Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Temporal dynamics of SARS-CoV-2 genome mutations that occurred in vivo on an aircraft

Yaqing He, Shengyuan Dang, Wentai Ma, Long Chen, Renli Zhang, Shujiang Mei, Xinyi Wei, Qiuying Lv, Bo Peng, Ying Sun, Dongfeng Kong, Jiancheng Chen, Shimin Li, Xiujuan Tang, Qingju Lu, Can Zhu, Zhigao Chen, Jia Wan, Xuan Zou, Mingkun Li, Tiejiang Feng, Lili Ren, Jianwei Wang

1. Introduction

Since the first identification of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), over 1,000 lineages of the virus have been classified using Phylogenetic Assignment of Named Global Outbreak Lineages (Pangolin) software [1]. Based on the standard definitions of variants of interest (VOIs) and variants of concern (VOCs) established by the World Health Organization (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/), two currently circulating VOCs (Delta and Omicron), three previously circulating VOCs (Alpha, Beta, and Gamma), and eight previously circulating VOIs (Epsilon, Zeta, Eta, Theta, Lambda, and Mu) have been detected. They have been spread worldwide since 22 March 2022.

We analyzed variations in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome during a flight-related cluster outbreak of coronavirus disease 2019 (COVID-19) in Shenzhen, China, to explore the characteristics of SARS-CoV-2 transmission and intra-host single nucleotide variations (iSNVs) in a confined space. Thirty-three patients with COVID-19 were sampled, and 14 were resampled 3–31 days later. All 47 nasopharyngeal swabs were deep-sequenced. iSNVs and similarities in the consensus genome sequence were analyzed. Three SARS-CoV-2 variants of concern, Delta (n = 31), Beta (n = 1), and C.1.2 (n = 1), were detected among the 33 patients. The viral genome sequences from 30 Delta-positive patients had similar SNVs; 14 of these patients provided two successive samples. Overall, the 47 sequenced genomes contained 164 iSNVs. Of the 14 paired (successive) samples, the second samples (T2) contained more iSNVs (median: 3; 95% confidence interval [95% CI]: 2.77–10.22) than did the first samples (T1; median: 2; 95% CI: 1.63–3.74). Wilcoxon test, P = 0.021. 38 iSNVs were detected in T1 samples, and only seven were also detectable in T2 samples. Notably, T2 samples from two of the 14 paired samples had additional mutations than the T1 samples. The iSNVs of the SARS-CoV-2 genome exhibited rapid dynamic changes during a flight-related cluster outbreak event. Intra-host diversity increased gradually with time, and new site mutations occurred in vivo without a population transmission bottleneck. Therefore, we could not determine the generational relationship from the mutation site changes alone.

© 2023 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
What are the dynamic characteristics of intra-host single nucleotide variations (iSNVs) and single nucleotide polymorphisms (SNPs) during severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission?

Evidence before this study
SARS-CoV-2 replication can generate a large amount of quasispecies that carried iSNVs within infected individuals. These iSNVs and SNPs provide much more viral intra-host genetic diversities in individuals, which may be helpful to SARS-CoV-2 evolution and evade both host and vaccine immunity.

New findings
The iSNVs of the SARS-CoV-2 genome exhibited rapid dynamic changes during a flight-related cluster outbreak event. The intra-host diversity of SARS-CoV-2 increased gradually with time, and new mutations can occur directly in vivo without travelling through a transmission bottleneck. The generational relationship of SARS-CoV-2 cannot be determined from the mutation site changes alone.

Significance of the study
This study informs that the intra-host diversity of SARS-CoV-2-infected individuals increased rapidly with time. Mutation site changes alone are insufficient to determine a generational relationship.

[2,3]. SARS-CoV-2 has a proofreading mechanism that can reduce replication errors, the mutation rate is $9.80 \times 10^{-4}$ site/year [4–7]. During viral transmission, additional single nucleotide polymorphisms (SNPs) may appear in the viral genome. Whether such additional SNPs represent a new viral generation remains unclear.

