

## RESEARCH ARTICLE

## Molecular characterization and the mutation pattern of SARS-CoV-2 during first and second wave outbreaks in Hiroshima, Japan

Ko Ko<sup>1</sup>, Shintaro Nagashima<sup>1</sup>, Bunthen E.<sup>1,2</sup>, Serge Ouoba<sup>1,3</sup>, Tomoyuki Akita<sup>1</sup>, Aya Sugiyama<sup>1</sup>, Masayuki Ohisa<sup>1</sup>, Takemasa Sakaguchi<sup>4</sup>, Hidetoshi Tahara<sup>5</sup>, Hiroki Ohge<sup>6</sup>, Hideki Ohdan<sup>7</sup>, Tatsuhiko Kubo<sup>8</sup>, Eisaku Kishita<sup>9</sup>, Masao Kuwabara<sup>10</sup>, Kazuaki Takahashi<sup>1</sup>, Junko Tanaka<sup>1\*</sup>

**1** Department of Epidemiology, Infectious Disease Control and Prevention, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan, **2** Payment Certification Agency, Ministry of Health, Phnom Penh, Cambodia, **3** Unité de Recherche Clinique de Nanoro (URCN), Nanoro, Burkina Faso, **4** Department of Virology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan, **5** Department of Cellular and Molecular Biology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan, **6** Department of Infectious Diseases, Hiroshima University Hospital, Hiroshima, Japan, **7** Department of Gastroenterological and Transplant Surgery, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan, **8** Department of Public Health and Health Policy, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan, **9** Hiroshima Prefecture Health and Welfare Bureau, Hiroshima, Japan, **10** Hiroshima Prefectural Center for Disease Control and Prevention, Hiroshima, Japan

\* [jun-tanaka@hiroshima-u.ac.jp](mailto:jun-tanaka@hiroshima-u.ac.jp)



## OPEN ACCESS

**Citation:** Ko K, Nagashima S, E. B, Ouoba S, Akita T, Sugiyama A, et al. (2021) Molecular characterization and the mutation pattern of SARS-CoV-2 during first and second wave outbreaks in Hiroshima, Japan. PLoS ONE 16(2): e0246383. <https://doi.org/10.1371/journal.pone.0246383>

**Editor:** Dong-Yan Jin, University of Hong Kong, HONG KONG

**Received:** December 2, 2020

**Accepted:** January 18, 2021

**Published:** February 5, 2021

**Copyright:** © 2021 Ko et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data used in this study are fully described in the figure and tables. The eight full genomes sequence data of SARS-CoV-2 included in this study are deposited at GISAID (<https://www.gisaid.org>) with accession numbers of EPI\_ISL\_855345 to EPI\_ISL\_855352 and GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) with the accession numbers of LC594644 to LC594651, and if there is any trouble in accessing the data, the sequences are available from the corresponding author upon the reasonable request.

## Abstract

## Background

In this study, we performed molecular characterization of SARS-CoV-2 strains in Hiroshima and its mutation pattern between the first and second waves of the outbreak.

## Method

A total of 55 nasal swab samples from the first wave in Hiroshima and 13 from the second wave were examined quantitatively by RT-qPCR and qualitatively by nested PCR using specific primers. Four samples from each wave underwent next-generation sequencing and phylogenetic tree analysis including controls and all sequences retrieved in Japan from GISAID and GenBank. Subsequently, mutations were examined.

## Results

Viral load ranged  $7.85 \times 10^1$ – $1.42 \times 10^8$  copies/ml. Of 68 samples, one was Asian type-O, 65 were European type-GR, and 2 were undetectable. Phylogenetic tree analysis indicated that Japan was infected with various Asian strains (L, S, V, O) from January through April. By second week of March, European strains (G, GH, GR) had appeared, and GR strains became predominant after mid-March. The first case in Hiroshima was classified as Asian strain O, and the rest were GR strains. Then, second wave of GR strains appeared independently with 11–15 base mutations. Comparing the first- and second-wave GR strains, mutation rate was  $1.17$ – $1.36 \times 10^{-3}$  base substitutions per site per year; in addition, amino acid

**Funding:** This study was partly funded by a grant from the Government Academia Collaboration of Hiroshima Prefecture. The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report. There was no additional external funding received for this study.

**Competing interests:** The authors have declared that no competing interests exist.

changes occurred at S1361P and P3371S in ORF1a, A314V in ORF1b, and P151L in N. All seven GR strains were D614G variants with R202K and G203R mutations in N. A single-nucleotide insertion in *ORF8* that causes a defect in ORF8 protein was found in one isolate (S66) from the second wave.

## Conclusion

Our findings reveal the evolutionary hierarchy of SARS-CoV-2 in Japan. The predominant D614G variants and a new form of *ORF8* deletion in Hiroshima provide the clue for role of viral factor in local outbreaks of SARS-CoV-2.

## Background

Novel coronavirus, formerly named 2019-nCoV and later officially termed as SARS-CoV-2 by the International Committee on Taxonomy of Viruses (ICTV), is a member of genus *Betacoronavirus*, which also includes SARS-CoV-1 and MERS-CoV [1]. The novel coronavirus is a positive-sense single-strand RNA genome of approximately 30 kb, the largest genome among RNA viruses [2].

SARS-CoV-2 causes a lower respiratory tract infection that progresses to severe acute respiratory syndrome and multi-organ failure [3, 4]. More than 1 million people have lost their lives as a result of infection by this virus. After detection of SARS-CoV-2 in Wuhan, China in late December 2019, approximately 33 million confirmed cases and 996,309 deaths had been reported worldwide as of September 27, 2020 [5]. The World Health Organization (WHO) declared this infection as a pandemic on March 11, 2020 [6].

As of August 3, 2020, a total of 40,244 confirmed cases and 1,018 deaths had been reported in Japan [7]. The incidence rate differs across prefectures, ranging from 181.51 per 100,000 people in Tokyo to 2.81 per 100,000 people in Aomori prefecture [7]. The case-fatality ratio also differs between prefectures, ranging from 0% in Iwate and Akita prefectures to 1.48% in Tokyo.

The first wave of corona outbreak in Japan occurred between the last week of March and the first week of May 2020, with a peak of 701 newly confirmed cases per day. The second wave of the outbreak has been ongoing since the last week of June 2020, with a peak of 1,762 newly confirmed cases per day [7].

Hiroshima Prefecture, located in the southwest of Japan's main island, had an estimated population of 1.1 million in 2019 [8]. The first wave of SARS-CoV-2 outbreak in Hiroshima, which occurred between the last week of March and the first week of May, affected 162 people. The second wave, from the last week of June to the first week of August 2020, affected 204 people [9]. Overall, as of August 3, 2020, Hiroshima has had a total of 366 confirmed cases with three deaths. Although confirmed cases are reported daily in Hiroshima, there is no detailed study on the molecular epidemiology of SARS-CoV-2 circulating in the prefecture.

In this study, we performed a molecular characterization of SARS-CoV-2 strains found in Hiroshima and compared mutations of these strains between the first and second waves of the outbreak.

## Materials and methods

### Subjects and their clinical data

55 nasal swab samples from confirmed cases of SARS-CoV-2 during the first wave and 13 samples from the second wave of outbreak were provided from Hiroshima Prefecture Research

Institute after being used by government laboratory. The average age of patients in first wave was  $62.2 \pm 23.9$  years old whilst that from second wave was  $33.1 \pm 14.9$  years old. The proportion of male patients were 32.7% and 69.2% in first and second wave respectively. The anonymized clinical data were collected from the respective hospital. This information includes age, sex, date of onset, admission date, discharge date, symptoms, outcomes and the contact history.

### Nucleic acid extraction and quantitative measurement of SARS-CoV-2

Nasal swab samples (50  $\mu$ l) were subjected to nucleic acid extraction using SMITEST EX-R&D (MBL, USA). The final pellet was diluted with 25  $\mu$ l of RNase inhibitor-based water (Thermo Fisher Scientific, USA). For each sample, viral load was quantified by real-time reverse transcriptase polymerase chain reaction (RT-qPCR) (Thermo Fisher Scientific) using the nucleocapsid (N) gene-specific primers NIID\_2019-n-CoV-N-F2 (targeting nt29,125–nt29,144) and NIID\_2019-n-CoV-N-R2 (targeting nt29,299–nt29,280) and probe NIID\_2019-n-CoV-N-P2 (nt29,222–nt29,241) (Eurofins Genomics, Japan).

