

Natural transmission of bat-like SARS-CoV-2_{ΔPRRA} variants in COVID-19 patients

Yik Chun Wong^{1,2}, Siu Ying Lau^{2,3}, Kelvin Kai Wang To^{2,3,4,5}, Bobo Wing Yee Mok^{2,3}, Xin Li^{1,2}, Pui Wang^{2,3}, Shaofeng Deng^{2,3}, Kin Fai Woo^{1,2}, Zhenglong Du^{1,2}, Cun Li², Jie Zhou^{2,3}, Jasper Fuk Woo Chan^{2,3,4,5}, Kwok Yung Yuen^{2,3,4,5}, Honglin Chen^{2,3,4}, Zhiwei Chen^{1,2,3,4,*}

¹AIDS Institute, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China.

²Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China.

³State Key Laboratory for Emerging Infectious Diseases, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China.

⁴Carol Yu Centre for Infection, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China.

⁵Department of Clinical Microbiology and Infection, The University of Hong Kong-Shenzhen Hospital, Shenzhen, Guangdong Province, China.

*Corresponding author: Z Chen (zchenai@hku.hk). Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, L5-45, 21 Sassoon Road,

Pokfulam, Hong Kong SAR, People's Republic of China. Phone: 852-3917-9831; Fax: 852-3917-7805.

Summary:

Using a specific and highly sensitive digital PCR assay, we discovered the presence of SARS-CoV-2 viral variants carrying mutations upstream and at the S1/S2 cleavage site in COVID-19 patients. Importantly, Bat-like SARS-CoV-2_{ΔPRRA} variants naturally exist and remain transmissible in humans.

Accepted Manuscript

Abstract:*Background:*

SARS-CoV-2 contains the furin cleavage PRRA motif in the S1/S2 region, which enhances viral pathogenicity but is absent in closely related bat and pangolin coronaviruses. It remains unknown if bat-like coronaviral variants without PRRA (Δ PRRA) can establish natural infection in humans.

Methods:

Here, we developed a duplex digital PCR assay to examine Δ PRRA variants in Vero-E6-propagated isolates, human organoids, experimentally infected hamsters and COVID-19 patients.

Results:

We found that currently transmitting SARS-CoV-2 contained a quasispecies of wildtype, Δ PRRA variants and upstream variants that have mutations upstream the PRRA motif. Moreover, the Δ PRRA variants were readily detected despite at a low intra-host frequency in transmitted founder viruses in hamsters and in COVID-19 patients including acute cases and a family cluster with a prevalence rate of 52.9%.

Conclusions:

Our findings demonstrate that bat-like SARS-CoV-2 $_{\Delta$ PRRA not only naturally exists but remains transmissible in COVID-19 patients, which have significant implications to zoonotic origin and natural evolution of SARS-CoV-2.

Keywords:

COVID-19, SARS-CoV-2, viral variants, transmission, furin cleavage PRRA motif

Introduction:

A novel beta-coronavirus, now recognized as SARS-CoV-2, led to the global coronavirus disease 2019 (COVID-19) outbreak [1, 2]. The rapid spreading of SARS-CoV-2 is clearly due to aggressive person-to-person transmission with early evidence in hospital and family settings [3, 4]. Based on viral genome analysis, early studies have indicated that SARS-CoV-2 is similar to bat coronaviruses with 96% identity, but is relatively distant from SARS-CoV [5-8]. In spite of only 40% amino acid identity in the external subdomain of receptor binding domain (RBD), SARS-CoV-2 uses the same cellular receptor angiotensin-converting enzyme 2 (ACE2) as SARS-CoV to initiate infection [6, 9]. Since many bat coronaviruses use ACE2 as cellular receptor but have not easily caused human outbreaks [10, 11], other viral factors may contribute to more efficient zoonotic and person-to-person transmission besides ACE2 usage [7, 12].

One such viral factor has been associated with an insertion of the polybasic cleavage motif PRRA in the S1/S2 cleavage site of SARS-CoV-2 spike protein, which has not been found in either bat- or pangolin-derived coronaviruses [10, 11, 13-15]. Protease-mediated viral entry is one of the determinants of success in SARS-CoV infection [16]. Similarly, furin and serine protease TMPRSS2 are essential for SARS-CoV-2 infection of human target cells [9, 17]. Interestingly, our team has recently discovered that a series of variants, which contain animal-like PRRA deletions (Δ PRRA) in the S1/S2 cleavage region through plaque purification of SARS-CoV-2 in Vero-E6 cells [18]. Since conventional methods failed to detect these variants in clinical samples [18], we sought to develop a highly sensitive digital PCR assay to investigate this viral variant among experimentally infected animals and naturally infected COVID-19 patients. Although several studies have indicated that SARS-CoV-2 is a naturally occurring virus through zoonotic transmission, we aimed to investigate the critical missing link whether or not SARS-CoV-2 with bat- or pangolin-like PRRA deletion (SARS-CoV-2 $_{\Delta$ PRRA) can be found in COVID-19 patients.

Results

Design and evaluation of a digital PCR assay for the detection of S1/S2 cleavage site variants

To detect SARS-CoV-2_{ΔPRRA} variants amongst COVID19 patients, we established a duplex digital PCR assay that included the amplification of a 162bp sequence spanning the viral S1/S2 cleavage site and the detection of this sequence with two fluorescent oligonucleotide probes, namely PRRA and upstream probes (Figure 1A). The PRRA probe specifically targeted the PRRA motif within the S1/S2 cleavage region. This probe would bind to amplicons generated from wildtype SARS-CoV-2 sequence, but not from ΔPRRA variants, including Del-Mut-1 and Del-Mut-2 isolates, which carry deletion in the S1/S2 site [18-20]. For viral variants containing deletion directly at the 5'-flank of the PRRA site, such as Del-Mut-3 [18-20], the PRRA probe would still detect their sequences. The probe upstream the PRRA site acted as a reference probe for the identification of SARS-CoV-2 sequence. In general, wildtype viral sequences should be detected by both the upstream and PRRA probes while ΔPRRA variants should be detected by the upstream probe only, but not the PRRA probe.