Aggregated SARS-CoV-2 infection events have been observed globally, especially in confined spaces, such as prisons, ships, and nursing homes [8–10]. Such confined spaces can provide ideal scenarios to analyze the characteristics of SARS-CoV-2 genome mutations during viral transmission [11]. In this study, we analyzed the temporal changes among intra-host single nucleotide variations (iSNVs) and SNPs in the SARS-CoV-2 genome using samples from 33 patients with COVID-19 who were clustered in an aircraft. In addition, we aimed to explore whether additional site mutations indicated viral generation during transmission.

2. Methods

2.1. Participants and samples

Nasopharyngeal swabs were collected from passengers on flight CA868 from Johannesburg, South Africa, to Shenzhen, China, arriving on 10 June 2021, and from residents who had close contact with the passengers. All samples were tested using two rapid nucleic acid test kits (Daan Gene and BioGerm, China). A cycle threshold (Ct) value of $< 36$ was considered positive for both kits. Consecutive samples were collected from 14 patients with COVID-19 at an average interval of 15.23 (3–31) days.

2.2. RNA extraction, library construction, and deep sequencing

RNA was extracted from 200 μl of each sample using a High Pure Viral RNA Kit (Roche, Germany). In addition, a ULSEN® 2019-nCoV Whole Genome Capture Kit (Beijing MicroFuture, V-090418, Beijing, China) was used to capture the viral RNA. The RNA sequencing libraries were prepared using an Illumina Nextera® XT Library Prep Kit (Illumina, CA, USA) and sequenced on an Illumina MiSeq platform (Illumina) with a paired-end read length of 150 bp.

2.3. SARS-CoV-2 genome sequence analysis

Data quality control and adaptor trimming (–150 –cut_tail –cut_tail_mean_quality 20) were performed using FASTP (version 0.21.0) to remove reads of $< 50$ bp and an average base quality $< 20$ [12]. Taxonomic classification was performed using Kraken (version 2.0.8) [13]. Qualified clean reads classified as Coronavirinae were extracted and mapped to the SARS-CoV-2 reference genome (GenBank: MN908947.3) using BWA-MEM software (version 0.7.17) [14]. Minimap2 (version 2.1) was used to confirm deletions [15]. Reads were deduplicated using the Picard toolset (version 1.119) [16]. BamUtil trimBam (version 1.0.15) was used to remove ten bp from both ends of each mapped read. Mpileup files and read counts files were generated using SAMtools (version 1.14) and VarScan2 (version 2.4.3), respectively; the alignment mapQ and baseQ cut-offs were both set to 20 [17]. Consensus sequences and SNPs were identified using the following criteria: 1) coverage $\geq 5$ fold; 2) frequency of the main allele $\geq 70\%$, and 3) consensus bases mapped to both strands. Lineages were assigned using the Pangolin web server (version 3.1.11) [1].

2.4. Intra-host nucleotide variation analysis

Samples with average viral sequencing depths equal to or greater than $50 \times$ were used to evaluate intra-host diversity. Minor allele frequency (MAF) was calculated from the read counts file generated by VarScan2 (version 2.3.9), iSNVs were called when the MAF $\geq 5\%$, and the minor allele was supported by at least 2% of reads mapped to each strand at positions covered by at least 100 reads and ten uniquely mapped reads.

