medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Introductions and evolutions of SARS-CoV-2 strains in Japan

Reitaro Tokumasu[1], Dilhan Weeraratne[2], Jane Snowdon[2], Laxmi Parida[3], Michiharu Kudo[1], Takahiko Koyama[3]

[1] IBM Research- Tokyo, Tokyo, Japan

[2] IBM Watson Health, Cambridge, MA 02142, USA

[3] IBM TJ Watson Research Center, Yorktown Heights, NY 10598, USA

Correspondence to Takahiko Koyama (email: tkoyama@us.ibm.com, postal address: 1101 Kitchawan Rd, Yorktown Heights, NY 10598, USA

Key Words: SARS-COV-2, variant, COVID-19, Japan, quarantine, strains

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Abstract

COVID-19 caused by SARS-CoV-2 was first identified in Japan on January 15th, 2020, soon after the pandemic originated in Wuhan, China. Subsequently, Japan experienced three distinct waves of the outbreak in the span of a year and has been attributed to new exogenous strains and evolving existing strains. Japan engaged very early on in tracking different COVID-19 sub-strains and have sequenced approximately 5% of all confirmed cases. While Japan has enforced stringent airport surveillance on cross-border travelers and returnees, some carriers appear to have advanced through the quarantine stations undetected. In this study, 17112 genomes sampled in Japan were analyzed to understand the strains, heterogeneity and temporal evolution of different SARS-CoV-2 strains. We identified 11 discrete strains with a substantial number of cases with most strains possessing the spike (S) D614G and nucleocapsid (N) 203_204delinsKR mutations. Besides these variants, ORF1ab P3371S, A4815V, S1361P, and N P151L were also detected in nearly half the samples constituting the most common strain in Japan. 115 distinct strains have been introduced into Japan and 12 of them were introduced after strict quarantine policy was implemented. In particular, the B.1.1.7 strain, that emerged in the United Kingdom (UK) in September 2020, has been circulating in Japan since late 2020 after eluding cross-border quarantine stations. Similarly, the B.1.351 strain dubbed the South African variant, P.1 Brazilian strain and R.1 strain with the spike E484K mutation have been detected in Japan. At least four exogenous B.1.1.7 sub-strains have been independently introduced in Japan as of late January 2021, and these strains carry mutations that give selective advantage including N501Y, H69_V70del, and E484K that confer increased transmissibility, reduced efficacy to vaccines and possible increased virulence. It is imperative that the quarantine policy be revised, cross-border surveillance reinforced, and new public health measures implemented to mitigate further transmission of this deadly disease and to identify strains that may engender resistance to vaccines.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Introduction

SARS-CoV-2, the etiological agent of COVID-19 was first identified in Wuhan, China in late 2019 before rapid worldwide transmission in the first quarter of 2020. In just a year, the number of confirmed cases exceeded 111 million globally with 2.5 million deaths as of February 23rd, 2021(*1*) The actual number of infections are likely much higher by accounting for asymptomatic cases and mild disease that do not get tested and under reporting due to social stigma and discrimination associated with COVID-19 infections(*2*). The novel coronavirus quickly spread to neighboring Japan, with a confirmed case of an individual on Jan 15, 2020 with travel history to Wuhan(*3*). Soon after the identification of the first case, a major initial transmission catastrophe was averted when the government decreed the Diamond Princess cruise ship with infected patients to be anchored off the coast of Japan and mandated a 14-day quarantine for all passengers(*4*). However, the number of confirmed infections has surpassed 425,000 with over 7400 reported fatalities in Japan(*1*).