### Classification of GISAID clade

To classify the GISAID clade, the partial genomes of SARS-CoV-2 were amplified from all samples by polymerase chain reaction (PCR) with 7 sets of primers as shown in [Table 1A](#). The first round of PCR was done by Prime Script One-Step RT-PCR kit Ver.2 (Takara Bio Inc., Shiga, Japan) for 30 minutes at 50°C and 1 minute at 94°C. Then, the samples were exposed to 40 cycles each of 30 seconds at 94°C, 30 seconds at 55°C, 1 minute at 72°C and final elongation for 7 minutes at 72°C. Then, the second round of PCR was done by TaKaRa Ex Taq® Hot Start version (Takara Bio Inc., Shiga, Japan) to 30 cycles each of 10 seconds at 98°C, 30 seconds at 55°C and 1 minute at 72°C followed by final elongation for 7 minutes at 72°C. The final PCR products were undergone partial genomes sequencing with 3730xl DNA sequencer and BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) using the corresponding primer set as shown in [Table 1B](#).

In order to differentiate between GISAID Asia and European clade, we examined nt23403 where Adenine (A) for European clade and Guanine (G) for Asian clade. For Asian clade, we subsequently examined at nt8782 (Thymine (T) for Asian clade S, nt11083 (T for Asian clade V or O), nt26144 (T for Asian clade V) and none of them at nt8782, nt11083 and nt26144 as Asian clade O. For European clade, we subsequently examined at nt25563 (T for European clade GH), at nt28881 (A for European clade GR) and the rest as European clade G. ([Fig 1](#))

### Full-length genome sequencing

Full-length genome sequence was acquired by next-generation sequencing (NGS). First, template RNA was translated into complementary DNA (cDNA) using the SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific). The cDNA was then processed by multiplex polymerase chain reaction (PCR) using the Q5 HoT START DNA polymerase kit and primer pool stocks 1 and 2. The PCR products in the pool 1 and 2 reactions for the same sample were mixed and purified with AmpureXP (1x volume), and full genome sequencing was performed on an Illumina MiSeq System using the QIAseq FX DNA Library Kit. The quality of the sequencing library was assessed using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific). The expected amplicon size was 250 bp. The resultant fragments were clustered using MiSeq Reagent Kit v3 and PhiX Control Kit v3, and sequences were analyzed using the MiSeq Control Software (MCS) v2.6.2.1, Real-Time Analysis (RTA) v1.18.54, and bc12fastq v2.17.

Table 1. SARS-CoV-2 specific primers used for partial sequences of particular region.

<b>(a) Primers used in Polymerase Chain Reaction</b>				
	<b>Stage polarity</b>	<b>Primer name</b>	<b>Nucleotide position</b>	<b>Nucleotide sequence (5'-3')</b>
<b>hCoV-A set</b>	PCR 1 <sup>st</sup> Sense	SC2-A-1-1	8561–8581	GTTAATAATTGGTTGAAGCAG
	PCR 1 <sup>st</sup> Antisense	SC2-A-1-2	8865–8884	TATCGTGCCAGGCCAAACCAG
	PCR 2nd Sense	SC2-A-2-1	8630–8650	TTAATAACACCTGTTTCATGTC
	PCR 2nd Antisense	SC2-A-2-2	8834–8854	ACCCACTTCTCTTGTATGAC
<b>hCoV-B set</b>	PCR 1st Sense	SC2-B-1-1	10922–10943	GAATTACACCTTTTGTATGTTG
	PCR 1st Antisense	SC2-B-1-2	11280–11300	GCTTAAACCAGACAAACTAG
	PCR 2nd Sense	SC2-B-2-1	10957–10977	AGGTGTTACTTTCCAAAGTGC
	PCR 2nd Antisense	SC2-B-2-2	11234–11254	AATACGCATCACCAACTAGC
<b>hCoV-C set</b>	PCR 1st Sense	SC2-C-1-1	14199–14218	AGAGTCACATGTTGACACTG
	PCR 1st Antisense	SC2-C-1-2	14524–14545	TAAACTAAGTCTAGAGCTATG
	PCR 2nd Sense	SC2-C-2-1	14297–14318	ATTTTAAATATTGGGATCAGAC
	PCR 2nd Antisense	SC2-C-2-2	14500–14521	AGTTTACATCCTGATTATGTAC
<b>hCoV-D set</b>	PCR 1st Sense	SC2-D-1-1	23249–23270	CAACAATTTGGCAGAGACATTG
	PCR 1st Antisense	SC2-D-1-2	23496–23516	CTATTAACAGCCTGCACGTG
	PCR 2nd Sense	SC2-D-2-1	23278–23297	TACTGATGCTGTCCGTGATC
	PCR 2nd Antisense	SC2-D-2-2	23453–23472	GAATAAACACGCCAAGTAGG
<b>hCoV-E set</b>	PCR 1st Sense	SC2-E-1-1	25359–25380	AAGGAGTCAAATTACATTACAC
	PCR 1st Antisense	SC2-E-1-2	25658–25678	GCAAAGGTGTGAGTAAACTG
	PCR 2nd Sense	SC2-E-2-1	25394–25416	TGGATTTGTTTATGAGAATCTTC
	PCR 2nd Antisense	SC2-E-2-2	25634–25654	CAAACAACACAGCAAGTTGC
<b>hCoV-F set</b>	PCR 1st Sense	SC2-F-1-1	25878–25898	TGAACATGACTACCAGATTGG
	PCR 1st Antisense	SC2-F-1-2	26297–26317	AGTACGCACACAATCGAAGCG
	PCR 1st and 2nd Sense	SC2-F-2-1	25919–25939	AATCTGGAGTAAAAGACTGTG
	PCR 1st and 2nd Antisense	SC2-F-2-2	26259–26279	GTAAC TAGCAAGAATACCACG
	PCR 2nd Sense	SC2-F-3-1	25948–25968	CACAGTTACTTCACTTCAGAC
	PCR 2nd Antisense	SC2-F-3-2	26216–26237	CTATTAACTATTAACGTACCTG
<b>hCoV-G set</b>	PCR 1st Sense	SC2-G-1-1	28612–28631	ATGGGTTGCAACTGAGGGAG
	PCR 1st Antisense	SC2-G-1-2	28903–28923	GACATTTTGCTCTCAAGCTGG
	PCR 2nd Sense	SC2-G-2-1	28709–28729	GGAACAACATTGCCAAAAGGC
	PCR 2nd Antisense	SC2-G-2-2	28874–28893	TCAAGCAGCAGCAAAGCAAG
<b>hCoV-ORF8 set</b>	PCR 1st Sense	ORF8-1-1	27645–27667	TAAACTGTTTCATCAGACAAGAGG
	PCR 1st Antisense	SC2-H-1-2	28314–28332	GGTCCACCAACGTAATGC
	PCR 1st and 2nd Sense	ORF8-2-1	27706–27727	GCGGCAATAGTGTTTATAACAC
	PCR 2nd Sense	ORF8-3-1	27779–27800	CTTCTATTTGTGCTTTTATAGCC
PCR 2nd Antisense	ORF8-2-2	28276–28297	ATTTTGGGGTCCATTATCAGAC	
<b>(b) Primers used in region-specific partial sequencing</b>				
	<b>Primer name</b>	<b>Nucleotide position</b>	<b>Nucleotide sequence (5'-3')</b>	<b>Target Nucleotide position</b>
<b>hCoV-A set</b>	SC2-A-2-1	8630–8650	TTAATAACACCTGTTTCATGTC	nt8782
	SC2-A-2-2	8834–8854	ACCCACTTCTCTTGTATGAC	
<b>hCoV-B set</b>	SC2-B-2-1	10957–10977	AGGTGTTACTTTCCAAAGTGC	nt11083
	SC2-B-2-2	11234–11254	AATACGCATCACCAACTAGC	
<b>hCoV-C set</b>	SC2-C-2-1	14297–14318	ATTTTAAATATTGGGATCAGAC	nt14408
	SC2-C-2-2	14500–14521	AGTTTACATCCTGATTATGTAC	
<b>hCoV-D set</b>	SC2-D-2-1	23278–23297	TACTGATGCTGTCCGTGATC	nt23403
	SC2-D-2-2	23453–23472	GAATAAACACGCCAAGTAGG	
<b>hCoV-E set</b>	SC2-E-2-1	25394–25416	TGGATTTGTTTATGAGAATCTTC	nt25563