3. Results

3.1. Positive detections and SARS-CoV-2 lineages

An outbreak of SARS-CoV-2 occurred on flight CA868 and involved 39 passengers and seven residents in Shenzhen, China, who had close contact with passengers [18]. They were diagnosed as SARS-CoV-2-positive following the guidelines 2–10 days after arriving in Shenzhen [19]. Complete viral genomes were obtained from 33 patients with high viral RNA loads (Table S1). All viral genomes obtained from the COVID-19-positive participants had point mutations compared with the reference sequence (MN908947.3). We identified three SARS-CoV-2 variants, Delta (n = 31), Beta (n = 1), and C.1.2 (n = 1), among the samples based on the similarity analysis and sequence alignment. Delta strains from 30 samples had similar iSNVs contained 36 site mutations and were named Delta-I. Another Delta strain from one sample had 38 site mutations and was named Delta-II. 26 of these mutations differed from those of Delta strain, and 3 consensus bases mapped to both strands. The second round of nasopharyngeal swabs was collected from 14 of the 33 patients who were COVID-19-positive and simultaneously Delta-I-positive. The swabs taken on arrival at Shenzhen were labeled T1, and those from the second round were labeled T2. The two sets of
samples allowed investigation of the in vivo evolution of SARS-CoV-2. From the 47 T1 and T2 samples, we identified 164 iSNVs, most of which (n = 96, 62.3%) had MAFs > 10% (Tables S2 and S3, Fig. 1A, B).

We found 1–2 and 3–15 iSNVs in 28 (59.6%) and 18 (38.3%) of the 47 samples. The T2 sample from one patient (55E) had 24 iSNVs. The 14 T2 samples contained significantly more iSNVs (median: 3; 95% confidence interval [95% CI]: 2.77–10.22) than did the 33 T1 samples (median: 2; 95% CI: 1.63–3.74; Wilcoxon test, P = 0.021; Fig. 2A), indicating that the iSNV diversity increased with infection time, as reported previously [21]. We also paired the T1 and T2 samples from the 14 patients (Table S1) and found significantly lower Ct values in the tested genes of the T1 samples (medians: 21.27 for N and 21.35 for orf1ab) than in those of the T2 samples (medians: 27.11 for N and 28.46 for orf1ab; Wilcoxon test, P < 0.001). However, no correlation was found between Ct values and the number of iSNVs per sample (r = 0.22, P = 0.14), indicating that the intra-host diversities were unlikely to have been caused by potential sampling bias. The iSNV frequencies varied considerably during the infection in the paired T1 and T2 samples. Thirty-one out of 38 (81.6%) of the iSNVs in the T1 sam-

