AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE
Ecological and Evolutionary Science

Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2

Yvonne C. F. Su,[a] Danielle E. Anderson,[a] Barnaby E. Young,[b,c,d] Martin Linster,[a] Feng Zhu,[a] Jayanthi Jayakumar,[a] Yan Zhuang,[a] Shirin Kalimuddin,[a,e] Jenny G. H. Low,[a,e] Chee Wah Tan,[a] Wan Ni Chia,[a] Tze Minn Mak,[b] Sophie Octavia,[b] Jean-Marc Chavatte,[b] Raphael T. C. Lee,[f] Surinder Pada,[g] Seow Yen Tan,[h] Louisa Sun,[i] Gabriel Z. Yan,[j] Sebastian Maurer-Stroh,[f,k] Ian H. Mendenhall,[a,m] Yee-Sin Leo,[b,c,d,l,n] David Chien Lye,[b,c,d,n] Lin-Fa Wang,[a,m,o] Gavin J. D. Smith[a,m,o]

[a]Programme in Emerging Infectious Diseases, Duke-NUS Medical School, Singapore
[b]National Centre for Infectious Diseases, Singapore
[c]Tan Tock Seng Hospital, Singapore
[d]Lee Kong Chian School of Medicine, Singapore
[e]Singapore General Hospital, Singapore
[f]Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore
[g]Ng Teng Fong General Hospital, Singapore
[h]Changi General Hospital, Singapore
[i]Alexandra Hospital, Singapore
[j]National University Hospital, Singapore
[k]Department of Biological Sciences, National University of Singapore, Singapore
[l]Saw Swee Hock School of Public Health, National University of Singapore, Singapore
[m]SingHealth Duke-NUS Global Health Institute, Singapore
[n]Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[o]Duke Global Health Institute, Duke University, North Carolina, USA

Yvonne C. F. Su and Danielle E. Anderson contributed equally to this article. Author order was determined by length of time actively engaged on the project.

Barnaby E. Young and Martin Linster contributed equally to this article. Author order was determined by discussion among the coauthors.

ABSTRACT To date, limited genetic changes in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome have been described. Here, we report a 382-nucleotide (nt) deletion in SARS-CoV-2 that truncates open reading frame 7b (ORF7b) and ORF8, removing the ORF8 transcription regulatory sequence (TRS) and eliminating ORF8 transcription. The earliest 382-nt deletion variant was detected in Singapore on 29 January 2020, with the deletion viruses circulating in the country and accounting for 23.6% (45/191) of SARS-CoV-2 samples screened in this study. SARS-CoV-2 with the same deletion has since been detected in Taiwan, and other ORF7b/8 deletions of various lengths, ranging from 62 nt to 345 nt, have been observed in other geographic locations, including Australia, Bangladesh, and Spain. Mutations or deletions in ORF8 of SARS-CoV have been associated with reduced replicative fitness and virus attenuation. In contrast, the SARS-CoV-2 382-nt deletion viruses showed significantly higher replicative fitness *in vitro* than the wild type, while no difference was observed in patient viral load, indicating that the deletion variant viruses retained their replicative fitness. A robust antibody response to ORF8 has been observed in SARS-CoV-2 infection, suggesting that the emergence of ORF8 deletions may be due to immune-driven selection and that further deletion variants may emerge during the sustained transmission of SARS-CoV-2 in humans.

IMPORTANCE During the SARS epidemic in 2003/2004, a number of deletions were observed in ORF8 of SARS-CoV, and eventually deletion variants became predomi-

Citation Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, Zhuang Y, Kalimuddin S, Low JGH, Tan CW, Chia WN, Mak TM, Octavia S, Chavatte J-M, Lee RTC, Pada S, Tan SY, Sun L, Yan GZ, Maurer-Stroh S, Mendenhall IH, Leo Y-S, Lye DC, Wang L-F, Smith GJD. 2020. Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. mBio 11:e01610-20. https://doi.org/10.1128/mBio.01610-20.

Editor Stacey Schultz-Cherry, St. Jude Children's Research Hospital

Address correspondence to Lin-Fa Wang, linfa.wang@duke-nus.edu.sg, or Gavin J. D. Smith, gavin.smith@duke-nus.edu.sg.

Received 16 June 2020
Accepted 23 June 2020
Published 21 July 2020

nant, leading to the hypothesis that ORF8 was an evolutionary hot spot for adaptation of SARS-CoV to humans. However, due to the successful control of the SARS epidemic, the importance of these deletions for the epidemiological fitness of SARS-CoV in humans could not be established. The emergence of multiple SARS-CoV-2 strains with ORF8 deletions, combined with evidence of a robust immune response to ORF8, suggests that the lack of ORF8 may assist with host immune evasion. In addition to providing a key insight into the evolutionary behavior of SARS-CoV-2 as the virus adapts to its new human hosts, the emergence of ORF8 deletion variants may also impact vaccination strategies.

KEYWORDS COVID-19, ORF8, natural selection, phylogeny, vaccines

In December 2019, a novel coronavirus (CoV) emerged from Hubei province in China and infected people visiting the Huanan seafood market in Wuhan (1). The virus demonstrated efficient human-to-human transmission within mainland China and subsequently spread across many countries. The virus was soon identified as novel CoV 2019 (2019-nCoV), more recently designated severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), while the associated disease is referred to as coronavirus disease 2019 (COVID-19). On 30 January 2020, the World Health Organization declared a Public Health Emergency of International Concern. As of 15 June 2020, there were almost 8 million confirmed COVID-19 cases and 431,541 deaths globally (2). In Singapore, the first case of SARS-CoV-2 was reported on 23 January 2020, in a 66-year-old man that had visited Wuhan. By 7 February, the number of cases had increased to 29, including individuals without travel history to China, prompting Singaporean authorities to implement control measures aimed at reducing the community spread of the virus (3). Successful zoonotic viral transmission from animals to humans is often associated with the ability of viruses to adapt to a new host, via genetic mutation events, and cause sustained transmission (4–6). A variety of genomic changes, including mutations, deletions, and recombinations, have been frequently observed in the other two documented zoonotic coronaviruses, SARS-CoV and Middle East respiratory syndrome-coronavirus (MERS-CoV) (7, 8).