We firstly validated the specificity of the digital PCR assay using cDNAs generated from plaque-purified wildtype and Del-Mut-1 isolates. Our assay successfully identified the presence of 100% ΔPRRA in the Del-Mut-1 isolate, without detection of wildtype sequence (Figure 1B). In comparison, the majority of viral cDNAs from the wildtype isolate contained the wildtype PRRA cleavage site. Unexpectedly, our assay also detected a low frequency of ΔPRRA variants (1.4% of total detected viral copies) within the wildtype isolate, which had undergone plaque purification in Vero-E6 cells (Figure 1B). Furthermore, another unexpected population of viral sequences from the wildtype isolate was readily detected by the PRRA probe, but not the upstream probe. These findings suggested that the wildtype isolate from Vero-E6 cells actually contained a mixture of wildtype, ΔPRRA and upstream variants.

We next examined the sensitivity of our digital PCR assay using serially diluted cDNA samples generated from the wildtype isolate. Focusing on the upstream probe, our assay consistently produced measurable signals from the diluted cDNA samples equivalent to 0.15 plaque forming unit (Figure 1C). We then evaluated the ability of the digital PCR assay to distinguish Δ PRRA variants from wildtype sequences using cDNA mixtures containing fixed amount of wildtype viral cDNAs with serially diluted Del-Mut-1 cDNAs (Figure 1D). Our assay readily detected the presence of a higher frequency of Δ PRRA mutants (averaged $2.96\% \pm 0.09\%$ SD) in the cDNA sample containing 1:0.01 ratio of wildtype to Del-Mut-1 cDNAs, as compared to the wildtype only sample (averaged $1.88\% \pm 0.46\%$ SD). Overall, our assay is sensitive for detecting the presence of low abundance Δ PRRA mutants.

The presence of S1/S2 cleavage site variants in human organoids and hamsters infected with wildtype and bat-like SARS-CoV-2 _{Δ PRRA} isolates

Our team recently demonstrated that SARS-CoV-2 replicates in human intestinal organoids [21]. We sought to determine if wildtype SARS-CoV-2 and SARS-CoV-2 _{Δ PRRA} could establish infection in human intestinal organoids using plaque-purified wildtype or Del-Mut-1 isolates. The presence of S1/S2 cleavage site variants were tested from culture supernatants at 48 hours post-infection. Although both upstream and Δ PRRA variants could be detected from the wildtype viral inoculum, as shown in Figure 1B, the presence of these variants was suppressed to minimal level after propagation in the intestinal organoids (Figure 2A). A single copy of Δ PRRA variant was detected from one of the triplicate organoid samples. In organoids infected with Del-Mut-1, only Δ PRRA variants were detected (Figure 2A). These findings suggest that the viral genomic region of the S1/S2 cleavage site remains stable after propagation in human intestinal organoids. Furthermore, due to the low Δ PRRA frequency in organoids infected with the wildtype isolate, Δ PRRA variants have no

growth advantage to outcompete wildtype virions in human organoids, as opposed in Vero-E6 cells [18].

To directly test for the transmissibility of SARS-CoV-2_{ΔPRRA} *in vivo*, the digital PCR assay was conducted with various airway tissue samples from hamsters at 4 days post-infection with either plaque-purified wildtype or Del-Mut-1 isolates (Figure 2B). There were no detectable levels of both upstream and ΔPRRA variants within the nasal turbinates and lung tissues from hamsters infected with the wildtype viral isolate. Both upstream and ΔPRRA variants, however, could be identified within the tracheal tissues. The frequencies of these two variants within tracheal tissues were lower than those in the viral inoculum. In hamsters infected with Del-Mut-1, only ΔPRRA variants could be detected from all airway tissues examined, with no evidence of wildtype virus (Figure 2B). The finding of a mixture of wildtype, ΔPRRA and upstream variants in tracheal tissues of both hamsters challenged with wildtype SARS-CoV-2 isolate at 4 days post-infection indicates a lack of genetic bottleneck for single virus mucosal transmission. The absence of ΔPRRA and upstream variants in nasal turbinates and lungs suggested that they probably have a reduced *in vivo* fitness when compared to the wildtype virus. Moreover, the lack of wildtype sequences in hamsters infected with Del-Mut-1 demonstrated that the ΔPRRA-to-wildtype reversion did not happen during the acute 4-day infection period.

Identification of ΔPRRA and upstream variants in COVID19 patients by digital PCR

We next determined the prevalence and frequencies of S1/S2 cleavage site viral variants in clinical samples collected from COVID-19 patients. In the digital PCR assay, 51 patients' samples showed positive signals for the presence of viral sequences. The wildtype viral sequence was the most abundant in all samples (Figure 3A-B), representing an averaged 98.6% (±1.61% SD) of detected viral molecules in individual samples (Figure 3C). Although 52.9% of the positive clinical

samples contained the Δ PRRA variant molecules, this mutant population represented a very minor viral population (averaged $0.33\% \pm 1.17\%$ SD) in intra-host level (Figure 3D). In comparison, we detected a higher prevalence of the upstream variants (82.4%) from the clinical samples, and this variant population comprised $1.12\% (\pm 1.21\%$ SD) of all detected viral copies in individual samples. In clinical samples containing either or both variants, the majority of them (93.2%) possessed a higher frequency of upstream variants than Δ PRRA mutants (Figure 3D). Two of the clinical samples contained only Δ PRRA variants. In the clinical samples containing both variants, the ratio of upstream/ Δ PRRA variants was all ≥ 1 (Figure 3E). Moreover, the higher frequency of upstream variants in patients was not associated with the severity of the COVID-19 symptoms because mild and severe patients showed similar patterns (Figure 3F). Similarly, on average, a higher frequency of upstream variants over Δ PRRA variants was detected in all four sample types (Figure 3G).