The COVID-19 pandemic has exerted an unprecedented stress on global health systems and has created a ripple effect touching every rubric of human life. Particularly the impact on healthcare (including mental health) and the underlying social, political, psychological and economic disruptions have had profound ramifications. In Japan, the dichotomy between the public health safety and societal and economic dynamics forced the postponement of the much-awaited summer Olympics and Paralympics games in Tokyo. Nonetheless, Japan implemented a stringent surveillance process at airports and seaports at the very early stages of the outbreak to monitor and quarantine travelers and repatriates with COVID-19 infection. In short, the rigorous surveillance process required a negative COVID-19 test prior to boarding, saliva antigen testing at the cross-border port-of-entry and, if positive, polymerase chain reaction (PCR) confirmation and sequencing at the National Institute of Infectious Diseases. While the surveillance process has been largely successful with 2145 patients detected and intercepted at quarantine stations by end of January 2021(*5*), there appear to be some patients harboring exogenous strains with different haplotypes of the virus who were undetected at the port-of-entry. Japan has been vested and engaged from the beginning of the pandemic to monitor genomic changes in SARS-CoV-2 and has sequenced remarkable 17112 genomes which are approximately 5% of all confirmed cases. Notwithstanding the rigor of the public health measures, three discrete waves of the disease have been observed in Japan and it's plausible that different viral strains may have contributed to each spike.

Mutations are inevitable as viruses evolve as a mechanism to cope with selective pressure and confer selective advantage. COVID-19 is the first pandemic to occur after inexpensive sequencing technologies became widely available. SARS-CoV-2 accumulates mutations at the rate of 1.0 x 10-3 (per site/genome/year), that corresponds to 2.5 mutations in a month(*6*). However, as number of the infections increases, the number of variants increase proportionally. SARS-CoV-2 has accumulated and established multiple mutations within a year from the first published report of D614G in early April 2020(*7*). New strains with spike N501Y mutation such as B.1.1.7 from England, B.1.351 from South Africa, P.1 from Brazil and R.1 have recently emerged(*8*). Previously, N501Y has been functionally characterized and was reported to cause higher binding with ACE2(*9*) and another study reported that B.1.1.7 carrying H69_V70del in addition to N501Y possess a higher infectivity rate of 75%(*10*). Furthermore, B.1.351, P.1

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

and R.1 strains contain spike E484K and appear to confer reduced efficacy to the currently available vaccines(*11, 12*).

In this work we have evaluated the publicly available SARS-CoV-2 genomes in Japan to elucidate different viral lineages that were exogenously introduced, to understand community transmission patterns and to delineate founder strains that further evolve within a community.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Methods

Genomes were downloaded from the publicly available viral sequencing repositories, Global Initiative on Sharing All Influenza Data (GISAID)(*13*) and National Center for Biotechnology Information (NCBI) on January 30th, 2021; of the 442061 genomes downloaded, 17112 were from specimens collected in Japan. Low quality genomes with gaps and ambiguous bases over 50 base pairs in length (excluding at the start and the end of the genomes) were discarded. After completing quality control, variant analysis was performed on 296053 including 14461 Japanese genomes with the method described previously(*6*). In brief, genomes were first aligned to the reference genome NC_045512 using The European Molecular Biology Open Software Suite (EMBOSS) needle(*14*) with open gap penalty of 100 to filter out fictitious frameshifts. Next, differences with the reference genome were extracted and annotated using gene definitions of SARS-CoV-2 as different variant types including missense, synonymous, non-coding and indels. Top recurrent variants observed in Japanese samples were summarized in Table 1. Subsequently, hierarchical clustering was performed to organize strains with similar haplotypes to construct a variant graph (Figure 1B).

Haplotype defining major strains were extracted from the variant graph in Figure 1B and list of top variants in Table 1. Numbers of confirmed cases and deaths in Japan obtained from Our World in Data(*15*) were shown in Figure 2A. Monthly occurrences of each strain defined by the haplotype are plotted in Figure 2B. Similarly, active periods and evolutionary relationships are illustrated in Figure 2C.

A candidate parent of a particular strain was identified by the maximum variant approach using the haplotype of the query strain(*16*). A parental strain should have a haplotype which is a subset of the query strain. Among ancestors obtained in the previous step, the closest ancestor is the one with maximal number of matches between haplotypes. Subsequently, a possible transmission route was extrapolated after omitting nodes without descendants for simplicity (Figure 3).

