VIRUS EVOLUTION

Virus Evolution, 2022, **8(1)**, 1–8
DOI: https://doi.org/10.1093/ve/veac034
Advance access publication date: 7 April 2022
Research Article

Molecular epidemiological features of SARS-CoV-2 in Japan, 2020–1

Hirotaka Ode,[1] Yoshihiro Nakata,[1,2] Mami Nagashima,[3] Masaki Hayashi,[3] Takako Yamazaki,[3] Hiroyuki Asakura,[3] Jun Suzuki,[3] Mai Kubota,[1] Kazuhiro Matsuoka,[1] Masakazu Matsuda,[1] Mikiko Mori,[1,2,†] Atsuko Sugimoto,[1] Mayumi Imahashi,[1] Yoshiyuki Yokomaku,[1] Kenji Sadamasu,[3] and Yasumasa Iwatani[1,2,*,‡]

[1]Clinical Research Center, National Hospital Organization Nagoya Medical Center, 4-1-1 Sannomaru, Naka-ku, Nagoya 460-0001, Japan, [2]Division of Basic Medicine, Nagoya University Graduate School of Medicine, 65 Tsurumi-cho, Showa-ku, Nagoya 466-8550, Japan and [3]Department of Microbiology, Tokyo Metropolitan Institute of Public Health, 3-24-1 Hyakunin-cho, Shinjuku-ku, Tokyo 169-0073, Japan
[†]https://orcid.org/0000-0001-7946-2045
[‡]https://orcid.org/0000-0001-9269-4828
*Corresponding author: E-mail: iwatani.yasumasa.cp@mail.hosp.go.jp

Abstract

There were five epidemic waves of coronavirus disease 2019 in Japan between 2020 and 2021. It remains unclear how the domestic waves arose and abated. To better understand this, we analyzed the pangenomic sequences of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and characterized the molecular epidemiological features of the five epidemic waves in Japan. In this study, we performed deep sequencing to determine the pangenomic SARS-CoV-2 sequences of 1,286 samples collected in two cities far from each other, Tokyo Metropolis and Nagoya. Then, the spatiotemporal genetic changes of the obtained sequences were compared with the sequences available in the Global Initiative on Sharing All Influenza Data (GISAID) database. A total of 873 genotypes carrying different sets of mutations were identified in the five epidemic waves. Phylogenetic analysis demonstrated that sharp displacements of lineages and genotypes occurred between consecutive waves over the 2 years. In addition, a wide variety of genotypes were observed in the early half of each wave, whereas a few genotypes were detected across Japan during an entire wave. Phylogenetically, putative descendant genotypes observed late in each wave displayed regional clustering and evolution in Japan. The genetic diversity of SARS-CoV-2 displayed uneven dynamics during each epidemic wave in Japan. Our findings provide an important molecular epidemiological basis to aid in controlling future SARS-CoV-2 epidemics.

Key words: SARS-CoV-2; viral genome sequencing; transmission dynamics; phylogenetic analysis; molecular epidemiology; Japan.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first discovered in patients suffering from viral pneumonia in Wuhan, China, at the end of 2019 (Zhu et al. 2020). Since then, the virus has spread worldwide (Ritchie et al. 2020). To date, >250 million people have been infected with SARS-CoV-2, >5 million of whom have been killed by the corresponding coronavirus disease 2019 (COVID-19). In Japan, five SARS-CoV-2 epidemic waves were encountered up to the end of 2021. Each wave consisted of a sharp surge and subsequent decline in new infection cases. By the end of 2021, the number of new infection cases had fallen to ∼100 cases per day.

The difficulty in controlling SARS-CoV-2 transmission is likely due to some unique characteristics of the virus. First, there are asymptomatic infected people with viral loads similar to those of symptomatic patients (Rothe et al. 2020; Zou et al. 2020; Yang et al. 2021). Second, in symptomatic patients, viral shedding may begin before symptom onset (He et al. 2020). It has been estimated that a substantial proportion of all infections are caused by stealth transmission from patients in asymptomatic states (Li et al. 2020b; Johansson et al. 2021; Subramanian, He, and Pascual 2021).

Another concern is that various variants of SARS-CoV-2 have evolved through mutations (Elbe and Buckland-Merrett 2017; Rambaut et al. 2020), albeit at a slow speed because of its proofreading machinery. Additionally, recombination events frequently observed in *Coronaviridae* promote its evolution (V'Kovski et al. 2021). In particular, emerging variants with advantageous mutations and/or insertions/deletions for (1) viral transmission such as D614G (Korber et al. 2020; Ozono et al. 2021), N501Y (Liu et al. 2021a), P681R (Saito et al. 2021) in the spike protein and R203K/G204R (Wu et al. 2021) and R203M (Syed et al. 2021) in the nucleocapsid; and/or (2) immune escape such as H69/V70 deletion, K417T, and E484K in spike protein (Chen et al. 2021; Cho et al. 2021; Gaebler et al. 2021; Wang et al. 2021) have been disseminated over time owing to their efficient transmission (Harvey et al. 2021; Tao et al. 2021).

To date, the World Health Organization has designated five lineages of variants that potentially cause severe outbreaks as variants of concern and several other lineages as variants of interest or monitoring (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/).

To track the epidemics of these variants and the emergence of new variants, viral genome surveillance has been performed in countries around the world (Elbe and Buckland-Merrett 2017; Rambaut et al. 2020). SARS-CoV-2 epidemics have been reported in Japan (Nagashima et al. 2020; Sekizuka et al. 2020, 2021; Hirotsu and Omata 2021; Ko et al. 2021; Nakata et al. 2021; Shimura et al. 2021; Tani-Sassa et al. 2021). Lineages of dominantly epidemic variants have been exchanged at every new wave in Japan (Elbe and Buckland-Merrett 2017; Rambaut et al. 2020). However, there is too little documentation to understand the structures of our domestic epidemic waves. In this study, to probe the mechanism of Japan's SARS-CoV-2 waves in detail, we characterized the molecular epidemiological features found in Japan based on SARS-CoV-2 pangenomic sequence analysis. First, we determined pangenomic sequences of epidemic viruses from patient samples collected at the Tokyo Metropolitan Institute of Public Health (TMIPH) in Tokyo Metropolis and Nagoya Medical Center (NMC) in Nagoya, which are located far from each other. Second, by adding publicly available sequences reported in other places in Japan between 2020 and 2021, we investigated the extent of the geographic and temporal distribution, and the genotype of each virus carrying a unique set of mutations that had been detected in Japan.

