Zeng *et al. Zoonoses* (2024) 4:8
DOI 10.15212/ZOONOSES-2023-0032

ORIGINAL ARTICLE

# The Intra-Host Evolution of SARS-CoV-2 After Neutralizing Antibody Therapy, Revealed by Nanopore Sequencing

**Hong-Xiang Zeng[1,4,#], Wen-Hong Zu[1,4,#], Hai-Yan Wang[1,4], Jing Yuan[2], Lin Cheng[1,4], Gang Xu[1,4], Yi-Gan Huang[1,4], Yang Liu[1,4], Shu-Ye Zhang[3,*] and Zheng Zhang[1,4,*]**

## Abstract

**Objective:** In the context of two Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) outbreaks involving local transmission and an international flight, we used meta-transcriptome and multi-amplicon sequencing to successfully acquire the complete viral genome sequences from clinical samples with varying viral loads.

**Methods:** To enhance viral transcript presence, we used a primer pool for reverse transcription and sequenced the samples with nanopore sequencing, and successfully acquired the entire genomic sequence of the virus within less than 4 hours. In a substantial sample size of approximately 800 clinical specimens, we thoroughly examined and compared different sequencing methods.

**Results:** Meta-transcriptome sequencing was effective for samples with viral reverse transcription polymerase chain reaction (RT-PCR) threshold cycle (Ct) values below 22, whereas multi-amplicon sequencing was effective across a wide Ct range. Additionally, enriched nanopore sequencing was valuable in capturing the complete genome sequence when rapid results are required.

**Conclusion:** Through monitoring the viral quasi-species in individual patients, we observed ongoing viral evolution during neutralizing antibody therapy and found evidence that vaccine administration may affect the development of viral quasi-species. Overall, our findings highlight the potential of this viral sequencing strategy for both outbreak control and patient treatment.

**Keywords:** SARS-CoV-2, Mutation, Vaccine, Antibody

[#]These authors contributed equally to this work.
***Corresponding authors:**
E-mail: zhangzheng1975@aliyun.com (Z.Z); shuye_zhang@fudan.edu.cn (SYZ)

[1]Institute for Hepatology, National Clinical Research Center for Infectious Disease, Shenzhen Third People's Hospital, The Second Affiliated Hospital, School of Medicine, Southern University of Science and Technology, Shenzhen 518112, Guangdong Province, China
[2]Department of Infectious Diseases, National Clinical Research Center for Infectious Disease, Shenzhen Third People's Hospital, The Second Affiliated Hospital, School of Medicine, Southern University of Science and Technology, Shenzhen 518112, Guangdong Province, China
[3]Clinical Center for Biotherapy at Zhongshan Hospital, Fudan University, Shanghai 200031, China
[4]Shenzhen Research Center for Communicable Disease Diagnosis and Treatment, Chinese Academy of Medical Sciences, Shenzhen, People's Republic of China

Received: July 22 2023
Revised: November 2 2023
Accepted: January 3 2024
Published Online: February 1 2024

## INTRODUCTION

The COVID-19 pandemic, caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection, continues to threaten health and the economy globally. The RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2 is widely recognized to have low fidelity, because of its lack of error correction mechanisms, thus leading to the emergence of viral quasi-species [1] characterized by a multitude of genetic variants within the population [2,3]. The presence of quasi-species has been observed in various RNA viruses, including the Ebola virus [4], Zika virus [5], influenza A virus [6], human immunodeficiency virus [7,8], and SARS-CoV [9] and SARS-CoV-2 [10]. The formation of viral quasi-species facilitates viral adaptation to dynamic environments [11],

thereby enabling evasion from immune defenses [12,13]. Additionally, the presence of viral quasi-species modulates disease severity in infected hosts [14]. The characterization of viral quasi-species composition has potential in monitoring the transmission of viral infections [15]. Moreover, the pool of variants within viral quasi-species has served as a source for the emergence of epidemic variants of concern for SARS-CoV-2 [16].

The process of evolution of quasi-species of SARS-CoV-2 during antiviral therapy remains unclear, as do the effects of neutralizing antibodies and vaccines on these viral quasi-species. SARS-CoV-2 has exhibited rapid evolution, thus leading to the emergence of new variants of concern, such as Alpha [17], Delta [18], and Omicron [19]. Certain variants, such as the D614G substitution variant, have enhanced binding ability to angiotensin-converting enzyme-2 (ACE2) [20], thereby increasing viral replication and infectivity [21]. Some variants with the N501Y substitution exhibit enhanced infection and transmission [22]. The L452R variants demonstrate enhanced infectivity through evasion of cellular immunity [23]. Mutant variants characterized by elevated viral load and fatality rates [24] have impeded ongoing vaccine efforts and neutralizing antibody-based therapies. Notably, several variants carrying mutations such as K417N/T, E484K, and N501Y have demonstrated high resistance to immune responses elicited by BNT162b2 or mRNA-1273 vaccines [25], as well as the Pfizer or AstraZeneca vaccines [26]. The B.1.351 and P.1 variants have been observed to evade therapeutic neutralizing antibodies [27], whereas the recently emerged Omicron variant has shown extraordinary capacity to escape neutralizing antibody responses [28,29].

Investigation of the evolution of SARS-CoV-2 quasi-species has been limited, primarily because of challenges in obtaining adequate numbers of viral reads from clinical samples with low viral load [30]. To address this challenge, we collected swab samples from patients treated at our hospital between May 2021 and July 2021. These samples included those from the cluster outbreak at Yantian Port in Yantian District, Shenzhen City [31], as well as an outbreak that occurred during international flight CA868, originating from South Africa and arriving in Shenzhen [32]. To investigate the evolutionary patterns of SARS-CoV-2 quasi-species, we gathered swab samples longitudinally from individuals within this cohort, spanning from the initial stages of infection to the ultimate recovery phase. Notably, this particular group included individuals both with and without SARS-CoV-2 vaccination, and a subset that underwent treatment with BRII-196 and BRII-198 neutralizing antibodies [33], thus providing a unique opportunity to examine the effects of vaccines and neutralizing antibodies on the evolutionary trajectory of viral quasi-species. Furthermore, in response to the emergent SARS-CoV-2 outbreak, we devised contingency measures to expedite the sequencing of the SARS-CoV-2 genome and acquire its genomic sequences from clinical RNA samples within 4 to 28 hours. Through in-depth analysis of the genomic data, we observed that antibodies exert a discernible influence on the evolutionary dynamics of SARS-CoV-2 quasi-species within hosts during therapeutic interventions, thereby offering novel perspectives for the development of antibody-based interventions.

## MATERIALS AND METHODS

### Ethical declaration

This study was conducted according to the principles of the Declaration of Helsinki. Ethical approval was obtained from the Research Ethics Committee of Shenzhen Third People's Hospital (2020-192). Written consent was obtained from patients or their guardians when samples were collected. Patients were informed of the surveillance before providing written consent, and data directly associated with disease control were collected and anonymized for analysis.

### RNA extraction and qRT-PCR

Nasal swab, throat swab, and urine samples were collected from the patients. Supernatants from cultured cells infected with body fluid from patients infected by SARS-CoV-2 were also collected. Total RNA from these samples was extracted with a QIAamp Viral RNA Mini Kit (Qiagen, Cat. No. 52904). Quantitative reverse transcription polymerase chain reaction detection of ORF-1a/b and the N gene of SARS-CoV-2 was conducted with a commercial kit (GeneoDX Co., Ltd., Shanghai, China), according to the manufacturer's protocol. The RNA samples confirmed to be SARS-CoV-2 positive and with Ct values less than 35 were further used for viral clade typing.

### ARMS qPCR identification of variants

We obtained primers specifically designed to detect the mutations HV69-70del, K417N, K417T, L452R, E484K, E484Q, N501Y, and A570D in the Spike protein. These primers were included in a commercially available kit (BGI, PGI030019) for identifying and differentiating between wild-type and mutant strains of SARS-CoV-2. Eight mutations were detected and identified with primers that contained specific mutations labeled with different fluorophores in three separate polymerase chain reaction (PCR) reaction tubes. The first reaction tube contained the amplification primers to detect the wild-type ORF1ab gene and its corresponding fluorescent probe (FAM), primers to amplify the N501 region and K417 region of Spike, and their corresponding probes to detect N501Y (VIC) and K417N (ROX) mutations. The second reaction tube contained the amplification primers and mutation detection probes for HV69-70del (VIC/HEX), K417T (ROX), E484K (FAM), and A570D (CY5). The third reaction tube included the amplification primers and mutation detection probes for L452R (ROX) and E484Q (VIC). All three reaction tubes contained the reagents for PCR amplification and an internal reference. The presence of mutations was confirmed by comparison between the ΔCt value of each mutant gene and the Ct

value of the ORF1ab gene. A mutation was confirmed if ΔCt≤6 and was considered not detected if ΔCt>6.

## Designing the primer pool for SARS-CoV-2 enrichment

The reverse complement sequence of SARS-CoV-2's RNA genome was obtained with WH-01 (MN908947.3) as the reference genome. Tiling primers of 15 nt were selected for every 15 nt starting from the 34th base. Finally, a total of 963 tiling primers were obtained. The library of primers was synthesized by Sangon Biotech (Shanghai, China).