Data on monthly COVID-19 positive cases at airport quarantine centers in Japan and number of passengers though immigration were obtained from a report released by the Japanese Government(*17, 18*) (Figure 4A). In Figure 4B, the monthly number of released genomes obtained at quarantine stations is shown. For each strain, we sought to identify a more likely parent including exogenous strains, which share more common variants than any domestic strains. Number of exogenous strains for each month was given in Figure 4C. Among the exogenous strains, strains whose all probably ancestors are found after April 15th are considered to have advanced through quarantine stations undetected (Table 3).

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Results

17112 genomes obtained between January 2020 and January 2021 were used for this analysis, accounting for about 5% of the confirmed cases in Japan. Frequently observed variants are shown in Table 1. Spike protein D614G followed by ORF1ab P4715L (RdRp P323L), 5'-UTR 241C>T, nucleocapsid protein (N) 203_204delinsKR and ORF1ab L16L (313C>T) are the most common variants among strains in Japan, representing over 90% of all variants. L16L is unique to Japanese D614G_KR and its derivative strains while other three variants are common to all global D614G clade derived strains. Among the top non-synonymous variants, ORF1ab P3371S, A4815V, S1261P, P5968L, R7085I, N P151L and spike M153T, are highly specific to Japanese strains.

By and large, strains circulating in Japan belong to one of the 11 strains as shown in Table 2. The most frequently observed strain is D614G_KR_P3371S, which harbors S1361P, P3371S, A4815V, and N: P151L and most likely evolved from D614G_KR in Japan in April. As shown in Figure 2, the Wuhan strain (wild type strain) and F6302L strain were introduced into Japan in January 2020 followed by L84S, and L3606F strains in February. Also, L84S_E972K strain evolved from L84S strain in Japan. In March, D614G related strains D614G, D614G_KR, D614G_Q57H were introduced into Japan, and D614G_KR_V70I strain evolved from D614G_KR. In April, a surge in D614G_KR occurred along with the emergence of D614G_KR_P3771S and D614G_KR_M234I strains. In May, D614G_KR_P3771S_M153T emerged from the parental strain D614G_KR_P3771S and started gaining momentum during the summer. Although D614G_KR_P3371S_M153T was first observed in May, transmission route as shown in Figure 3 does not originate from a single founder. Therefore, the actual emergence appears to be at an earlier time. Significant decrease in numbers were observed in June marking the end of the first wave. In July, Japan experienced the second wave of the pandemic with three predominant strains detected: D614G_KR_P3771S, its offspring D614G_KR_P3771S_M153T, and D614G_KR_M234I. Although infection rates plateaued in October, Japan entered the third wave in November. While there is a paucity of available genomes in November, D614G_KR_M234I appears to be the dominant strain in the third wave.

2145 positive cases were detected and intercepted at airport quarantine stations by the end of January 2021 since its inception in March 2020; of which 724 samples were sequenced. In total, we have identified that 115 distinct strains have been introduced into Japan. However, the number of potentially undetected cases that evaded the surveillance scrutiny during the same time period was 12 after eliminating ambiguous cases as shown in Table 3. Most of the exogenous strains appears to have been introduced into Japan before strict airport surveillance policy was implemented as shown in Figure 4C. Among the 12 cases, four independent B.1.1.7 strains advanced through the quarantine checkpoints while 30 individuals carrying the strains were intercepted and quarantined. Besides spike N501Y variant, it is quite concerning that spike E484K mutation has been reported to reduce vaccine efficacy(*12*). Furthermore, strains harboring spike E484K, namely, B.1.351, the South African strain, P.1, the Brazilian strain, and recently identified R.1 strain have been introduced into Japan. The R.1 strains were first observed in Niigata Prefecture in November (GISAID: EPI_ISL_895012) as shown in Table 3. We were not able to identify the progenitor R.1 strain in any of the high-quality genomes in the cohort. However, we discovered a probable ancestor in a genome from United Kingdom (UK) (GISAID: EPI_ISL_675164) collected in November 14 which had long undetermined bases beyond the quality threshold of this study. As UK has high volume of

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

SARS CoV-2 sequencing, and no other R.1 strain was uncovered in November, it is likely that the R.1 strain may have originated elsewhere.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 1. Top recurrent variants of SARS-CoV-2 in Japan between January 2020 and January 2021.