2. Materials and methods

2.1 Determination of viral genome sequences

We determined viral genome sequences by deep sequencing with independent protocols between TMIPH and NMC as described previously (Nagashima et al. 2020; Nakata et al. 2021). The details are also described in the **Supplementary Data**. The study was approved by the institutional ethics review boards, according to the Declaration of Helsinki 2013 (approval numbers: 31kenkenken-2007 and 3kenkenken-466 in TMIPH, and 2020-059 in NMC). All of the sequences that were determined and analyzed in this study (Supplementary Table S1) have been deposited in the GISAID EpiCoV database.

2.2 Sequence dataset

The number of daily cases of SARS-CoV-2 infection was obtained from the governmental websites of Tokyo Metropolitan, Aichi Prefecture, and Nagoya city. The data for Japan, Europe, and India were downloaded from 'Our World in Data' (Ritchie et al. 2020).

A total of 5,014,375 sequences of SARS-CoV-2, including the sequences determined here, were downloaded from the GISAID EpiCoV database on 10 November 2021. Then, we extracted 1,756,168 pangenomic sequences [>29,000 nucleotides (nt)] that had no ambiguous bases from the downloaded sequences that had been deposited from other countries, and we extracted 160,844 sequences determined in Japan regardless of the existence of ambiguous bases. Given that the total number of infection cases was ∼1.7 million in Japan, the sequences covered approximately 9 per cent of the newly infected cases. In this study, each mixed base, represented by the letter R, Y, M, K, S, or W, within the sequences from Japan was switched to a nucleotide base that introduced a substitution compared to the reference one (Wuhan strain; EPI_ISL_402124). Meta-information on SARS-CoV-2 lineages, sample collection dates, and sampling locations corresponding to each sequence were also obtained from the database. Of note, 'AY.29' includes its descendant lineage 'AY.29.1', whereas 'other Delta' consists of lineages of B.1.617.2 and AY.*x* (*x* is a number other than 29).

2.3 Sequence data analysis

The extracted sequences were mapped onto the reference sequence (Wuhan strain; EPI_ISL_402124) with the minimap2 program (ver. 2.17-r974; Li and Birol 2018) with the options of '-a -A 2 -O 24,24 -E 2,2'. To exclude any bias resulting from different sequence lengths, the sequences between the first nucleotide of *open reading frame 1a* (*orf1a*) and the end of *orf10* (266–29,674 nt, according to the position numbering of the reference) were extracted by using in-house programs (provided upon request). The resultant sequences were analyzed further here.

Maximum likelihood trees were estimated from bootstrap tests of 1,000 replicates by the IQ-TREE2 program (Minh et al. 2020) with options 'GTR+F+I+G4'. To draw the tree, FigTree v.1.4.2 software (http://tree.bio.ed.ac.uk/software/figtree/) was used. Comparative analyses of each sequence were performed by in-house scripts (provided upon request).

3. Results

3.1 Chronological changes in viral lineages in Japan

Japan has had five epidemic waves of SARS-CoV-2 infection: 2020/2–2020/5, 2020/7–2020/9, 2020/11–2021/2, 2021/3–2021/6, and 2021/7–2021/10. These epidemic wave periods are defined in this study according to the reports by the Ministry of Health, Labour, and Welfare, Japan. These waves were observed at the same months in two remote areas (∼260 km as the crow flies), Tokyo Metropolis and Nagoya city, although on different scales and slight delays in Nagoya (Fig. 1). Each peak of daily new infection cases reached approximately 700, 1,600, 8,000, 7,200 and 26,000, respectively, in Japan as a whole.

We determined 1,148 and 138 pangenomic sequences in samples collected from SARS-CoV-2-infected patients in TMIPH and NMC, respectively. Although the two areas are distant, identical lineages were identified from samples at every time point of all five waves (Fig. 1). In the first wave, sequences of prototype Wuhan-strain-related lineages (B, A, or A.16) as well as B.1.1 (carrying both spike D614G and nucleocapsid R203K/G204R mutations) were identified. Lineages of the determined sequences comprised B.1.1.284 and, to a lesser degree, B.1.1.214 in the second wave, whereas the ratio of B.1.1.284 to B.1.1.214 was inverted in the third wave. The fourth and fifth waves consisted mainly of the B.1.1.7/Alpha and AY.29/Delta lineages, respectively. In addition, the R.1 lineage was identified as a minor population of the third and fourth waves, as previously reported (Sekizuka et al. 2021; Tani-Sassa et al. 2021). The lineage trends observed in each wave were consistent with previous reports (Sekizuka et al. 2020, 2021; Nakata et al. 2021; Tani-Sassa et al. 2021). However, it is not known how the genetic diversity of each lineage changed in the individual epidemic waves in Japan.

3.2 Genetic diversity within the viral genotypes

We performed an in-depth analysis of all 1,286 sequences we determined. First, when we compared individual sequences, 873 genotypes carrying different sets of mutations were found (Fig. 2). A reconstructed phylogenetic tree of the sequences revealed