### Table 1

| Case no. | Disease severity | Viral nucleic acid test positive date | Symptom onset date | Viral nucleic acid test positive date | Viral lineage |
|----------|------------------|--------------------------------------|-------------------|--------------------------------------|---------------|
| 36A      | Moderate         | 2021/6/14                            | 2021/6/19         | 2021/7/17                            | Delta-I((B.1.617.2) |
| 36J      | Mild             | 2021/6/18                            | 2021/6/18         | 2021/6/30                            | Delta-I((B.1.617.2) |
| 36K      | Moderate         | 2021/6/14                            | 2021/6/14         | 2021/7/15                            | Delta-I((B.1.617.2) |
| 39A      | Mild             | 2021/6/10                            | 2021/6/10         | 2021/7/16                            | Delta-II((B.1.617.2) |
| 46K      | Mild             | 2021/6/18                            | 2021/6/18         | 2021/7/22                            | Delta-I((B.1.617.2) |
| 47L      | Moderate         | 2021/6/14                            | 2021/6/14         | 2021/7/24                            | Delta-I((B.1.617.2) |
| 48A      | Moderate         | 2021/6/16                            | 2021/6/15         | 2021/7/23                            | Delta-II((B.1.617.2) |
| 48C      | Moderate         | 2021/6/13                            | 2021/6/13         | 2021/7/18                            | Delta-I((B.1.617.2) |
| 49A      | Mild             | 2021/6/15                            | 2021/6/14         | 2021/7/18                            | Delta-I((B.1.617.2) |
| 49E      | Moderate         | 2021/6/14                            | 2021/6/18         | 2021/8/18                            | Beta((B.1.351) |
| 49L      | Mild             | 2021/6/10                            | 2021/6/14         | 2021/7/16                            | Delta-I((B.1.617.2) |
| 50A      | Mild             | 2021/6/14                            | 2021/6/13         | 2021/7/16                            | Delta-I((B.1.617.2) |
| 50B      | Mild             | 2021/6/14                            | 2021/6/13         | 2021/7/19                            | Delta-I((B.1.617.2) |
| 50C      | Mild             | 2021/6/13                            | 2021/6/14         | 2021/7/5                             | Delta-I((B.1.617.2) |
| 51A      | Mild             | 2021/6/14                            | 2021/6/14         | 2021/7/20                            | Delta-I((B.1.617.2) |
| 51B      | Moderate         | 2021/6/14                            | 2021/6/14         | 2021/7/16                            | Delta-I((B.1.617.2) |
| 51C      | Moderate         | 2021/6/14                            | 2021/6/16         | 2021/7/24                            | Delta-I((B.1.617.2) |
| 51E      | Moderate         | 2021/6/14                            | 2021/6/14         | 2021/7/16                            | Delta-I((B.1.617.2) |
| 52E      | Moderate         | 2021/6/17                            | 2021/6/17         | 2021/7/19                            | Delta-I((B.1.617.2) |
| 53D      | Moderate         | 2021/6/17                            | 2021/6/23         | 2021/7/22                            | Delta-I((B.1.617.2) |
| 53E      | Mild             | 2021/6/14                            | 2021/6/14         | 2021/7/20                            | Delta-I((B.1.617.2) |
| 53L      | Mild             | 2021/6/14                            | 2021/6/14         | 2021/7/7                             | Delta-I((B.1.617.2) |
| 55E      | Moderate         | 2021/6/14                            | 2021/6/14         | 2021/8/11                            | Delta-I((B.1.617.2) |
| 58H      | Moderate         | 2021/6/14                            | 2021/6/18         | 2021/8/15                            | Delta-I((B.1.617.2) |
| 57C      | Moderate         | 2021/6/14                            | 2021/6/15         | 2021/7/18                            | C.1.2 |
| 57H      | Moderate         | 2021/6/14                            | 2021/6/14         | 2021/7/16                            | Delta-I((B.1.617.2) |
| 58E2     | Mild             | 2021/6/17                            | 2021/6/16         | 2021/7/24                            | Delta-I((B.1.617.2) |
| LOC1     | Mild             | 2021/6/14                            | 2021/6/14         | 2021/7/18                            | Delta-I((B.1.617.2) |
| LOC2     | Moderate         | 2021/6/21                            | 2021/6/21         | 2021/7/18                            | Delta-I((B.1.617.2) |
| LOC3     | Moderate         | 2021/6/20                            | 2021/6/20         | 2021/7/18                            | Delta-I((B.1.617.2) |
| LOC4     | Moderate         | 2021/6/18                            | 2021/6/12         | 2021/7/18                            | Delta-I((B.1.617.2) |
| LOC5     | Mild             | 2021/6/17                            | 2021/6/17         | 2021/7/18                            | Delta-I((B.1.617.2) |

*Disease severity was determined following the guideline (National Health Commission of the People's Republic of China, 2020). *The date format is year/month/day.

Fig. 1. Intra-host single nucleotide variation (iSNV) distribution of all 47 positive samples. A) Minor allele frequency (MAF) distribution of all iSNVs identified. B) iSNV number identified from 50 samples.
ples were not observed in T2, whereas 80 new iSNVs were detected in T2 (Fig. 2C), indicating a dramatic frequency change between two time points (Fig. 2C). The trend was also similar for higher-frequency iSNVs. Among 96 iSNVs with MAFs > 10% at either time point, the MAF of 90 (93.8%) iSNVs changed significantly between the two time points (Fisher’s exact test, P < 0.05, Fig. 3). Only three (out of 19) iSNVs with MAFs ≥ 10% in T1 samples were detected at T2 time points (MAF > 5%).