## Building a nanopore sequencing library

RNA (13 μl) was incubated at 65°C for 5 min in a thermal cycler, then placed on ice for 2 min to denature the RNA. The RNA sample was then treated with 2 μl of 5× gDNA wiper Mix (Vazyme, China) and incubated at 42°C for 2 min to remove genomic DNA contamination. Then 2 μl 10× RT mix, 2 μl HiScript III Enzyme Mix (R323, Vazyme, China), and 5 μl primer pool targeting SARS-CoV-2 (2.5 μM) were added to the previous reaction, and incubation was performed at 25°C for 5 min, 37°C for 45 min, and 85°C for 5 s to synthesize the first-strand complementary DNA (cDNA) (RNA:cDNA hybrid). Subsequently, 20 μl of the first-strand cDNA synthesis reaction was collected, and 26 μl Second Strand Buffer, 1 μl RNase H, and 3 μl T4 DNA polymerase from a Second Strand cDNA Synthesis Kit (D7172, Beyotime Biotech) were added; the reaction was incubated at 16°C for 60 min to obtain double-stranded DNA (dsDNA). The dsDNA was purified with Agencourt AMPure XP beads (A63881, Beckman Coulter) according to the manufacturer's instructions. Finally, the dsDNA was eluted from beads with 21 μl nuclease-free water. The dsDNA was quantified with a Qubit instrument in 1 μl eluate.

To 200 ng cDNA in 20 μl nuclease-free water, 30 μl nuclease-free water, 7 μl Ultra II End-prep reaction buffer, and 3 μl Ultra II End-prep enzyme mix (E7546L, NEB) were added. The reaction was incubated at 20°C for 20 min, then 65°C for 10 min, with a thermal cycler. Purification was performed with AMPure XP beads. The end-repaired dsDNA was eluted with 23 μl nuclease-free water, then quantified with a Qubit instrument. To 20 μl end-repaired DNA with a concentration of 200 fmol, 25 μl Blunt/TA Ligation Master Mix (M0367L, NEB), 2.5 μl nuclease-free water, and 2.5 μl Adapter Mix (SQK-LSK109, Oxford Nanopore, UK) were added and gently mixed by flicking of the tube. The reaction was incubated for 10 minutes at room temperature with rotation. The reaction was then purified with AMPure XP beads, and Short Strand Buffer (SQK-LSK109, Oxford Nanopore, UK) was used to wash the beads instead of 75% ethanol, to avoid inactivating motor proteins bound to the sequencing adaptors. The library on the beads was eluted with 22 μl Elution Buffer (SQK-LSK109, Oxford Nanopore, UK) and quantified with a Qubit instrument. Subsequently, 37.5 μl Sequencing Buffer, 25.5 μl Loading Beads, and 12 μl DNA library were mixed and loaded into the Oxford Nanopore sequencer.

## Nanopore data analysis

The electrical signals stored in fast5 format were used for base-calling with the GPU version of guppy (V4.4.4) with the parameters of --device "cuda: all:100%" -min_qscore 7 and other default parameters. The fastq files generated by base-calling were combined to trim the adaptors with Porechop (V0.2.4), and reads with length less than 100 nt were filtered with NanoFilt (V2.7.1). Trimfq from the seqtk package (V1.3) was used to trim 15 nt off both the 5′ and 3′ ends; this step was critical for removing the regions of primers, because we used SARS-CoV-2 genome-specific primers (GSPs) for the first strand cDNA synthesis. Clean reads were then mapped to the reference genome of SARS-CoV-2 (MN908947.3) with minimap2 (V2.17) with the parameters of "-a -x map-ont" and other default parameters. The output in Sequence Alignment/Map format was transferred to Binary Alignment/Map format with samtools (V1.11), then indexed with samtools. The indexed bam files were processed with pysamstats (V1.1.2) to count the base composition at each position. A homemade script was used to distinguish the mutations and call the regions with deletion from results of pysamstats (V1.1.2).

## Meta-transcriptome sequencing

An MGIEasy RNA Library preparation Kit v2 (1000005953, MGI) was used to construct the sequencing library. DNase I (M0303S, NEB) was used to digest the residual DNA in the RNA sample. To 10 μl of reaction, 4 μl fragmentation buffer was added, and the reaction was incubated at 87°C for 6 min, then immediately placed on ice for 2 min to fragment the RNA. Subsequently, 5 μl reverse transcription (RT) Buffer and 1 μl RT Enzyme Mix were added, and the reaction was incubated at 25°C for 10 min, 42°C for 30 min, and 70°C for 15 min in a thermal cycler to obtain the first-strand cDNA. Then 26 μl Second Strand Buffer and 4 μl Second Strand Enzyme Mix were added, and the reactions were incubated at 16°C for 60 min. Purification was performed with DNA Clean Beads. Subsequently, 2.9 μl end-repair and A-tailing (ERAT) Enzyme Mix and 7.1 μl ERAT Buffer were added to the purified dsDNA and incubated at 37°C for 30 min and 65°C for 15 min to prepare the ends for adaptor ligation. Then 1 μl sequencing adaptor, 9 μl tris-ethylene diamine tetraacetic acid buffer, 23.4 μl ligation buffer, and 1.6 μl DNA ligase were added, and the reaction was incubated at 23°C for 30 min. Bead purification and 18-cycle PCR amplification were performed. Finally, purified dsDNA was sequenced by the BGI-Shenzhen genome sequencing team with a single end 50 nt strategy generating 20 M reads for each sample, on average.

## Amplicon sequencing

The first-strand cDNA synthesis followed the same steps as those for building the nanopore library. Multi-panel

amplification of the SARS-CoV-2 genome was separately performed with primer pool T1 (106 amplicons) and T2 (108 amplicons) from a commercial kit (A186XV6, iGeneTech). The PCR products of primer pool T1 and T2 were pooled and purified with AMPure XP beads. The primer used for the first round of PCR consisted of sequences that were the reverse complement of the reference SARS-CoV-2 genome and a partial sequence of Illumina's sequencing adaptor. The second round PCR reaction was performed with IGT-I5 and IGT-I7 primers, which bound the 3′ ends of the first round PCR products, to index the previously purified dsDNA and enable the formation of full-length sequencing adaptors for the Illumina platform. The final PCR products were purified with AMPure XP beads and then sequenced on the Illumina platform with a paired-end 150 nt strategy generating 2 G reads for each sample, on average.

### Amplicon data analysis

Removal of low-quality bases and residual sequencing adapters from the raw reads was performed with trimmomatic (V0.39) in paired-end mode with the parameter of LEADING:3 TRAILING:3 MINLEN:50. Clean reads were then aligned to the reference SARS-CoV-2 genome (MN908947.3) with bwa mem with default parameters. Cut_Multi_Primer.py from SARS-CoV-2_Multi-PCR_v1.0 tools (https://github.com/MGI-tech-bioinformatics/SARS-CoV-2_Multi-PCR_v1.0) was used to modify the sequencing quality value of the bases belonging to the multiple amplification primers to 0. Trimmomatic was used again to remove bases with low quality, i.e., those belonging to the multiplex primers. The trimmed reads were re-aligned to the reference SARS-CoV-2 genome (MN908947.3) with bwa mem. The files storing the mutation information were created with freebayes (V0.9.21) with the parameters of --ploidy 1 --min-base-quality 20 --min-mapping-quality 60 --min-coverage 30. The generated vcf files were annotated with snpEff (V4.3) with a home-made annotation file. Only nucleotide positions with greater than 100× sequencing depth were kept for calling variants.

## RESULTS

### A strategy for rapid typing and sequencing of SARS-CoV-2

We developed a sequencing strategy capable of handling clinical samples with a wide range of Ct values. This strategy allowed us to obtain the complete viral genome sequence in only 4 hours. After obtaining clinical RNA sample, we used multiple sequencing methods tailored to the specific range of Ct values present in the samples.

After patient admission, the type of virus strains can be quickly determined to avoid cross-infection by different strains during hospitalization. First, we used amplification refractory mutation system (ARMS) PCR to specifically identify and distinguish wild-type strain and variants with a Ct value of 35 or less (Fig 1D). Primers used for ARMS PCR were labeled with different fluorophores enabling simultaneous detection of the HV69-70del, K417N, K417T, L452R, E484Q, E484K, N501Y, and A570D mutations. Through analysis of the combination of detected mutations, the sequences were classified into distinct clades, such as Alpha (B 1.1.7), Delta (B 1.617.2), Beta (B 1.351), Gamma (P.1), and Omicron (B 1.1.529), as shown in S1 Table. Our primary focus was on the identification of the L452R, N501Y, and K417N mutations, because of their prevalence in the most common SARS-CoV-2 variant strains, Alpha, Delta, and Omicron. For example, identification of the L452R mutation indicated that the virus belongs to the Delta clade, because this mutation is unique to this specific clade (notably, as of June 2021, this mutation was found exclusively in the Delta variant) (S1 Fig). Consequently, the possibility of the virus belonging to the Omicron clade could be confidently excluded. Detection of the N501Y mutation suggested that the virus might belong to either the Alpha or Omicron clade. Finally, if both the N501Y and K417N mutations were identified concurrently, the virus was considered highly likely to be associated with the Omicron clade. The ARMS PCR-based method was able to quickly and approximately identify viral types within a timeframe of 3 hours, thus providing timely assistance in clinical management.