To further understand the person-to-person transmission of SARS-CoV-2 variants, we focused on clinical samples from patients derived from a COVID-19 family cluster [n=11; 22]. Because of the limited specimen availability, our cohort contained samples from 4 patients of this infection cluster. Consistently, 3 of the patients demonstrated the presence of a higher frequency of upstream variants than Δ PRRA variants (Suppl. Figure 1). Although the remaining sample showed the presence of Δ PRRA variants without any upstream variants, it should be taken into account that only a low viral copy level was detected from this sample.

Overall, while three types of viruses have been transmitted in humans, the wildtype virus has greater advantages than upstream and Δ PRRA mutants for infection at both inter- and intra-host levels.

Discussion

We here demonstrated that bat-like SARS-CoV-2_{ΔPRRA} and upstream variants exist naturally and are currently transmitting in COVID-19 patients, as revealed by our duplex digital PCR assay. Although these variants only consisted of a very small fraction in the wildtype viral challenge stock, they were consistently detected in intranasally inoculated hamsters. ΔPRRA and upstream variants were also readily detected among acute patients, including a family cluster. These results indicate that person-to-person mucosal transmission of SARS-CoV-2 is unlikely a genetic bottleneck allowing infection only by single transmitted founder viruses, but rather by viral quasispecies. ΔPRRA is unlikely an overwhelming restriction factor for human transmission by zoonotic bat-like SARS-CoV-2_{ΔPRRA}, which may support zoonotic origin and evolution of SARS-CoV-2 in humans. It is, therefore, necessary to implement stringent measures to prevent human infection by animal SARS-CoV-2_{ΔPRRA} variants including handling field and laboratory specimens derived from wild bats and pangolins.

Due to mutations and ability to undergo genomic recombination [23], genetic variations of different coronaviruses, including SARS-CoV [24, 25], MERS-CoV [26, 27], and other animal coronaviruses [28], are readily identified at both population and intra-host levels. SARS-CoV-2 variants have also been reported among COVID-19 patients, mainly by next generation sequencing methods [2, 19, 29, 30]. In particular, viral variants carrying ΔPRRA in SARS-CoV-2 isolates passaged *in vitro* in Vero-E6 cells have been recently reported by our team [18] and others [19, 20]. Using the duplex digital PCR assay, we detected ΔPRRA and upstream variants from clinical isolate propagated in Vero-E6 cells, suggesting that this genomic region is instable and dispensable in SARS-CoV-2 during viral replication in Vero-E6 cells. An initial minor viral variant population carrying ΔPRRA mutations would be generated and selected for, becoming the dominant strain after further propagation in Vero-E6 cells [18, 19]. The spike protein of coronaviruses needs to be activated via

sequential proteolysis at the S1/S2 and S2' sites for viral entry into target cells [31]. For MERS-CoV and SARS-CoV-2, spike protein activation is achieved either via initial cleavage at the S1/S2 site by furin, followed by S2' site cleavage by TMPRSS2 on cell surfaces or via proteolytic processing by cathepsin B/L after endocytosis [9, 17, 32]. Vero cells express low level of furin [33] but high level of cathepsin B/L, and they support furin-independent MERS-CoV and SARS-CoV-2 viral entry [17, 32]. This might create a selection pressure against the furin cleavage site. Moreover, Vero cells are interferon defective [34]. Δ PRRA variants might have enhanced fitness in cell lines with suboptimal innate immune responses to SARS-CoV-2 infection. This is supported by our previous findings which demonstrated that the Δ PRRA Del-Mut-1 isolate grew well in Vero-E6, but replicated poorly in interferon competent cells [18]. Interestingly, another study showed that propagation of a SARS-CoV-2 isolate in another Vero-derived cell line, Vero-76, did not lead to PRRA site mutation [30]. Differences in cell lines used might influence the selective pressures for adaptive mutations.

Acquiring the PRRA insertion in SARS-CoV-2 might enhance its host adaptation with increased growth capacity and pathogenicity. One hypothesis is that after zoonotic transmission, SARS-CoV-2 with an intact PRRA motif at the S1/S2 cleavage site has probably selected from bat- or pangolin-derived SARS-CoV-2 _{Δ PRRA} to become the most prevalent viral type in both inter- and intra-host levels in patients. Although Δ PRRA variants could be detected in about 52.9% of the patients in this study, these variants were only present in very low frequency in individual patients. Experimentally, in the hamsters challenged with the wildtype SARS-CoV-2 strain, a lower frequency of Δ PRRA variants were always detected in the airway tissues than from the challenge inoculum. Thus, in contrast to a better replicative fitness in Vero-E6 cells *in vitro*, Δ PRRA variants were likely selected against *in vivo* during the natural course of viral evolution.

We identified a new type of viral variants containing mutations at the 5' upstream region of the S1/S2 site. These variants were more prevalent than the Δ PRRA variants in both intra- and inter-host levels in our COVID-19 patient cohort. It is common for the fixation of coronavirus variants in a

global epidemiological scale. Founder effect might be involved in selecting orf8 deletion during the early stage of SARS-CoV human transmission [35]. Mutations in viral lineages specific to certain geological regions have also been documented for SARS-CoV and MERs-CoV infections [25, 36, 37]. Using whole genome phylogenetic analysis, a recent report identified three SARS-CoV-2 variant lineages clustered in distinct geological regions and suggested a possibility of founder events or selective pressures for viral variants specific to certain environmental and/or host backgrounds [2]. In our patient cohort, SARS-CoV-2 with intact PRRA motif remained the dominant strain in both intra- and inter-host levels. Wildtype virions have a better selection advantage *in vivo* than S1/S2 cleavage site variants, arguing against the possibility for the fixation of S1/S2 cleavage site variants in a population level.