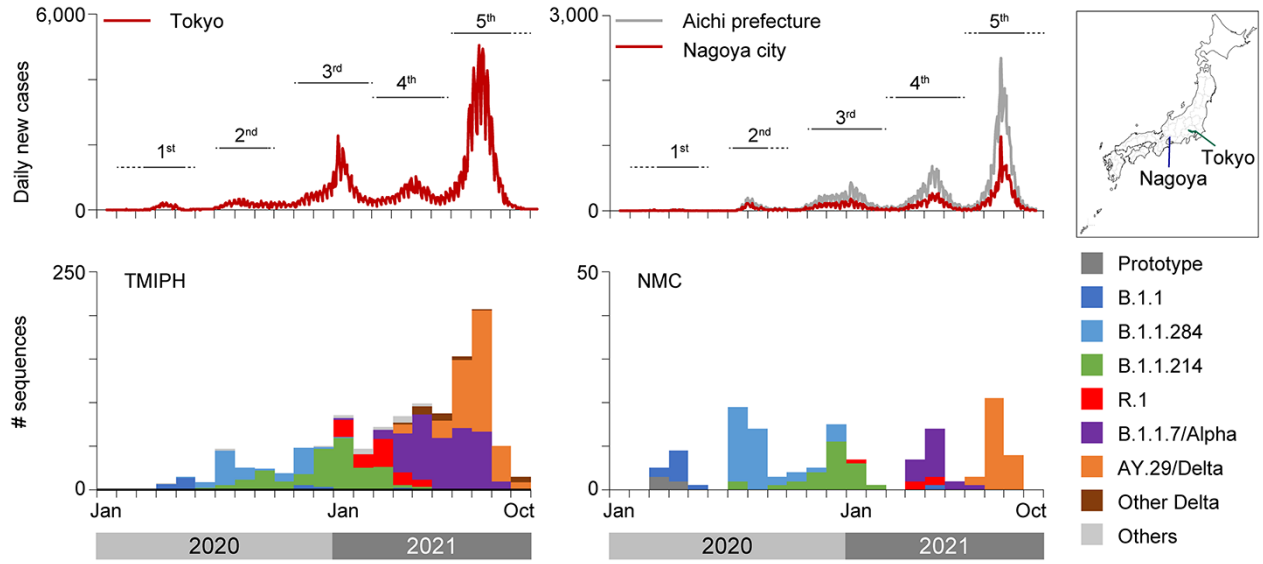

Figure 1. The genetic characteristics of SARS-CoV-2 identified in Tokyo Metropolis and Nagoya city, Japan. The upper panels show the numbers of daily new cases of SARS-CoV-2 infections in Tokyo Metropolis and Nagoya city. The bottom panels represent the numbers of viral genome sequences for clinical samples taken at TMIPH in Tokyo ($n = 1{,}148$) and NMC in Nagoya ($n = 138$) per month. 'AY.29/Delta' includes its descendent lineage, AY.29.1, whereas 'other Delta' consists of B.1.617.2 and AY.x ($x \neq 29$). The prototype lineage contains B, A, and A.16.

clustering of all the sequences into each relevant lineage, except upstream lineage sequences such as B.1.1. This confirmed the validity of our phylogenetic tree reconstruction protocol. In addition, the tree indicated wide genetic diversity between the genotypes in every lineage (Fig. 2A). In particular, maximum numbers of genotypes were observed at the wave peaks (Fig. 2B). Moreover, for each B.1.1, B.1.1.284, B.1.1.214, B.1.1.7/Alpha, and AY.29/Delta, each of the parental sequences that corresponded to the parental node of the lineage on the tree (blue closed circles in Fig. 2A) was assigned to the sequence carrying the minimal number of mutations within its corresponding lineage. In contrast, the sequences branched off from the putative parental sequence and they had unique incremental mutations over time (Fig. 2A, C). Interestingly, 29 genotypes, including those of the parental sequences, were continuously isolated for 2–3 months in TMIPH or NMC (blue and pink arrows in Fig. 2A–B), 7 of which were detected at both sites (purple arrows in Fig. 2A–B). In contrast, the other 844 genotypes did not last for more than 1 month. These results suggest that various genotypes of the same SARS-CoV-2 lineage were detected in each epidemic wave, showing different geographical and temporal distributions, while a few genotypes were continuously found from the beginning through the end of any wave in two distant areas of Japan.

3.3 Geographical distribution of the viral genotypes within Japan

To confirm whether this trend is applicable across all of Japan, we further analyzed the sequences, including all domestic sequences, in the GISAID EpiCoV dataset (Fig. 3). Herein, we divided the whole area of Japan into seven regions. When we first examined the geographical distribution of the seven aforementioned genotypes that we detected in both TMIPH and NMC (purple arrows), six of them were identified in samples collected in all seven regions (Fig. 3). However, most of the genotypes found only at TMIPH were identified only from sequences in the Kanto region, including Tokyo (the third row from the top of each heatmap panel). Similarly, the genotypes determined at NMC were predominantly found in the Chubu area, in the center of which is Nagoya. These results indicate that most SARS-CoV-2 genotypes were uniquely detected in limited areas, although a small number of genotypes were detected throughout Japan.

3.4 Temporal distribution of the viral genotypes

To clarify the chronological distribution of the genotypes in Japan, we focused on 90 prevalent major genotypes, defined as those that had more than 25 identical sequences deposited in the GISAID EpiCoV database (Fig. 4A–B). Because they contained the parental sequences of B.1.1, B.1.1.284, B.1.1.7/Alpha, and AY.29/Delta, but not B.1.1.214, we added B.1.1.214 to our analyses. The phylodynamic analysis of the sequences along with their sample collection dates showed that the parental sequences for B.1.1, B.1.1.284, B.1.1.214, B.1.1.7/Alpha, and AY.29/Delta were identified on parental nodes of the respective lineages throughout all of Japan (highlighted with red dotted lines in Fig. 4A), as observed in Tokyo and Nagoya (Fig. 2A). In addition, the parental genotypes of B.1.1, B.1.1.284, B.1.1.214, B.1.1.7/Alpha, and AY.29/Delta were detected for several months from the early phase of the epidemic wave through the time immediately before the next wave in each of the first, second, fourth, and fifth waves (red arrowheads in Fig. 4C, D). Similar trends were also observed in a few early descendant genotypes (blue arrowheads in Fig. 4C–D) that were major genotypes at the epidemic peak times. As an exception, the parental genotype of B.1.1.214 was transiently found in the second wave, although it did not last for a long time, and a mutated genotype of B.1.1.214 appeared in the third wave. In contrast, genotypes found only in local areas were collected at relatively late phases of the waves (Fig. 4B–D). These results support the notion that locally detected genotypes were basically descendants of certain prevalent genotypes on the phylogenetic tree (Fig. 4A). The results also suggest that there was a unique epidemiological tendency in which a few genotypes, including the parental genotypes, were detected across Japan during the entire wave, whereas the majority of their descendant genotypes were