(Continued)

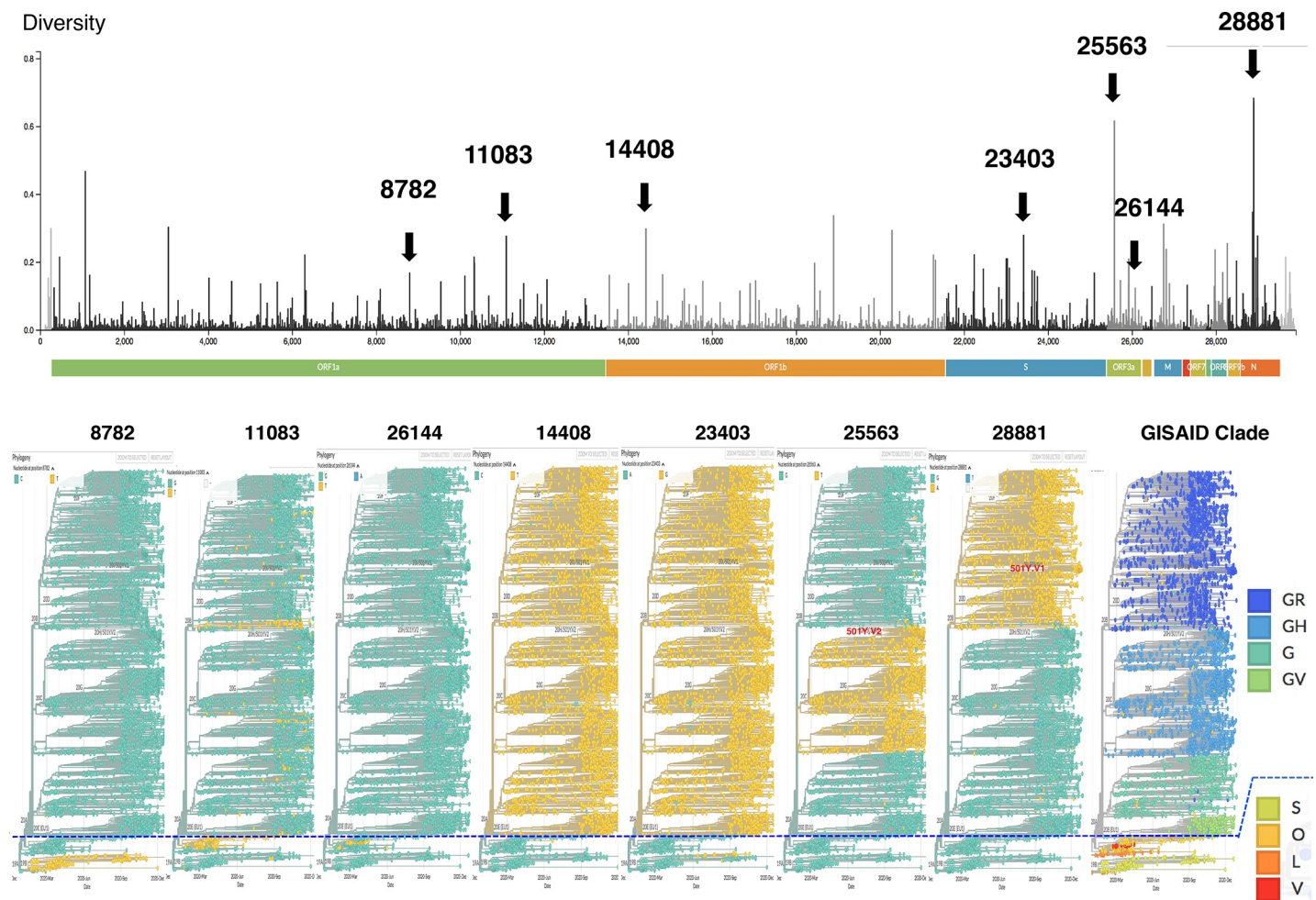
Table 1. (Continued)

	SC2-E-2-2	25634–25654	CAAACAACAACAGCAAGTTGC	
hCoV-F set	SC2-F-3-1	25948–25968	CACAGTTACTTCACCTCAGAC	nt26144
	SC2-F-3-2	26216–26237	CTATTAACTATTAACGTACCTG	
hCoV-G set	SC2-G-2-1	28709–28729	GGAACAACATTGCCAAAAGGC	nt28881
	SC2-G-2-2	28874–28893	TCAAGCAGCAGCAAAGCAAG	
hCoV-ORF8 set	ORF8-3-1	27779–27800	CTTCTATTGTGCTTTTATAGCC	nt27906-nt27907

<https://doi.org/10.1371/journal.pone.0246383.t001>

### Phylogenetic tree analysis

A total of 651 SARS-CoV-2 whole-genome sequences were collected from the GenBank (<http://www.ncbi.nlm.gov/genbank/>) and Global Initiative on Sharing All Influenza Data (GISAID: <http://www.gisaid.org>). Sequences duplicated between sources were checked and excluded manually. Sequences with many ambiguous sites (N) were removed before analysis. Using 52 SARS-CoV-2 whole-genome sequences from around the world as controls, 522



**Fig 1. Distinctive variation of nucleotide in specific position for each GISAID clade.** This figure shows the spectrum of different nucleotide at their specific position which can be used to classify the GISAID clade. Each color represents the aforementioned base (A: Adenine, G: Guanine, T: Thymine) Asian clade includes S, V, O, L and European clade includes G, GH, GR and GV.

<https://doi.org/10.1371/journal.pone.0246383.g001>



SARS-CoV-2 whole-genome sequences from throughout Japan and eight full-length genomes isolated in this study were subjected to evolutionary analysis by the neighbor-joining method using Molecular Evolutionary Genetics Analysis (MEGA) version 7.0 [10].

### Examination of molecular characterization and mutation of SARS-CoV-2 strains from Hiroshima

Using the reference strains (MN908947) from GenBank, eight SARS-CoV-2 whole-genome sequences obtained in this study were aligned using MUSCLE multiple sequence alignment algorithms in MEGA v7.0, and then transformed into amino acid sequences. In all whole-genome sequences, the mutations were compared between the reference strain and SARS-CoV-2 in each fragment of *ORF1a*, *ORF1b*, *S*, *ORF3a*, *E*, *M*, *ORF6*, *ORF7a*, *ORF8*, *N*, and *ORF10*. The mutation rate was calculated based according to the sample collection date, and then transformed into units of mutations per year and per 1kb.

### Screening, identification and relativeness of ORF8 deletion in our study with the strains in whole Japan retrieved from GISAID

We used the aforementioned similar partial sequencing method to identify and screen the remaining 60 SARS-CoV-2 samples of Hiroshima using the primer hCoV-ORF8 set (Table 1) targeting the ORF8 region of all samples. Thereafter, we searched through GISAID for the similar ORF8 deletion mutants in the whole Japan and then investigate the molecular pattern between our study and the reported strains.

### Ethic consideration

This study was approved by the Ethics Committee of Hiroshima University (E2124). All participants provided the written informed consent at the beginning of the study. All procedures strictly adhered to the guidelines and the Declaration of Helsinki.

