Emergence of a Severe Acute Respiratory Syndrome Coronavirus 2 virus variant with novel genomic architecture in Hong Kong

Herman Tse[1], David Christopher Lung[1,2], Sally Cheuk-Ying Wong[1], Ka-Fai Ip[1], Tak-Chiu Wu[3], Kelvin Kai-Wang To[4,5], Kin-Hang Kok[4,5], Kwok-Yung Yuen[4,5,6], Garnet Kwan-Yue Choi[1,6,*]

[1] Department of Pathology, Hong Kong Children's Hospital, Hong Kong Special Administrative Region, China.

[2] Department of Pathology, Queen Elizabeth Hospital, Hong Kong Special Administrative Region, China.

[3] Department of Medicine, Queen Elizabeth Hospital, Hong Kong Special Administrative Region, China.

[4] State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China.

[5] Department of Microbiology, The University of Hong Kong, Hong Kong Special Administrative Region, China.

[6] These authors contributed equally to this manuscript.

[*] Corresponding author: Garnet Kwan-Yue Choi (cky603@ha.org.hk), Department of Pathology, Hong Kong Children's Hospital, Hong Kong Special Administrative Region, China. Tel: 852-35133109. Fax: 852-35127515.

Abstract

Throughout the COVID-19 pandemic, divergent SARS-CoV-2 lineages have emerged continuously, mostly through the genomic accumulation of substitutions. We report the discovery of a SARS-CoV-2 variant with a novel genomic architecture characterized by absent ORF7a, ORF7b and ORF8, and a C-terminally modified ORF6 product resulting from partial 5'-UTR duplication and transposition.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the novel causative agent of coronavirus disease 2019 (COVID-19). As of end January 2021, the number of confirmed COVID-19 cases exceeded 100 million globally. While limited genetic diversity of SARS-CoV-2 was observed early in the pandemic, numerous distinct clades and lineages have since emerged [1]. More recently in late 2020, several "variants of concern", namely the lineages B.1.1.7, B.1.351, and P.1, attracted global attention due to potential associations with increased transmissibility, disease severity and/or resistance to vaccine-induced immunity. The last possibility is particularly concerning, as large-scale rollout of effective COVID-19 vaccines has just begun in hopes to end the pandemic.

Most studies on the genomic evolution of SARS-CoV-2 have focused on single nucleotide variants, as this remains the primary source of genetic diversity in sequenced genomes. In comparison, deletion variants are uncommon, with small deletions seen in ≤3% of viral genomes [2]. Notably, VOC lineage B.1.1.7 variants carried a Δ69/70 deletion in the spike protein, which enabled reverse transcriptase polymerase chain reaction (RT-PCR) assays targeting the spike gene to identify such variants based on the spike gene dropout. Large deletions are rarer, as they often cause significant disturbance of genomic architecture and/or functional loss of gene products. In January 2020, an ORF8 Δ382 SARS-CoV-2 variant was detected in three clusters in Singapore [3], and was reportedly associated with a favorable clinical outcome. Here, we report three cases of infection caused by a SARS-CoV-2 variant with a novel genomic architecture resulting from deletion of an 882-nucleotide region spanning ORF6 to ORF8 together with duplication and transposition of a 57-nucleotide segment from the 5'-untranslated region (UTR).

Patient 1 was a 66-year old male local resident who presented with productive cough in late December 2020. He had untreated diabetes mellitus (HbA1c 8.2%) and no history of recent travel. On presentation, he was afebrile and did not report any dyspnea. A combined throat and nasopharyngeal swab (cT/NS) was performed, and tested positive (threshold cycle (C_t) = 15.7) for SARS-CoV-2 RNA by real-time RT-PCR using the Xpert Xpress SARS-CoV-2/Flu/RSV assay (Cepheid). Around day 10 from symptom onset, the patient developed fever and dyspnea with increasing peripheral haziness on serial chest X-rays. Complete blood count showed lymphopenia and thrombocytopenia with nadirs of 0.3×10^9/L and 113×10^9/L respectively. Routine biochemistry testing showed elevated liver parenchymal enzymes and lactate dehydrogenase (LDH) levels. C-reactive protein (CRP) progressively increased to a peak of 106 mg/L. He was given a 10-day course of dexamethasone with good clinical response, and was discharged after 18 days of in-patient management.