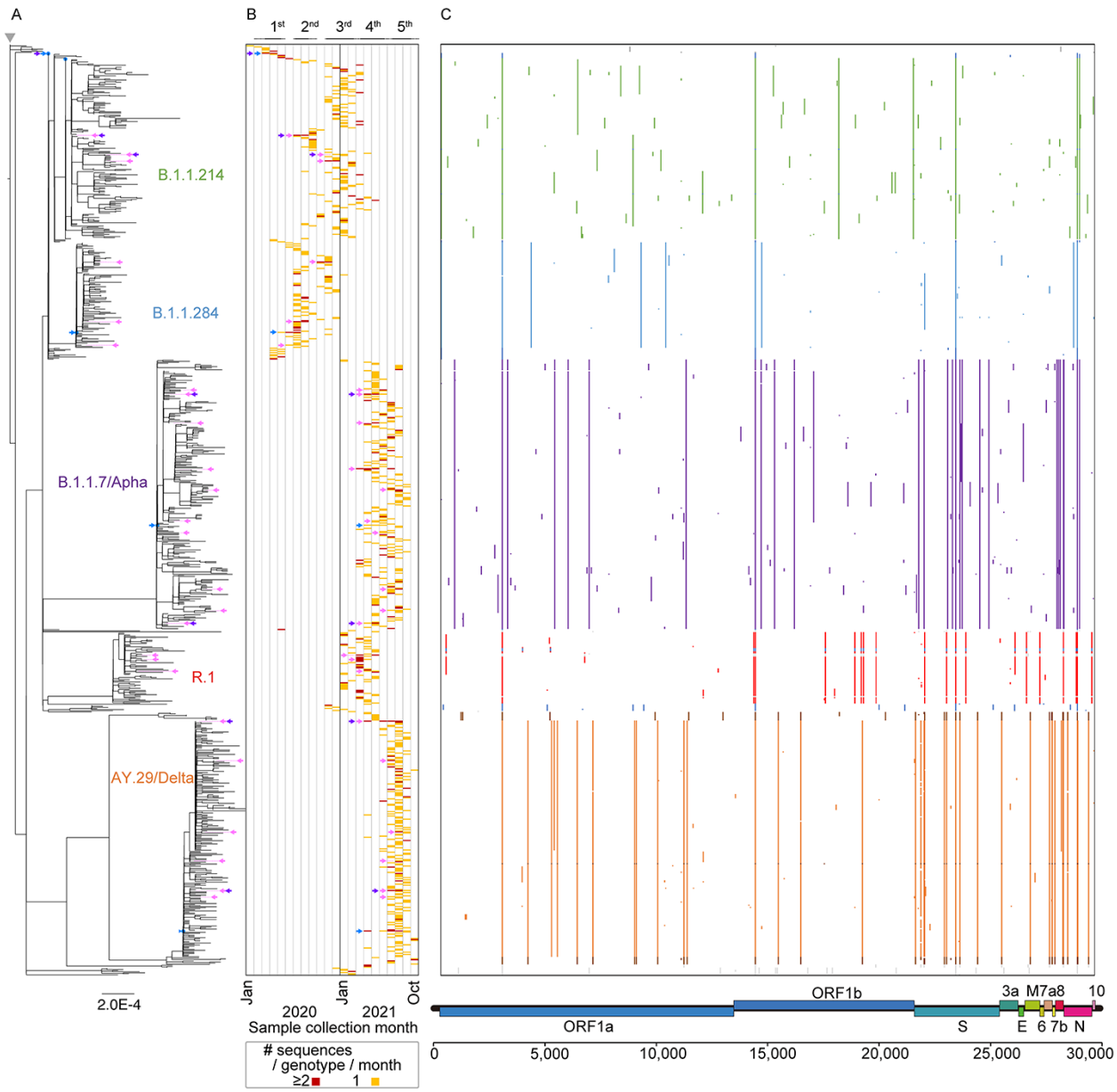

Figure 2. Genetic diversities of the SARS-CoV-2 genotypes detected at TMIPH and NMC. We examined the sequences between the first nucleotide of orf1a and the end of orf10 [positions 266–29,674 according to the reference sequence (Wuhan variant; EPI_ISL_402124)] and detected 873 genotypes. (A) Maximum likelihood tree of the sequences representing each genotype. The gray arrowhead highlights the reference sequence. (B) Heatmaps of the sequence number per sampling month in TMIPM or NMC. Genotypes determined from ≥2 and 1 sequences are shown. In panels (A) and (B), parental sequences corresponding to the parental nodes in the tree for each lineage are highlighted by closed blue circles. Note that a parental sequence for the R.1 lineage is not shown because it was not identified in our analysis. The sequences that were detected in samples taken ≥2-month intervals are colored with blue (parental sequences) and pink arrows (others). The genotypes found in both TMIPH and NMC are indicated with purple arrows. (C) Nucleotide mutation signatures between the sequences representing each genotype. The mutations that were observed at >0.5 per cent prevalence are highlighted with different colors for each lineage, as shown in Fig. 1. 'AY.29/Delta' includes its descending lineage, AY.29.1.

identified in local regions, briefly, late in a wave. In the early phases of the third and fifth waves, the preceding genotypes, i.e. B.1.1.284 and B.1.1.7/Alpha, respectively, were limited to a few isolated areas, and they were later displaced by two other major genotypes, new B.1.1.214 and AY.29/Delta.

Next, we analyzed the phylogenetic relationships and sampling dates of infrequent genotypes, defined as those that had 3–25 sequences in the database. We tested whether the trend observed among the prevalent genotypes (Fig. 4) was also applicable to these infrequent genotypes (Supplementary Fig. S1). As expected, the results showed that infrequent genotypes carrying incremental mutations were also found in samples with relatively later collection dates within each lineage (Fig. 4). When we compared the domestic genotypes with those detected outside of Japan (Fig. 4A, E), most of the genotypes were hardly identified in the database. As two exceptions, the genotypes of the parental sequences of B.1.1 and B.1.1.7/Alpha in Japan were identified at earlier times in other countries than in Japan, suggesting inbound events bringing these lineages into Japan. However, the other preceding genotypes of lineages, B.1.1.284, B.1.1.214,