## Results

### Quantitative measurement of SARS-CoV-2

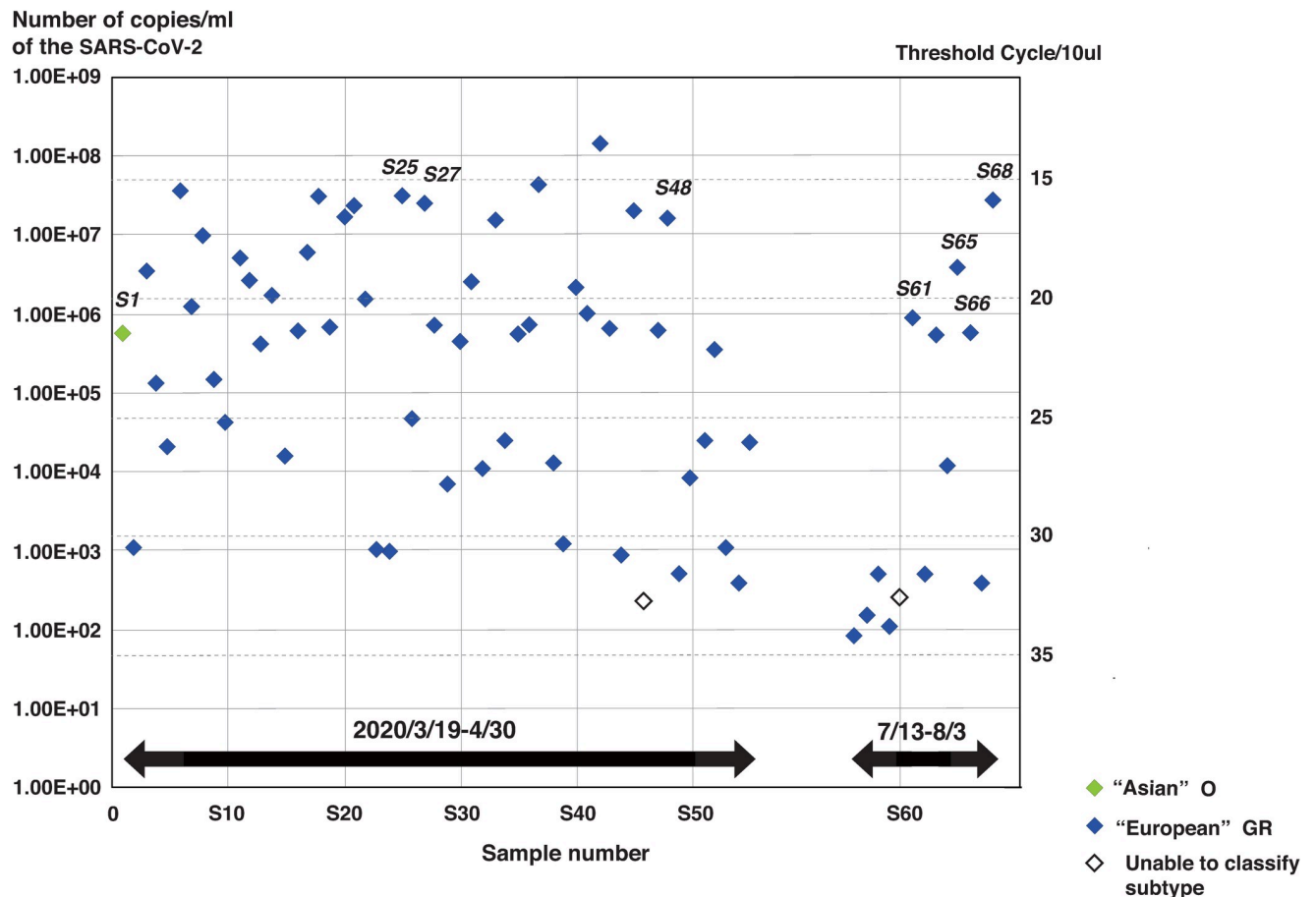
Using RT-qPCR, the Asian type yielded a viral load of  $5.45 \times 10^5$  copies/ml. European strains from the first wave had viral loads of  $2.18 \times 10^2$ – $1.42 \times 10^8$  copies/ml, and those from the second wave had viral loads of  $7.85 \times 10^1$ – $9.16 \times 10^6$  copies/ml (Fig 2).

### GISAID clade (genotype) of SARS-CoV-2 in Hiroshima

Sixty-six (66) of 68 samples were available for partial sequencing to classify GISAID clade (97.1%), whereas two samples were undetectable (2.9%). Both of those undetectable samples had low viral loads,  $2.42 \times 10^2$  and  $2.18 \times 10^2$  copies/ml, respectively (Fig 2). Among two samples collected 1 week apart in March, the first isolate was Asian type O and the latter was European type GR. Thereafter, all remaining samples were collected between April 2020 and August 2020, covering the first and second waves; all were European type GR.

### Phylogenetic tree analysis

Using all isolates confined to Japan and controls retrieved from GISAID and GenBank, a phylogenetic tree was constructed and categorized by region of Japan: from North to South, Hokkaido, Tohoku, Kanto, Chubu, Kansai, Chushikoku (originally Chugoku and Shikoku separately, but combined in our study), and Kyushu (Fig 3). In Japan, the Asian type of SARS-CoV-2 spread



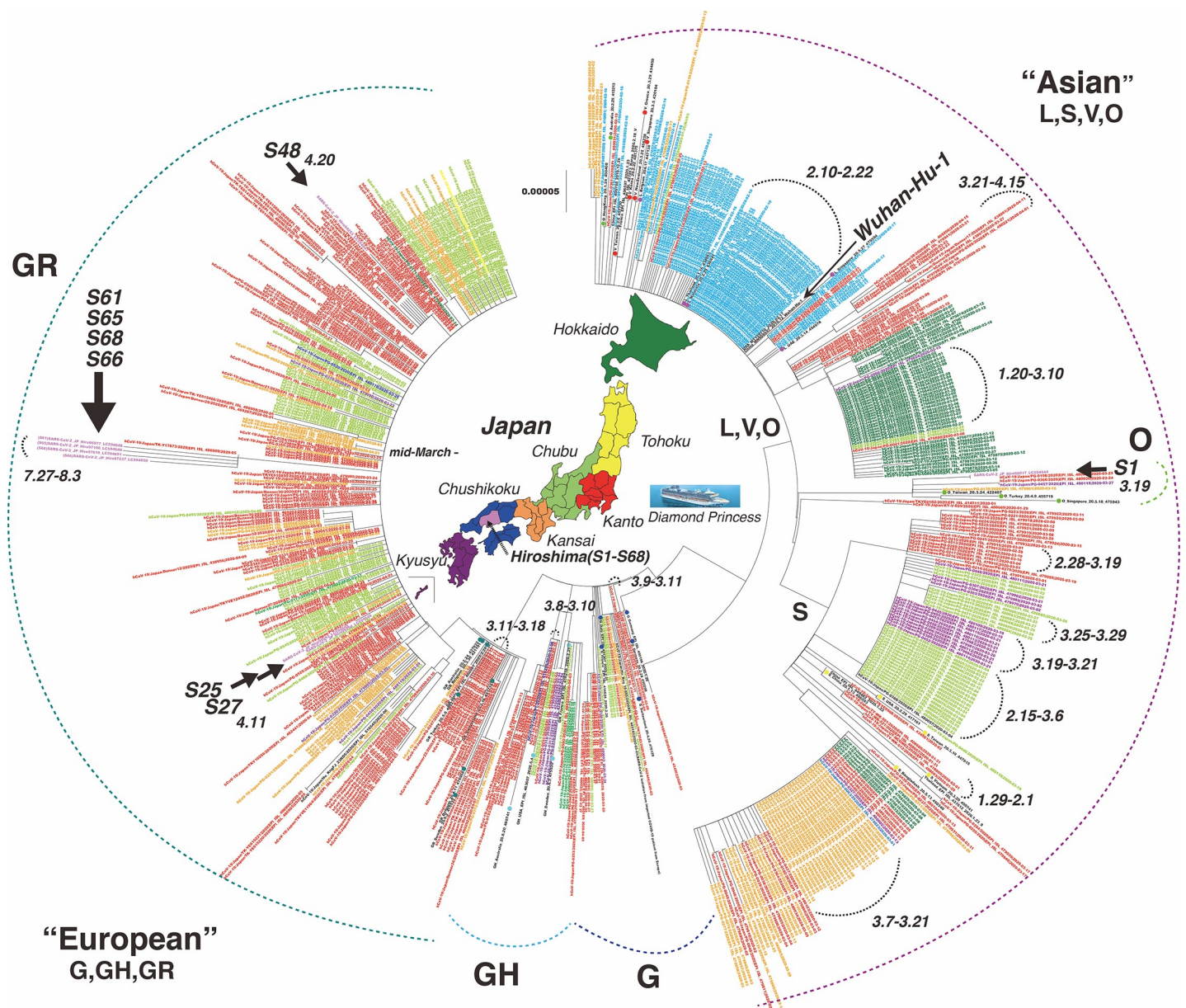
**Fig 2. Variation in viral titers of SARS-CoV-2 in Hiroshima.** Quantitative measurement of SARS-CoV-2 in 68 samples classified by GISAID clades. Two could not be classified as GISAID clade subtype G, GH, or GR; one was Asian strain O (green dot); and 65 were European strain GR (shown by blue dots). Titers were plotted against cycle threshold for each sample.