RESULTS

During the early epidemic in January to March 2020, we collected clinical specimens (including nasopharyngeal swab, endotracheal aspirate, urine, and stool samples) from 28 hospitalized patients that tested positive for SARS-CoV-2 in Singapore. Specimens were subjected to metagenomic next-generation sequencing (NGS), with and without passaging in Vero-E6 cells, and 21 full genomes were recovered (see Table S1 in the supplemental material). Whole-genome sequencing of early samples showed that SARS-CoV-2 from Singapore shared high similarity with viruses from Wuhan, which was expected due to the importation of cases from China. Sequencing of later patient samples showed a large deletion toward the 3′ end in 6 of the 21 virus genomes (Fig. 1A). To verify this observation, specific PCR primers flanking the deleted region were designed and Sanger sequencing confirmed a 382-nucleotide (nt) deletion corresponding to positions 27,848 to 28,229 of the SARS-CoV-2 genome (see Fig. S1 in the supplemental material). Interrogation of the NGS assemblies of these 382-nt deletion variants (referred to here as Δ382) indicated that the virus populations were homogenous. Apart from the 382-nt deletion, the genome organization of the Δ382 viruses is identical to that of other SARS-CoV-2 isolates (Fig. 1A). Closer examination indicated that the deletion spans an area of open reading frame 7b (ORF7b) and ORF8 (Fig. 1B). Specifically, there is a 40-nt deletion at the 3′ end of ORF7b, a deletion of the 6-nt intergenic region of ORF7b/8, and a 336-nt deletion from the 5′ end of ORF8 that eliminates the ORF8 transcription regulatory sequence (TRS). The 382-nt deletion also results in a predicted ORF7b-ORF8 fusion protein composed of a truncated ORF7b in which its C-terminal 12 amino acids (aa), including the stop codon, are replaced with the last 5 aa from the remaining C terminus of ORF8 (Fig. S2).

