Specific re-distribution of SARS-CoV-2 variants in the respiratory system and intestinal tract

Pengcheng Du[1,2#], Chuan Song[1,2#], Rui Li[1,2#], Yangzi Song[1,2#], Jiarui Li[1,2], Nan Ding[1,2], Ju Zhang[1,2], Rui Song[4], Junyan Han[1,2], Guiju Gao[4], Jinglin Yue[1,2], Ang Duan[1,2], Yanyi Huang[5], Jing An[6], Jianbin Wang[3]*, Fujie Zhang[4]*, Chen Chen[1,2]*, and Hui Zeng[1,2]*

1. Institute of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing, P. R. China
2. Beijing Key Laboratory of Emerging Infectious Diseases, Beijing, P. R. China
3. School of Life Sciences, Beijing Advanced Innovation Center for Structural Biology (ICSB), Chinese Institute for Brain Research (CIBR), Tsinghua University, Beijing, P. R. China
4. Clinical and Research Center of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing, P. R. China
5. Biomedical Pioneering Innovation Center (BIOPIC), Beijing Advanced Innovation Center for Genomics (ICG), College of Chemistry and Molecular Engineering, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, P. R. China
6. Depart of Microbiology, School of Basic Medical Science, Capital Medical University, Beijing, P. R. China

[#] Pengcheng Du, Chuan Song, Rui Li and Yangzi Song contributed equally to this manuscript.

* Hui Zeng, Chen Chen, Fujie Zhang, and Jianbin Wang contributed equally to this manuscript.

Corresponding author:

Prof. Hui Zeng,

Institute of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University and Beijing Key Laboratory of Emerging Infectious Diseases

No.8 Jingshundongjie, Beijing 100015, China

E-mail: zenghui@ccmu.edu.cn Tel.: 00861084322621 Fax: 00861084322606

Abstract

Intra-host analysis of SARS-CoV-2 genomic sequences identified two viral haplotypes that comprised of three genetically linked mutations from the respiratory and intestinal tracts of a patient with COVID-19. Spatiotemporal data suggest that this patient initially had dual-infection of two SARS-CoV-2 variants, which subsequently re-distributed into the two systems.

Keywords: Severe acute respiratory syndrome coronavirus 2; dual-infection; organ specific re-distribution

Introduction

The coronavirus disease 2019 (COVID-19) pandemic has engendered a growing global health crisis with unpredictable long-term consequences for human health. [1, 2] The causative agent of COVID-19, SARS-CoV-2, is transmitted by direct contact and exhaled respiratory droplets, with increasing evidences of potential fecal-oral transmission. [3] SARS-CoV-2 RNA has been detected in respiratory tract specimens and feces for weeks after acute infection. Live virus has been cultured from multiple sample types, including feces. [4, 5] In addition to airway epithelial cells, human intestinal epithelium has been shown to be a productive site of SARS-CoV-2 replication, suggesting that the intestine may be an alternative target organ for SARS-CoV-2. [6]

Genetic variation driven by SARS-CoV-2 adaptions occurring between and within human hosts may have important effects on the transmissibility and pathogenicity of the virus. To date, studies have documented over 19,000 SARS-CoV-2 mutations in the public databases. [7] The genomic heterogeneity of the virus also existed within individual patients, [8] although it is still unclear whether these genetic variations represent specific adaptations to different organs. Here, by using deep sequencing, we identified two distinct SARS-CoV-2 variants in a COVID-19 patient distinguished by three genetically linked mutations and observed changes in the distribution of these variants in specimens collected from the respiratory and intestinal tracts of the patient during the course of their infection.