<https://doi.org/10.1371/journal.pone.0246383.g002>

from January to April 2020, and many small clusters of cases occurred transiently in the Chubu, Kansai, and Hokkaido regions, as well as the on the Diamond Princess Cruise. Various types of Asian clades were circulating in these areas. Later, starting in the first week of March, clusters of European types G and GH appeared. Since mid-March, a number of European GR clusters have been observed simultaneously across the country; subsequently, this strain became the predominant type in Japan. In Hiroshima, analyses of eight full genomes revealed that one (S1: SARS-CoV-2\_JP\_Hiro66017\_LC584644) isolate belonged to Asian type O; the rest were European type GR, in which S25 (SARS-CoV-2\_JP\_Hiro99262\_LC594645), S27 (SARS-CoV-2\_JP\_Hiro99268\_LC594646), and S48 (SARS-CoV-2\_JP\_Hiro94253\_LC594647) were circulating in the first wave and S61 (SARS-CoV-2\_JP\_Hiro96977\_LC594648), S65 (SARS-CoV-2\_JP\_Hiro97166\_LC594649), S66 (SARS-CoV-2\_JP\_Hiro97237\_LC594650), and S68 (SARS-CoV-2\_JP\_Hiro97618\_LC594651) were circulating in the second wave (Fig 3). The second wave GR strains appeared independently with 11–15-base mutations.

### Mutations and the subsequent amino acid changes in SARS-CoV-2 strains in Hiroshima

Relative to reference strains (MN908947), seven different types of mutation were found at 44 different sites among the eight isolates of SARS-CoV-2; 59.1% (26/44) were nonsynonymous



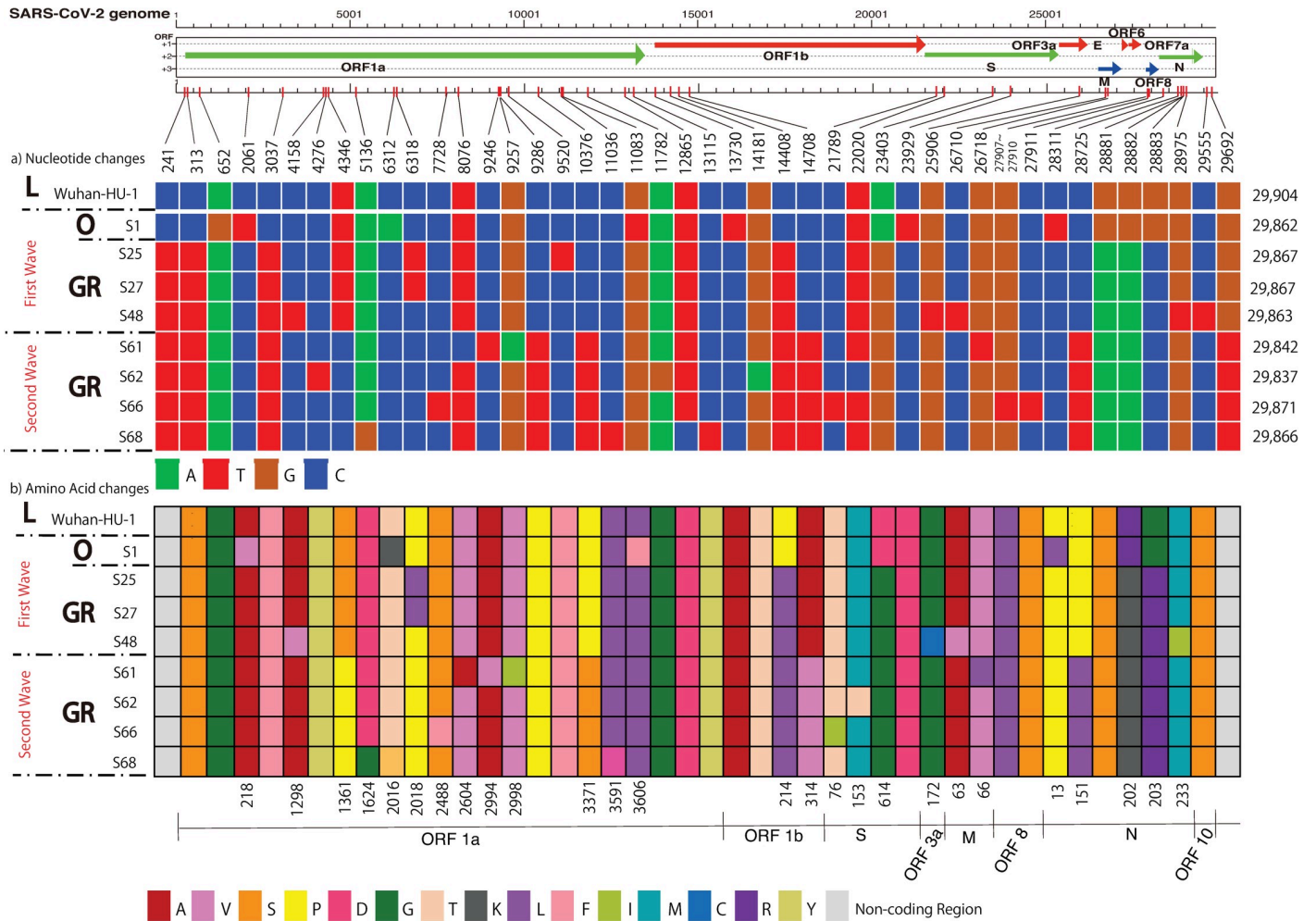
**Fig 3. Phylogenetic tree of SARS-CoV-2 in Japan, including eight SARS-CoV-2 isolates from Hiroshima.** This phylogenetic tree includes 52 control strains from around the world, 522 strains from Japan retrieved from GISAID and GenBank, and 8 SARS-CoV-2 strains isolated in this study. Controls are shown as colored dots. The strains from Japan are indicated by colored letters corresponding to the colors shown in the map of Japan, in which the country is divided into seven main regions. The sky-blue color represents the strains from the Diamond Princess cruise. Our eight SARS-CoV-2 strains are shown in purple and are also indicated by sample number. The tree was performed by the neighbor-joining method using MEGA7. There were total of 29038 positions in the final dataset.

<https://doi.org/10.1371/journal.pone.0246383.g003>

mutations, and the rest (40.9%) were synonymous mutations. Among these 44 sites, *ORF1a* and *1b* contained 27 changes (61.4%), nucleocapsid (*N*) contained five changes (11.4%), and spike (*S*) had four changes (9.1%) (Fig 4).

Seven nucleotide mutations were found in the Asian O strain, amongst which five were located in *ORF1a*; one each was in *S* or *N*. All European type GR strains in the first wave had similar mutations at eight different sites: one in a non-coding region (C241T), two in *ORF1a* (C313T, C3037T), one in *ORF1b* (C14408T), one in *S* (A23403G), and three in *N* (G28881A, G28882A, G28883C); half were nonsynonymous. Additionally, several discrete mutations





**Fig 4. Nucleotide mutations and amino acid replacements in the whole genomes of SARS-CoV-2 isolates from Hiroshima.** This figure shows a) nucleotide mutations and b) resultant amino acid changes identified in eight SARS-CoV-2 isolates from Hiroshima. Colors represent nucleotides or amino acids, according to the legend below each panel. Genomic regions are indicated by the whole-genome model at the top, and point mutations are indicated by the color changes.