To obtain the complete genomic sequence of SARS-CoV-2, we propose three potential approaches considering the Ct ranges and the urgency of obtaining the results. When the Ct value of the sample is equal to or below 20, an abundance of SARS-CoV-2 transcripts is present and can be directly sequenced with an Oxford Nanopore sequencer. To facilitate this approach, we designed a set of inverse complementary primers to the SARS-CoV-2 genome, thus enabling the enrichment and generation of viral cDNA through reverse transcription (Fig 1A). After completion of the second-strand synthesis end repair, the adaptors attached to the nanopore motor protein are ligated to the prepared double-stranded DNA, thereby finalizing the construction of the sequencing library. Subsequently, the library is sequenced with a Nanopore sequencer; 30–60 minutes is required for electronic signal data collection, and an additional 30 minutes is required for base calling and bioinformatic analysis. The complete genome sequence of SARS-CoV-2 can be obtained in this manner. In contrast to the ARTIC protocol, our method achieves acquisition of the entire viral genomic sequence within 4 hours, because PCR amplification is not required.

When the Ct value is equal to or below 22, the meta-transcriptome sequencing method can be used to acquire the genomic sequence of SARS-CoV-2. In this approach, the isolated RNA serves as the template for reverse transcription using random hexamers (Fig 1B). Subsequently, second-strand synthesis and end repair are conducted, and followed by ligation of universal adaptors

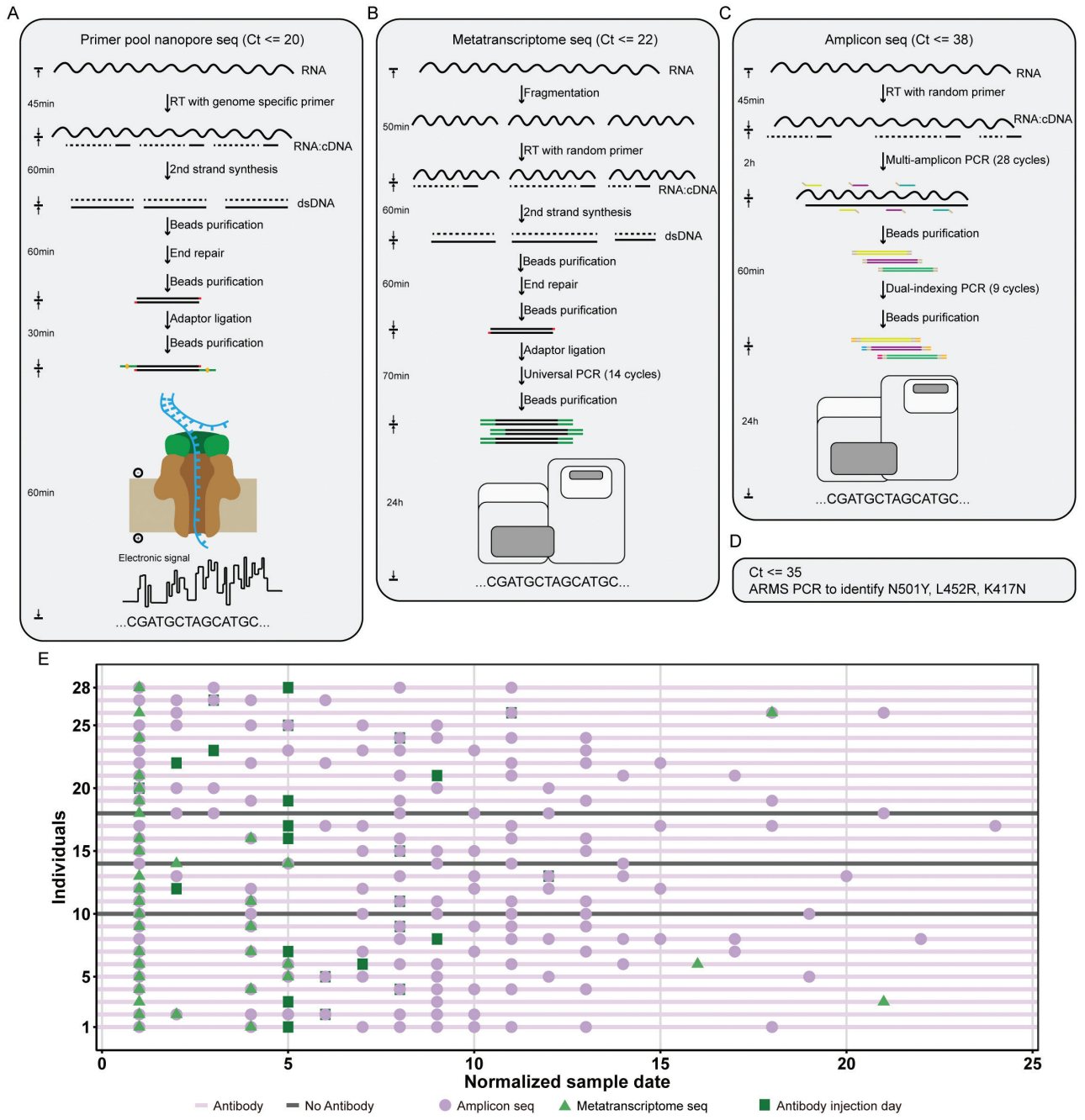


**FIGURE 1** | Strategies for rapid whole genome sequencing of SARS-CoV-2. (A) Primer pool enrichment pipeline for rapid nanopore sequencing. (B) Workflow for meta-transcriptome sequencing. (C) Workflow for amplicon sequencing. (D) Rapid identification of key mutations defining the SARS-CoV-2 clade by ARMS PCR when Ct≤35. (E) Schematic representation of SARS-CoV-2 infected patient samples. Dots and triangles indicate sample collection time points, with different sequencing methods applied to process the samples. Of the 28 patients, six had received vaccination. Antibody treatment refers to administration of a combination of BRII-196 and BRII-198 neutralizing antibodies targeting non-competing epitopes on the RBD of spike, belonging to class 1 and class 3 mAbs, respectively.

to the dsDNA ends, which overhang because of A-tailing. Subsequently, PCR with 14 cycles is performed to amplify the library, which is then subjected to deep sequencing. The entire process, encompassing library construction (4 hours) and sequencing (approximately 24 hours), requires approximately 28 hours to complete.

When the Ct value surpasses 22, obtaining the complete genomic sequence through meta-transcriptome sequencing or primer-enriched nanopore sequencing is challenging. Simply increasing the sequencing depth does not effectively enhance the efficiency of capturing the SARS-CoV-2 genome [10]. Therefore, PCR amplification of the SARS-CoV-2 genome can be performed with customized multiplex primers that target distinct genomic regions. A second dual-indexing PCR is then conducted to incorporate the sequencing index and

finalize the construction of the sequencing library. The complete genome sequence is then obtained by assembling the amplified fragments, as depicted in Fig 1C. The time required to obtain the complete genomic sequence of SARS-CoV-2 is similar for both amplicon sequencing and metagenomic sequencing methods, requiring approximately 30 hours.

## Primer pool enriched nanopore sequencing to capture SARS-CoV-2's whole genome

For nanopore sequencing, we used GSPs targeting the SARS-CoV-2 genome for cDNA synthesis. The resulting library was then subjected to nanopore sequencing. To assess the effects of primer enrichment, we conducted cDNA synthesis with random hexamers at an equivalent concentration to that of the GSPs. The resulting library was then subjected to both nanopore sequencing and deep sequencing. The data obtained from these three methods were subsequently gathered and individually aligned to the reference SARS-CoV-2 genome. The use of GSPs for reverse transcription resulted in a successful mapping ratio of 40.6% of the reads to the reference genome. In contrast, when random primers were used, the mapping ratio was markedly lower, at 3.4% (Fig 2A). We also calculated the ratio of the bases aligned to the reference genome to the total number of bases, and found a similar mapping ratio. These findings highlighted the 10-fold enrichment efficiency achieved through the utilization of SARS-CoV-2 GSPs for reverse transcription. Additionally, when samples were reverse transcribed with random hexamers and sequenced on the MGI 2000 platform, we observed a lower ratio of reads successfully mapped to the SARS-CoV-2 reference genome (1.2%) than observed with the use of GSPs (3.4%, Fig 2A). This difference in mapping ratios might potentially be attributable to the amplification bias introduced by the meta-transcriptome sequencing pipeline, which uses a universal PCR amplified library.

We then randomly sampled an equal number of nanopore sequencing reads with GSPs or random hexamers, and aligned them to the reference SARS-CoV-2 genome. The samples subjected to reverse transcription with GSPs exhibited clearly greater genome coverage (Fig 2B and 2D) and significantly higher sequencing depth (Fig 2C). In contrast, the samples subjected to reverse transcription with random hexamers displayed longer read lengths (Fig 2E).

Our method is applicable to obtaining the complete genome sequence of SARS-CoV-2 not only from cultured virus but also from clinical specimens with low Ct values. Furthermore, the use of nanopore technology enables the generation of long reads, thereby enabling effective investigation of the sub-genomes of SARS-CoV-2 under various experimental conditions. In comparison to direct RNA sequencing, our method offers the potential to acquire more sequencing reads, through use of cDNA as the input, thus offering greater stability than RNA.