Results

Clinical course

A 30-year-old man was tested positive for SARS-CoV-2 RNA by quantitative RT-PCR (qRT-PCR), and admitted to Beijing Ditan hospital on February 29, 2020. He had contact with a confirmed COVID-19 patient on February 18. Prior to the admission, he exhibited mild symptoms (sore throat, dry cough and runny nose), but no fever or diarrhea (Figure S1). During hospitalization, chest CT scans taken on hospital days 1 and 14 were normal. He was treated with alpha-interferon, but no other antiviral medication. The patient's pharyngeal swabs, sputum and feces were collected at 3-day intervals and tested by qRT-PCR to monitor

virus load; the last positive test was carried on hospital day 33 (Figure S1). Additionally, serum specimens collected on hospital days 10, 18 and 26 were tested for SARS-CoV-2-specific antibodies. IgM antibodies were negative at all three time points and IgG antibodies were positive on days 18 and 26 (Figure S1). The patient was discharged on day 37 in accordance with the Chinese Diagnosis and Treatment Protocols (no clinical symptoms and two consecutive negative qRT-PCR tests).

Genomic sequences and single nucleotide polymorphisms (SNPs)

SARS-CoV-2 genome sequencing was conducted using two strategies to confirm sequence data: (i) whole metatranscriptome sequencing using a TNBSEQ-T7 platform and (ii) targeted SARS-CoV-2 sequencing on an Illumina NextSeq500 (Supplementary methods). After patient admission and consenting, the parallel sequencing using the two strategies was performed on three samples (pharyngeal swab, sputum and feces) collected on March 3, 2020 (hospital day 4, Table S1 and S2). Sequence comparison to the reference genome (Wuhan-Hu-1, accession number: NC_045512.2) revealed six high quality SNPs (C884T, G1397A, G8653T, G11083T, T28688C and G29742T) in all three samples (Figure 1A). The first four SNPs were located in ORF1a and led to non-synonymous mutations (R207C, V378I, M2796I and L3606F); the other two synonymous SNPs were located in the N gene (T28688C) and the 3'-UTR region (G29742T), respectively (Figure 1A). The six unique SNPs were not detected in other patients in our hospital during the same period.

Haplotype identification based on intra-host single nucleotide variations (iSNVs)

We further analyzed the sequence data for the presence of iSNVs in the three samples. We detected four iSNVs (C8481T, C22000A, T26975G and G28812T - Figure 1B) in the pharyngeal swab and sputum samples, three (C8481T, C22000A, and G28812T) of which have been identified as SNP sites previously. [7] These four iSNVs were located in the ORF1a, S, M and N genes, respectively, and led to nonsynonymous mutations (Figure 1B). In the pharyngeal swab sample, the mutation allele frequencies (MuAFs) at these four sites (C8481T, C22000A, T26975G and G28812T) were 0%, 6.25%, 8.2% and 4.69%. In marked contrast, the MuAFs at these four sites in sputum were 74.29%, 75.35%, 76.49% and 79.30%. Frequencies of all four mutant iSNV alleles were significantly correlated across the two

samples ($p<0.001$) (Figure S2, S3, S4). Haplotype analysis of these iSNVs identified three positions that were closely linked and located near the 3'-end of the viral genome at positions 22,000, 26,975 and 28,812 bp. Using these positions, two haplotypes were defined corresponding to "C-T-G" (Variant I) and "A-G-T" (Variant II) at (Figure 1C). Variant I made up ~90% of the viral sequences in the pharyngeal swab, while Variant II was predominant (~70%) in the sputum sample (Table S3). In the fecal sample, we detected 1,088 T nucleotides at position 26,975, and 289 G nucleotides at position 28,812 of the genome. We did not find any alternative allele at either of these two sites, which suggested that Variant II was extremely low in feces.

Dynamic alteration of two variants in the respiratory and digestive systems

To investigate any spatiotemporal changes in the two genetic variants over the course of infection, we performed a follow-up sequencing analysis on pharyngeal swab, sputum and feces samples collected on hospital days 10 or 11. We confirmed the presence of the same six SNPs originally identified on hospital day 4. However, the distribution of iSNV MuAFs was significantly different (Figure S5, Table S3). Variant I became overwhelmingly dominant (>99%) in the pharyngeal swab and sputum, though we still observed a few reads supporting continued presence of Variant II in pharyngeal swab (0.07%) and sputum (0.04%, Figure 1D). In feces, the variant distribution was profoundly reversed, with Variant II predominating and only 0.04% of Variant I, despite the fact that this was the only variant detectable in the fecal sample on hospital day 4. This might be caused by stochastic sampling of the extremely low proportion minor variants in feces at the two time points (Figure S6). These results based on MuAFs at the three iSNV sites revealed dynamic changes in distribution of the variant types in respiratory and gastrointestinal systems.