<https://doi.org/10.1371/journal.pone.0246383.g004>

were also found at various sites in particular strains, as shown in Fig 3. Overall, there were 32 mutations at 15 sites in European type strains of the first wave: 50% (16/32) were in *ORF1ab*, 31.3% (10/32) in *N*, 9.4% (3/32) in *S*, and 3.1% (1/32) each in *ORF3a*, *ORF10*, and *M* (Fig 4).

The aforementioned eight similar mutations also occurred in all European strains of the second wave. Additionally, another similar pattern of mutations was shared among these strains at six different sites: three (T4346C, C9286T, C10376T), one (C14708T), one (C28725T), and one (G29692T) in *ORF1a*, *ORF1b*, *N*, and non-coding regions, respectively; four of them were nonsynonymous. Moreover, 16 discrete nucleotide changes were also found at various sites in particular strains (Fig 3). Thus, there were 72 mutations at 30 sites in European strains of the second wave: 59.7% (43/72) were in *ORF1ab*, 22.2% (16/72) in *N*, 8.3% (6/72) in *S*, 5.5% (4/72) in a non-specific region, 2.8% (2/72) in *ORF8*, and 1.4% (1/72) in *M* (Fig 3). Notably, all seven European strains shared a triple mutation at G28881A, G28882A, and G28883C in the *N* gene. The estimated mutation rate was 38.4 nucleotides per year and  $1.16\text{--}1.87 \times 10^{-3}$  base substitutions per site per year. The number of mutations within the first wave ranged from one to seven changes, but in the second wave, it was eight changes.

Four amino acid substitutions were found in Asian strain O: three (218: Arginine to Glycine, 2016: Threonine to Lysine, 3606: Leucine to Phenylalanine) were in ORF1a and one (13: Proline to Leucine) was in N. Four similar patterns of amino acid substitutions were found in all European strains from both the first and second waves: ORF1b had one (214: Proline to Leucine), spike had one (614: Aspartic acid to Glycine), and nucleocapsid had two (2020: Arginine to Lysine and 203: Glycine to Arginine). Notably, all four European strains from the second wave shared four similar patterns of amino acid substitutions (1361: Serine to Proline and 3371: Proline to Serine in ORF1a, 314: Alanine to Valine in ORF1b, and 151: Proline to Lysine in nucleocapsid). Moreover, the rest of the amino acid substitutions were individually distributed at various sites in particular European strains. The majority of amino acid substitutions were found in non-structural proteins (ORF1a and 1b), accounting for 57.7% of total substitutions. Among structural proteins, N (15.4%) and S (11.5%) exhibited the most frequent amino acid substitutions sites in all European strains, and other substitutions were also found in ORF3a and M (Fig 4).

### The D614G variants of European strain GR

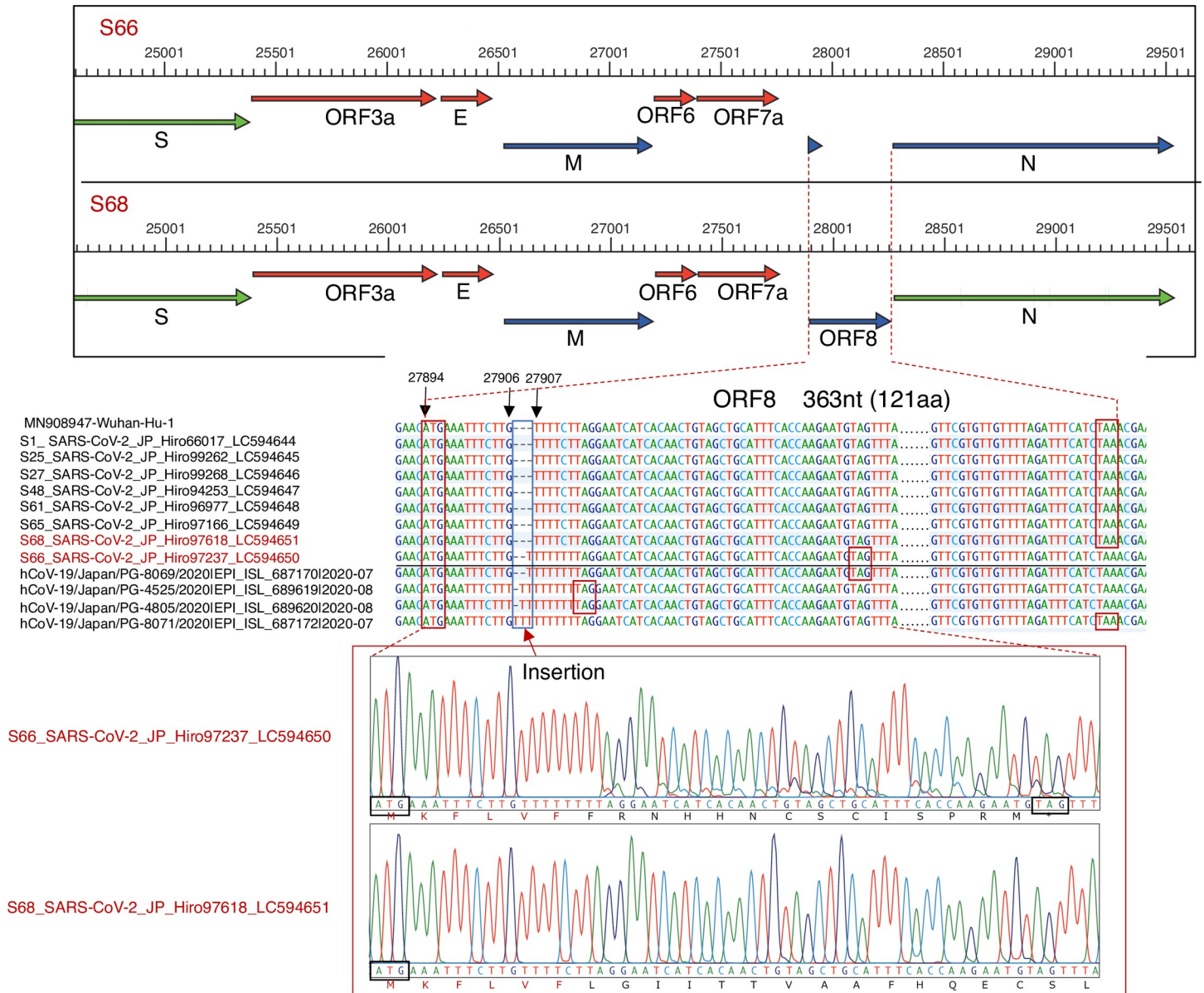
The D614G amino acid change was found in the spike region of European strain GR, and this amino acid change was accompanied by the silent mutation of C241T in a non-coding region, C313T and C3037T in *ORF1a*, the nonsynonymous mutation at C14408T (P214L) in *ORF1b*, and triple mutation of G28881A, G28882A, and G28883C resulting in amino acid changes (R202K and G203R) in N. Notably, the spike region contained four nucleotide mutations resulting in three amino acid changes, as shown in Fig 3. Additionally, the D614G variant of S65 had the T22020C mutation in the same region, and S66 had the C21789T mutation in the spike region; both of these mutations resulted in amino acid changes (153: Methionine to Threonine and 76: Threonine to Isoleucine) (Fig 4).