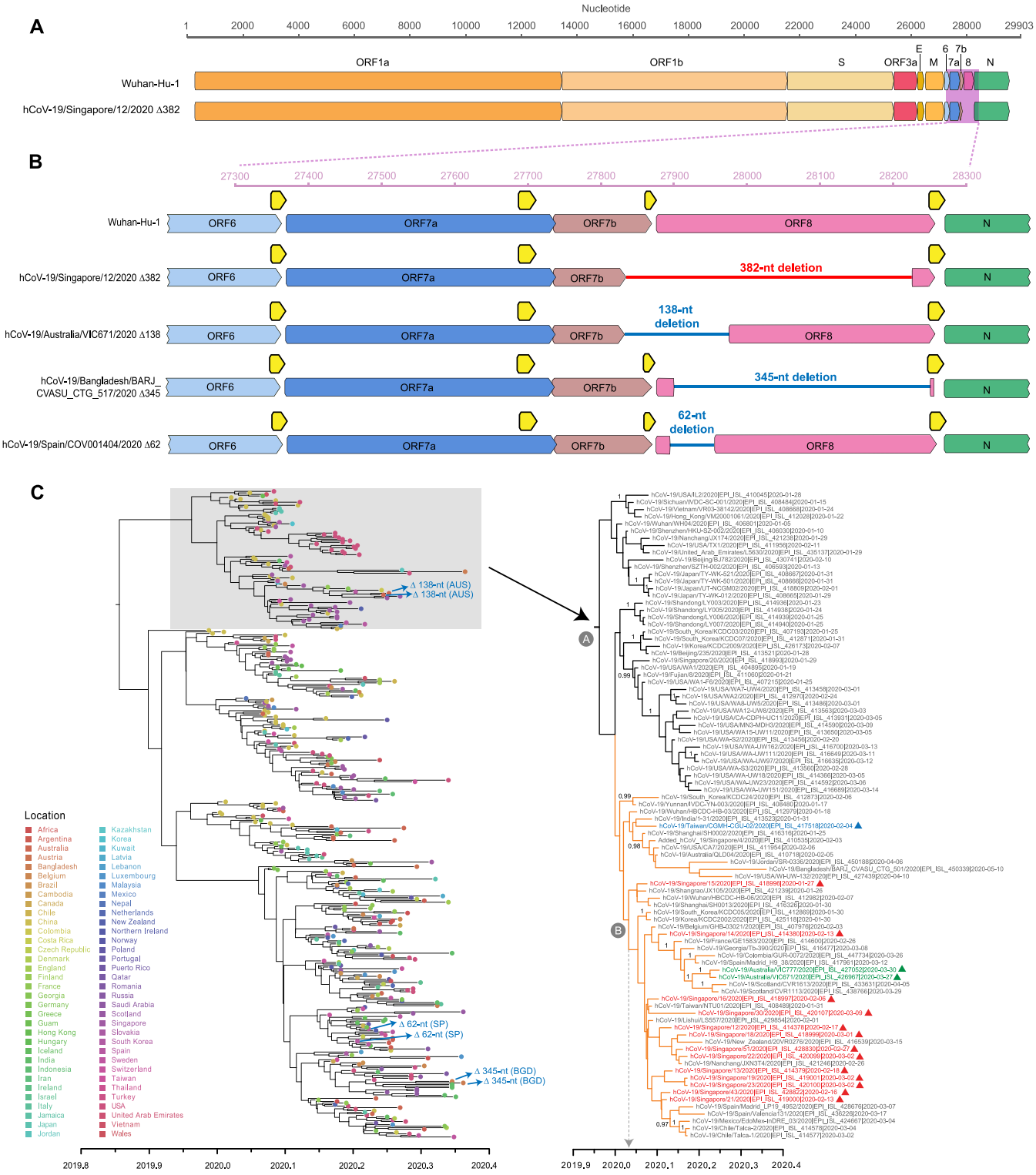

FIG 1 Genomic organization and evolutionary relationships of human SARS-CoV-2 and SARS-CoV-2 Δ382. (A) Schematic diagram of the genomes of SARS-CoV-2 isolates Wuhan-Hu-1 (GenBank accession no. MN908947) and human CoV-19/Singapore/12/2020 (hCoV-19/Singapore/12/2020) Δ382 (GISAID: EPI_ISL_414378). (B) Magnification of genomic region (pink box in panel A) showing the 382-nt deletion in ORF7b and ORF8 (indicated by red line) in hCoV-19/Singapore/12/2020. Other ORF7b/8 deletions are indicated by blue lines as follows: a 138-nt deletion in hCoV-19/Australia/VIC671/2020 (EPI_ISL_426967), a 345-nt deletion in hCoV-19/Bangladesh/BARJ_CVASU_CTG_517/2020 (EPI_ISL_450344), and a 62-nt deletion in hCoV-19/Spain/COV001404/2020 (EPI_ISL_452497). Horizontal axes indicate the nucleotide position relative to Wuhan-Hu-1; open reading frames (ORFs) are indicated by solid colored arrows. Transcription regulatory sequences (TRSs) are indicated by yellow arrows. (C) Temporal phylogeny of 419 complete genomes inferred using an uncorrelated lognormal relaxed clock model in BEAST. Colored circles at the tips represent geographic locations of virus sampling. Colored triangles represent ORF7b/8 deletion variants. A fully labeled tree with Bayesian posterior probabilities indicated is presented in Fig. S4. Red isolate names indicate SARS-CoV-2 from Singapore with a 382-nt deletion

(Continued on next page)

To investigate the possible origin and evolutionary relationships of these Δ382 viruses, a maximum likelihood (ML) phylogeny of SARS-CoV-2 full genomes was inferred from serially sampled data sets. The global evolutionary tree shows the cocirculation of multiple lineages (Fig. S3), consistent with published topologies (9), indicating lineage diversification of this pandemic virus following zoonotic transmission. The ML tree indicates that all Δ382 viruses from Singapore ($n = 11$), plus an additional Δ382 virus genome from Taiwan ($n = 1$), formed a monophyletic clade (Fig. S3), although there is a lack of statistical support, reflecting that these Δ382 viruses share a high (99.9%) level of nucleotide similarity (Table S2). We then inferred time-scaled phylogenies of SARS-CoV-2 using a relaxed-molecular-clock model in BEAST. The dated tree demonstrates the intra- and intercontinental dissemination of the wild-type (WT) viruses, whereas all Δ382 viruses from Singapore (marked by red triangles in Fig. 1C) and Taiwan (marked by a green triangle) are closely related; however, there is a lack of statistical support. While it is not possible to determine the direction of transmission based on the phylogeny, our results suggest that the introduction of Δ382 viruses likely arose from a single source rather than from multiple introductions of variants into Singapore.

Our date estimates indicate the Δ382 viruses emerged between the middle of December 2019 (node A time to most common ancestor [TMRCA] 95% highest posterior density [HPD], 2019.90 to 2020.00) and early January 2020 (node B TMRCA 95% HPD, 2019.98 to 2020.04) (Fig. 1C; see also Table S3), suggesting rapid mutation of SARS-CoV-2 following its emergence. Consistent with previous analyses, the estimated rate of nucleotide substitutions among SARS-CoV-2 viruses is approximately 0.91×10^{-3} substitutions per site per year (95% HPD, 0.79×10^{-3} to 1.03×10^{-3}) and dating estimates indicate that the introduction of SARS-CoV-2 into humans occurred early November 2019 (TMRCA 95% HPD, 2019.76 to 2019.92) (Table S3), suggesting that the viruses were present in humans approximately 1 month before the outbreak was detected.

In addition to Δ382 viruses, other ORF7b and/or ORF8 deletions, from 62 nt to 345 nt, have been observed in Australia ($n = 2$), Bangladesh ($n = 2$), and Spain ($n = 2$) (Fig. 1B; see also Table S4). Singapore and Taiwan Δ382 viruses are grouped with Australia Δ138 viruses within the same lineage (lineage A), whereas Bangladesh Δ345 viruses and Spain Δ62 viruses fall in lineage B (Fig. S3). The Δ138 viruses from Australia form a strongly supported monophyletic group (Bayesian posterior probability = 1.00) and appear to be closely related to viruses circulating from Europe (Fig. 1C; see also Fig. S3). The 138-nt deletion occurs from nucleotide position 27,846 to position 27983 of the SARS-CoV-2 genomes, resembling Singapore Δ382 viruses with nucleotide deletions across ORF7b and ORF8 regions (Fig. 1B), although the remaining ORF8 of Δ138 variants remains intact.

