

An epidemiological study of the SARS-CoV-2 genome in Japan (updated: April 16, 2020)

■ Active, nationwide epidemiological surveillance of COVID-19 clusters in Japan

In late December 2019, the initial outbreak of the “coronavirus disease 2019” (COVID-19) occurred in Wuhan, China. It was caused by a new strain of beta coronavirus, known as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) ¹⁻³. After the identification of the first COVID-19 patient in Japan on January 15, 2020, multiple local COVID-19 clusters were identified nationwide by the end of February. The Japanese government focused on identifying and mitigating emerging COVID-19 clusters before they could spread further. In an effort to contain these clusters and limit the number of new cases, active, nationwide epidemiological surveillance of each cluster was conducted in order to identify the close contacts of existing COVID-19 patients.

Although some of the COVID-19 clusters were successfully contained, the number of cases continued to increase. The Japanese government declared a nationwide state of emergency in view of the worsening spread. To support the ongoing epidemiological surveillance, we collaborated with the local public health institutes in Japan (listed below), and conducted whole genome sequencing of SARS-CoV-2. Our goal was to apply genomic epidemiology to predict potential routes of infection within or between clusters ⁴.

■ Improvement of the ARTIC Network’s multiplex polymerase chain reaction (PCR) for SARS-CoV-2 genome sequencing

Whole genome sequences of SARS-CoV-2 were obtained by means of the PrimalSeq protocol for enriching the cDNA of the SARS-CoV-2 genome using multiplex RT-PCR, as proposed by the Wellcome Trust ARTIC Network ⁵. We found two amplicons that regularly showed low to zero coverage due to primer dimerization as described by Itokawa et al. ⁶. Therefore, we used a modified primer for the multiplex PCR amplification ⁶. The next-generation sequencing reads were mapped to the SARS-CoV-2 Wuhan-Hu-1 reference genome sequence (29.9 kb ss-RNA; GenBank ID: MN908947), followed by *de novo* assembly to obtain the full genome

sequence.

■ Genomic epidemiology based on the full genome sequence of SARS-CoV-2

The nearly full-length genome sequences (≥ 29 kb, 4,511 entries) of SARS-CoV-2 were retrieved from the GISAID EpiCoV database (updated on April 16, 2020) ⁷. We determined the full genome sequences using 562 clinical specimens in Japan, and performed a haplotype network analysis using genome-wide single nucleotide variations (SNVs) to trace potential infection routes based on genealogy (Fig. 1). Thus far, the mutation rate and frequency has been estimated at 25.9 SNVs/genome/year using the molecular clock technique ⁸, suggesting that throughout the whole genome sequence at least 9 SNVs may have emerged during the four months since the end of 2019.

Some of the primary clusters identified through January and February in Japan (red circle ● in Fig 1) have been identified around the Wuhan-Hu-1-related isolates from China. These clusters appeared to decrease and are the most containable clusters till date.

Passengers aboard the Diamond Princess (DP) cruise ship were quarantined shortly after its return to Yokohama, Japan on February 3, 2020, because an 80-year-old passenger who disembarked in Hong Kong was confirmed to have COVID-19 on February 1, 2020. We determined the SARS-CoV-2 genome sequences from 70 individuals aboard the DP cruise ship (148 positive individuals among 895 passengers) ⁹ and most of the isolates showed one SNV (G₁₁₀₈₃T) (magenta circle ● in Fig. 1). These DP-related isolates have not yet been detected in Japan, and have not spread throughout the country.

The aforementioned primary wave of infections originating in China appeared to have been mostly contained through implementation of active surveillance in Japan. In contrast, the number of COVID-19 cases increased rapidly across the Europe and North America during early March, indicative of a pandemic. Concurrently, many sporadic COVID-19 cases were detected in Japan from the end of March through early April. The haplotype network analysis demonstrated that an additional large cluster (red circle ● at the upper left corner in Fig. 1) around Europe-related isolates (originating in China or other countries) was predominantly observed in Japan during early April, suggesting that they could have caused additional domestic spread within Japan.