Nucleotide change	Gene	Amino acid change	Number of samples
23403A>G	S	D614G	14083
3037C>T	ORF1ab	F924F	14076
14408C>T	ORF1ab	P4715L	14072
241C>T	5'-UTR		14059
28881_28883delinsAAC	N	203_204delinsKR	13500
313C>T	ORF1ab	L16L	12977
10376C>T	ORF1ab	P3371S	6269
9286C>T	ORF1ab	N3007N	6268
14708C>T	ORF1ab	A4815V	6261
4346T>C	ORF1ab	S1361P	6261
28725C>T	N	P151L	6260
29692G>T	3'-UTR		6198
28975G>T	N	M234I	3736
18167C>T	ORF1ab	P5968L	3540
21518G>T	ORF1ab	R7085I	3535
22020T>C	S	M153T	2536
8076T>C	ORF1ab	V2604A	2415
8917C>T	ORF1ab	F2884F	2375
29679C>T	3'-UTR		1387
12049C>T	ORF1ab	N3928N	702
6433C>T	ORF1ab	T2056T	621
7728C>T	ORF1ab	S2488F	564
29353C>T	N	Y360Y	525
23587G>T	S	Q675H	522
28541G>A	N	A90T	508
6592A>G	ORF1ab	K2109K	506
1738G>A	ORF1ab	V491V	501
6990C>T	ORF1ab	S2242F	487
11083G>T	ORF1ab	L3606F	447
8031A>G	ORF1ab	K2589R	396
26730G>A	M	V70I	373
29764G>T	3'-UTR		373
25563G>T	ORF3a	Q57H	358
17502C>T	ORF1ab	F5746F	341
6380C>T	ORF1ab	L2039F	335
19002A>G	ORF1ab	L6246L	327
6235T>C	ORF1ab	I1990I	312
25673T>C	ORF3a	L94P	308
2167G>T	ORF1ab	K634N	259
10156C>T	ORF1ab	D3297D	252

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 2. Major strains of SARS-CoV-2 circulating in Japan.