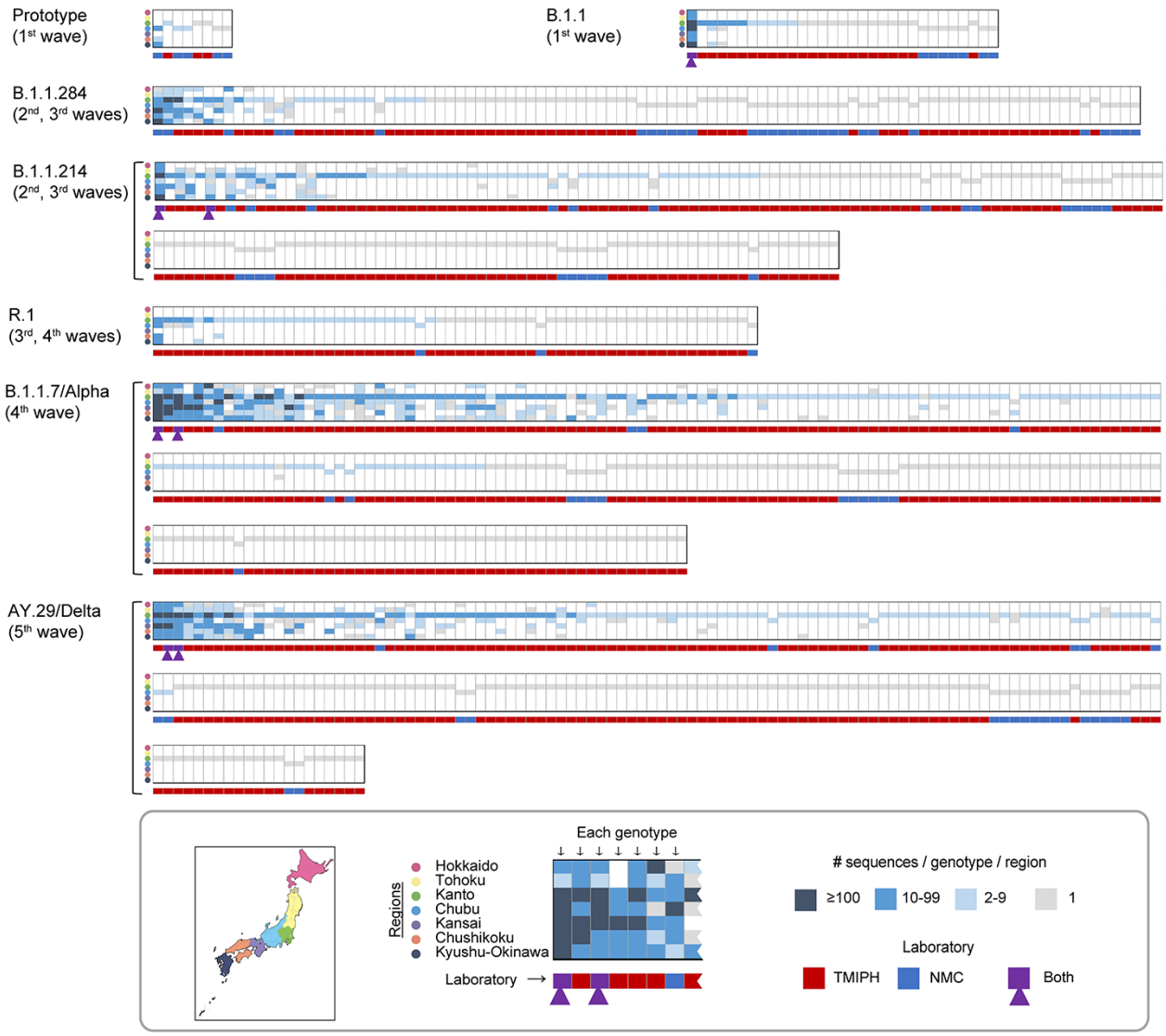

Figure 3. Geographical distributions of SARS-CoV-2 genotypes in Japan. Based on the dataset filled out by laboratories in Japan (160,844 sequences in total), the number of sequences supporting each genotype was counted in every sampling area and was plotted in heatmaps with different colors according to the bottom legend. The vertical and horizontal axes of the heatmaps correspond to the seven independent regions of Japan and each genotype we identified in the respective lineages. On the horizontal axes, the genotypes were sorted in a descending order of the sequence number. Below each heatmap, the bars represent the sequence numbers determined in either TMIPH or NMC. 'AY.29/Delta' includes its descending lineage, AY.29.1. The prototype lineage contains B, A, and A.16.

and AY.29/Delta, could not be identified among the database sequences.

4. Discussion

In Japan, five waves of SARS-CoV-2 epidemics had been encountered until November 2021. More patients were infected in the later waves. As reported previously (Sekizuka et al. 2020, 2021; Tani-Sassa et al. 2021), the major lineages changed with every new wave (Fig. 1), and the lineages of the later waves were presumably selected because of their more advantageous transmissibility and/or immune escape potential (Mlcochova et al. 2021; Planas et al. 2021; Volz et al. 2021). The genetic features of the lineages detected in Japan had not been well characterized, so we performed an in-depth analysis of SARS-CoV-2 sequences and found three common features of the spatiotemporal changes in the SARS-CoV-2 genotypes in Japan. First, sharp lineage and genotype displacements between two consecutive waves were clearly observed (Fig. 4), which displayed so-called founder/bottleneck effects in the selection of epidemic variants in Japan. Second, only a few genotypes were detected across Japan during the entirety of any wave, whereas many genotypes were observed between the early and peak phases (Figs 2–4). These findings suggest that these limited types of SARS-CoV-2 variants that differentiated from the parental nodes in a phylogenetic tree were persistently transmitted from the early dates of a given wave and caused epidemic transmission from urban to local areas over a span of months. Third, their descendant genotypes, varying regionally in Japan, were identified in the late phase of each wave. Such descendant genotypes form long-term phylogenetic clusters with unique incremental mutations over time.