There are some limitations in this study. We could not determine whether or not the event of acquiring the PRRA motif took place in humans. Interestingly, a partial insertion of a similar polybasic motif as SARS-CoV-2 has been recently reported in a newly identified bat coronavirus strain at the S1/S2 cleavage site [7]. This type of viral strains might also serve as ancestral viruses for SARS-CoV-2. The analytical power of our digital PCR assay would be enhanced by coupling with sequencing analysis. It should also be noted that variants with a small deletion at the direct 5'-upstream of the PRRA site, the same deletion as of Del-Mut-3 [18], have been reported in another patient cohort, with a prevalence rate of 4.4% [19]. Our assay does not distinguish this variant from wildtype sequence. Lastly, it remains to be investigated if bat-like SARS-CoV-2_{ΔPRRA} would modulate host immune responses for different clinical outcomes.

Materials and Methods

Virus, cells and organoids

The SARS-CoV-2 wildtype and Del-Mut-1 viral isolates were plaque purified and passaged in Vero-E6 cells obtained from ATCC, as described previously [18]. Human intestinal organoids derived from normal human small intestinal tissues, collected according to the ethical approval by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW13-364), were generated and maintained, as previously described [21]. Organoids were infected with SARS-CoV-2 wildtype or Del-Mut-1 viral isolates using a similar protocol recently published [21]. All viral culture experiments were performed in biosafety level-3 facilities.

Specimen collection from COVID-19 patients

This study included clinical specimens, including saliva, nasopharyngeal secretions, throat swabs, or endotracheal aspirate, from 51 COVID-19 patients (Table 1), according to the ethical approval by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW13-372). These patients were confirmed by the Public Health Laboratory Services Branch of the Centre for Health Protection in Hong Kong using SARS-CoV-2-specific reverse-transcription PCR [38]. Patients who required supplemental oxygen, admission to the intensive care unit, or death were defined as severe patients.

Nucleic acid isolation and cDNA synthesis

cDNA generated from plaque purified wildtype and Del-Mut-1 viral isolates were obtained from our previous study [18]. RNA from infected organoids and tissues from infected hamsters were extracted using MiniBEST Viral RNA/DNA Extraction Kit (TaKaRa) and QIAamp Viral RNA Mini Kit (Qiagen), respectively. Total nucleic acids were extracted from clinical samples using NucliSENS easyMAG kit (BioMerieux) [38]. cDNA was then generated from the nucleic acid samples using PrimeScript 1st strand cDNA Synthesis Kit with random hexamer oligos (TaKaRa).

Digital PCR

Primers, 5'-GCAGGCTGTTTAATAGGGGC-3' and 5'-ACCAAGTGACATAGTGTAGGCA-3', and TaqMan fluorescent probes, 5'-6-FAM-ATTCTCCTC-ZEN-GGCGGGCACGT-Iowa Black FQ-3' and 5'-HEX-CCTGCACCA-ZEN-ATGGGTATGTCACACTC-Iowa Black FQ-3' (Figure 1A), were synthesised by ThermoFisher and Integrated DNA Technologies, respectively. Digital PCR was performed using the QuantStudio 3D Digital PCR System (ThermoFisher). Reactions were carried out in a 14.5µl reaction mix with 1X QuantStudio 3D Digital PCR Master Mix, 1.24µM of each primer, 138nM of each fluorescent probe, and 1-3µl of cDNA. Positive (cDNA from wildtype and Del-Mut-1 isolates) and negative (H₂O alone) controls were included in each run. Reaction mixes were loaded into QuantStudio 3D Digital PCR 20K chip using a QuantStudio 3D Digital PCR Chip loader. PCR was then conducted in the following cycling conditions: 10 min denaturation at 96°C, 39 cycles of 2 min at 57°C for annealing and elongation, and 30 sec at 98°C for denaturation, and final elongation at 57°C for 2 min. Fluorescence signals from individual microwells on the chips were determined using the QuantStudio 3D Digital PCR instrument and the data was analysed using the QuantStudio3D AnalysisSuite Cloud Software. Samples with more than 25 microwells showing either or both FAM and HEX signals after digital PCR were considered positive and were included for analysis.

Statistical analysis

Statistical analyses were performed using SPSS version 26 (IBM) or Prism version 7 (GraphPad). For patient characteristics shown in Table 1, categorical variables were compared using the Fisher's exact test while continuous variables were compared using the Mann-Whitney test. Comparisons of frequencies of viral variants from individual samples were performed with Wilcoxon matched-pairs signed rank test. A p -value less than 0.05 was considered statistically significant.

Accepted Manuscript

NOTES:**Acknowledgements:**

Y.C.W. and Z.C. designed and supervised the study. S.Y.L., B.W.Y.M, X.L., P.W., S.D., K.F.W., Z.D. C.L. and J.Z. performed the experiments. Y.C.W, X.L., Z.D. and Z.C. analysed the data. K.K.W.T., J.F.W.C. and K.Y.Y. provided the clinical specimens. W.Y.C., K.K.W.T., J.Z., H.C. and Z.C. wrote the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Disclaimer:

The funding sources had no role in study design, data collection, analysis, interpretation, or writing of the report. The authors declare no competing interests.

Funding:

This work was supported by Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Diseases and Research Capability on Antimicrobial Resistance for Department of Health of the Hong Kong, the Theme-Based Research Scheme (T11-706/18-N) of the Hong Kong Research Grants Council, University Development Fund, and Li Ka Shing Faculty of Medicine Matching Fund from HKU to AIDS Institute.