We then tested residual diagnostic samples from COVID-19 patients in Singapore for the presence of Δ382 viruses to understand the epidemiological fitness of Δ382 viruses. Using in-house PCR primers and Sanger sequencing, we screened a total of 191 individual samples collected from 23 January 2020 to 25 March 2020, among which 45 (23.6%) contained the 382-nt deletion (Fig. 2A), suggesting that these viruses retain the ability to efficiently infect and transmit between humans. This is further supported by evaluation of the growth of Δ382 SARS-CoV-2 viruses *in vitro* and by patient viral load data. We compared two Singapore Δ382 isolates with the wild type using Vero-E6 cells. While Δ382 SARS-CoV-2 displayed replication kinetics similar to the wild-type kinetics at 24 h postinfection (hpi), titers of the Δ382 viruses were significantly higher at later time points, even though the cytopathic effects were similar (Fig. 2B and C). The viral

FIG 1 Legend (Continued)

as described in this study. Blue isolate names indicate a 382-nt deletion from Taiwan, whereas green isolate names indicate a 138-nt deletion from Australia. Node A represents the time to most common ancestor (TMRCA) for the lineage containing Δ382 viruses from Singapore and Taiwan, while node B represents the TMRCA of the clade containing Δ382 viruses from Singapore. Bayesian posterior probabilities of ≥0.95 are indicated at nodes. Scale bar represents time in years. Abbreviations: AUS, Australia; SP, Spain; BGD, Bangladesh.

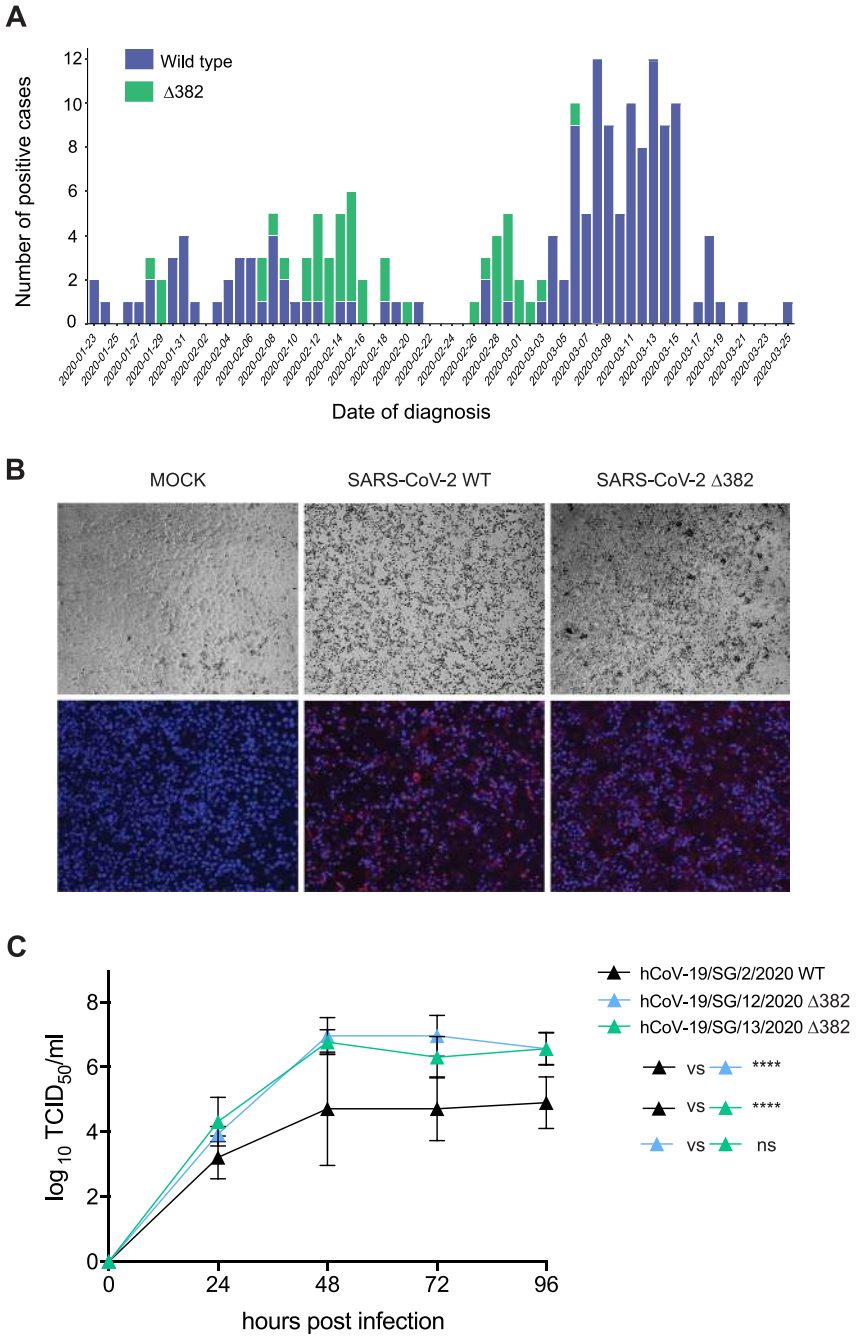

FIG 2 Prevalence, viral antigen staining, and growth kinetics of human SARS-CoV-2 and SARS-CoV-2 Δ382. (A) Daily number of wild-type and Δ382 deletion variants of SARS-CoV-2 detected in Singapore based on PCR screening of 191 patient specimens. (B) Cellular and viral fluorescence staining of SARS-CoV-2 wild-type and Δ382 strains in Vero-E6 cells. Bright-field images were captured to visualize cytopathic effect. SARS-CoV-2 antigen was stained with COVID-19 convalescent-phase serum. Red, viral proteins; blue, DAPI (nuclei). (C) Growth curves of SARS-CoV-2 wild-type and Δ382 strains at a multiplicity of infection (MOI) of 0.01 in Vero-E6 cells. Virus titers were expressed as 50% tissue culture infectious doses ($TCID_{50}$)/ml and were plotted as means of results from three independent replicates and standard deviations. Error bars indicate standard errors of means. Significance was calculated by two-way ANOVA with Tukey's multiple-comparison test.