Strain	Number of samples	Common variants/haplotype	Other major variants	Associated PANGO lineages
D614G_KR_P3371S	3747	5'-UTR: 241C>T, ORF1ab: L16L, F924F, S1361P, N3007N, P3371S, P4715L A4815V, S: D614G, N: P151L, 203_204delinsKR,3'-UTR: 29692G>T	ORF1ab: K2589R (374), V2604A (2412), S2242F (482), S2488F (560)	B.1.1.284
D614G_KR_M234I	3693	5'-UTR: 241C>T, ORF1ab: L16L, F924F, P4715L, P5968L, R7085I, S: D614G, N: 203_204delinsKR, M234I	ORF1ab: V491V (499), F2884F (2275), T2056T (615), K2109K (504), N3928N (697), S: Q675H (515), N: A90T(504), Y360Y (515), 3'-UTR: 29679C>T (1381)	B.1.1.214, B.1.1.282, B.1.1.46, B.1.1.64 B.1.1.101
D614G_KR	2810	5'-UTR: 241C>T, ORF1ab: L16L, F924F,P4715L, S:D614G, N:203_204delinsKR	3'-UTR: 29764G>T (369)	B.1.1.282, B.1.1.216, B.1.1.119, B.1.1.283
D614G_KR_P3371S_M153T	2518	5'-UTR: 241C>T, ORF1ab: L16L, F924F, S1361P, P3371S, N3007N, P4715L A4815V, S: M153T, D614G, N: P151L, 203_204delinsKR, 3'-UTR: 29692G>T	ORF1ab: K634N (259)	B.1.1.284
D614G_KR_V70I	373	5'-UTR 241C>T, ORF1ab:L16L, F924F, P4715L, S:D614G, M: V70I, N:203_204delinsKR	ORF1ab: T1242I (59), G3846S (37), D4695G (52), T5195I (37), D5225D (59)	B.1.1.48, B.1.1.289
D614G	148	5'-UTR: 241C>T, ORF1ab: F924, P4715L, S:D614G	ORF1ab: T265I(16), Y2968H (16), P5319P (25), Q5855H (25), L6688L (86), S: D796Y (16), M: D3G (26), L120L (16), V143V (16), N: S194L (17)	B.1
L84S	120	ORF1ab: S2839S, ORF8: L84S	3'-UTR: 29742G>A (81), ORF1ab: L1379L (23), L1599F (23), V5373I (84), T5871T (84), N: S202N (84)	A
D614G_Q57H	91	5'-UTR: 241C>T, ORF1ab:F924F, P4715L, S:D614G, ORF3a: Q57H	ORF1ab: T265I (68), T4640T (24), L6205L (16), M: Y71Y (18)	B.1
F6302L	68	ORF1ab: L3829L, L6302F	ORF1ab: L642F (9), N: T296I (9), K387R (9)	B.12
L84S_E972K	60	ORF1ab: E972K, C2239C, S2839S, ORF8: L84S	M: C64F (7), 3'-UTR 29739C>T (7)	A.16
L3606F	46	ORF1ab: L3606F	ORF1ab: T2016K (10), A4489V (10), S: &789Y (10), N: P13L (10), ORF10: Y26Y (6)	B
Others	65	N/A		N/A
Airport/Seaport	713	N/A		N/A

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 3 List of strains which are likely introduced into Japan after quarantines were reinforced in March 2020. EPI_ISL_675164 does not meet our quality criteria with long undetermined bases.

accession	Collection date	ancestors in global samples	description
EPI_ISL_687718	2020-06	EPI_ISL_553452 etc. (United Kingdom, 2020-06-11)	D614G_KR (B.1.1.1)
EPI_ISL_686881	2020-07	EPI_ISL_553921 etc. (United Kingdom, 2020-06-09)	D614G_KR (B.1.1.1)
EPI_ISL_902575	2020-09	EPI_ISL_734689 (Belgium, 2020-05-01)	D614G_Q57H (B.1.36)
EPI_ISL_898269	2020-10	EPI_ISL_437723 etc. (Saudi Arabia, 2020-04-16), EPI_ISL_516980 (India, 2020-05-25), EPI_ISL_558041 (United Kingdom, 2020-07-4)	D614G_Q57H (B.1.36)
EPI_ISL_900803	2020-10	EPI_ISL_481126 (India, 2020-05-22)	D614G_KR (B.1.1.74)
EPI_ISL_895012	2020-11	EPI_ISL_675164 (United Kingdom, 2020-11-14)*	D614G_KR (R.1) with spike E484K
EPI_ISL_892452	2020-12	EPI_ISL_728652 (Netherlands, 2020-12-07)	D614G (B.1.221)
EPI_ISL_892701	2020-12	EPI_ISL_811496 etc. (United Kingdom, 2021-01-02)	D614G_KR (B.1.1.7)
EPI_ISL_892703	2020-12	EPI_ISL_629440 etc. (United Kingdom, 2020-10-22)	D614G_KR (B.1.1.7)
EPI_ISL_893239	2020-12	EPI_ISL_777283 (United Kingdom, 2020-12-17)	D614G_KR (B.1.1.7)
EPI_ISL_893241	2020-12	EPI_ISL_735562 etc. (United Kingdom, 2020-12-13)	D614G_KR (B.1.1.7)
EPI_ISL_896006	2020-12	EPI_ISL_481126 (India, 2020-05-22)	D614G_KR (B.1.1.74)

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Discussion

In this work we have identified multiple independently introduced discrete strains of SARS-CoV-2 in Japan which may have contributed to community spread. We also observed further evolution of exogenous founder strains after their introduction into Japan. It is quite intriguing that D614G_KR is dominant despite the fact that comparable numbers of D614G and D614G_Q57H strains were introduced into Japan. For unknown reasons, strains other than D614G_KR are short lived in Japan(*19*).