Potential conflicts:

Dr. Yik Chun Wong reports grants from Theme-Based Research Scheme of Hong Kong Research Grants Council during the conduct of the study. Dr. Kelvin Kai Wang To reports grants from Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Diseases and Research Capability on Antimicrobial Resistance for Department of Health of the Hong Kong during

the conduct of the Study. Dr. Jasper Fuk Woo Chan reports grants from Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Diseases and Research Capability on Antimicrobial Resistance for Department of Health of the Hong Kong during the conduct of the study. Dr. Kwok Yung Yuen reports grants from Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Diseases and Research Capability on Antimicrobial Resistance for Department of Health of the Hong Kong during the conduct of the study. Dr. Honglin Chen reports grants from Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Diseases and Research Capability on Antimicrobial Resistance for Department of Health of the Hong Kong during the conduct of the study. Dr. Zhiwei Chen reports grants from Theme-Based Research Scheme of Hong Kong Research Grants Council during the conduct of the study. All other authors have no conflicts

Accepted Manuscript

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: 727-33.
2. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences* **2020**; 117(17): 9241-3.
3. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **2020**; 395: 514-23.
4. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients With 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* **2020**; 323: 1061-9.
5. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**; 579: 270-3.
6. Chan JF, Kok KH, Zhu Z, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* **2020**; 9(1): 221-36.
7. Zhou H, Chen X, Hu T, et al. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Current Biology* **2020**; 30(11): 2196-203.e3.
8. Wu A, Peng Y, Huang B, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* **2020**; 27: 325-8.
9. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **2020**; 181(2): 271-80.e8.
10. Menachery VD, Yount BL, Debbink K, et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Medicine* **2015**; 21(12): 1508-13.
11. Ge X-Y, Li J-L, Yang X-L, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **2013**; 503(7477): 535-8.
12. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**; 395: 497-506.
13. Lam TT-Y, Shum MH-H, Zhu H-C, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* **2020**; <https://doi.org/10.1038/s41586-020-2169-0>.
14. Xiao K, Zhai J, Feng Y, et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **2020**; <https://doi.org/10.1038/s41586-020-2313-x>.
15. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Current Biology* **2020**; 30(7): 1346-51.e2.
16. Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences* **2009**; 106(14): 5871-6.
17. Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Molecular Cell* **2020**; 78(4): 779-84.e5.
18. Lau S-Y, Wang P, Mok BW-Y, et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerging Microbes & Infections* **2020**; 9: 837-42.

19. Liu Z, Zheng H, Yuan R, et al. Identification of a common deletion in the spike protein of SARS-CoV-2. *bioRxiv* **2020**; Preprint at <https://www.biorxiv.org/content/10.1101/2020.03.31.015941v1>.
20. Ogando NS, Dalebout TJ, Zevenhoven-Dobbe JC, et al. SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *bioRxiv* **2020**; Preprint at <https://www.biorxiv.org/content/10.1101/2020.04.20.049924v1>.
21. Zhou J, Li C, Liu X, et al. Infection of bat and human intestinal organoids by SARS-CoV-2. *Nature Medicine* **2020**; <https://doi.org/10.1038/s41591-020-0912-6>.
22. Leung KS-S, Ng TT-L, Wu AK-L, et al. A territory-wide study of early COVID-19 outbreak in Hong Kong community: A clinical, epidemiological and phylogenomic investigation. *medRxiv* **2020**; Preprint at <https://doi.org/10.1101/2020.03.30.20045740>.
23. Su S, Wong G, Shi W, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends in Microbiology* **2016**; 24(6): 490-502.
24. Xu D, Zhang Z, Wang F-S. SARS-associated coronavirus quasispecies in individual patients. *New England Journal of Medicine* **2004**; 350(13): 1366-7.
25. Ruan Y, Wei CL, Ling AE, et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *The Lancet* **2003**; 361(9371): 1779-85.
26. Cotten M, Lam TT, Watson SJ, et al. Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. *Emerg Infect Dis* **2013**; 19(5): 736-42B.
27. Briese T, Mishra N, Jain K, et al. Middle East respiratory syndrome coronavirus quasispecies that include homologues of human isolates revealed through whole-genome analysis and virus cultured from dromedary camels in Saudi Arabia. *mBio* **2014**; 5(3): e01146-14.
28. Woo PCY, Lau SKP, Tsang C-C, et al. Coronavirus HKU15 in respiratory tract of pigs and first discovery of coronavirus quasispecies in 5'-untranslated region. *Emerging Microbes & Infections* **2017**; 6(1): 1-7.
29. Karamitros T, Papadopoulou G, Bousali M, Mexias A, Tsiodras S, Mentis A. SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies. *bioRxiv* **2020**; Preprint at <https://www.biorxiv.org/content/10.1101/2020.03.27.009480v1>.
30. Moreno GK, Braun KM, Halfmann PJ, et al. Limited SARS-CoV-2 diversity within hosts and following passage in cell culture. *bioRxiv* **2020**; Preprint at <https://www.biorxiv.org/content/10.1101/2020.04.20.051011v1>.
31. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology* **2016**; 3(1): 237-61.
32. Park J-E, Li K, Barlan A, et al. Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. *Proceedings of the National Academy of Sciences* **2016**; 113(43): 12262-7.
33. Mukherjee S, Sirohi D, Dowd KA, et al. Enhancing dengue virus maturation using a stable furin over-expressing cell line. *Virology* **2016**; 497: 33-40.
34. Emeny JM, Morgan MJ. Regulation of the interferon system: evidence that Vero cells have a genetic defect in interferon production. *Journal of General Virology* **1979**; 43(1): 247-52.
35. Muth D, Corman VM, Roth H, et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Scientific Reports* **2018**; 8(1): 15177.