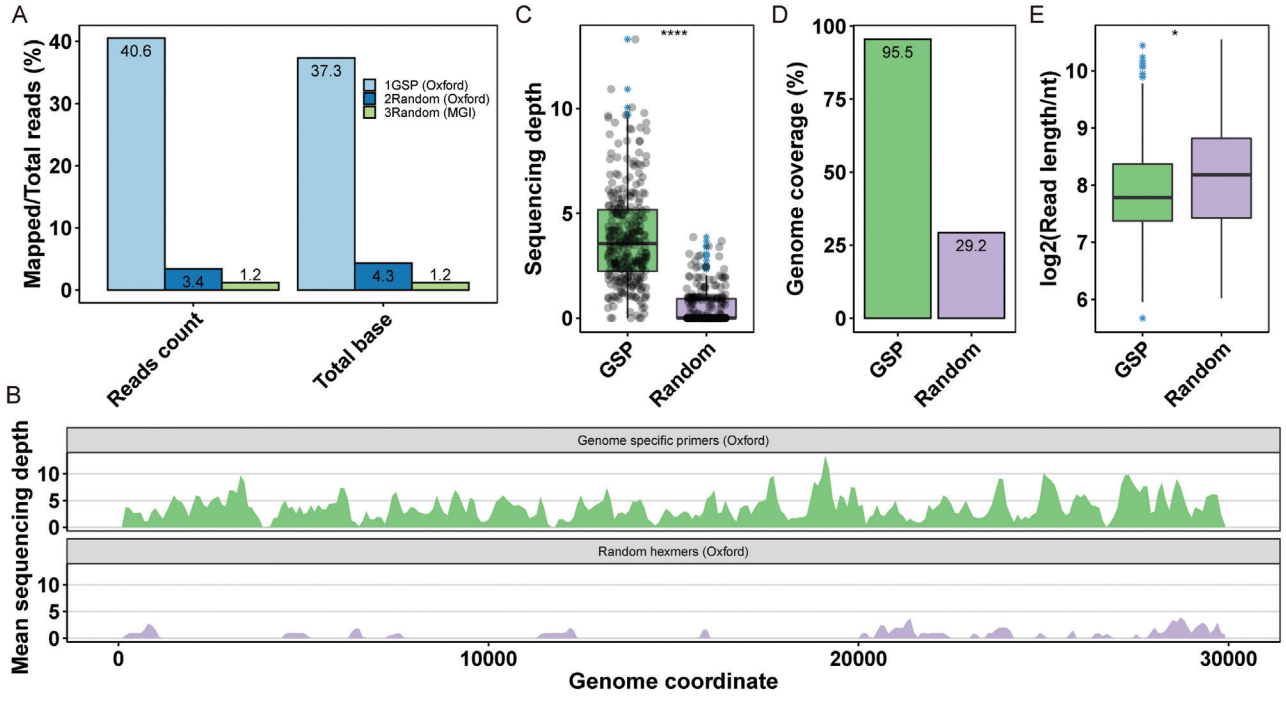


**FIGURE 2** | Characteristics of the primer pool enriched rapid nanopore sequencing workflow. (A) cDNA was synthesized with GSP and sequenced with the Nanopore platform. cDNA was synthesized with random primers and sequenced with the Nanopore and MGI 2000 platforms. (B) The same number of reads, from which the cDNA was synthesized by GSP and random primers and sequenced with the Nanopore platform, were sampled and aligned to the reference SARS-CoV-2 genome. (C) Sequencing depth, (D) genome coverage, and (E) read length for the two methods in (B).

## Meta-transcriptome sequencing

Because deep sequencing is the gold standard for studying intra-host evolution, we used meta-transcriptome sequencing to capture the complete genomic sequence of SARS-CoV-2 in selected samples. Our analysis revealed a linear decrease in the proportion of reads aligned to the reference SARS-CoV-2 genome as the Ct value of the sample increased (Fig 3A and 3G). In addition, the Ct values of the viral culture samples were lower than those of the clinical samples, in agreement with the higher viral load in the viral culture (Fig 3A and 3H). Moreover, the proportion of reads aligning to the SARS-CoV-2 reference varied from 8/10,000 to nearly 0 (Fig 3G). Specifically, for a sample with a Ct value of approximately 11, the proportion of reads belonging to SARS-CoV-2 ranged from 2.5/10,000 to 7.5/10,000 (Fig 3G).

Our findings indicated that achieving nearly 100% genome coverage was feasible when the number of aligned bases was ten times the length of the SARS-CoV-2 genome, which is approximately 300 k (0.3 M) (Fig 3B). Thus, to achieve a genomic coverage of almost 100% for a sample with a Ct value of 11, sequencing of approximately 600 M (0.3 M/0.0005) bases from the sequencer is necessary, on the basis of the ratio of 5/10,000 (Fig 3G) for the 0.3 M total SARS-CoV-2 reads. In contrast, for a sample with a Ct value of 20, the amount of data required from the sequencer would be 6000 M (0.3 M/0.00005) and would be cost prohibitive. The average sequencing depth of SARS-CoV-2 increased linearly as the proportion of SARS-CoV-2 reads increased (Fig 3C). We also found a negative linear correlation between the average sequencing depth of SARS-CoV-2 and the sample Ct values (Fig 3D). A positive relationship was observed between sequencing

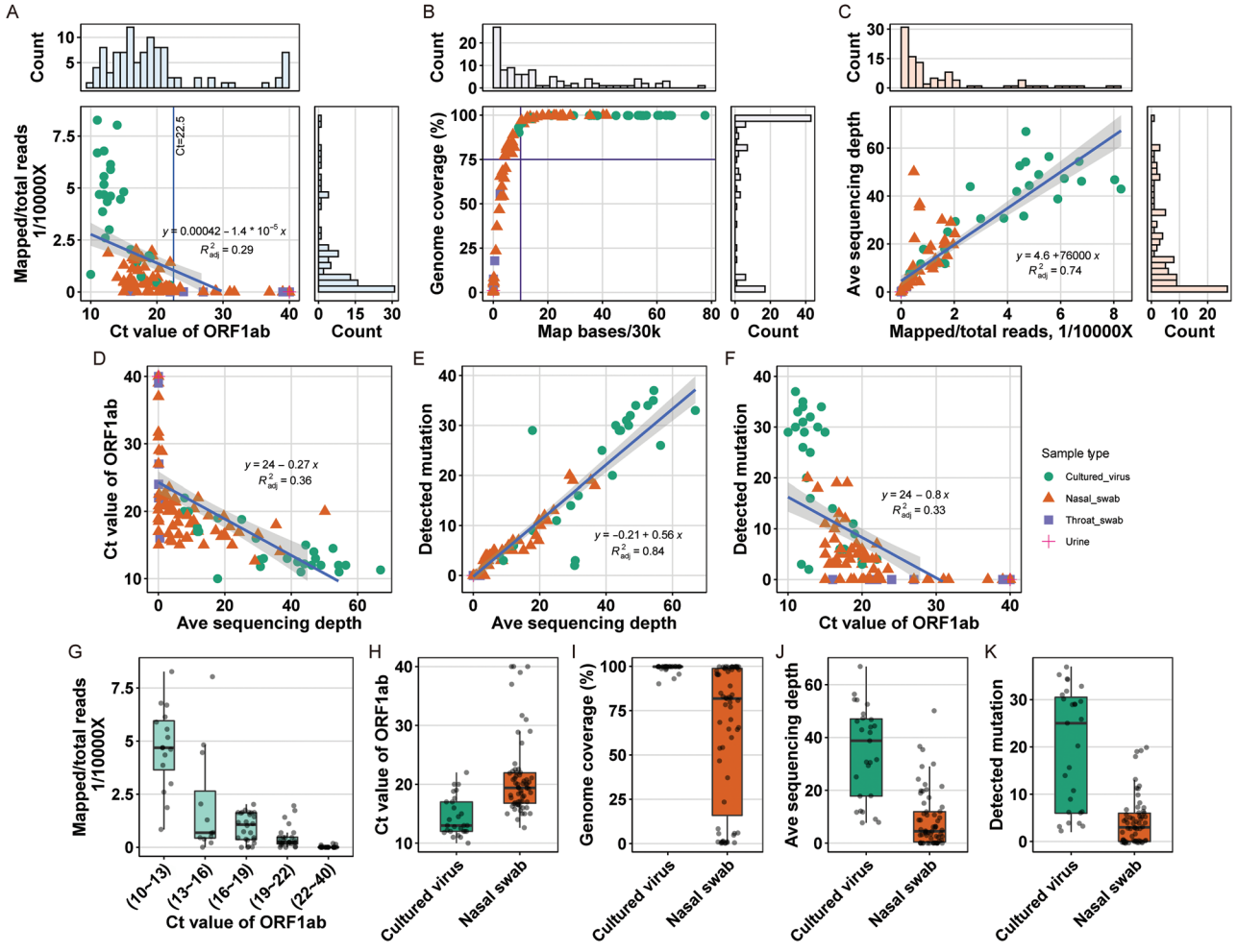


**FIGURE 3** | Technical characteristics of meta-transcriptome sequencing. (A) Correlation between the Ct value and the proportion of reads belonging to SARS-CoV-2. (B) Correlation between the number of bases that aligned to the SARS-CoV-2 genome and the genome coverage. (C) Correlation between the number of reads that aligned to the SARS-CoV-2 genome and the average sequencing depth. (D) Correlation between the average sequencing depth of SARS-CoV-2 and the Ct value of the sample. (E) Correlation between the average sequencing depth of SARS-CoV-2 and the number of detected mutations. (F) Correlation between the Ct value of the sample and the number of detected mutations. (G) The proportion of reads belonging to SARS-CoV-2 for different ranges of Ct values. Ct values (H), genome coverage (I), average sequencing depth (J), and number of detected mutations (K) of SARS-CoV-2 for the supernatants of SARS-CoV-2 infected cell cultures and nasal swabs.

depth and the number of mutations detected (Fig 3E). These findings suggested that low sequencing depth in meta-transcriptome sequencing may lead to underrepresentation of rare mutations. Furthermore, we observed a correlation between the number of identified mutations and the Ct values of the samples (Fig 3F), thereby suggesting that the abundance of SARS-CoV-2 reads influences the number of mutations detected. Specifically, samples with lower Ct values have a higher proportion of SARS-CoV-2 reads and result in a greater likelihood of detection of viral variants.