The timings of all five epidemic waves in Japan were synchronized with those elsewhere in the world. Therefore, we attempted to identify preceding SARS-CoV-2 sequences in the GISAID EpiCoV database, which may be linked to inbound transmission at the five

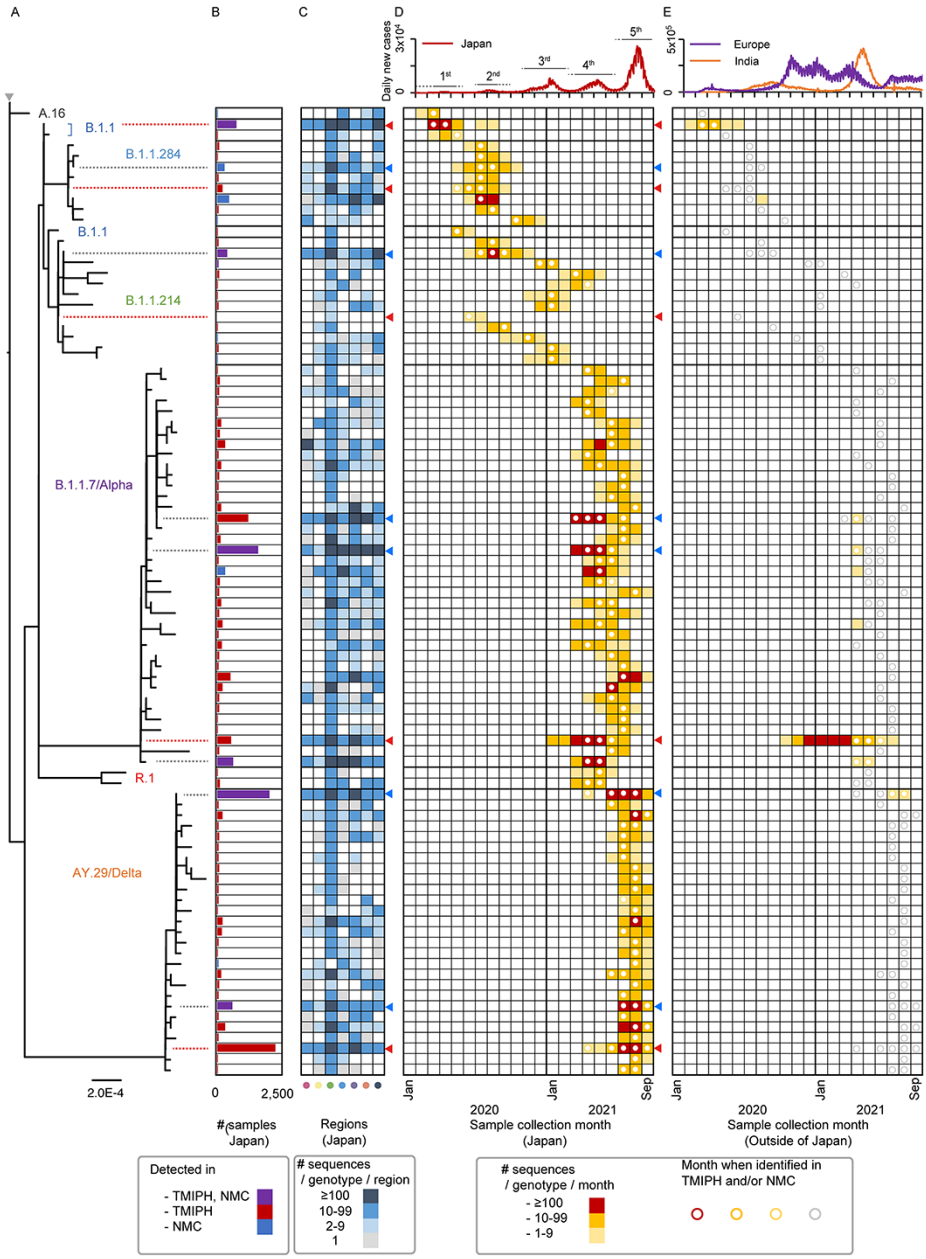

Figure 4. Temporal distribution of SARS-CoV-2 genotypes. Ninety genotypes identified from >25 sequences in Japan were analyzed. (A) Maximum likelihood tree of the representative sequences for each genotype. The Wuhan variant (EPI_ISL_402124) was used as a reference (gray arrowhead). The sequences detected throughout Japan are shown with dotted lines. The putative parental sequences for each lineage are colored red. (B) Bar graphs of the total number of sequences supporting each genotype in Japan. The sequences determined at TMIPH or NMC are shown. (C) Heatmaps of the numbers of sequences supporting each genotype at the respective sampling locations. The regions Hokkaido, Tohoku, Kanto, Chubu, Kansai, Chushikoku, and Kyushu–Okinawa are shown from left to right. (D), (E) Heatmaps of the numbers of sequences supporting each genotype per sampling month in Japan (D) and outside of Japan (E). Open circles represent the collection date (month) at TMIPH or NMC. At the top of the heatmaps, changes in new infection cases over time in Japan, the U.K., or India are shown. In panels (C) and (D), the genotypes of the parental sequences of B.1.1, B.1.1.284, B.1.1.214, B.1.1.7/Alpha, and AY.29/Delta are highlighted with red arrowheads. Their early descendant genotypes are shown with blue arrowheads.

epidemic waves in Japan. However, only two identical genotypes of the presumably parental lineages that were registered slightly earlier from outside of Japan: the B.1.1 lineage in the first wave and the B.1.1.7/Alpha lineage in the fourth wave. The majority of the genotypes detected here were hardly identified at all outside of Japan (Fig. 4). In addition, because travel restrictions and border control were tightened before the second wave, it appears that the variants of B.1.1.284 and B.1.1.214 lineages had originally evolved from B.1.1 in the second and third waves in Japan (Nakata et al. 2021; Sekizuka et al. 2021). Moreover, newer generations of descendant genotypes for the two domestic lineages were detected over time, even during the B.1.1.7/Alpha epidemic period of the fourth wave. These results suggest that not only inbound transmission but also domestically persistent transmission of the B.1.1.284 and B.1.1.214 lineages, although low, could have been potential triggers of the next epidemic waves. In this study, we could not find any potential parental sequences of the two lineage variants, R.1 and AY.29, that were predominantly identified

in Japan (Fig. 4E). The B.1.617.2/Delta variant, which was first identified in India, is an ancestral variant of the AY.29/Delta lineage, although their genetic distances seem to be relatively great. Therefore, the presumed variants might have been disseminated in a population with little viral genome surveillance.