36. Kim Y, Cheon S, Min C-K, et al. Spread of mutant Middle East respiratory syndrome coronavirus with reduced affinity to human CD26 during the South Korean outbreak. *mBio* **2016**; 7(2): e00019-16.
37. Mart ML, Raj VS, Mah'd S, et al. Deletion variants of Middle East respiratory syndrome coronavirus from humans, Jordan, 2015. *Emerging Infectious Disease journal* **2016**; 22(4): 716.
38. To KK-W, Tsang OT-Y, Yip CC-Y, et al. Consistent detection of 2019 novel coronavirus in saliva. *Clinical Infectious Diseases* **2020**; ciaa149: <https://doi.org/10.1093/cid/ciaa449>.

Accepted Manuscript

Figure 1: A duplex digital PCR assay for sensitive detection of SARS-CoV-2 S1/S2 cleavage site viral variants. (A) Design of primers and fluorescent probes used in the digital PCR assay. Multiple alignment of the nucleotide sequences spanning the S1/S2 cleavage site of the SARS-CoV-2 wildtype virus, three recently identified SARS-CoV-2 S1/S2 deletion mutant isolates (Del-Mut-1, Del-Mut-2, and Del-Mut-3 [18-20]), bat coronavirus RATG13 [5] , pangolin coronavirus GX/P4L [13], and SARS-CoV Urbani strain (GenBank accession no. AY278741). Digital PCR primers and probes were shown underneath the alignment. (B-D) To validate the digital PCR assay, cDNA was firstly generated from viral RNA from wildtype (WT) and Del-Mut-1 viral isolates plaque-purified from Vero-E6 cells and were then subjected to the digital PCR assay for the detection of S1/S2 cleavage site variants. (B) Representative plots showing the fluorescence signals of the HEX-tagged upstream probe (x-axis) and FAM-tagged PRRA probe (y-axis) detected from individual wells of digital PCR chips loaded with cDNA generated from samples indicated. (C) Sensitivity of the upstream probe to detect SARS-CoV-2 sequence in serial diluted cDNA samples from the wildtype isolate. (D) Frequencies of PRRA variants, upstream variants, and WT sequences in all detected viral copies from cDNA generated from WT or Del-Mut-1 isolates, and various mixing ratios of their cDNA.

Figure 2: The presence of S1/S2 cleavage site variants in human organoids and hamsters infected with wildtype and bat-like SARS-CoV-2_{ΔPRRA} isolates. Frequencies of PRRA variants, upstream variants, and WT sequences in (A) infected small intestinal organoids ($n=3$) at 48 hours post-infection, and in (B) nasal, tracheal, and lung tissues collected from hamsters ($n=2$) at 4 days post-infection with SARS-CoV-2 wildtype (WT) or Del-Mut-1 isolates. Data shown as means + ranges.

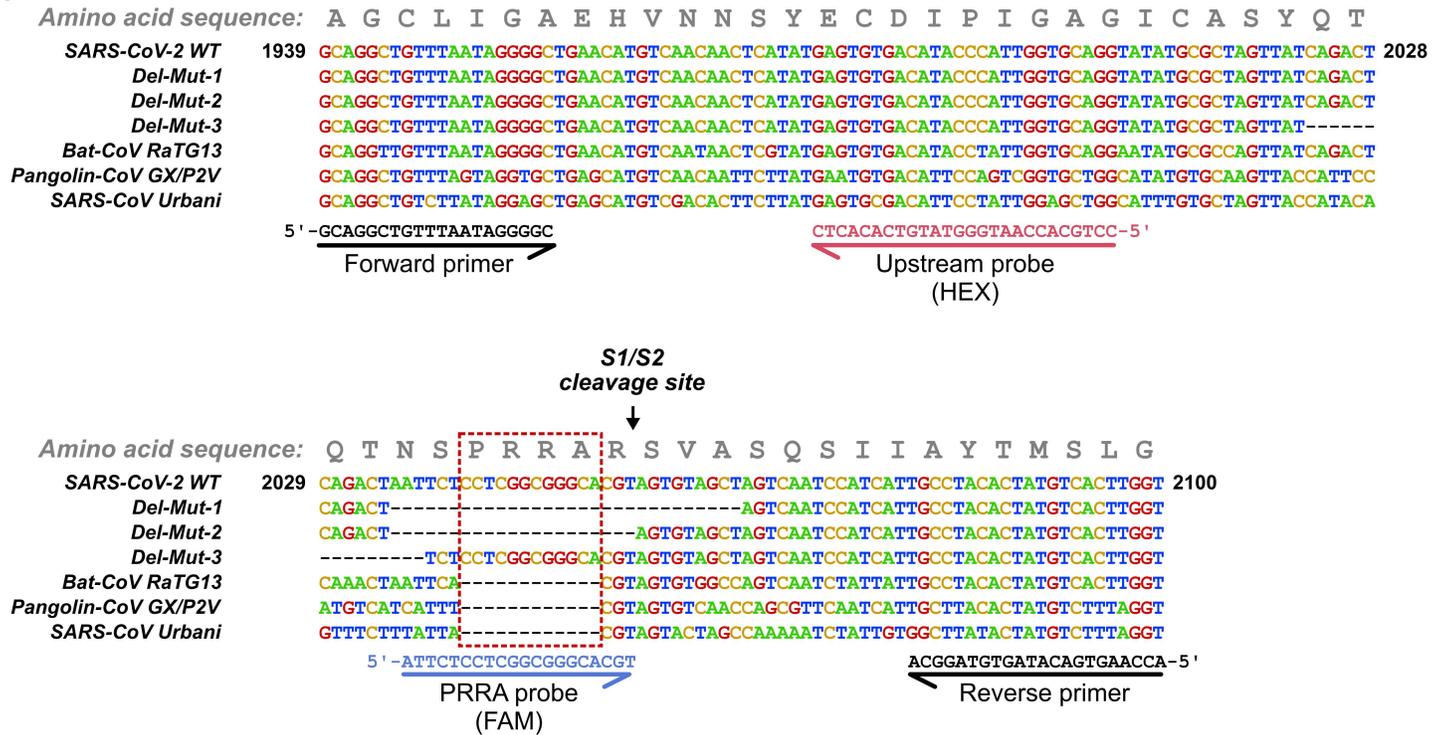
Figure 3: Identification of S1/S2 cleavage site Δ PRRA and upstream variants in COVID19 patients by digital PCR. Clinical specimens were subjected to the digital PCR assay to determine the presence of S1/S2 cleavage site variants. (A) The left pie chart illustrates the prevalence of wildtype, upstream variants, and/or Δ PRRA variants within the clinical specimens that showed positive signals in the digital PCR assay. The pie chart on the right shows the frequency of clinical samples with both upstream and Δ PRRA variants that possessed a higher level of upstream variants than Δ PRRA variants. (B) A representative plot showing the digital PCR fluorescence signals from a cDNA sample of a COVID-19 patient. (C) The frequency of wildtype viral sequences in all detected viral copies in individual specimens. Mean \pm ranges is shown. (D) Comparison of frequencies of the upstream and Δ PRRA variants from clinical specimens carrying either or both variants. (E) The ratio of the frequencies of upstream to Δ PRRA variants in clinical samples carrying both mutants. (F) Frequencies of upstream and Δ PRRA variants in mild and severe COVID19 patients. (G) Frequencies of upstream and Δ PRRA variants in different sample types. Statistical differences in *D*, *F*, and *G* were determined using Wilcoxon matched-pairs signed rank test.