Patient 2 was a 61-year old female local resident hospitalized for low grade fever and productive cough in late December 2020. She had history of diabetes mellitus (HbA1c 7.3%), hypertension, hyperlipidemia and vitamin B12 deficiency. There was no history of recent travel. Her cT/NS tested positive for SARS-CoV-2 RNA by RT-PCR (C_t = 22.3) using the Xpert Xpress SARS-CoV-2/Flu/RSV assay. As the patient showed clinical and radiological progression of disease on day 6 from symptom onset, she was given a regimen of interferon beta-1b (3 doses) and ribavirin (7 days) on day 7, followed by dexamethasone (12 days) and tocilizumab (1 dose) starting on days 11 and 14, respectively. Complete blood count showed lymphopenia with a nadir of 0.8×10^9/L. LDH level peaked at 417 U/L but alanine transaminase level remained within normal limits. CRP levels peaked at 58 mg/L on day 10 from symptom onset. She recovered fully and was discharged after 33 days of hospitalization.

Patient 3 was a 46-year old female local nursing home staff who tested positive for SARS-CoV-2 RNA (C_t = 30.6) in mid-January 2021 during a compulsory COVID-19 testing exercise organized by the government. She had good past health, and her symptoms were limited to mild cough and running nose. Complete blood counts were normal throughout. There was mild elevation of LDH and CRP levels to 234 U/L and 12 mg/L respectively. The patient was keen to receive early specific treatment for COVID-19, and was given a course of interferon beta-1b (5 doses). She had an uneventful hospital stay and was discharged 26 days after admission.

All 3 patients were tested for antibodies against SARS-CoV-2 before discharge. Detection of IgG against the nucleocapsid protein (NP) was performed using the Architect SARS-CoV-2 IgG assay (Abbott), while testing for anti-Orf8 IgG was performed as described previously [4]. The sera samples from all 3 patients were positive for anti-NP IgG but not anti-Orf8, consistent with infection by a SARS-CoV-2 strain with absent ORF8 expression.

We performed ongoing genomic surveillance of SARS-CoV-2 to guide local public health and infection control measures. The study protocol was approved by the Research Ethics Committee (Kowloon Central/ East) of the Hospital Authority (KC/KE-20-0321/ER2). Detailed methods on SARS-CoV-2 whole genome sequencing and sequence analysis are provided in the Supplementary Materials. SARS-CoV-2 genome sequences were obtained for a cT/NS specimen from patient 1 (hCoV-19/Hong_Kong/HKCH_789/2020; GISAID accession EPI_ISL_1046886), and saliva specimens from patients 2 (hCoV-19/Hong_Kong/HKCH_906/2020; GISAID accession EPI_ISL_1046887) and 3 (hCoV-19/Hong_Kong/HKCH_914/2021; GISAID accession EPI_ISL_1046888). Comparative analysis showed that they were genetically identical. Curiously, there was an absence of reads mapping to the reference genome from nucleotide position 27373 to 28254, which suggested

the presence of a major deletion, but some reads around this region also partially mapped to the 5'-UTR of the viral genome. We performed long-range PCR and Sanger sequencing on the cDNA to confirm the deletion (see Supplementary Methods). The sequencing results confirmed that the 882-nucleotide genomic region described above was indeed replaced by a 57-nucleotide segment that is completely identical to a segment of the 5'-UTR (nucleotide position 8 to 64) (Figure 1a). Hence, ORF7a, ORF7b, ORF8 and the last 12 nucleotides from the 3'-end of ORF6 were lost. The remaining partial ORF6 was extended by the duplicated 5'-UTR sequence and the non-coding region preceding the nucleocapsid gene ORF to produce a novel ORF6x (Figure 1b). Since the C-terminus of the ORF6 protein is important for antagonizing interferon signaling [5], we speculate that Orf6x might be non-functional in this respect. There is no significant sequence similarity between the deleted and the transposed segments, or between their 5' flanking regions. A short 10-nucleotide segment of identical sequence was noted at the 3' flanking regions (Figure 1b) due to the presence of the conserved transcription regulatory sequence.

