Key Medical Discipline (Specialty) Construction Project (No. TJYXZDXK-050A). This study's genome sequences and metadata are available on GISAID.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Author contributions

Xin Gao: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Ming Zou:** Investigation. **Yue Lei:** Funding acquisition, Project administration, Resources, Supervision, Investigation. **Zhaolin Tan:** Investigation. **Zhichao Zhuang:** Investigation. **Baolu Zheng:** Investigation. **Aiping Yu:** Investigation. **Yanzhen Han:** Investigation. **Xiaohui Lu:** Investigation. **Xiaochang Liu:** Investigation. **Ying Wang:** Investigation. **Yuan Wang:** Investigation. **Liru Guo:** Investigation. **Guangwen Liu:** Investigation. **Wen Li:** Investigation. **Yang Liu:** Investigation. **Likun Lv:** Investigation. **Peiyong Ning:** Investigation. **Xiaoyan Li:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision.

References

[1] C. Bai, Q. Zhong, G.F. Gao, Overview of SARS-CoV-2 genome-encoded proteins, Sci. China Life Sci. 65 (2022) 280–294, https://doi.org/10.1007/s11427-021-1964-4.

[2] J.W. Lin, C. Tang, H.C. Wei, B. Du, C. Chen, M. Wang, Y. Zhou, M.X. Yu, L. Cheng, S. Kuivanen, et al., Genomic monitoring of SARS-CoV-2 uncovers an Nsp1 deletion variant that modulates type I interferon response, Cell Host Microbe 29 (2021) 489–502.e8, https://doi.org/10.1016/j.chom.2021.01.015.

[3] R. Viana, S. Moyo, D.G. Amoako, H. Tegally, C. Scheepers, C.L. Althaus, U.J. Anyaneji, P.A. Bester, M.F. Boni, M. Chand, et al., Rapid epidemic expansion of the SARS-CoV-2 omicron variant in southern Africa, Nature 603 (2022) 679–686, https://doi.org/10.1038/s41586-022-04411-y.

[4] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, et al., A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269, https://doi.org/10.1038/s41586-020-2008-3.

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

[6] Á. O'Toole, E. Scher, A. Underwood, B. Jackson, V. Hill, J.T. McCrone, R. Colquhoun, C. Ruis, K. Abu-Dahab, B. Taylor, et al., Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool, Virus Evol. 7 (2021) veab064, https://doi.org/10.1093/ve/veab064.

[7] J. Hadfield, C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R.A. Neher, Nextstrain: real-time tracking of pathogen evolution, Bioinformatics 34 (2018) 4121–4123, https://doi.org/10.1093/bioinformatics/bty407.

[8] K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform, Nucleic Acids Res. 30 (2002) 3059–3066, https://doi.org/10.1093/nar/gkf436.

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

[10] P. Sagulenko, V. Puller, R.A. Neher, TreeTime: maximum-likelihood phylodynamic analysis, Virus Evol. 4 (2018) vex042, https://doi.org/10.1093/ve/vex042.

[11] G. Yu, D. Smith, H. Zhu, Y. Guan, T. Lam, Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data, Methods Ecol. Evol. 8 (2017) 9, https://doi.org/10.1111/2041-210X.12628.

[12] Z. Tan, Z. Chen, A. Yu, X. Li, Y. Feng, X. Zhao, W. Xu, X. Su, The first two imported cases of SARS-CoV-2 omicron variant - Tianjin municipality, China, December 13, 2021, China CDC Wkly 4 (2022) 76–77, https://doi.org/10.46234/ccdcw2021.266.

[13] Y. Pan, L. Wang, Z. Feng, H. Xu, F. Li, Y. Shen, D. Zhang, W.J. Liu, G.F. Gao, Q. Wang, Characterisation of SARS-CoV-2 variants in Beijing during 2022: an epidemiological and phylogenetic analysis, Lancet 401 (2023) (2022) 664–672, https://doi.org/10.1016/s0140-6736(23)00129-0.

[14] Z.J. Cheng, Z. Zhan, M. Xue, P. Zheng, J. Lyu, J. Ma, X.D. Zhang, W. Luo, H. Huang, Y. Zhang, et al., Public health measures and the control of COVID-19 in China, Clin Rev Allergy Immunol 64 (2023) 1–16, https://doi.org/10.1007/s12016-021-08900-2.

[15] S. Chatterjee, M. Bhattacharya, S. Nag, K. Dhama, C. Chakraborty, A detailed overview of SARS-CoV-2 omicron: its sub-variants, mutations and pathophysiology, clinical characteristics, immunological landscape, immune escape, and therapies, Viruses 15 (2023) 167, https://doi.org/10.3390/v15010167.

[16] Y. Cao, J. Wang, F. Jian, T. Xiao, W. Song, A. Yisimayi, W. Huang, Q. Li, P. Wang, R. An, et al., Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies, Nature 602 (2022) 657–663, https://doi.org/10.1038/s41586-021-04385-3.

[17] Y. Cao, F. Jian, J. Wang, Y. Yu, W. Song, A. Yisimayi, J. Wang, R. An, X. Chen, N. Zhang, et al., Imprinted SARS-CoV-2 humoral immunity induces convergent omicron RBD evolution, Nature 614 (2023) 521–529, https://doi.org/10.1038/s41586-022-05644-7.

[18] Y. Sun, M. Wang, W. Lin, W. Dong, J., Xu, evolutionary analysis of omicron variant BF.7 and BA.5.2 pandemic in China, J. Biosaf. Biosecur. 5 (2023) 14–20, https://doi.org/10.1016/j.jobb.2023.01.002.

[19] Y. Cao, A. Yisimayi, F. Jian, W. Song, T. Xiao, L. Wang, S. Du, J. Wang, Q. Li, X. Chen, et al., BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by omicron infection, Nature 608 (2022) 593–602, https://doi.org/10.1038/s41586-022-04980-y.

[20] X. Li, X. Gao, M. Zou, Z. Zhuang, Z. Tan, B. Zheng, A. Yu, X. Su, SARS-CoV-2 genomic sequencing revealed N501Y and L452R mutants of s/a lineage in Tianjin municipality, China, Virol Sin. 36 (2021) 1228–1231, https://doi.org/10.1007/s12250-021-00432-5.

[21] H. Tegally, M. Moir, J. Everatt, M. Giovanetti, C. Scheepers, E. Wilkinson, K. Subramoney, Z. Makatini, S. Moyo, D.G. Amoako, et al., Emergence of SARS-CoV-2 omicron lineages BA.4 and BA.5 in South Africa, Nat. Med. 28 (2022) 1785–1790, https://doi.org/10.1038/s41591-022-01911-2.

[22] F. Benedetti, G.A. Snyder, M. Giovanetti, S. Angeletti, R.C. Gallo, M. Ciccozzi, D. Zella, Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1, J. Transl. Med. 18 (2020) 329, https://doi.org/10.1186/s12967-020-02507-5.

[23] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, et al., Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, Acta Pharm. Sin. B 10 (2020) 766–788, https://doi.org/10.1016/j.apsb.2020.02.008.