### Single-base insertion in SARS-CoV-2\_JP\_Hiro97327\_LC594650

A single-nucleotide insertion was found in sample S66, collected during the second wave of the outbreak in Hiroshima. The nucleotide insertion was assumed to have occurred between nt27906 and nt27907 of ORF8, which contains 363 nucleotides (121 amino acid). This single-nucleotide insertion caused a frameshift and consequently the disappearance of ORF8, as shown in Fig 4. Instead of the usual length of 121 amino acids in the wild type, ORF8 of this mutant variant was shortened to 20 amino acids. As a result, only initial six amino acids were the same as the wild type, and the rest differed due to the frameshift. (Fig 5) Among 9,844 SARS-CoV-2 strains of the whole Japan submitted at GISAID as of January 12, 2021, we found only 4 mutant strains having similar nucleotide insertion (one has single base insertion, two have double base insertion and one has triple base insertion) with subsequent frameshift and early appearance of stop codon as shown in Fig 5. As per clinical profile, this patient, 60 years old man, infected with ORF8 deletion mutant, had mild symptoms only and cured after 13 days of hospitalization.

### Discussion

This study revealed the molecular characteristics and mutation pattern of SARS-CoV-2 strains circulating in Hiroshima from the day of first confirmed case to August 3, 2020. At that time, the total number of cases in Hiroshima Prefecture was 366 while the whole country has reported 40,244 new confirmed cases. It means Hiroshima had 0.9% of total cases in Japan at that time. Among them, 68 samples were provided from Hiroshima Prefecture Research Institute after being used by government laboratory and we performed real-time PCR for



**Fig 5. Single-nucleotide insertions in open reading frame (ORF8) resulting in frameshift and ORF8 deletion.** The framed box at the top compares ORF8 between wild-type S68 and mutant variant S66, using whole-genome sequences obtained by next-generation sequences (NGS). Disappearance of ORF8 was observed in the sample. Additionally, 4 SARS-CoV-2 strains of Japan having the similar mutation pattern retrieved from GISAID were also included in this figure. This small peak results from mixed cloning of SARS-CoV-2 mutants with single and double Thymine (T) nucleotide insertion between nt27906 and nt27907 whereas the large dominant wave represent the dominant mutants single Thymine insertion and small peak represents the recessive mutants with double Thymine insertion.

<https://doi.org/10.1371/journal.pone.0246383.g005>

quantitative measurement of virus and region specific-partial sequencing for GISAID classification as the first step. Then, we randomly selected 8 samples to undergo NGS full genome sequencing. As of January 14, 2021, 9,885 strains (3.3%) out of 298,172 confirmed cases in Japan are able to register at GISAID. Our study could report 8 (2.2%) out of 366 almost the same as the whole Japan. Interestingly, among 68 samples, only one strain collected during the first wave was of the Asian type; the remaining strains from both the first and second wave of the outbreak were GISAID clade GR. Our phylogenetic tree analysis explained the evolution of SARS-CoV-2 in Japan, although the origin and cascade of transmission were difficult to

determine. From January through April 2020, a large number of case clusters occurred on the Diamond Princess Cruise and in the Hokkaido, Kanto, Kansai, Chubu, and Kyushu regions; these clusters shared various Asian strains (L, S, V, and O) and were confined to particular areas. In the second week of March, the European strains G, GH, and GR entered Japan and were mainly found in the Kanto region. After mid-March, European strain GR was predominant throughout the whole of Japan. The full-length genomes obtained in our study revealed that all seven European strains were genotype GR (specifically, D614G variants). Japan had an urgent response to the outbreak, and the alert level was upgraded in late March. Our study showed that the Asian strains were less likely to cause further outbreaks after the strict measures against coronavirus, and were no longer reported in Japan. This implies Asian strain had a low level of replication and was less transmissible than the European strains. Therefore, the strict public health measures could interrupt the chain of transmission. By contrast, the European strains have higher infectivity, virulence, and replication rate than the Asian type, especially the D614G mutant variants [11]. Our study also indicated that European strains had higher viral loads than Asian strain, suggesting higher replication efficiency, and all 7 European strains of type GR had high mutation rates. Under strict public health measures, new cases decreased temporarily, but evolution of SARS-CoV-2 occurred during this period. Subsequently, new episodes of outbreak was occurred by newly emerged mutant variants with differences in virulence, infectivity, replication, and transmissibility.

Among all proteins in SARS-CoV-2 strains in Hiroshima, both Asian and European strains had amino acid substitutions as a result of specific single-base mutations. Because SARS-CoV-2 is an RNA virus, the mutation rate is high, leading to continued evolution over time that might affect replication, infectivity, transmissibility, virulence, and immunogenicity [12]. Our study also showed that the mutation rate between the first and second waves was 34.9–40.6 nucleotides per year, corresponding to  $1.17\text{--}1.36 \times 10^{-3}$  base substitutions per site per year. This mutation rate is lower than that of hepatitis C virus [13]. A recent study reported that the mutation rate of SARS-CoV-2 in variable sites was  $3.5 \times 10^{-3}$  changes per site per year, higher than the value we measured [14]. Interestingly, all European strains shared similar patterns of mutation at specific sites, and the accompanied amino acid substitutions were mostly found in polyprotein ORF1ab rather than in structural proteins. Md Rafiqul Islam *et al.* also reported that most nucleotide mutations occurred in the polyprotein, with ORF1a being the most frequently mutated region [15]. Our findings are consistent with previous reports. Mutations in ORF1a and 1b might influence the replication efficiency of the virus because they are essential for the main protease (Mpro) and replicase involved in processing of polyproteins and control of viral replication [16, 17].

Notably, we found a mutation in the nucleocapsid protein in all seven GR strains from Hiroshima: a triple mutation (G28881A, G28882A, and G2888C) resulting in an amino acid change. Moreover, four GR strains from the second wave had the same mutation at C28725T. The nucleocapsid proteins play important roles in RNA replication, transcription, and genome packaging [18, 19]. The high frequency of mutation in the genomes, particularly in the structural proteins, promotes the emergence of new strains of the virus with differences in morphological structure, virulence, infectivity capacity, and replication efficiency. This phenomenon complicates the development of vaccine candidates and drugs. Additionally, we identified a predominant type of D614G variants with the concomitant silent mutations C241T, C313T, and C3037T in ORF1a; the P214L amino acid substitution in S; and double mutation of R202K and G203R in N. A recent study suggested that D614G variants became the predominant type circulating worldwide; these strains account for 78% of total distribution as of May 2020 and are associated with increased infectivity [20] and a high case-fatality rate [11]. In light of the alterations in infectivity and transmissibility of D614G variants, our findings provide



important insights into clinical management of SARS-CoV-2, as well as how we might modify preventive and control strategies in the future. In addition to the D614G mutation, the spike region contained two different types of mutations, M153T in S65 and T76I in S66. The spike proteins are key surface glycoproteins that play essential roles in viral entry and interact with host cell receptors [21, 22]. Hence, a high mutation in the spike region might interfere with the development of vaccines or antibody-based therapeutic agents.

Another noteworthy finding is the detection of a single-nucleotide insertion between nt27906 and 27907, resulting in deletion of the ORF8 protein. A SARS-CoV-2 variant with a 382-nucleotide deletion was reported in Singapore, Bangladesh, Australia, and Spain [23–25]. This type of deletion removes an ORF8 regulatory sequence, eliminating ORF8 transcription; the variant is associated with milder infectivity and lower levels of cytokine release during the acute phase [25]. Although the role of ORF8 of SARS-CoV-2 remains unclear, one study suggested that it disrupts antigen presentation and decreases the recognition and elimination of virus-infected cells by cytotoxic T lymphocytes (CTLs) [26]. In our study, we suggested that single base insertion in ORF8 causes the subsequent amino acid frameshift and early appearance of the stop codon, with concomitant disappearance of ORF8. We found only one case among eight full-length genome and 60 partial genomes. Additionally, only 4 similar mutants strains were reported among 9844 strains of the whole Japan submitted at GISAID. Therefore, we suggested that the transmission of ORF8 deletion mutants was temporarily occurred during the second wave showing the less longevity, less viral replication, less transmission and less infectivity. The finding is consistent with the recently published report [24].