To examine the evolutionary relationships of these viruses, a maximum likelihood (ML) phylogenetic tree was constructed from SARS-CoV-2 genome sequences in our surveillance dataset (see Supplementary Methods). The genome sequences were also submitted to the PANGOLIN web application (https://pangolin.cog-uk.io/) for lineage assignment. Both the PANGOLIN results and the ML tree (Figure 1c) indicated that these novel variants belonged to lineage B.1.36.27. This lineage had been predominant among local COVID-19 cases in Hong Kong since September 2020 and was rarely reported elsewhere, suggesting that the present variants had emerged locally.

While small deletion variants are uncommon in SARS-CoV-2, those with selective advantage can achieve rapid spread and predominance, as exemplified by the lineage B.1.1.7 variants. Large deletion variants are rarer, though major deletions around ORF7a, ORF7b and ORF8 had been reported [3,6–8]. Nonetheless, the presently described variant is significant in various ways. Firstly, it contained the largest genetic deletion that had occurred naturally in SARS-CoV-2 to date. The accessory proteins ORF6, ORF7a, ORF7b, and ORF8 are non-essential for viral replication and infection in vitro, but are suggested to have important functions like immune evasion [5,8–10]. While we refrain from extrapolating its clinical significance in this report, it is somewhat remarkable that the present variant achieved effective transmission and caused disease in at least 2 out of 3 cases.

Another feature of note is the duplication and transposition of the partial 5'-UTR in place of the deletion. According to the copy-choice model of genomic recombination in positive-sense single-stranded RNA viruses,

this genomic rearrangement could be achieved by the dissociation of a replication complex from the negative-sense genomic RNA template and re-association with a negative-sense subgenomic RNA (sgRNA) template for the nucleocapsid gene. However, the re-association would have to occur near the 3'-end of the sgRNA template without any homologous segment for anchoring. Such an occurrence should be very rare, since the probability of successful template-switching correlates with the local sequence similarity between the donor and acceptor templates [11]. It is uncertain whether the present genomic rearrangement was the result of an accidental one-time occurrence or an alternative mechanism [11,12].

The present discovery highlights the considerable evolutionary potential of SARS-CoV-2. In this case, the tolerance of a major and unusual genomic arrangement led to the formation of ORF6x, which involved recruitment of typically non-coding sequences into the open reading frame. These events greatly increased the genetic diversity of SARS-CoV-2, thereby accelerating the evolution of new potentially advantageous genotypes and compensating for the relatively low mutation rates in coronaviruses. On a more immediate note, caution is warranted when using diagnostic assays targeting only accessory genes or proteins such as Orf8 [4], given the risk of false-negative results from the sporadic emergence of major deletion variants.

NOTES

Conflict of Interest

All authors have no conflict of interest.

References

1.   Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol **2020**; 5:1403–1407.

2.   Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. Front Microbiol **2020**; 11:1800.

3.   Su YCF, Anderson DE, Young BE, et al. Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. mBio **2020**; 11:e01610-20.

4.   Wang X, Lam J-Y, Wong W-M, et al. Accurate diagnosis of COVID-19 by a novel immunogenic secreted SARS-CoV-2 Orf8 protein. mBio **2020**; 11:e02431-20.

5.   Miorin L, Kehrer T, Sanchez-Aparicio MT, et al. SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. Proc Natl Acad Sci **2020**; 117:28344–28354.

6.   Addetia A, Xie H, Roychoudhury P, et al. Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. J Clin Virol **2020**; 129:104523.

7.   Holland LA, Kaelin EA, Maqsood R, et al. An 81-nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (January to March 2020). J Virol **2020**; 94:e00711-20.

8.   Zinzula L. Lost in deletion: The enigmatic ORF8 protein of SARS-CoV-2. Biochem Biophys Res Commun **2020**; :S0006291X20319628.