## Amplicon sequencing

For amplicon sequencing, hundreds of primer pairs were used for viral genome amplification. However, importantly, the amplified segments contained regions occupied by the primers, which needed to be removed to avoid missing important information generated from these primer regions. The primers used for amplification encompassed 20% of the SARS-CoV-2 genome, whereas the remaining 80% of the genome was amplified via DNA polymerase. We used amplicon sequencing to analyze numerous clinical samples, because of their high Ct values (>22), given that both the primer-enriched nanopore sequencing and meta-transcriptome sequencing were insufficient in generating complete genomic sequences of SARS-CoV-2.

Achieving nearly 100% coverage of viral genomes was more feasible when the Ct values were below 30 rather than above 30 (Fig 4A and 4D). Additionally, the proportion of bases remaining after primer region trimming was higher in samples with Ct values below 30 but was lower (less than 20%) in samples with Ct values above 30 (Fig 4A and 4E). This finding might be attributable to the low viral load in samples with a Ct value greater than 30, thus making amplification of the SARS-CoV-2 genome with PCR more challenging.

Complete genomic coverage was not achieved for some number of samples with the remaining proportion of bases below 20% (Fig 4B). To reach complete genome coverage with amplicon sequencing, we determined that 30 million reads would be required from the sequencer, corresponding to $30\times1000\times2^{10}$ mapped bases (Fig 4B). Additionally, the number of detected mutations did not increase beyond an average sequencing depth of $2^{12.5}$ (Fig 4C), thus suggesting that the range of bases to be collected from the sequencer could range from $2^{7.5}$ to $2^{12.5}$. Samples with lower Ct values had fewer mutations and less variability than samples with higher Ct values (Fig 4C and 4F). We also compared the number of mutations detected by amplicon sequencing and meta-transcriptome sequencing, and found that amplicon sequencing had a greater ability to detect mutations (Fig 4G).

## Quasi-species evolution of SARS-CoV-2 during treatment

To track the evolution of mutations in the same patient during hospitalization, we continuously sampled and sequenced the composition of SARS-CoV-2 with various methods (Fig 1E). Throughout the 20-day period, we observed a change in composition of C13019T, a synonymous mutation (L122L) in the nsp9 gene. Initially, 13019C was present in almost 100% of the viral population but gradually decreased to approximately 25% (Fig 5A). Furthermore, we observed a similar pattern with the D614G mutation in the spike gene. The 614D variant was initially dominant, then transitioned to the 614G variant, and 614D reached a composition of approximately 60% during the treatment process (Fig 5B). When mutations within a single patient were summarized, we observed that certain mutations remained relatively stable in composition. However, most mutations continued to decrease during treatment, thus indicating diverse responses to treatment (Fig 5C). To evaluate the rates at which the quasi-species composition declined, we calculated the slope of the quasi-species composition according to the duration of treatment. Our analysis revealed that 79.5% of the mutation composition continued to decline, whereas 20.5% increased (Fig 5D). Of these mutations, T181I of nsp6, L112L of nsp9, G142D and D614G of Spike, T61fs of ORF7a, and certain mutations in the 3′ UTR showed a significant decrease in their composition. These findings suggested that these mutations may have notable effects on the viral life cycle, because they were quickly eliminated during the treatment. Apart from the well-studied D614G mutation [34], we did not examine other mutation sites' potential effects on the replication and survival of SARS-CoV-2. Among the Spike mutations, an increased frequency of the A222V mutation was observed. Moreover, we analyzed all patients infected with SARS-CoV-2 and found that the quasi-species ratios of most mutations decreased (Fig 5E). Notably, Spike mutations, such as T19R, G142D, T478K, L452R, and D614G, exhibited a declining trend. Mutations in other genes, including I292T in nsp4, T181I in nsp6, and L112L in nsp9, and G662S in nsp12, also decreased. The results in Fig 5E were consistent with those in Fig 5D.

## Effects of vaccination and neutralizing antibody therapy on the evolution of SARS-CoV-2

By binding the viral Spike protein, neutralizing antibodies block the binding of SARS-CoV-2 to the ACE2 receptor on the cell surface, and serve as an important treatment for SARS-CoV-2 infections. Because of their direct interaction with the Spike protein, antibodies play critical roles in shaping the evolution of the virus. In this study, because a subset of patients received treatment with BRII-196 and BRII-198 antibodies, we were able to examine the effects of antibody therapy on the in vivo evolution of SARS-CoV-2. By comparing the mutations present in samples before and after antibody treatment from the same patients, we inferred potential mutation sites that might indicate interactions with the administered antibodies (Fig 6A). Notably, the K417N mutation was observed in the treated samples from two patients but

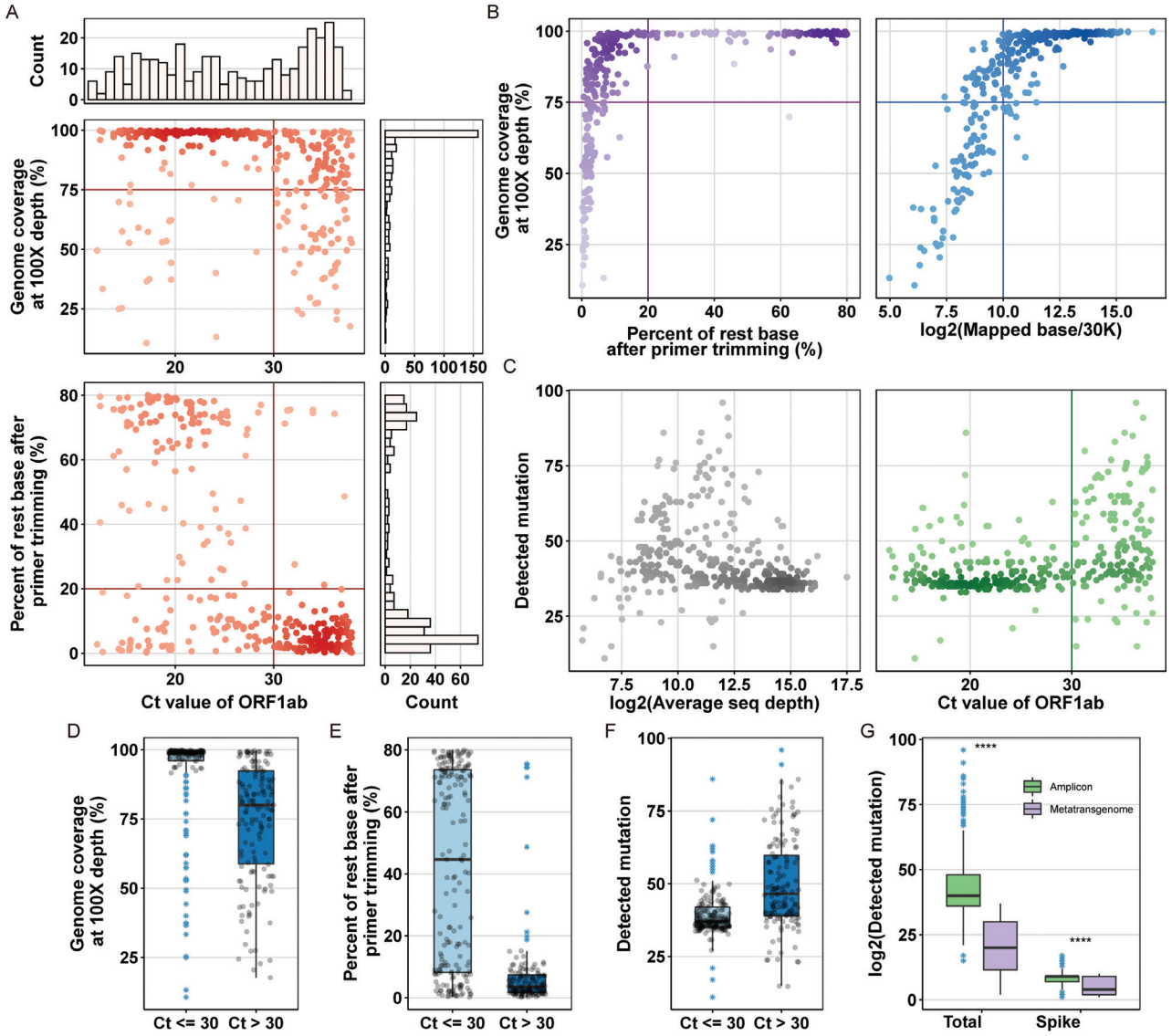


**FIGURE 4** | Technical characteristics of amplicon sequencing. (A) Correlation between the Ct values of the samples and the genome coverage of SARS-CoV-2, with a sequencing depth of at least 100, and between the Ct values of the samples with the percentage of remaining bases after primer trimming. (B) Correlations between SARS-CoV-2 genome coverage and sequencing depth of at least 100 and the percentage of remaining bases after primer trimming, and the number of aligned bases. (C) Associations between average sequencing depth and the number of detected mutations; associations between the Ct values of the samples and average sequencing depth. Genome coverage (D), percentage of remaining bases after primer trimming (E), and number of detected mutations (F) in samples with different ranges of Ct values. (G) Comparison of the number of mutations detected with amplicon sequencing and meta-transcriptome sequencing.

was detected in only one patient from the untreated group (Fig 6B). We also observed two de novo mutants: F456L and F456S (Fig 6B). We defined samples from individuals before antibody therapy as the untreated group and samples after therapy as the treated group, and identified distinct mutations in both groups (Fig 6D). F456, regardless of whether it was mutated to L or S, was effectively neutralized (Fig 6E), in agreement with the findings in Fig 6B. Because of poor statistical power, no significance was observed; however, the general trend indicated the presence of the K417R mutation, which affected antibody neutralization, in three individuals (Fig 6F), thereby suggesting a direct link between Spike K417 and the antibody treatment. Moreover, the observation of V539V and V539C mutations in untreated samples (Fig 6F) suggested a possible interaction between Spike V539 and the antibody. Therefore, through continuous monitoring of in vivo mutations, we experimentally revealed the involvement of the Spike K417 site, which caused the variant to escape neutralization by most class 1 neutralizing antibodies [35]. Additionally, we discovered that specific mutations at Spike sites might potentially affect the efficacy of antibody treatment in individual patients, although further confirmation is required in future studies.