Interestingly, at the beginning of the fifth wave, multiple genotypes of Delta lineages, B.1.617.2 and AY.*x* ($x \neq 29$), were detected in Japan, although the AY.29/Delta variants outcompeted the other Delta variants at the wave peak (Supplementary Fig. S2). Similarly, all the parental genotypes at every wave were commonly displaced by their descendant genotypes. These phenomena might be explained by two things. First, the transmissibility of the parental genotypes differs from that of subsequently prevalent variants, even in the same lineage, and the viral adaptation of each parental genotype may have occurred early in the epidemic waves in Japan. As there is a unique mutation(s) in the descendants—nsp13 E261D and nsp2 F163L in B.1.1.7/Alpha, nsp3 V932A and ORF8 P93S in AY.29/Delta—such unique mutations might help accelerate transmission between people living in Japan. Second, there might have been social behavior effects that arose by chance or other factors, such as superspreading events, including from patients infected with these variants (Yang et al. 2021). More stringent infection control, including self-defensive measures, might help reduce the transmission of parental variants at the early phases of epidemic waves.

In early 2022, a growing number of people have been fully vaccinated against COVID-19 and/or have been naturally infected with SARS-CoV-2 worldwide. Therefore, there will be more chances of the emergence of variants carrying convergent mutations associated with immune escape (Weisblum et al. 2020; Li et al. 2020a; Francino-Urdaniz et al. 2021; Greaney et al. 2021a, 2021b; Liu et al. 2021b). B.1.1.529/Omicron, with significantly enhanced immune evasion, has been rapidly spreading worldwide since its first isolation in South Africa (Callaway and Ledford 2021). Although any putative intermediate genotypes that phylogenetically gave rise to B.1.1.529/Omicron have not been identified, Omicron might have evolved among certain local populations where SARS-CoV-2 surveillance was weak. In this study, we observed that the descendant genotypes that were identified at the late phase of each wave varied regionally, even though the number of newly infected cases was lower than that in other countries in every epidemic wave. In fact, we detected B.1.1.7/Alpha (six samples) or AY.29/Delta (two samples) genotypes carrying mutations associated with immune escape [i.e. L455F, S477G/I (Greaney et al. 2021b; Liu et al. 2021b)] in our eight samples. These regional genotypes may have also been spread and transmitted globally since new variants carry certain mutations advantageous for transmissibility and/or immune evasion. Continuous global surveillance of SARS-CoV-2 is required to track newly emergent variants in the future.

Potential limitations of this study include a limited sample size and no complete information for patients. Regarding the sample size, there were two major factors that limited our sample size: (1) a limited number of samples were collected from the newly infected patients (2) and all our collected samples were not analyzed, because of the technical difficulty of viral genome amplification by polymerase chain reaction (PCR), followed by deep sequencing, for specimens with low viral RNA contents (cycle threshold [Ct] values $\geq$30). Therefore, different ancestral genotypes and/or descendant genotypes may have not been identified even though the sequences analyzed in this study covered approximately 9 per cent of the newly infected cases reported in Japan. With respect to patients' information, there were difficulties in collecting their personal information, such as records of natural infection, immunization, and traveling, which are likely to have an impact on viral genetic diversification. In Japan, vaccination coverage (two doses) levels were precipitously elevated from June to October 2021, which corresponds to a period of the fifth epidemic wave, being dominated by the AY.29/Delta variants. A sharp decline in the newly infected cases was observed between September and October 2021. These epidemical situations may reflect a vaccination effect on viral transmission, possibly virus evolution, although it is difficult to assess the effect(s) of immunization on viral genetic diversification in Japan. Another critical factor that potentially affects viral genetic diversity is travel (or travel policies), especially restrictions on overseas and domestic travel. For example, the founder effects might be driven by patient's travel from epidemic to non-endemic areas, whereas limited virus transmission in local areas generated region-specific variants (Fig. 4). Further analyses are required to clarify these effects on spatiotemporal genetic changes.

In summary, we analyzed the spatiotemporal distributions of SARS-CoV-2 genotypes in Japan. Three common features of the genetic changes were shared by Japan's five epidemic waves: (1) sharp genotype displacements were observed between any two consecutive waves and such drastic changes were distinctive of Japan's waves; (2) only a few genotypes were detected across Japan during an entire wave, although many genotypes were observed from the early to peak phases; and (3) region-specific genotypes descending from their parental genotypes were identified in the late phase of each wave. These local genotypes of SARS-CoV-2 seemed to be generated not only in Japan but also worldwide and would become genotype sources of the next epidemic wave. Therefore, worldwide surveillance information of the SARS-CoV-2 genome will be important to understand future epidemics. Such information, including from the present study, may help us control future epidemics of SARS-CoV-2.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Funding

This work was supported by the Japan Agency for Medical Research and Development (AMED) Projects on COVID-19 (JP20fk0108536 led by Prof. Kenjiro Kosaki, Keio University, Japan and 20nk0101627h001 led by Dr. Wataru Akahata, VLP therapeutics Japan) and the Aichi Health Promotion Foundation.

Conflict of interest: None declared.

Author contributions

Y.I. designed the research;
Y.N., M.K., M.N., K.M., M. Matsuda, M. Mori, and Y.I. performed the experiments;
H.O., Y.N., A.S., M.I., Y.Y., K.S., and Y.I. analyzed the data;
H.O., K.S., and Y.I. wrote and prepared the manuscript.
All authors reviewed the manuscript.

References

Callaway, E., and Ledford, H. (2021) 'How Bad Is Omicron? What Scientists Know So Far', *Nature*, 600: 197–9.

Chen, R. E. et al. (2021) 'Resistance of SARS-CoV-2 Variants to Neutralization by Monoclonal and Serum-derived Polyclonal Antibodies', *Nature Medicine*, 27: 717–26.

Cho, A. et al. (2021) 'Anti-SARS-CoV-2 Receptor-Binding Domain Antibody Evolution after mRNA Vaccination', *Nature*, 600: 517–22.

Elbe, S., and Buckland-Merrett, G. (2017) 'Data, Disease and Diplomacy: GISAID's Innovative Contribution to Global Health', *Global Challenges*, 1: 33–46.

Francino-Urdaniz, I. M. et al. (2021) 'One-Shot Identification of SARS-CoV-2 S RBD Escape Mutants Using Yeast Screening', *Cell Reports*, 36: 109627.

Gaebler, C. et al. (2021) 'Evolution of Antibody Immunity to SARS-CoV-2', *Nature*, 591: 639–44.

Greaney, A. J. et al. (2021a) 'Comprehensive Mapping of Mutations in the SARS-CoV-2 Receptor-Binding Domain that Affect Recognition by Polyclonal Human Plasma Antibodies', *Cell Host & Microbe*, 29: 463–76 e6.