Discussion

Here, we present spatiotemporal dynamic of molecular data demonstrating that two SARS-CoV-2 variants might initially co-infect the respiratory tract of a single patient. The case presented here is a young adult with mild symptoms, and without being subjected to specific anti-viral therapy, except for atomized alpha-interferon. Our results were supported by sequencing data from both whole metatranscriptome sequencing and targeted SARS-CoV-2

sequencing. The ultra-deep sequencing provides high quality unbiased data, and the targeted approach extended the coverage and sequencing depth of the viral genome (Table S2, Figure S2). Spatiotemporal analysis of the iSNVs provided concrete evidences to support the notion that a dual SARS-CoV-2 infection occurred in this patient. For instance, three iSNVs were highly genetically linked in both pharyngeal swabs and sputum in the early stage of infection (hospital day 4), presented as two haplotypes. Moreover, only haplotype Variant I was identified in samples from the patient's respiratory tract, whereas only Variant II was found in intestinal specimens approximately one week later. We provide concrete evidence to support the notion of presence of two distinct variants in a patient. First, the three iSNVs in Variant II are highly correlated in both pharyngeal swabs and sputum in the early stage of infection. Second, the MuAF of the three iSNVs vary in a coordinated pattern across different samples. Finally, near homogenous populations were identified in the samples from respiratory system (Variant I) and digestive systems (Variant II, see supplementary materials). The dual infection in this patient might be ascribable to two possibilities: (i) the two distinct variants evolved previously, then were transmitted to this patient; (ii) a single variant infected this patient and then evolved into two distinct variants due to compensatory genetic changes at these sites.

It is known that SARS-CoV-2 infection is not confined to the lung, but can involve other parts of the respiratory tract and digestive system. [9] Multiplicity of RNA virus infection has been widely studied. The dynamic distribution of SARS-CoV-2 variants within hosts have been observed in the course of infection, as well as at different anatomical sites. [10] Moreover, the composition of different variants could also alter dynamically: minor haplotypes could become dominant during inter-host transmission. [11] Here, within an individual infected with SARS-CoV-2, we observed the spatiotemporal alteration of the dominant haplotypes in the organs, which possibly hinted at distinct trophism or adaptive advantages of the two variants under immune selective pressures in different organs and might allow immunological escape of the virus. [12] Alternatively, this phenomenon could simply be the random and variable outcome of the heterologous replication of virus within the interconnected respiratory and intestinal tracts. Given that three mutations causing amino acid alternations were identified in three important genes (S, M, and N), more investigations on organ specific distribution of the variants and functional impacts of the mutations are urgently needed, particularly by the intra-host studies of polymorphic viral variants or

quasispecies. Finally, the dual-infection of SARS-CoV-2 variants augments the viral complexity in each individual. As the existence of multiple variants *in vivo* and associated pathogenic changes has implications for disease progression and therapeutics, more work is necessary to resolve whether increasing complexity in patients will increase difficulties of vaccine and drug development, and make the virus even harder to be controlled at the population level.

NOTES

Acknowledgments

We thank all health care workers involved in the diagnosis and treatment of the COVID-19 in Beijing Ditan Hospital, and Professor Joseph D. Tucker and and Dr. Jonathan Julino from the University of North Carolina, and Professor Mark Edwards from the University of Oxford for the kindly assistances in polishing the English writing.

Funding

None.