We collected samples from both vaccinated and unvaccinated patients, which enabled us to investigate

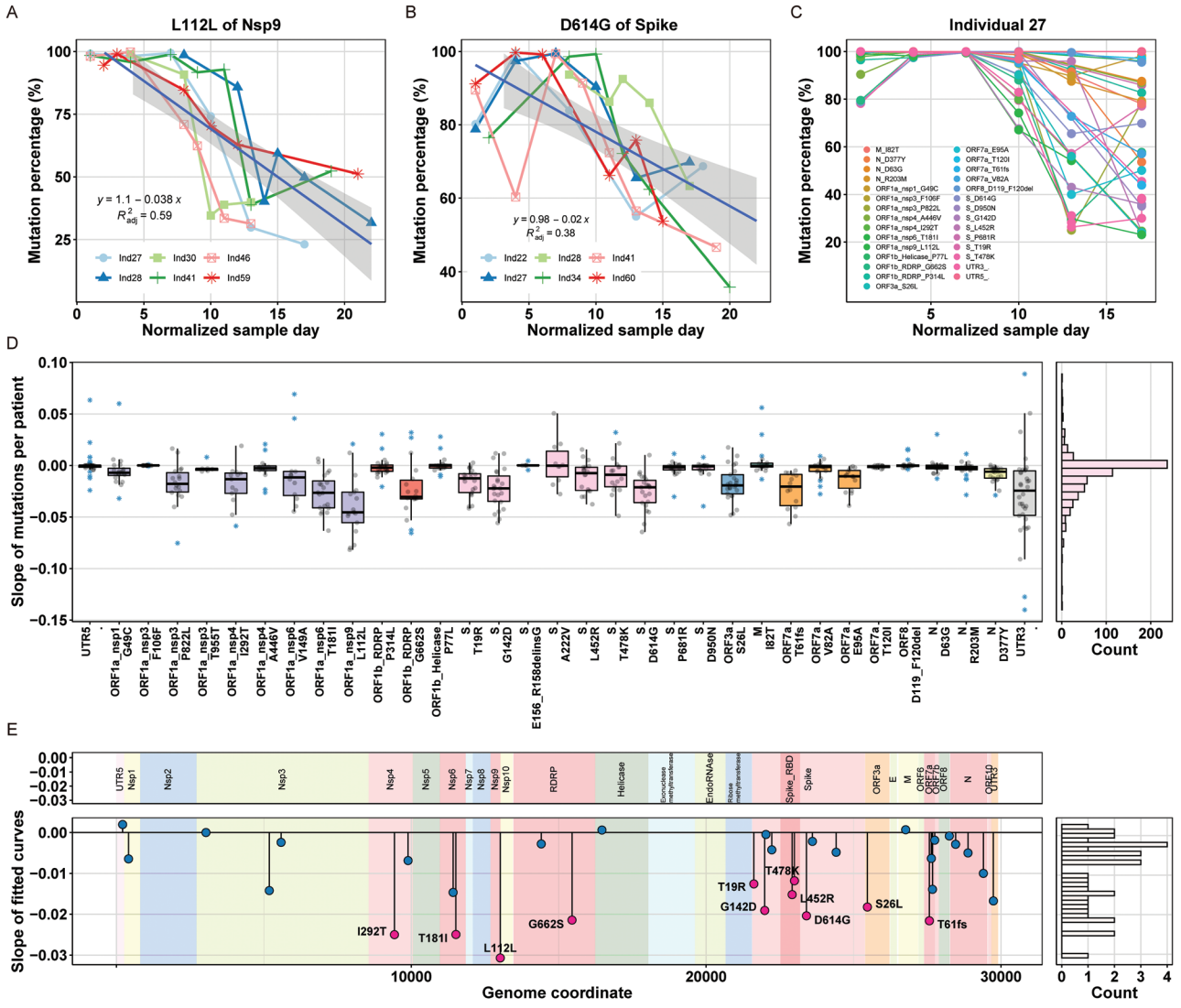


**FIGURE 5** | Changes in the quasi-species composition of SARS-CoV-2 during treatment. The composition of (A) L112/C13019 (Nsp9) and (B) D614 (Spike) decreased over the course of treatment. (C) Fluctuations in all detected mutations in individual 27. (D) Slopes of the fluctuation trend in composition for each mutation across all patients, plotted against sampling time. Only mutations with at least 5 days of sample collection were included. (E) Slopes of the change trends in mutation at a specific site, in the data from multiple individuals. Only mutations with at least 5 days of sample collection were retained. Note: The trend in mutation site changes refers to the transition from the amino acid before the numerical designation of the site to the amino acid after the numerical designation. For example, D614G indicates a change from D614 to G614.

the influence of vaccination on the evolution of SARS-CoV-2. Comparison of the mutations identified in these two groups of samples indicated no statistically significant difference in the overall number of mutations observed in the Spike gene and the receptor-binding domain (RBD) region, when the entire genome was considered (Fig 6H). However, after categorizing the Ct values of consecutively sampled patients into 3-day intervals, we observed that during the second and third 3-day intervals, the Ct values were significantly higher in the vaccinated group than the unvaccinated group (Fig 6I). These findings suggested that vaccinated patients experienced faster clearance of SARS-CoV-2. To compare the evolving quasi-species, we categorized the changing slopes into three groups: fast drop, slow drop, and unchanged slope. The vaccinated group showed significantly higher proportions of both fast decrease and slow decrease than the unvaccinated group (Fig 6J). This finding suggested that vaccination accelerates the rate of quasi-species decline.

## DISCUSSION

Viral genome analysis is a crucial approach for studying infectious disease outbreaks. However, current methods typically focus on assembling a single genomic sequence for each infected individual. Recent studies have recognized the importance of nucleotide polymorphisms in viral sequences within hosts. Through analysis of these polymorphisms, the distribution characteristics of viral