loads seen in nasopharyngeal samples from patients infected with SARS-CoV-2 WT virus ($n = 8$, 22.48 ± 3.97) were not significantly different ($P = 0.175$) from those obtained from Δ382 virus-infected individuals ($n = 11$, 27.00 ± 6.31). This suggests that despite the loss of partial ORF7b and ORF8 regions in the genome, SARS-CoV-2 Δ382 viruses

retain replicative fitness *in vitro* and *in vivo*. The earliest Δ382 viruses were detected on 28 and 29 January 2020 from individuals that had recently traveled from Wuhan, China, coinciding with our estimated date of introduction and indicating the likely origin of the Δ382 virus. The most recent Δ382 viruses were detected in Singapore on 6 March 2020, suggesting that these viruses are no longer circulating in Singapore, likely due to the aggressive contact tracing and isolation/quarantine that had been enacted in the country. Our data, together with the detection of viruses with an identical 382-nt deletion in Taiwan, indicates that Δ382 viruses are fit and capable of transmission within humans, raising the possibility that more Δ382 viruses may be discovered when additional genomic data become available.

DISCUSSION

A number of genomic deletions in SARS-CoV were observed during the course of the 2003/2004 SARS epidemic (7, 10–15). A 29-nt ORF8 deletion occurred in all SARS-CoVs in the middle and late phases, while complete or nearly complete ORF8 deletions were observed toward the end of the outbreak (7, 13, 15). The gradual occurrence of these deletions and their eventual predominance led to the hypothesis that ORF8 was an evolutionary hot spot for adaptation of SARS-CoV to humans (7, 11, 13). Experimental studies have since shown that ORF8 of SARS-CoV plays a functional role in virus replicative fitness *in vitro*, with partial or full deletions of ORF8 demonstrating reduced replication compared to the wild type (16). However, full-length ORF8 of human SARS-CoV is only distantly related to that of SAR-CoV-2 (55.4% nt similarity), and the genomic features are correspondingly divergent. For example, ORF8 of SARS-CoV-2 lacks a functional motif (VLVVL) present in SARS-CoV ORF8b (17) that is associated with induction of cell stress pathways and activation of macrophages during SARS-CoV infection (18). Further investigation is required to determine whether the predicted ORF7b-ORF8 fusion protein (see Fig. S2 in the supplemental material) is translated and whether it has any associated virus phenotype, although human SARS-CoV (Frankfurt-1 strain) with an ORF7b deletion shows higher growth *in vitro* than viruses with the full-length ORF7b (19). A comparison of subgenomic RNA reads predicted from the sequence data (see the supplemental material) suggests that Δ382 viruses may have altered levels of transcription compared to wild-type viruses (Fig. S5), including those of the ORF6 and N genes which are known SARS-CoV interferon (IFN) antagonists (20–23), raising the possibility that infection with Δ382 viruses might result in an altered innate immune response. Due to the successful control of the SARS epidemic, the importance of these deletions for the epidemiological fitness of SARS-CoV in humans remains unknown, and experimental studies are required to assess any virus phenotypic changes in SARS-CoV-2 due to the 382-nt and other ORF8 deletions.

In this report, we describe a major evolutionary event of the SARS-CoV-2 virus following its emergence in humans. Although the biological consequences of this deletion remain largely unknown, the observed replication differences and previously described immunological consequences of SARS-CoV genome deletions suggest potential phenotypic changes in Δ382 viruses. The robust immune response to ORF8 during SARS-CoV-2 infection (24) also suggests that the emergence of ORF8 deletions may be due to immune-driven selection. Given that genetic variants will continue to arise driven by random mutation and natural selection, it is likely that we will see further deletion variants emerge with the sustained transmission of SARS-CoV-2 in humans. Although metagenomics can provide advanced tools to track the changing dynamics of SARS-CoV-2, the complex mechanisms underpinning pathogenicity, epidemiological behavior, transmission patterns, and host immunity must be examined to provide a more comprehensive understanding of this unfolding disease outbreak.

MATERIALS AND METHODS

Ethics statement. This study was undertaken as part of the national disease outbreak, and the response and the protocols were approved by the ethics committee of the National Healthcare Group. Patient samples were collected under the guidelines provided by PROTECT (2012/00917), a multicentered prospective study to detect novel pathogens and characterize emerging infections. Work under-

taken at the Duke-NUS Medical School animal biological safety level 3 (ABSL-3) laboratory was approved by the Duke-NUS ABSL3 Biosafety Committee, the National University of Singapore, and the Ministry of Health Singapore.

Virus culture, RNA extraction, and sequencing. Clinical samples from SARS-CoV-2-positive patients were collected at public hospitals in Singapore from January through February 2020. Clinical samples were used to inoculate Vero-E6 cells (ATCC CRL-1586). Total RNA was extracted using E.Z.N.A. total RNA kit I (Omega Bio-Tek) according to the manufacturer's instructions, and the samples were analyzed by real-time quantitative reverse transcription-PCR for the detection of SARS-CoV-2 as previously described (25). Whole-genome sequencing was performed using next-generation sequencing (NGS) methodology. The cDNA libraries were constructed using a TruSeq RNA library prep kit (Illumina) according to the manufacturer's instructions and sequenced on an Illumina MiSeq system. Raw NGS reads were trimmed by the use of Trimmomatic v0.39 (26) to remove adaptors and low-quality bases. Genome sequences were assembled and consensus sequences obtained using the BWA-MEM algorithm in UGENE v.33. To verify the presence of the deletion in the SARS-CoV-2 genome, we designed two specific PCR primers (primer F [5′-TGTTAGAGGTACAACAGTACTTT-3′] and primer R [5′-GGTAGTAGAAATACCATCTTGGA-3′]) targeting the ORF7-to-ORF8 regions. For samples with low cycle threshold (C_T) values, a second heminested PCR was performed with primers 5′-TGTTTATAACACTTTGCTTCACA-3′ and 5′-GGTAGTAGA AATACCATCTTGGA-3′. The PCR mixture contained the cDNA, primers (10 μM each), 10× *Pfu* reaction buffer (Promega), *Pfu* DNA polymerase (Promega), and deoxynucleoside triphosphate (dNTP) mix (Thermo Scientific) (10 mM). The PCR was carried out under the following conditions: 95°C for 2 min; 35 cycles at 95°C for 1 min, 52°C for 30 s, and 72°C for 1 min; and a final extension at 72°C for 10 min in a thermal cycler (Applied Biosystems Veriti). Deletions in the PCR products were visualized by gel electrophoresis and confirmed by Sanger sequencing. Full complete genomes of SARS-CoV-2 wild-type and Δ382 viruses generated in Singapore were deposited in the GISAID database (see Table S1 in the supplemental material).