Conflict of Interest

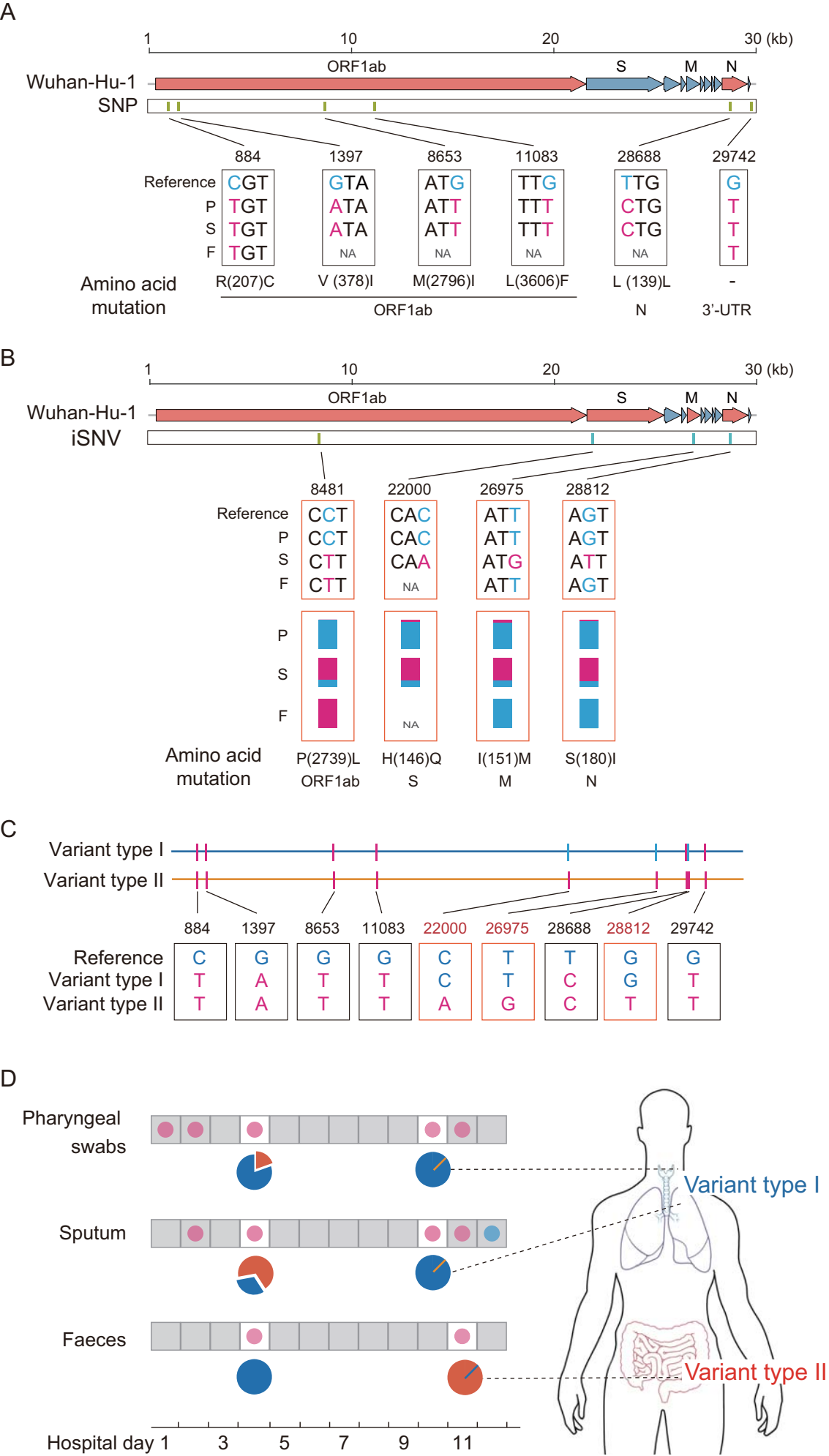
None declared.