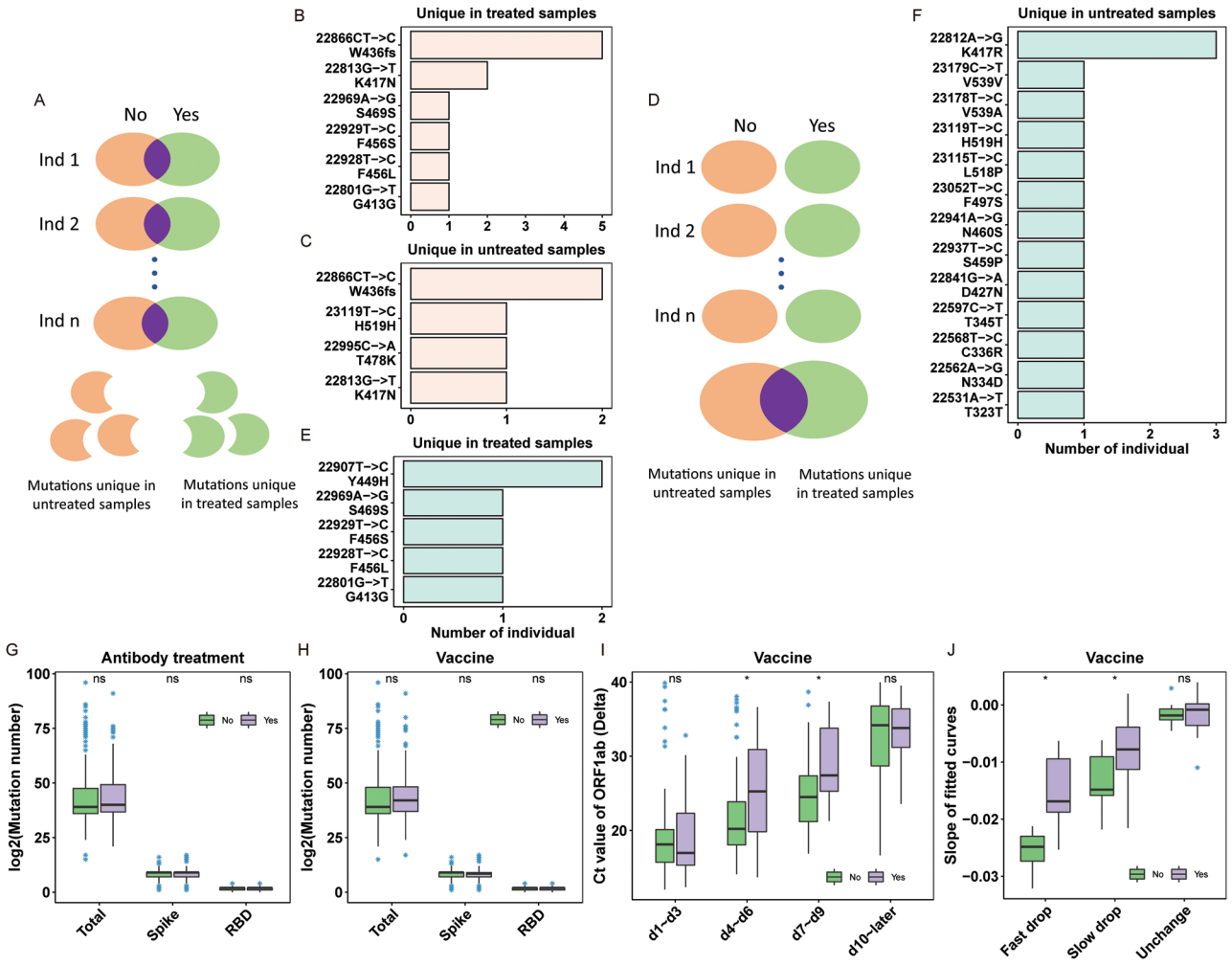


**FIGURE 6** | Effects of vaccination and neutralizing antibody therapy on the evolution of SARS-CoV-2. (A) Strategy to analyze mutations associated with antibody interaction by merging unique mutations in antibody treated and untreated samples for each patient, and identifying mutations occurring exclusively in antibody-treated (B) or untreated (C) samples. (D) Strategy to analyze mutations associated with antibody interaction by combining mutations detected in antibody treated and untreated samples for each patient, then identifying unique mutations occurring exclusively in pretreated (F) or antibody-treated (E) samples. Effects of (G) antibody and (H) vaccine administration on the number of detectable mutations. Effects of vaccine administration on (I) the rate of SARS-CoV-2 clearance in the body and on (J) the rate of change in quasi-species composition.

quasi-species within the host can be understood, and viral transmission pathways can potentially be traced [36].

Understanding of the evolution of quasi-species composition during the transition from early latent infection to the subsequent clearance of SARS-CoV-2 is limited. Many published studies have lacked continuous sampling from a single patient or have not captured the earliest stages of viral infection. To address this knowledge gap, we performed comprehensive sampling and sequencing starting from the onset of SARS-CoV-2 infection and continuing throughout the entire course of the infection until hospital discharge. Our approach offers a substantial advantage over other methods, because it effectively identifies viral mutations. By using diverse sequencing methods, we successfully obtained comprehensive genome sequences of SARS-CoV-2, thereby facilitating close monitoring of the evolutionary patterns of quasi-species in different patients. Additionally, our investigation assessed the effects of antibodies and vaccines on the evolution of SARS-CoV-2 quasi-species.

We established a comprehensive sequencing strategy, including amplicon sequencing, meta-transcriptome sequencing, and primer pool-based nanopore sequencing. These techniques enabled us to obtain the complete genome sequence of SARS-CoV-2 from clinical samples with Ct values ranging from 10 to 38. We used multiple sequencing methods, then cross-validated and compared the results obtained from each approach. The amplicon sequencing method uses PCR amplification to obtain the whole-genome sequence of SARS-CoV-2, even from samples with high Ct values. We successfully obtained the entire genome sequence from samples with Ct values as high as 38. Our findings indicated the feasibility of obtaining the complete genome from any SARS-CoV-2 positive sample. Moreover, amplicon sequencing provides a cost-effective alternative to meta-transcriptome

sequencing, because it requires fewer sequencing resources and offers advantages in library construction cost. However, the sensitivity of PCR amplification notably increases the risk of false-positive results caused by minor aerosol contamination. To address this concern, precautions, such as the use of pipette tips with filters and minimizing the number of PCR cycles, are necessary to mitigate the introduction of aerosol contamination.

Studies have been conducted to obtain the intra-host variant of SARS-CoV-2 through amplicon sequencing [30,37]. Amplicon sequencing involves amplifying short DNA segments through PCR, which are subsequently subjected to deep sequencing. Other sequencing studies have used Oxford Nanopore sequencing, such as the ARTIC protocol. In contrast to the standard ARTIC pipeline, which uses 400 bp amplicons to obtain the entire genome through Oxford Nanopore sequencing, our approach uses shorter amplicons ranging from 200 bp to 250 bp. Consequently, both deep sequencing and nanopore sequencing techniques can be used to sequence the same subset of samples. In practical operation, after amplifying the library into short fragments, we first perform sequencing with the Oxford Nanopore Technologies platform to quickly obtain mutation results of the viral strain and report them to the hospital in the shortest possible time. The same library is reported to the disease control center and released to the public only after validation through second-generation sequencing. However, importantly, nanopore sequencing has lower base-calling accuracy than other sequencing technologies. This limitation makes accurately distinguishing between sequencing errors and original mutations challenging. Because the use of nanopore sequencing for analyzing quasi-species is constrained by the inherent inaccuracy associated with single-base errors and indels, nanopore sequencing is used primarily for viral typing, genome assembly, and sub-genome analysis. To enhance the reliability of single-base sequencing, a double index located at the two ends could be used [38]. Herein, we developed a sequencing method that enriches SARS-CoV-2 transcripts with a primer pool and subsequent sequencing on an Oxford Nanopore sequencer. By eliminating the need for PCR amplification, we decreased the time required by at least 4 hours and were able to obtain the complete genome sequence of SARS-CoV-2 in less than 4 hours for samples with Ct values below 20. This rapid sequencing method not only allows for the identification of strain types and tracing of viral infection sources, but also outperforms the standard ARTIC pipeline in terms of efficiency. Furthermore, our method, because of its ability to generate long reads, enables the analysis of sub-genomic transcripts of SARS-CoV-2. In summary, the primer pool enriched nanopore sequencing approach allows for acquisition of the complete genome within a short 4-hour timeframe for samples with Ct values below 20. For samples with Ct values below 22, an alternative option is meta-transcriptome sequencing, which requires approximately 28 hours. However, amplicon sequencing is recommended for analysis of substantial sample sizes or samples with Ct values exceeding 22.

In this study, we sought to observe the ongoing evolution of SARS-CoV-2 quasi-species by conducting continuous sampling on a substantial number of infected individuals. Previous studies have often sampled limited numbers of patients at various time points, thus potentially introducing bias [39,40]. By analyzing the changes in quasi-species composition throughout the treatment, we observed a significant shift in the proportion of variants carrying mutations in the spike gene, which is known to affect viral infectivity, such as the K417 site. We also identified variants with mutations in the Nsp4, Nsp6, Nsp9, RDRP, ORF3a, and ORF7a genes, which have not been extensively explored in previous studies. Nsp4 interacts with Nsp3 and induces the rearrangement of host-derived membranes, thereby facilitating the replication of SARS-CoV [41]. Nsp6 facilitates coronavirus infection by restricting autophagosome expansion [42]. Nsp9 is essential for the replication of SARS-CoV-2, because of its ability to bind viral genomic RNA [43,44]. RdRp is responsible for replicating and transcribing the SARS-CoV-2 genome [45]. ORF3a and ORF7a are accessory genes adapted by SARS-CoV-2 to modify the cellular environment to benefit viral survival. ORF3a functions as a viroporin that may promote virus release [46], whereas ORF7a functions as an immunomodulating factor that suppresses the immune response [47,48]. Further studies are needed to address the functions of these viral variants carrying mutations.

The development of primary neutralizing antibodies for the treatment of SARS-CoV-2 infection involved use of the RBD of the Spike protein as bait [49]. According to the "arms race" theory [50,51], these antibodies may exert selection pressure on SARS-CoV-2, thus leading to accelerated mutation and potential evasion [52,53]. In our previous study, we solved the crystal structures of BRII-196 and BRII-198 antibodies, and identified various interaction sites by directly examining these structures [54]. Additionally, through experimental investigation, we have successfully identified interactions at certain sites [55]. Here, we used the quasi-species analysis method to uncover potential virus-antibody interaction sites. Notably, these newly discovered sites overlapped with existing data, thereby indirectly validating the reliability of our methods. The examination of these interaction sites will be valuable in anticipating treatment-induced viral mutations in future endeavors.

**ACKNOWLEDGEMENTS**

This work was supported by grants from the National Natural Science Foundation of China (82151212 and 2021B1212030010), the emergency grants for prevention and control of SARS-CoV-2 of Guangdong province (2020B1111340036), the Guangzhou Lab Foundation (EKPG21-29), and Shenzhen Science and Technology Innovation Committee (JCYJ20220818103017036 and JCYJ20190809112205541). The funders had no role in study design, data collection, data analysis, data interpretation, or writing

of the report. We are very grateful to the clinicians and nurses in our hospital for their assistance with sample collection. This work was funded by Chinese Academy of Medical Sciences Clinical and Translational Medicine Research Project (grant number 2022-I2M-C&T-B-113); the Guangdong Science and Technology Plan Project, construction of high-level biosafety laboratories (2021B1212030010 to Z.Z.); the Shenzhen Science and Technology Program (ZDSYS20210623091810030 to Z.Z. JCYJ20220818103017036 to Z.Z., KQTD20200909113758004 to Z.Z.).

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest regarding the content of this article.