Table 1: Characteristics of the COVID-19 patients, by disease severity

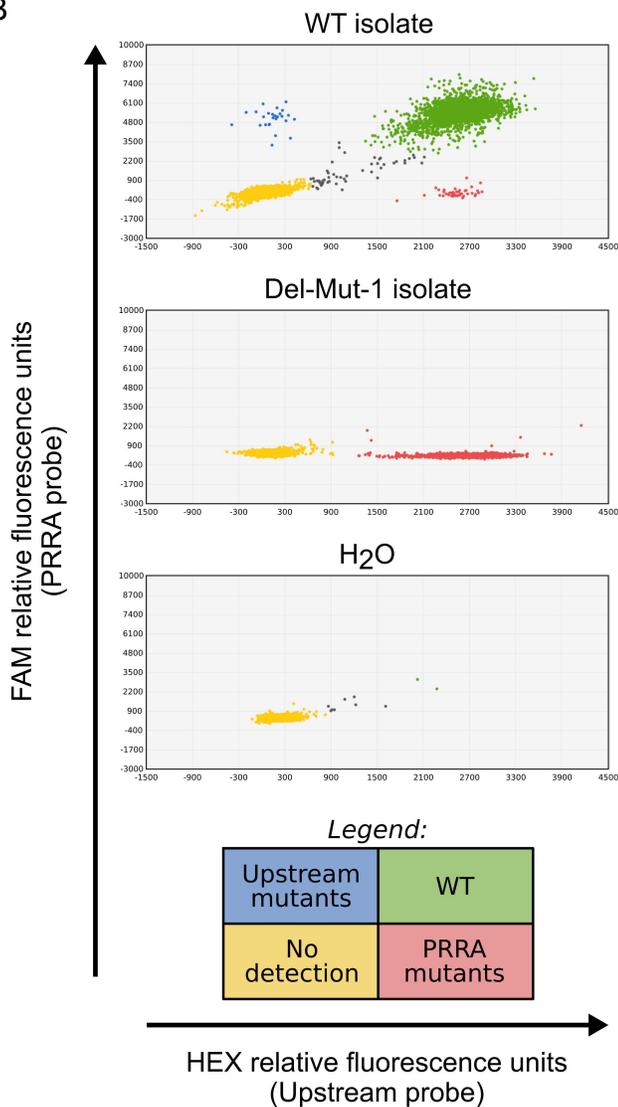
Characteristic	All	Mild patients	Severe patients	P value
n (%)	51	38 (74.5%)	13 (25.5%)	
Age, median (range), yr	51 (18-91)	51 (18-91)	64 (39-75)	0.0009 ^a
Sex				0.52 ^b
Female, n (%)	25 (49.0%)	20 (80%)	5 (20%)	
Male, n (%)	26 (51.0%)	18 (69.2%)	8 (30.8%)	
Sample type				0.034 ^b
Nasopharyngeal secretions, n (%)	26 (51.0%)	22 (57.9%)	4 (30.8%)	
Nasopharyngeal secretions + Throat swabs, n (%)	4 (7.8%)	2 (5.3%)	2 (15.4%)	
Saliva, n (%)	19 (37.3%)	14 (36.8%)	5 (38.5%)	
Endotracheal aspirates, n (%)	2 (3.9%)	0 (0%)	2 (15.4%)	

Note: For statistical analyses, ^aMann-Whitney U test and ^bFisher's exact test were performed accordingly.