—— et al. (2021b) 'Mapping Mutations to the SARS-CoV-2 RBD that Escape Binding by Different Classes of Antibodies', *Nature Communications*, 12: 4196.

Harvey, W. T. et al. (2021) 'SARS-CoV-2 Variants, Spike Mutations and Immune Escape', *Nature Reviews Microbiology*, 19: 409–24.

He, X. et al. (2020) 'Temporal Dynamics in Viral Shedding and Transmissibility of COVID-19', *Nature Medicine*, 26: 672–5.

Hirotsu, Y., and Omata, M. (2021) 'Detection of R.1 Lineage Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) with Spike Protein W152L/E484K/G769V Mutations in Japan', *PLOS Pathogens*, 17: e1009619.

Johansson, M. A. et al. (2021) 'SARS-CoV-2 Transmission from People without COVID-19 Symptoms', *JAMA Network Open*, 4: e2035057.

Ko, K. 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: e0246383.

Korber, B. et al. (2020) 'Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus', *Cell*, 182: 812–27 e19.

Li, H., and Birol, I. (2018) 'Minimap2: Pairwise Alignment for Nucleotide Sequences', *Bioinformatics*, 34: 3094–100.

Li, Q. et al. (2020a) 'The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity', *Cell*, 182: 1284–94 e9.

Li, R. et al. (2020b) 'Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2)', *Science*, 368: 489–93.

Liu, Y. et al. (2021a) 'The N501Y Spike Substitution Enhances SARS-CoV-2 Infection and Transmission', *Nature*, 602: 294–9.

Liu, Z. et al. (2021b) 'Identification of SARS-CoV-2 Spike Mutations that Attenuate Monoclonal and Serum Antibody Neutralization', *Cell Host & Microbe*, 29: 477–88 e4.

Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.

Mlcochova, P. et al. (2021) 'SARS-CoV-2 B.1.617.2 Delta Variant Replication and Immune Evasion', *Nature*, 599: 114–9.

Nagashima, M. et al. (2020) 'Characteristics of SARS-CoV-2 Isolated from Asymptomatic Carriers in Tokyo', *Japanese Journal of Infectious Diseases*, 73: 320–2.

Nakata, Y. et al. (2021) 'Molecular Epidemiological Insights into Transmission Trends in Nagoya Area Based on SARS-CoV-2 Genome Sequencing (Mar-Oct, 2020), *Kansenshogaku Zasshi*', *The Journal of the Japanese Association for Infectious Diseases*, 95: 293–300.

Ozono, S. et al. (2021) 'SARS-CoV-2 D614G Spike Mutation Increases Entry Efficiency with Enhanced ACE2-Binding Affinity', *Nature Communications*, 12: 848.

Planas, D. et al. (2021) 'Reduced Sensitivity of SARS-CoV-2 Variant Delta to Antibody Neutralization', *Nature*, 596: 276–80.

Rambaut, A. et al. (2020) 'A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology', *Nature Microbiology*, 5: 1403–7.

Ritchie, H. et al. (2020), *Coronavirus Pandemic (COVID-19)*. Published online at OurWorldInData.org, <https://ourworldindata.org/coronavirus> accessed 1 Apr 2022.

Rothe, C. et al. (2020) 'Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany', *New England Journal of Medicine*, 382: 970–1.

Saito, A. et al. (2021) 'Enhanced Fusogenicity and Pathogenicity of SARS-CoV-2 Delta P681R Mutation', *Nature*, 602: 300–6.

Sekizuka, T. et al. (2020) 'A Genome Epidemiological Study of SARS-CoV-2 Introduction into Japan', *mSphere*, 5: e00786–20.

—— et al. (2021) 'A Discernable Increase in the Severe Acute Respiratory Syndrome Coronavirus 2 R.1 Lineage Carrying an E484K Spike Protein Mutation in Japan', *Infection, Genetics and Evolution*, 94: 105013.

Shimura, T. et al. (2021) 'Multiple Introductions of SARS-CoV-2 B.1.1.214 Lineages from Mainland Japan Preceded the Third Wave of the COVID-19 Epidemic in Hokkaido', *Travel Medicine and Infectious Disease*, 44: 102210.

Subramanian, R., He, Q., and Pascual, M. (2021) 'Quantifying Asymptomatic Infection and Transmission of COVID-19 in New York City Using Observed Cases, Serology, and Testing Capacity', *Proceedings of the National Academy of Sciences of the United States of America*, 118: e2019716118.

Syed, A. M. et al. (2021) 'Rapid Assessment of SARS-CoV-2 Evolved Variants Using Virus-like Particles', *Science*, 374: 1626–32.

Tani-Sassa, C. et al. (2021) 'Viral Loads and Profile of the Patients Infected with SARS-CoV-2 Delta, Alpha or R.1 Variants in Tokyo', *Journal of Medical Virology*, 94: 1707–10.

Tao, K. et al. (2021) 'The Biological and Clinical Significance of Emerging SARS-CoV-2 Variants', *Nature Reviews. Genetics*, 22: 757–73.

V'Kovski, P. et al. (2021) 'Coronavirus Biology and Replication: Implications for SARS-CoV-2'', *Nature Reviews Microbiology*, 19: 155–70.

Volz, E. et al. (2021) 'Assessing Transmissibility of SARS-CoV-2 Lineage B.1.1.7 In England', *Nature*, 593: 266–9.

Wang, P. et al. (2021) 'Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7', *Nature*, 593: 130–5.

Weisblum, Y. et al. (2020) 'Escape from Neutralizing Antibodies by SARS-CoV-2 Spike Protein Variants', *Elife*, 9: e61312.

Wu, H. et al. (2021) "Nucleocapsid Mutations R203K/G204R Increase the Infectivity, Fitness, and Virulence of SARS-CoV-2', *Cell Host & Microbe*, 1788-801: e6.

Yang, Q. et al. (2021) 'Just 2% of SARS-CoV-2-Positive Individuals Carry 90% of the Virus Circulating in Communities', *Proceedings of the National Academy of Sciences*, 118: e2104547118.

Zhu, N. et al. (2020) "A Novel Coronavirus from Patients with Pneumonia in China, 2019', *New England Journal of Medicine*, 382: 727–33.

Zou, L. et al. (2020) 'SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients', *New England Journal of Medicine*, 382: 1177–9.