## REFERENCES

1. Bordería AV, Rozen-Gagnon K, Vignuzzi M. Fidelity variants and RNA quasispecies. In: *Quasispecies: From Theory to Experimental Systems*. Edited by Domingo E, Schuster P. Cham: Springer International Publishing; 2016:303-322.
2. Domingo E, Perales C. Viral quasispecies. PLoS Genet. 2019;15(10):e1008271.
3. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiol Mol Biol Rev. 2012;76(2):159.
4. Ni M, Chen C, Qian J, Xiao HX, Shi WF, Luo Y, et al. Intra-host dynamics of Ebola virus during 2014. Nat Microbiol. 2016;1(11):16151.
5. van Boheemen S, Tas A, Anvar SY, van Grootveld R, Albulescu IC, Bauer MP, et al. Quasispecies composition and evolution of a typical Zika virus clinical isolate from Suriname. Sci Rep. 2017;7(1):2368.
6. Lackenby A, Democratis J, Siqueira MM, Zambon MC. Rapid quantitation of neuraminidase inhibitor drug resistance in influenza virus quasispecies. Antivir Ther. 2008;13(6):809-820.
7. Ode H, Matsuda M, Matsuoka K, Hachiya A, Hattori J, Kito Y, et al. Quasispecies analyses of the HIV-1 near-full-length genome with illumina MiSeq. Front Microbiol. 2015;6:1258.
8. Goodenow M, Huet T, Saurin W, Kwok S, Sninsky J, Wain-Hobson S. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. J Acquir Immune Defic Syndr (1988). 1989;2(4):344-352.
9. Xu D, Zhang Z, Wang FS. SARS-associated coronavirus quasispecies in individual patients. N Engl J Med. 2004;350(13):1366-1367.
10. Gao R, Zu W, Liu Y, Li J, Li Z, Wen Y, et al. Quasispecies of SARS-CoV-2 revealed by single nucleotide polymorphisms (SNPs) analysis. Virulence. 2021;12(1):1209-1226.
11. Stapleford KA, Coffey LL, Lay S, Bordería AV, Duong V, Isakov O, et al. Emergence and transmission of arbovirus evolutionary intermediates with epidemic potential. Cell Host Microbe. 2014;15(6):706-716.
12. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. PLoS Pathog. 2012;8(3):e1002529.
13. Parameswaran P, Wang C, Trivedi SB, Eswarappa M, Montoya M, Balmaseda A, et al. Intrahost selection pressures drive rapid dengue virus microevolution in acute human infections. Cell Host Microbe. 2017;22(3):400-410.e5.
14. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature. 2006;439(7074):344-348.
15. Skums P, Zelikovsky A, Singh R, Gussler W, Dimitrova Z, Knyazev S, et al. QUENTIN: reconstruction of disease transmissions from viral quasispecies genomic data. Bioinformatics. 2018;34(1):163-170.
16. Sun F, Wang X, Tan S, Dan Y, Lu Y, Zhang J, et al. SARS-CoV-2 quasispecies provides an advantage mutation pool for the epidemic variants. Microbiol Spectr. 2021;9(1):e0026121.
17. Garvey MI, McMurray C, Casey AL, Ratcliffe L, Stockton J, Wilkinson MAC, et al. Observations of SARS-CoV-2 variant of concern B.1.1.7 at the UK's largest hospital trust. J Infect. 2021;83(4):e21-e23.
18. Moelling K. Within-host and between-host evolution in SARS-CoV-2—new variant's source. Viruses. 2021;13(5):751.
19. He X, Hong W, Pan X, Lu G, Wei X. SARS-CoV-2 omicron variant: characteristics and prevention. MedComm. 2021;2(4):838-845.
20. Benton DJ, Wrobel AG, Roustan C, Borg A, Xu P, Martin SR, et al. The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. Proc Natl Acad Sci USA. 2021;118(9):e2022586118.
21. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature. 2021;592(7852):116-121.
22. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. Nature. 2022;602(7896):294-299.
23. Motozono C, Toyoda M, Zahradnik J, Saito A, Nasser H, Tan TS, et al. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. Cell Host Microbe. 2021;29(7):1124-1136.e11.
24. Levine-Tiefenbrun M, Yelin I, Alapi H, Katz R, Herzel E, Kuint J, et al. Viral loads of Delta-variant SARS-CoV-2 breakthrough infections after vaccination and booster with BNT162b2. Nat Med. 2021;27(12):2108-2110.
25. Garcia-Beltran WF, Lam EC, St Denis K, Nitido AD, Garcia ZH, Hauser BM, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell. 2021;184(9):2372-2383.e9.
26. Zhou D, Dejnirattisai W, Supasa P, Liu C, Mentzer AJ, Ginn HM, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell. 2021;184(9):2348-2361.e6.
27. Hoffmann M, Arora P, Groß R, Seidel A, Hörnich BF, Hahn AS, et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. Cell. 2021;184(9):2384-2393.e12.
28. Li M, Lou F, Fan H. SARS-CoV-2 variant Omicron: currently the most complete "escapee" from neutralization by antibodies and vaccines. Signal Transduct Target Ther. 2022;7(1):28.
29. Dejnirattisai W, Huo J, Zhou D, Zahradník J, Supasa P, Liu C, et al. SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. Cell. 2022;185(3):467-484.e15.
30. Xiao M, Liu X, Ji J, Li M, Li J, Yang L, et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. Genome Med. 2020;12(1):57.
31. He Y, Niu P, Peng B, Sun Y, Lyu Z, Zhang R, et al. Genome characterization of COVID-19 lineage B.1.1.7 detected in the first six patients of a cluster outbreak - Shenzhen City, Guangdong Province, China, May 2021. China CDC Wkly. 2021;3(25):541-543.
32. Lv Q, Kong D, He Y, Lu Y, Chen L, Zhao J, et al. A SARS-CoV-2 Delta variant outbreak on airplane: vaccinated air passengers are more protected than unvaccinated. J Travel Med. 2021;28(8):taab161.
33. Wang R, Zhang Q, Ge J, Ren W, Zhang R, Lan J, et al. Analysis of SARS-CoV-2 variant mutations reveals neutralization escape mechanisms and the ability to use ACE2 receptors from additional species. Immunity. 2021;54(7):1611-1621 e5.
34. Fratev F. N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human-derived antibody: a free energy of perturbation retrospective study. J Chem Inf Model. 2021;61(12):6079-6084.

35. Zhang Q, Ju B, Ge J, Chan JFW, Cheng L, Wang R, et al. Potent and protective IGHV3-53/3-66 public antibodies and their shared escape mutant on the spike of SARS-CoV-2. Nat Commun. 2021;12(1):4210.
36. Worby CJ, Lipsitch M, Hanage WP. Shared genomic variants: identification of transmission routes using pathogen deep-sequence data. Am J Epidemiol. 2017;186(10):1209-1216.
37. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20(1):8.
38. Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nat Methods. 2021;18(2):165-169.
39. Choi B, Choudhary MC, Regan J, Sparks JA, Padera RF, Qiu X, et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. N Engl J Med. 2020;383(23):2291-2293.
40. Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature. 2021;592(7853):277-282.
41. Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. Virology. 2017;510:165-174.
42. Cottam EM, Whelband MC, Wileman T. Coronavirus NSP6 restricts autophagosome expansion. Autophagy. 2014;10(8):1426-1441.
43. El-Kamand S, Du Plessis MD, Breen N, Johnson L, Beard S, Kwan AH, et al. A distinct ssDNA/RNA binding interface in the Nsp9 protein from SARS-CoV-2. Proteins. 2022;90(1):176-185.
44. Frieman M, Yount B, Agnihothram S, Page C, Donaldson E, Roberts A, et al. Molecular determinants of severe acute respiratory syndrome coronavirus pathogenesis and virulence in young and aged mouse models of human disease. J Virol. 2012;86(2):884-897.
45. Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P, et al. Structure of replicating SARS-CoV-2 polymerase. Nature. 2020;584(7819):154-156.
46. Lu W, Zheng BJ, Xu K, Schwarz W, Du L, Wong CK, et al. Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release. Proc Natl Acad Sci U S A. 2006;103(33):12540-12545.
47. Nemudryi A, Nemudraia A, Wiegand T, Nichols J, Snyder DT, Hedges JF, et al. SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. Cell Rep. 2021;35(9):109197.
48. Cao Z, Xia H, Rajsbaum R, Xia X, Wang H, Shi PY. Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response. Cell Mol Immunol. 2021;18(3):746-748.
49. Ju B, Zhang Q, Ge J, Wang R, Sun J, Ge X, et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. Nature. 2020;584(7819):115-119.
50. Duggal NK, Emerman M. Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol. 2012;12(10):687-695.
51. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet. 2012;46:677-700.
52. Boehm E, Kronig I, Neher RA, Eckerle I, Vetter P, Kaiser L; Geneva Centre for Emerging Viral Diseases. Novel SARS-CoV-2 variants: the pandemics within the pandemic. Clin Microbiol Infect. 2021;27(8):1109-1117.
53. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol. 2021;19(7):409-424.
54. Ge J, Wang R, Ju B, Zhang Q, Sun J, Chen P, et al. Antibody neutralization of SARS-CoV-2 through ACE2 receptor mimicry. Nat Commun. 2021;12(1):250-250.
55. Wang R, Zhang Q, Ge J, Ren W, Zhang R, Lan J, et al. Analysis of SARS-CoV-2 variant mutations reveals neutralization escape mechanisms and the ability to use ACE2 receptors from additional species. Immunity. 2021;54(7):1611-1621.e5.