Genomic characterization. To characterize and map the deletion regions of SARS-CoV-2 viruses, we compared viral genome organizations of Wuhan-Hu-1 (GenBank accession number MN908947) and Singapore SARS-CoV-2 (Singapore/2/2020: EPI_ISL_407987). The genomes comprised the following gene order and lengths: ORF1ab (open-reading frame) replicase (21,291 nt), spike (S: 3,822 nt), ORF3 (828 nt), envelope (E: 228 nt), membrane (M: 669 nt), ORF6 (186 nt), ORF7ab (498 nt), ORF8 (366 nt), nucleocapsid (N: 1,260 nt), and ORF10 (117 nt).

Phylogenetic analyses. All available genomes of SARS-CoV-2 with associated virus sampling dates were downloaded from the GISAID database. To reduce bias from locations with higher virus sampling and genome availability, data sets were subsampled randomly based on geographical location and collection month using in-house scripts. Genome sequence alignment was performed using MAFFT (27) in Geneious R9.0.3 software (Biomatters Ltd.) followed by manual alignment. Maximum likelihood phylogenies of 1,038 complete genomes were reconstructed using RAxML with 200 bootstrap replicates (28). Any sequence outliers were removed from subsequent analyses. Lineage circumscription of SARS-CoV-2 was conducted using pangolin software (9). To reconstruct a time-scaled phylogeny, we analyzed serially sampled data sets of 419 complete genomes (Table S5) using an uncorrelated lognormal relaxed-clock (UCLN) model with an exponential growth coalescent prior and the HKY85+Γ substitution model in BEAST program v1.10.4 (29) to simultaneously estimate phylogenies, divergence times, and rates of nucleotide substitution. The strict and UCLN models were recently tested by Duchene et al. (30), who showed that the UCLN model is preferred over a strict clock for analyzing large (>122) SARS-CoV-2 genome data sets. At least four independent Markov chain Monte Carlo (MCMC) runs of 100 million generations were performed with sampling every 10,000 generations. The runs were checked for convergence in Tracer v1.7 (31), and the effective sampling size (ESS) values of all parameters were >200. The resulting log and tree files were combined after removal of appropriate burn-in values using LogCombiner (29), and a maximum clade credibility (MCC) tree was subsequently generated using TreeAnnotator (29).

Replication kinetics. Vero-E6 cells were infected with wild-type and Δ382 viruses and fixed at 48 and 72 hpi with 4% paraformaldehyde for 30 min at room temperature. Cells were washed with phosphate-buffered saline (PBS), and SARS-CoV-2 viral antigens were detected using a COVID-19 convalescent human serum at a 1:400 dilution in PBS for 30 min at 37°C. Phycoerythrin (PE)-conjugated goat anti-human IgG polyclonal antibody (eBioscience) was added at a 1:400 dilution and incubated for 30 min at 37°C. For nuclear visualization, cells were stained with 0.01% 4′,6-diamidino-2-phenylindole (DAPI; Abcam). Images were captured by the use of an Eclipse Ti-U fluorescence microscope (Nikon). For virus kinetics analyses, Vero-E6 cells were infected at a multiplicity of infection (MOI) of 0.01. Supernatant was harvested daily for 6 days following infection and 50% tissue culture infective doses ($TCID_{50}$) titers were determined. Statistical analysis was performed using a two-way analysis of variance (ANOVA) with Tukey's multiple-comparison test. Patient viral load was measured by diagnostic quantitative PCR (qPCR) to determine cycle threshold (C_T) values as previously described (32), and the data were compared using a two-tailed *t* test.

Subgenomic RNA analysis. Unambiguous reads which uniquely mapped to the specific joint leader and transcription regulatory sequences (TRS) were counted for each gene from independent NGS library preparations taken from the same RNA extractions of wild-type ($n = 3$) and Δ382 mutant ($n = 2$) SARS-CoV-2 strains collected from two patients. Coding DNA sequence (CDS) and leader and TRS sequence annotations were generated in Geneious and followed published SARS-CoV studies (33). To characterize the differential levels of TRS for each gene, 70 nt of leader sequence and 230 nt downstream of each TRS sequence were annotated individually to form a 300-bp leader-TRS transcript for the

splicing-aware aligners in R package rtracklayer (v 1.44). NGS raw fastq reads were then mapped to the reference genome by the use of the Geneious RNA-Seq mapper with the annotation of splice junctions for the leader-TRS. For each gene, the total number of transcripts per million (TPM) of the leader-TRS was calculated in Geneious by excluding ambiguous reads which might have come from other TRS. The resulting TPM data were plotted using ggplot 2 (v3.2.0) in R v3.6.1. Wilcoxon one-sided tests were performed in R to test for significant differences between wild-type and Δ382 samples.