9.   Xia H, Cao Z, Xie X, et al. Evasion of type I interferon by SARS-CoV-2. Cell Rep **2020**; 33:108234.

10.   Tan Y, Schneider T, Leong M, Aravind L, Zhang D. Novel immunoglobulin domain proteins provide insights into evolution and pathogenesis of SARS-CoV-2-related viruses. mBio **2020**; 11:e00760-20.

11.   Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? Nat Rev Microbiol **2011**; 9:617–626.

12.   Gallei A, Pankraz A, Thiel H-J, Becher P. RNA recombination in vivo in the absence of viral replication. J Virol **2004**; 78:6271–6281.
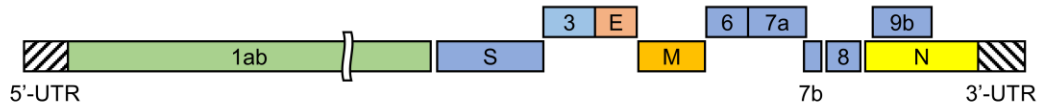
Figure legend

Figure 1. (a) Schematic diagram showing the differences in genomic organization between the reference SARS-CoV-2 genome and the present variant. (b) Sequence alignment of ORF6, ORF6x, and partial 5'-untranslated region (UTR), showing ORF6x as a fusion product of partial ORF6, partial 5'-UTR, and part of the non-coding region preceding the nucleocapsid (N) gene. The conserved transcription regulatory sequence, AACGAAC, is underlined. Standard IUPAC nucleic acid and amino acid one-letter notations are shown. (c) Maximum likelihood tree of SARS-CoV-2 genome sequences showing the phylogenetic position of the present variant relative to other local COVID-19 cases during the same period. Ultrafast bootstrap support values are shown in italics. PANGO lineages (version 2021-02-06) are shown in boxes. Samples 789, 906 and 914 correspond to patients 1, 2 and 3 respectively.
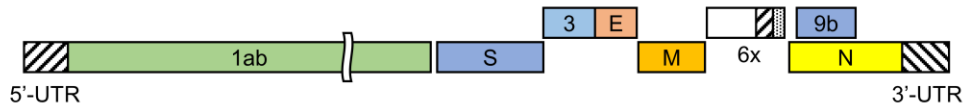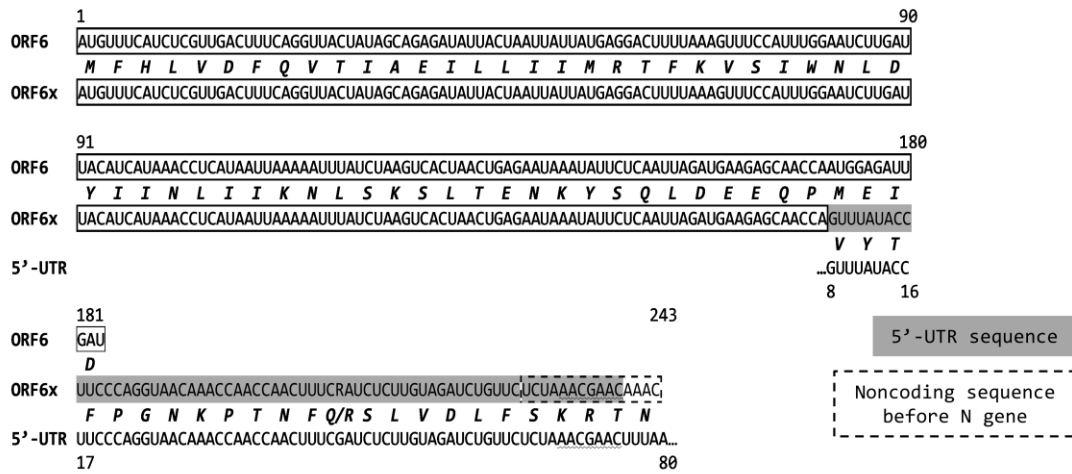
Figure 1

a

SARS-CoV-2 (Wuhan-Hu-1)

SARS-CoV-2 (HKCH_789)

b

c