■ Genomic epidemiology to support active epidemiological surveillance

Analysis of whole genome sequences provides information that is useful for tracing back infection using genome-wide SNVs among clusters. During the primary wave, local public health centers were able to identify likely COVID-19 patients and their close contacts within location-specific clusters by conducting active epidemiological surveillance using the keywords “Wuhan, Hubei, and China.” Just as the situation had begun to improve around mid-March, a large number of new COVID-19 patients were diagnosed, many of with unclear infection routes. Tracing the infection routes was difficult because some Japanese cases had no recent history of travel to China or any country outside Japan. The SARS-CoV-2 haplotype network analysis suggested that the second wave of COVID-19 cases could have been imported via the returnees from Europe, North America, or other countries.

This report suggests that haplotype network analysis using genome-wide SNVs can support epidemiological field investigations and highlight potential infection linkages. A project to determine viral evolution and its spread across the community is being conducted by researchers in the US and UK ^{10,11}. To mitigate the next wave of COVID-19 in Japan, it is necessary to formulate a more efficient containment strategy.

Acknowledgements:

We sincerely thank the COVID-19 working group in Japan listed below for the collection and transportation of clinical specimens. We are particularly grateful to the staff of the National Institute of Infectious Diseases, Field Epidemiology Training Program (FETP) team, the Ministry of Health, Labor and Welfare, and local governments for their assistance with administrative matters, field investigations, data collection, and laboratory testing.

We would like to thank all the researchers who have kindly deposited and shared genomic data on GISAID.

This study was supported by a Grant-in Aid from the Japan Agency for Medical Research and Development (AMED) under Grant Numbers JP19fk0108104 and JP20fk0108103).

Correspondence:

Makoto Kuroda, PhD

Director, Pathogen Genomics Center, National Institute of Infectious Diseases.

1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan.