Data availability. Sequences generated in this study have been deposited in the GISAID database (see Table S1 for accession numbers). All data are available in the main text or the supplemental material.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.7 MB.

FIG S2, PDF file, 0.4 MB.

FIG S3, PDF file, 0.3 MB.

FIG S4, PDF file, 0.3 MB.

FIG S5, PDF file, 0.2 MB.

TABLE S1, DOCX file, 0.01 MB.

TABLE S2, DOCX file, 0.01 MB.

TABLE S3, DOCX file, 0.01 MB.

TABLE S4, DOCX file, 0.01 MB.

TABLE S5, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Randy Foo and Akshamal Gamage for processing clinical samples; Velraj Sivalingam, Adrian Kang, and Yilong Peng for assistance with virus growth kinetics; and Su Ting Tay, Ming Hui Lee, and Angie Tan of the Duke-NUS Medical School Genome Biology Facility and Bei Bei Chen of the National Public Health Laboratory for technical assistance.

This study was supported by the Duke-NUS Signature Research Programme funded by the Ministry of Health, Singapore; by the National Medical Research Council under its COVID-19 Research Fund (NMRC project no. COVID19RF-001 and COVID19RF-004); and by National Research Foundation Singapore grant NRF2016NRFNSFC002-013 (Combating the Next SARS- or MERS-Like Emerging Infectious Disease Outbreak by Improving Active Surveillance). R.T.C.L. and S.M.-S. were supported by A*STAR. Y.C.F.S. and G.J.D.S. are supported by contract HHSN272201400006C from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, U.S. Department of Health and Human Services.

We contributed to the work as follows. Y.C.F.S., D.E.A., L.-F.W., and G.J.D.S. designed and supervised research. B.E.Y., S.K., J.G.H.L., S.P., S.Y.T., L.S., G.Z.Y., Y.-S.L., and D.C.L. collected and provided samples. D.E.A., M.L., Y.Z., C.W.T., W.N.C., T.M.M., S.O., and J.-M.C. conducted experiments. Y.C.F.S., D.E.A., B.E.Y., M.L., F.Z., J.J., R.T.C.L., S.M.-S., I.H.M., and G.J.D.S. performed analyses. Y.C.F.S., M.L., L.-F.W., and G.J.D.S. wrote the paper. All of us reviewed and approved the manuscript.

We declare no competing financial interests.

REFERENCES

1. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395:565–574. https://doi.org/10.1016/S0140-6736(20)30251-8.
2. World Health Organization. 2020. Coronavirus disease (COVID-2019) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed 2 March 2020.
3. Pung R, Chiew CJ, Young BE, Chin S, Chen MI-C, Clapham HE, Cook AR, Maurer-Stroh S, Toh MPHS, Poh C, Low M, Lum J, Koh VTJ, Mak TM, Cui L, Lin RVTP, Heng D, Leo Y-S, Lye DC, Lee VJM, Singapore 2019 Novel Coronavirus Outbreak Research Team. 2020. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. Lancet 395:1039–1046. https://doi.org/10.1016/S0140-6736(20)30528-6.
4. Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev 72:457–470. https://doi.org/10.1128/MMBR.00004-08.
5. Geoghegan JL, Holmes EC. 2017. Predicting virus emergence amid evolutionary noise. Open Biol 7:170189. https://doi.org/10.1098/rsob.170189.
6. Warren CJ, Sawyer SL. 2019. How host genetics dictates successful viral zoonosis. PLoS Biol 17:e3000217. https://doi.org/10.1371/journal.pbio.3000217.
7. Chinese SARS Molecular Epidemiology Consortium. 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303:1666–1669. https://doi.org/10.1126/science.1092002.