The proportion of synonymous iSNVs increased from 25.6% (T1) to 31.1% (T2), and the proportion of nonsynonymous iSNVs decreased from 38.4% (T1) to 25.9% (T2; Fig. 2B). The dN/dS ratio for the iSNVs did not significantly change between two time points, and the ratio did not deviate from 1:1 at any time point, suggesting that the frequency change was likely caused by genetic drift rather than selection, which had also been proposed in other studies [21,22].

**Fig. 2.** Intra-host single nucleotide variation (iSNV) number and variation type between T1 and T2 of all sequenced cases. A) iSNV number and Ct value at the sampling time points. Samples from the same patient are connected by a gray line. The boxplot represents the iSNV number or Ct value (both **orf1ab** and N) distribution at each time point. B) Proportions of different iSNV types obtained from the samples collected at both T1 and T2. C) The three pie charts from left to right indicate the detailed proportion of iSNV types lost at T2 compared with T1, the detailed proportion of new iSNV types at T2 compared with T1, and the proportion of iSNV types at both T1 and T2. D) Site mutation (fraction ≥ 70%) number of the consensus genome of each Delta-I sample. Samples from the same case are connected by a black arc.
In addition to the iSNVs, we also analyzed the consensus sequences in each sample (Fig. 2D and Table S4). Compared with the T1 samples, two Delta-I-positive patients (52E, LOC3) had additional mutations in the T2 samples, including an AGTGGTTATT3350A deletion (orf1ab: SGYL1029I; patient 52E) and a C10605A substitution (orf1ab: P3447H; patient LOC3). The MAFs of the AGTGGTTATT3350A mutation increased from 0.0% to 87.3% from T1 to T2 (P < 0.001, Fisher’s exact test), and the MAFs of the C10605A decreased from 80.7% to 0.0% from T1 to T2 (P < 0.001, Fisher’s exact test, Supplementary Table S4). These mutations at the consensus-sequence level may be caused by genetic drift or selection pressure.

4. Discussion

We retrospectively analyzed iSNVs from SARS-CoV-2 genomes from a COVID-19 outbreak event on an aircraft. The Delta-I strain we identified had 36 site mutations and matched a strain detected mainly in the United Kingdom and accounted for < 0.02% of the total number of Delta sequences in the GISAID database [23]. The Delta-II strain had 38 mutations representing the canonical genome of the Delta variant, which was widely distributed in South Africa, India, the United States, and Europe. The different mutation site in the spike gene was C22227T (A222V) in Delta-I compared with the spike gene sequence in Delta-II, which is related to the antigenic activity of the virus [24]. Most of the passengers were infected with the Delta-I strain during the flight, suggesting a more vital transmission ability of Delta-I or these passengers infected with Delta-I were infected before boarding, or they were more like to get in close contact with other people on the plane.

The Delta-I strains from the T1 samples had the same site mutations in their genomes, indicating a single infectious source in the confined space of the aircraft. The number of iSNVs in the T2 samples increased in vivo with the infectious time compared with that in the T1 samples. Such a trend has been reported previously [25]. The frequency of some iSNVs increased rapidly over a short period (from T1 to T2). In three patients who provided T1 and T2 samples, three SNP changes were detected in the T2 viral genome sequence: two new mutations and one back mutation were generated during the infection. Notably, changes in iSNVs and SNPs cannot be used as evidence to determine SARS-CoV-2 transmission generations in one outbreak event. Viral generations, viral load, disease course, vaccination, and sampling time were all significant factors influencing the number of viral gene mutations [11,26–30].

In summary, we analyzed the site mutation characteristics of SARS-CoV-2 genomes from a COVID-19 outbreak that occurred on an aircraft. The intra-host diversity increased gradually with time, and new mutations occurred in vivo with no population transmission bottleneck. However, mutation site changes alone are insufficient to determine a generational relationship.