Phone: +[81]-3-5285-1111

makokuro@niid.go.jp

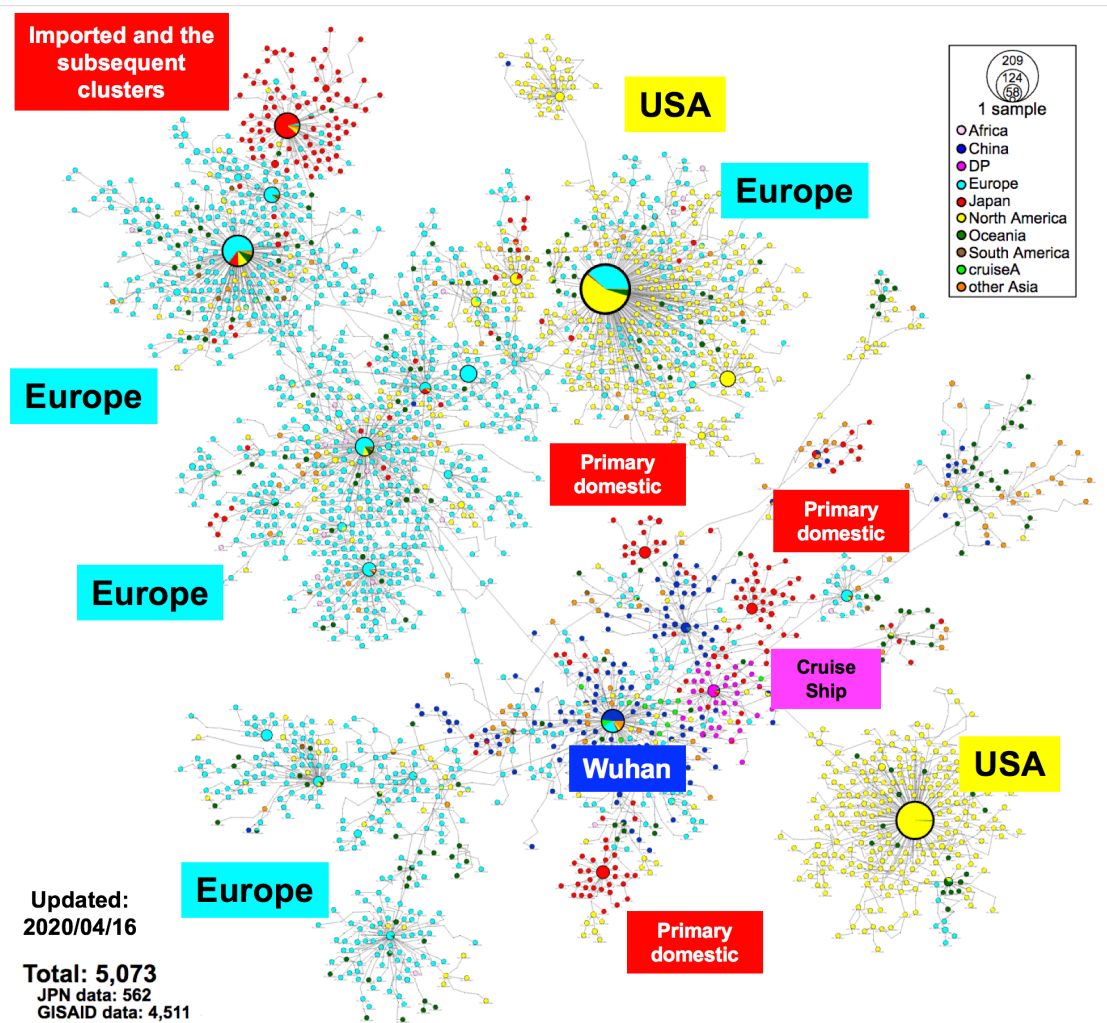


Fig. 1. Haplotype network analysis using genome-wide single-nucleotide variations (HN-GSNVs) of worldwide SARS-CoV-2 isolates.

Whole-genome sequences of SARS-CoV-2 isolates in Japan ($n = 562$) were compared with all GISAID-available SARS-CoV-2 genomes ($n = 4,511$, updated on April 16, 2020) using median-joining SNV network analysis. SARS-CoV-2 that disseminated from Wuhan City, China, at the end of December, 2019 (one of the potential origins of Wuhan-Hu-1) is plotted at the center of the haplotype network.

Collaboration with local public health institutes in Japan:

Tokyo Metropolitan Institute of Public Health

Kobe Institute of Health

Sakai City Public Health Research Institute

Gunma Prefectural Institute of Public Health and Environmental Sciences

Wakayama Prefectural Institute of Public Health and Environmental Science

Nagoya City Public Health Research Institute

Saitama Prefectural Institute of Public Health,

Chiba Prefectural Institute of Public Health

Oita Prefectural Institute of Public Health and Environmental Science

Gifu Prefectural Institute of Public Health and Environmental Sciences

Shizuoka City Public Health Research Institute

Ishikawa Prefectural Institute of Public Health and Environmental Science

Himeji City Public Health Research Institute

Sagamihara City Public Health Research Institute

Kumamoto City Public Health Research Institute

Ota-ku Public Health Center

Toyama Institute of Health

Fukui Prefectural Institute of Public Health and Environmental Science

Osaka Institute of Public Health

Nagasaki Prefectural Institute of Public Health and Environmental Science

Eighteen additional institutes

References

- 1 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature*, doi:10.1038/s41586-020-2008-3 (2020).
- 2 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, doi:10.1038/s41586-020-2012-7 (2020).
- 3 Coronaviridae Study Group of the International Committee on Taxonomy of, V. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*, doi:10.1038/s41564-020-0695-z (2020).
- 4 <https://www.niid.go.jp/niid/images/epi/corona/2019nCoV-02-200420.pdf>.

- 5 ARTIC Network protocol. <https://artic.network/ncov-2019>.
- 6 Itokawa, K., Sekizuka, T., Hashino, M., Tanaka, R. & Kuroda, M. A proposal of an alternative primer
for the ARTIC Network's multiplex PCR to improve coverage of SARS-CoV-2 genome sequencing.
bioRxiv, doi:10.1101/2020.03.10.985150 (2020).
- 7 <https://www.gisaid.org/epiflu-applications/next-hcov-19-app/>.
- 8 <https://nextstrain.org/ncov?!=unrooted>.
- 9 Sekizuka, T. *et al.* Haplotype networks of SARS-CoV-2 infections in the Diamond Princess cruise
ship outbreak <https://www.medrxiv.org/content/10.1101/2020.03.23.20041970v1> *medRxiv*,
doi:10.1101/2020.03.23.20041970 (2020).
- 10 <https://www.cogconsortium.uk/>.
- 11 <https://covidgenomics.org/>.