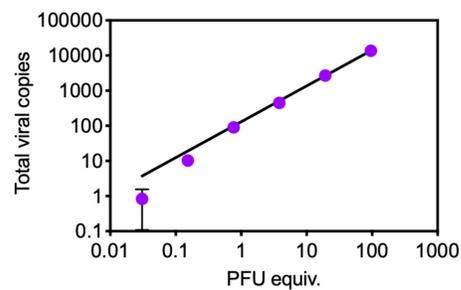
A



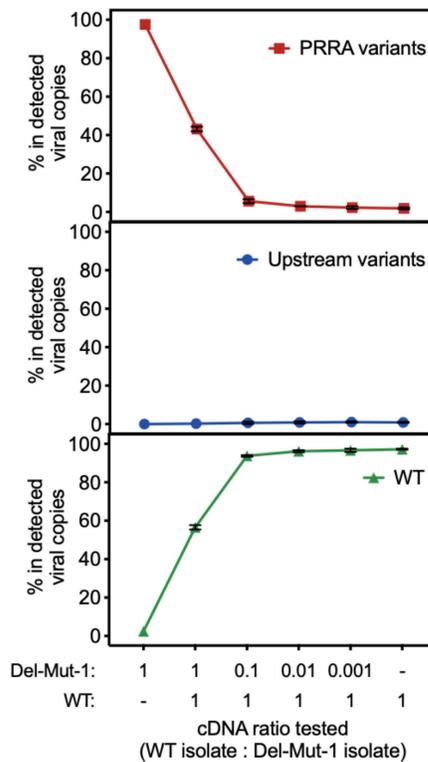
B



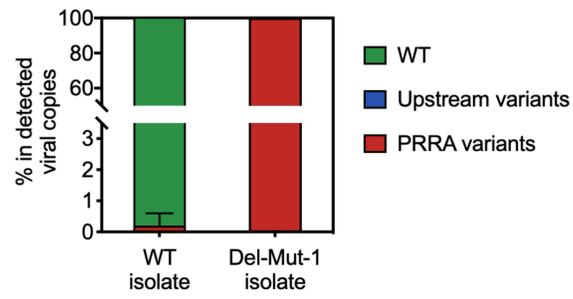
C



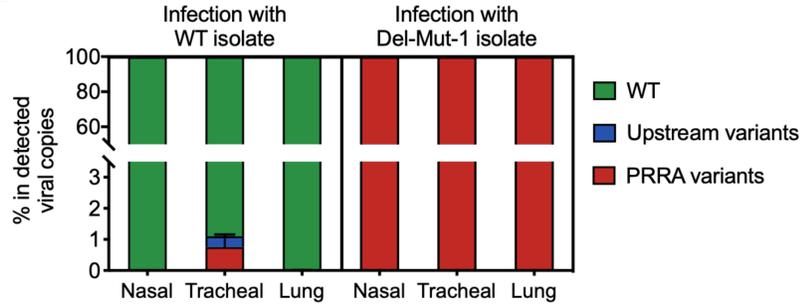
D



A



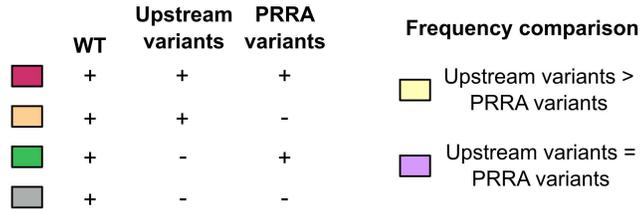
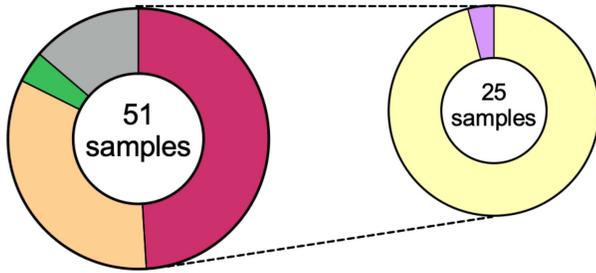
B



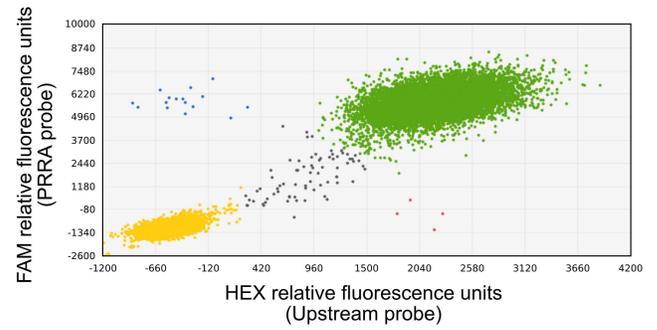
A

All clinical samples with positive digital PCR signals

Samples with both variants

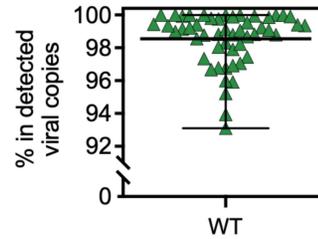


B



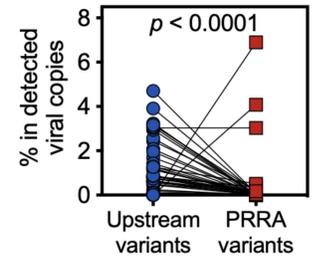
C

All clinical samples with positive digital PCR signals

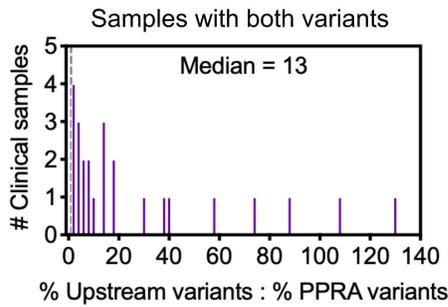


D

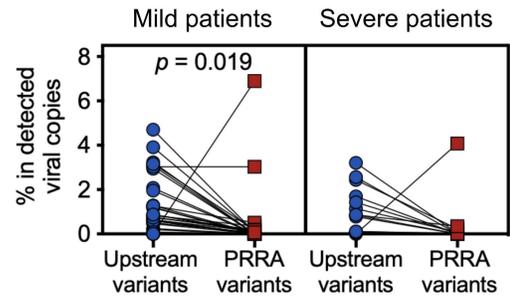
Samples with either or both variants



E



F



G