Our study had some limitations. Because of insufficient sample volume, it is impossible for all samples to do full-genomes sequencing. However, we decreased selection bias by random blinded selection of eight samples for whole-genome sequences analysis but the random error cannot be excluded. Although the sample size is small, this is the very first report about the full-genomes of SARS-CoV-2 strain in Hiroshima as well as the Japan, so that it is very precious data and findings which can be used as the baseline for the further molecular surveillance as well as the clue for development of vaccine, immunotherapy and diagnostic tool. Moreover, it is necessary to conduct the large study to correlate between mutation result and their clinical outcomes in patients, infectivity, virulence, or viral factors in the future. As our study focused on phylogeny, molecular characteristics and their mutation patterns, we did not examine detailed morphological changes or alterations of biological functions caused by the mutations we identified.

In conclusion, our findings revealed that the recently emerged strain of SARS-CoV-2 in Hiroshima was European clade GR. The transmission cascade and high mutation rate might be associated with alteration of infectivity and transmissibility of SARS-CoV-2 strains in this region. The predominant D614G variants and a new form of ORF8 deletion also provide the clues for the role of viral factor in local outbreak of SARS-CoV-2 in Hiroshima. Understanding the molecular characteristics and mutation patterns over time would be useful in clinical settings, in design of public health measures, and for development of vaccines and therapeutic drugs. Further study and molecular surveillance of circulating strains are required. In addition, it will be necessary to optimize clinical management of novel coronavirus infection.

## Acknowledgments

The authors acknowledge the public health and environment center of Hiroshima Prefectural Technology Research Institute for valuable support of this study. We also wish to acknowledge all of those who contributed and deposited sequences at GISAID and GenBank.

## Author Contributions

**Conceptualization:** Kazuaki Takahashi, Junko Tanaka.

**Data curation:** Tomoyuki Akita, Aya Sugiyama, Masayuki Ohisa, Junko Tanaka.

**Formal analysis:** Ko Ko, Tomoyuki Akita, Kazuaki Takahashi, Junko Tanaka.

**Funding acquisition:** Junko Tanaka.

**Investigation:** Ko Ko, Shintaro Nagashima, Bunthen E., Serge Ouoba.

**Methodology:** Ko Ko, Kazuaki Takahashi, Junko Tanaka.

**Project administration:** Ko Ko, Aya Sugiyama, Eisaku Kishita, Masao Kuwabara, Junko Tanaka.

**Resources:** Shintaro Nagashima, Bunthen E., Serge Ouoba, Eisaku Kishita, Masao Kuwabara.

**Software:** Bunthen E., Serge Ouoba, Tomoyuki Akita, Masayuki Ohisa.

**Supervision:** Shintaro Nagashima, Aya Sugiyama, Hidetoshi Tahara, Eisaku Kishita, Masao Kuwabara, Kazuaki Takahashi, Junko Tanaka.

**Validation:** Ko Ko, Shintaro Nagashima, Tomoyuki Akita, Aya Sugiyama, Takemasa Sakaguchi, Hidetoshi Tahara, Hiroki Ohge, Hideki Ohdan, Tatsuhiko Kubo, Kazuaki Takahashi.

**Visualization:** Bunthen E., Serge Ouoba, Takemasa Sakaguchi, Hiroki Ohge, Hideki Ohdan, Tatsuhiko Kubo, Kazuaki Takahashi, Junko Tanaka.

**Writing – original draft:** Ko Ko.

**Writing – review & editing:** Kazuaki Takahashi, Junko Tanaka.

## References

1. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*. 2020; 5(4):536–44. <https://doi.org/10.1038/s41564-020-0695-z> PMID: 32123347
2. Belouzard S, Millet JK, Licitra BN, Whittaker GR. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses*. 2012; 4(6):1011–33. <https://doi.org/10.3390/v4061011> PMID: 22816037
3. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. *Nature Reviews Immunology*. 2020; 20(6):363–74. <https://doi.org/10.1038/s41577-020-0311-8> PMID: 32346093
4. Ramani T, Auletta CS, Weinstock D, Mounho-Zamora B, Ryan PC, Salcedo TW, et al. Cytokines: The Good, the Bad, and the Deadly. *International Journal of Toxicology*. 2015; 34(4):355–65. <https://doi.org/10.1177/1091581815584918> PMID: 26015504
5. (WHO) WHO. Corona Virus disease (COVID-19) Situation Report. 2020.
6. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. *Acta bio-medica: Atenei Parmensis*. 2020; 91(1):157–60. <https://doi.org/10.23750/abm.v91i1.9397> PMID: 32191675
7. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*. 2020; 20(5):533–4. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1) PMID: 32087114
8. Office HP. Annual Report on Population Statistics from Basic Resident Registration. 2020.
9. Office HP. Report on Novel Corona Infection In Hiroshima. 2020.
10. Stecher G, Kumar S, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*. 2016; 33(7):1870–4. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
11. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of Human Genetics*. 2020. <https://doi.org/10.1038/s10038-020-0808-9> PMID: 32699345

12. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*. 2020; 18(1):179. <https://doi.org/10.1186/s12967-020-02344-6> PMID: 32321524
13. Yamamoto C, Nagashima S, Chuon C, Ko K, Huy Do S, Lim O, et al. Substitution of the CD81 Binding Site and  $\beta$ -Sandwich Area in E2 of HCV in Cambodia. *Viruses*. 2020; 12(5). <https://doi.org/10.3390/v12050551> PMID: 32429467
14. Salemi M, Fitch WM, Ciccozzi M, Ruiz-Alvarez MJ, Rezza G, Lewis MJ. Severe acute respiratory syndrome coronavirus sequence characteristics and evolutionary rate estimate from maximum likelihood analysis. *Journal of virology*. 2004; 78(3):1602–3. <https://doi.org/10.1128/jvi.78.3.1602-1603.2004> PMID: 14722315
15. Islam MR, Hoque MN, Rahman MS, Alam ASMRU, Akther M, Puspo JA, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Scientific Reports*. 2020; 10(1):14004. <https://doi.org/10.1038/s41598-020-70812-6> PMID: 32814791
16. Lindner HA, Fotouhi-Ardakani N, Lytvyn V, Lachance P, Sulea T, Ménard R. The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J Virol*. 2005; 79(24):15199–208. <https://doi.org/10.1128/JVI.79.24.15199-15208.2005> PMID: 16306591
17. Shimamoto Y, Hattori Y, Kobayashi K, Teruya K, Sanjoh A, Nakagawa A, et al. Fused-ring structure of decahydroisoquinolin as a novel scaffold for SARS 3CL protease inhibitors. *Bioorg Med Chem*. 2015; 23(4):876–90. <https://doi.org/10.1016/j.bmc.2014.12.028> PMID: 25614110
18. Hsin WC, Chang CH, Chang CY, Peng WH, Chien CL, Chang MF, et al. Nucleocapsid protein-dependent assembly of the RNA packaging signal of Middle East respiratory syndrome coronavirus. 2018; 25(1):47. <https://doi.org/10.1186/s12929-018-0449-x> PMID: 29793506
19. Masters PS. Coronavirus genomic RNA packaging. *Virology*. 2019; 537:198–207. <https://doi.org/10.1016/j.virol.2019.08.031> PMID: 31505321
20. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 2020; 182(4):812–27.e19. <https://doi.org/10.1016/j.cell.2020.06.043> PMID: 32697968
21. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. 2020; 367(6485):1444. <https://doi.org/10.1126/science.abb2762> PMID: 32132184
22. Tai W, He L, Zhang X, Pu J, Voronin D. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. 2020; 17(6):613–20. <https://doi.org/10.1038/s41423-020-0400-4> PMID: 32203189
23. Gong Y-N, Tsao K-C, Hsiao M-J, Huang C-G, Huang P-N, Huang P-W, et al. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. *Emerg Microbes Infect*. 2020; 9(1):1457–66. <https://doi.org/10.1080/22221751.2020.1782271> PMID: 32543353
24. Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, 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(4):e01610–20. <https://doi.org/10.1128/mBio.01610-20> PMID: 32694143
25. Young BE, Fong S-W, Chan Y-H, Mak T-M, Ang LW, Anderson DE, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet*. 2020; 396(10251):603–11.
26. Zhang Y, Zhang J, Chen Y, Luo B, Yuan Y, Huang F, et al. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. *bioRxiv*. 2020:2020.05.24.111823.