Ethics statement

Shenzhen Center for Disease Control and Prevention Ethics Committee approved this study (QS2021070047). Informed consent was obtained from all individual participants included in the study.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant number 82161148009), the Non-profit Central Research Institute Fund of the Chinese Academy of Medical Science (grant number APL211276910010201002008), the Shenzhen Science
and Technology Innovation Commission Key project (grant number JSGG20200225152648408), the Shenzhen Key Medical Discipline Construction Fund (grant number SZXX064), the Key Project of Shenzhen Science and Technology Innovation Commission (grant number KCXZF20200210061900), the First Fighting the Epidemic Project of Shenzhen (grant number JSGG20210901145004012) and the key project of Beijing Natural Science Foundation (grant number Z190017). We also thank the clinicians of Shenzhen Third People’s Hospital for their help in collecting the samples.

Conflict of interest statement
The authors declare that there are no conflicts of interest.

Author contributions

Mingkun Li: Conceptualization. Tiejian Fen: Conceptualization. Lili Ren: Conceptualization. Jianwei Wang: Conceptualization. Yaqing He: Methodology, Investigation. Shengyuan Dang: Methodology, Investigation, Data curation, Software, Writing - original draft. Wentai Ma: Data curation, Software, Writing - original draft. Long Chen: Investigation. Renli Zhang: Investigation. Shuangli Mei: Investigation. Xinyi Wei: Investigation. Qiuying Lv: Investigation. Bo Peng: Investigation. Ying Sun: Investigation. Dongfeng Kong: Investigation. Can Zhu: Investigation. Shimin Li: Investigation. Xiujuan Tang: Investigation. Qiang Li: Investigation. Jiancheng Chen: Investigation. Zhiqiao Chen: Investigation. Jia Wan: Investigation. Xuan Zou: Investigation.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bshelth.2022.10.004.

References

[1] A. Rambaut, E.C. Holmes, A. O’Toole, Y. Hill, J.T. McCrone, C. Ruis, L. du Plessis, O.G. Pybus, A dynamic.numerical proposal for SARS-CoV-2 lineages to assist genomic epidemiology, Nat. Microbiol. 5 (11) (2020) 1403-1407, https://doi.org/10.1038/s41564-020-0770-5.

[2] D. Petrovic, E. Dempsey, D.G. Doherty, D. Kelleher, A. Long, Hepatitis C virus–T-cell responses and viral escape mutations, Eur. J. Immunol. 42 (1) (2012) 17–26, https://doi.org/10.1002/eji.201141593.

[3] N. Gulzar, K.F. Copeland, CD8+ T-cells: function and response to HIV infection, Curr. HIV Res. 2 (1) (2004) 23–31, https://doi.org/10.2174/1566002043840877.

[4] E. Domingo, C. Perales, Quasispecies and virus, Eur. Biophys. J. 47 (4) (2018) 443–457, https://doi.org/10.1007/s00249-018-1284-6.

[5] R. Andino, E. Domingo, Viral quasispecies, Virology 479–480 (2015) 46–51, https://doi.org/10.1016/j.virology.2015.03.022.

[6] F. Robson, K.S. Khan, T.K. Le, C. Paris, S. Demirbag, P. Barfuss, P. Rocchi, W.L. Ng, Coronavirus RNA proofreading: Molecular basis and therapeutic targeting, Mol. Cell 79 (5) (2020) 710–727, https://doi.org/10.1016/j.molcel.2020.07.025.

[7] The Nextstrain team, Genomic epidemiology of SARS-CoV-2 with subsampling focused globally since pandemic start. https://nextstrain.org/covid/gisaid/global/all-time?c=clock, 2022 (accessed 28 September 2022).

[8] L.M. Hagan, S.P. Williams, A.C. Spaulding, R.L. Toblin, J. Figlenski, J. Ocampo, J. Wang, Y. Huang, F. Xiao, J. Yao, et al., Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients, Genome Med. 13 (1) (2021) 30, https://doi.org/10.1186/s13073-021-00847-5.