8. Lamers MM, Raj VS, Shafei M, Ali SS, Abdallh SM, Gazo M, Nofal S, Lu X, Erdman DD, Koopmans MP, Abdallat M, Haddadin A, Haagmans BL. 2016. Deletion variants of Middle East respiratory syndrome coronavirus from humans, Jordan, 2015. Emerg Infect Dis 22:716–719. https://doi.org/10.3201/eid2204.152065.
9. Rambaut A, Hill V, O'Toole Á, McCrone JT, Ruis C, du Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. bioRxiv https://doi.org/10.1101/2020.04.17.046086.
10. Wang L-F, Shi Z, Zhang S, Field H, Daszak P, Eaton BT. 2006. Review of bats and SARS. Emerg Infect Dis 12:1834–1840. https://doi.org/10.3201/eid1212.060401.
11. Lau SKP, Feng Y, Chen H, Luk HKH, Yang W-H, Li KSM, Zhang Y-Z, Huang Y, Song Z-Z, Chow W-N, Fan RYY, Ahmed SS, Yeung HC, Lam CSF, Cai J-P, Wong SSY, Chan JFW, Yuen K-Y, Zhang H-L, Woo PCY. 2015. Severe acute respiratory syndrome (SARS) coronavirus ORF8 protein is acquired from SARS-related coronavirus from greater horseshoe bats through recombination. J Virol 89:10532–10547. https://doi.org/10.1128/JVI.01048-15.
12. Xu L, Zhang F, Yang W, Jiang T, Lu G, He B, Li X, Hu T, Chen G, Feng Y, Zhang Y, Fan Q, Feng J, Zhang H, Tu C. 2016. Detection and characterization of diverse alpha- and betacoronaviruses from bats in China. Virol Sin 31:69–77. https://doi.org/10.1007/s12250-016-3727-3.
13. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JSM, Poon LLM. 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302:276–278. https://doi.org/10.1126/science.1087139.
14. Tang JW, Cheung JLK, Chu IMT, Sung JJY, Peiris M, Chan PKS. 2006. The large 386-nt deletion in SARS-associated coronavirus: evidence for quasispecies? J Infect Dis 194:808–813. https://doi.org/10.1086/507044.
15. Chiu RWK, Chim SSC, Tong Y-k, Fung KSC, Chan PKS, Zhao G-p, Lo YMD. 2005. Tracing SARS-coronavirus variant with large genomic deletion. Emerg Infect Dis 11:168–170. https://doi.org/10.3201/eid1101.040544.
16. Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, Gloza-Rausch F, Balboni A, Battilani M, Rihtarič D, Toplak I, Ameneiros RS, Pfeifer A, Thiel V, Drexler JF, Müller MA, Drosten C. 2018. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. Sci Rep 8:15177. https://doi.org/10.1038/s41598-018-33487-8.
17. Chan JF-W, Kok K-H, Zhu Z, Chu H, To KK-W, Yuan S, Yuen K-Y. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 9:221–236. https://doi.org/10.1080/22221751.2020.1719902.
18. Shi C-S, Nabar NR, Huang N-N, Kehrl JH. 2019. SARS-coronavirus open reading frame-8b triggers intracellular stress pathways and activates NLRP3 inflammasomes. Cell Death Discov 5:101. https://doi.org/10.1038/s41420-019-0181-7.
19. Pfefferle S, Krähling V, Ditt V, Grywna K, Mühlberger E, Drosten C. 2009. Reverse genetic characterization of the natural genomic deletion in SARS-coronavirus strain Frankfurt-1 open reading frame 7b reveals an attenuating function of the 7b protein in-vitro and in-vivo. Virol J 6:131. https://doi.org/10.1186/1743-422X-6-131.
20. Frieman M, Yount B, Heise M, Kopecky-Bromberg SA, Palese P, Baric RS. 2007. Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. J Virol 81:9812–9824. https://doi.org/10.1128/JVI.01012-07.
21. Kopecky-Bromberg SA, Martínez-Sobrido L, Frieman M, Baric RA, Palese P. 2007. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. J Virol 81:548–557. https://doi.org/10.1128/JVI.01782-06.
22. Sims AC, Tilton SC, Menachery VD, Gralinski LE, Schäfer A, Matzke MM, Webb-Robertson B-JM, Chang J, Luna ML, Long CE, Shukla AK, Bankhead AR, Burkett SE, Zornetzer G, Tseng C-TK, Metz TO, Pickles R, McWeeney S, Smith RD, Katze MG, Waters KM, Baric RS. 2013. Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells. J Virol 87:3885–3902. https://doi.org/10.1128/JVI.02520-12.
23. Hu Y, Li W, Gao T, Cui Y, Jin Y, Li P, Ma Q, Liu X, Cao C. 2017. The severe acute respiratory syndrome coronavirus nucleocapsid inhibits type I interferon production by interfering with TRIM25-mediated RIG-I ubiquitination. J Virol 91:e02143-16. https://doi.org/10.1128/JVI.02143-16.
24. Hachim A, Kavian N, Cohen CA, Chin AWH, Chu DKW, Mok CKP, Tsang OTY, Yeung YC, Perera RAPM, Poon LLM, Peiris MJS, Valkenburg SA. 2020. Beyond the spike: identification of viral targets of the antibody response to SARS-CoV-2 in COVID-19 patients. medRxiv https://doi.org/10.1101/2020.04.30.20085670.
25. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DG, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette J-L, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MP, Drosten C. 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance 25:2000045. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045.
26. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
27. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
28. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
29. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:vey016. https://doi.org/10.1093/ve/vey016.
30. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. bioRxiv https://doi.org/10.1101/2020.05.04.077735.
31. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901–904. https://doi.org/10.1093/sysbio/syy032.
32. Young BE, Ong SWX, Kalimuddin S, Low JG, Tan SY, Loh J, Ng O-T, Marimuthu K, Ang LW, Mak TM, Lau SK, Anderson DE, Chan KS, Tan TY, Ng TY, Cui L, Said Z, Kurupatham L, Chen MI-C, Chan M, Vasoo S, Wang L-F, Tan BH, Lin RTP, Lee VJM, Leo Y-S, Lye DC, Singapore 2019 Novel Coronavirus Outbreak Research Team. 2020. Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. JAMA 323:1488–1494. https://doi.org/10.1001/jama.2020.3204.
33. Hussain S, Pan J, Chen Y, Yang Y, Xu J, Peng Y, Wu Y, Li Z, Zhu Y, Tien P, Guo D. 2005. Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. J Virol 79:5288–5295. https://doi.org/10.1128/JVI.79.9.5288-5295.2005.

Author/s:
Su, YCF;Anderson, DE;Young, BE;Linster, M;Zhu, F;Jayakumar, J;Zhuang, Y;Kalimuddin, S;Low, JGH;Tan, CW;Chia, WN;Mak, TM;Octavia, S;Chavatte, J-M;Lee, RTC;Pada, S;Tan, SY;Sun, L;Yan, GZ;Maurer-Stroh, S;Mendenhall, IH;Leo, Y-S;Lye, DC;Wang, L-F;Smith, GJD

Title:
Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2

Date:
2020-07-01

Citation:
Su, Y. C. F., Anderson, D. E., Young, B. E., Linster, M., Zhu, F., Jayakumar, J., Zhuang, Y., Kalimuddin, S., Low, J. G. H., Tan, C. W., Chia, W. N., Mak, T. M., Octavia, S., Chavatte, J. -M., Lee, R. T. C., Pada, S., Tan, S. Y., Sun, L., Yan, G. Z. ,... Smith, G. J. D. (2020). Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. MBIO, 11 (4), https://doi.org/10.1128/mBio.01610-20.

Persistent Link:
http://hdl.handle.net/11343/277716