References

1. Zhu N, Zhang D, Wang W, *et al*. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; 382(8):727-33.
2. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 2020; 368(6493):860-8.
3. Zhang H, Kang Z, Gong H, *et al*. Digestive system is a potential route of COVID-19: an analysis of single-cell coexpression pattern of key proteins in viral entry process. *Gut* 2020; 69(6):1010.
4. Wang W, Xu Y, Gao R, *et al*. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *Jama* 2020; 323(18): 1843-4.
5. Sehmi P, Cheruiyot I. Presence of Live SARS-CoV-2 Virus in Feces of Coronavirus Disease 2019 (COVID-19) Patients: A Rapid Review. *medRxiv* 2020. https://doi.org/10.1101/2020.06.27.20105429
6. Lamers MM, Beumer J, van der Vaart J, *et al*. SARS-CoV-2 productively infects human gut enterocytes. *Science* 2020; 369(6499):50-4.
7. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Eurosurveillance* 2017; 22(13): 30494.
8. Shen Z, Xiao Y, Kang L, *et al*. Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients With Coronavirus Disease 2019. *Clin Infect Dis* 2020; 71(15):713-20.
9. Guan W-J, Ni Z-Y, Hu Y, *et al*. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020; 382(18):1708-20.
10. Jary A, Leducq V, Malet I, *et al*. Evolution of viral quasispecies during SARS-CoV-2 infection. *Clin Microbiol Infect* 2020.
11. Sun F, Wang X, Tan S, *et al*. SARS-CoV-2 Quasispecies provides insight into its genetic dynamics during infection. *BioRxiv* 2020. https://doi.org/10.1101/2020.08.20.258376
12. Karamitros T, Papadopoulou G, Bousali M, *et al*. SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies. *J Clin Virol* 2020; 131: 104585.

Figures

Figure 1. The distribution of SNPs and iSNVs and definition of the two variant types. (A) The six SNPs and their genomic locations and genes. The scale and arrow map on the top represent the genetic positions and genes of SARS-CoV-2 reference strain Wuhan-Hu-1. The green lines below mark the positions of the six SNPs. The characters below represent the nucleotide substitutions and corresponding amino acid substitutions (P: pharyngeal swab; S: sputum; F: feces). (B) The genomic position and composition of the four iSNVs identified. The characters display the nucleotide substitutions and the histograms below show the allele frequencies. The characters at the bottom display the corresponding mutations of amino acid residues. (C) The definition of the two variant types based on the three iSNVs. On the top are the genomic locations of the six SNPs and the three iSNVs used to define the two variant types, and the alleles are displayed at the bottom. (D) The distribution and dynamics of the two variant types in the respiratory and digestive systems. The proportion of the two variant types in pharynx, sputa and feces collected on hospital days 4 and 10/11, respectively. The pie charts present the proportion of the two variants (blue: variant type I; orange: variant type II). The purple and blue dots represent the positive or negative results of rRT-PCR around the two sequencing time points.


A
1
10
20
30 (kb)
Wuhan-Hu-1
ORF1ab
S
M
N
SNP
884
1397
8653
11083
28688
29742
Reference
CGT
GTA
ATG
TTG
TTG
G
P
TGT
ATA
ATT
TTT
CTG
T
S
TGT
ATA
ATT
TTT
CTG
T
F
TGT
NA
NA
NA
NA
T
Amino acid mutation
R(207)C
V (378)I
M(2796)I
L(3606)F
L (139)L
-
ORF1ab
N
3'-UTR
B
1
10
20
30 (kb)
Wuhan-Hu-1
ORF1ab
S
M
N
iSNV
8481
22000
26975
28812
Reference
CCT
CAC
ATT
AGT
P
CCT
CAC
ATT
AGT
S
CTT
CAA
ATG
ATT
F
CTT
NA
ATT
AGT
P
S
F
NA
Amino acid mutation
P(2739)L
ORF1ab
H(146)Q
S
I(151)M
M
S(180)I
N
C
Variant type I
Variant type II
884
1397
8653
11083
22000
26975
28688
28812
29742
Reference
C
G
G
G
C
T
T
G
G
Variant type I
T
A
T
T
C
T
C
G
T
Variant type II
T
A
T
T
A
G
C
T
T
D
Pharyngeal swabs
Sputum
Faeces
Hospital day 1
3
5
7
9
11
Variant type I
Variant type II