[9] H. Li, R. Zhang, Y. Sun, Z. Lyu, R. Zhang, X. Zhao, Y. Feng, X. Tang, 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 3 (25) (2021) 541–543, https://doi.org/10.46234/cckw2021.129.

[10] National Health Commission of the People’s Republic of China, COVID-19 diagnosis and treatment plan (Revised Eighth Edition). http://www.gov.cn/zhengce/zhengceku/2021-04/15/5599795/content_5599795.htm, (2021-04-15) 2021.

[11] J. Wu, L. Zhang, Y. Zhang, H. Wang, R. Ding, J. Nie, Q. Li, S. Liu, Y. Xu, Y. Yang, et al., The antigenicity of epidemic SARS-CoV-2 variants in the United Kingdom, Front. Immunol. 12 (2021) 687869, https://doi.org/10.3898/fimmu.2021.687869.

[12] Y. Wang, D. Zhang, L. Wang, Z. Chen, W. Zhang, C. Zhu, Y. Huang, F. Xiao, J. Yao, et al., Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients, Genome Med. 13 (1) (2021) 30, https://doi.org/10.1186/s13073-021-00847-5.

[13] D.E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2, Genome Biol. 20 (1) (2019) 257, https://doi.org/10.1186/s13059-019-1891-0.

[14] Y. He, P. Niu, B. Peng, Y. Sun, Z. Lyu, R. Zhang, Y. Chen, F. Hu, Y. Lan, Z. Yang, et al., Variations in SARS-CoV-2 lineage co-infection on a prolonged viral shedding COVID-19 case: Insights into clinical severity and disease duration, Microorganisms 9 (2) (2021) 300, https://doi.org/10.3390/mi9020300.

[15] H. Li, M. Hu, Q. Yu, B. Peng, Y. Sun, Z. Lyu, R. Zhang, X. Zhao, Y. Song, et al., Vaccinated air passengers are more protected than unvaccinated, J. Travel Med. (2021) taab161, https://doi.org/10.1093/jtm/taab161.

[16] L. Zhang, D. Wang, M. Wang, Q. Zhu, C. Yan, M. Meng, X. Tian, et al., Two-step fitness selection for intra-host variations in SARS-CoV-2, Cell Rep. 38 (2) (2022) 110205, https://doi.org/10.1016/j.celrep.2021.110205.

[17] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, et al., The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (9) (2010) 1297–1303, https://doi.org/10.1101/gr.107524.110.

[18] Y. He, P. Niu, B. Peng, Y. Sun, Z. Lyu, R. Zhang, Y. Chen, F. Hu, Y. Lan, Z. Yang, et al., Variations in SARS-CoV-2 lineage co-infection on a prolonged viral shedding COVID-19 case: Insights into clinical severity and disease duration, Microorganisms 9 (2) (2021) 300, https://doi.org/10.3390/mi9020300.

[19] N. Pedro, C.N. Silva, A.C. Magalhães, B. Cavadas, A.M. Rocha, A.C. Moreira, M.S. Santos, J.A. Sousa, M. Silva, et al., SARS-CoV-2 infection, clinical features and outcome of COVID-19 in United Kingdom nursing homes, J. Infect. 81 (3) (2020) 411–419, https://doi.org/10.1016/j.jinf.2020.05.073.

[20] Y. Wang, R. Chen, F. Hu, Y. Lan, Z. Yang, C. Zhan, J. Shi, X. Deng, M. Jiang, et al., Pathogenicity and immunity of the alpha variant of SARS-CoV-2, Viruses 13 (8) (2021) 1642, https://doi.org/10.3390/v13081642.

[21] N. Pedro, C.N. Silva, A.C. Magalhães, B. Cavadas, A.M. Rocha, A.C. Moreira, M.S. Santos, J.A. Sousa, M. Silva, et al., SARS-CoV-2 infection, clinical features and outcome of COVID-19 in United Kingdom nursing homes, J. Infect. 81 (3) (2020) 411–419, https://doi.org/10.1016/j.jinf.2020.05.073.