The airport quarantine stations have been intercepted 2145 individuals carrying SARS-CoV-2 strains. For landlocked countries with contiguous borders, such quarantine strategy is not easy to implement. While cross-border surveillance is easier to implement in an island such as Japan, the study identified 12 cases, which had advanced through airport quarantine checkpoints and introduced to the Japanese population (Table 3). Since only 5% of confirmed cases have been sequenced, the actual number of individuals and SARS-CoV-2 strains passing undetected at quarantine stations is likely to be larger. Airport quarantine checkpoints have been utilizing antigen testing on saliva samples to improve lead times and throughput since July 27th 2020. The number of positive cases detected at borders decreased in August in spite of the slight increase in number of people who entered Japan as seen in Figure 4A. It is more difficult to detect pre-symptomatic or asymptomatic patients with an antigen test on a saliva sample. It is concerning that the antigen test has low positive agreement rate of 55.2% with respect to qPCR resulting in high false negative rates for nasopharyngeal samples (*20*). In fact, four distinct B.1.1.7 strains have been identified and the number of people infected with B.1.1.7 has alarmingly augmented. Mandatory qPCR should be introduced to identify and intercept emerging strains such as B.1.1.7, B.1.351, P.1. and R.1. and any novel future strains. Additionally, global genomic data should be vigilantly monitored to identify new variants that potentially may be hard to discover with a traditional antigen or an existing qPCR test.

Besides test accuracy at quarantine stations, compliance among travelers is another possibility of cases being undetected. Public health offices are urging cross-border travelers to be monitored for 14 days but there is no guarantee that they have followed the proper self-quarantine protocols. In fact, approximately 20% of cross border travelers has been lost for follow-up (*21*). Furthermore, some travelers enter Japan from exempt countries viewed as low risk; similarly, airline employees had been exempted from testing. During 14-day self-quarantine period, one can transmit to his/her family members or cohabitants, who can further transmit to people outside the household. Thus, it would be ideal to conduct interim and exit testing to further reduce infected passengers from circulating exogenous strains(*22*).

There are several limitations of this study. While the genomes from airport quarantine checkpoints were released in a timely manner; the genomes from Japanese domestic samples lagged behind by months. Furthermore, more precise information on collection dates is needed; most of the Japanese samples only carry year and month but not the date. Furthermore, specimen collection locations are not available on most instances. Therefore, it is challenging to evaluate spatial propagation of the virus and to assess the effect of a 'go-to' travel campaign to encourage business and leisure travel in an attempt to rescue the ailing travel industry. Therefore, while the study could not reveal unambiguous evidence, the campaign may have contributed to the resurgence of the virus across Japan and exacerbated the situation(*23*).

It is imperative that the government mitigate the spread of new strains carrying the spike N501Y and E484K by reinforcing quarantine policies and further ramping up genome sequencing. In particular, wide

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

spread of E484K variant strains may undermine vaccination efforts and discourage Japanese citizens from getting the vaccines. More aggressive genome sampling would facilitate understanding transmission dynamics of the virus including origins, routes, and rates, which is critical to its containment. Additional genomic analysis with clinical data as well as employing animal models will provide insights on infectivity, susceptibility, treatment resistance, vaccine evasion as well as the inferred pathogenicity and virulence(*24, 25*).

The COVID-19 pandemic is not over yet. Most citizens are not expected to be immunized by summer 2021, a time when Japan is hoping to host the postponed Olympic and Paralympics games. We should take appropriate measures for the successful events while protecting Japanese citizens. The COVID-19 pandemic won't be the last pandemic to affect mankind; therefore, reflections on the lessons learned through this ordeal should be used to ameliorate the effects of the next pandemic.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Funding

The authors received no specific funding for this work.

Authors Contributions

TK conception of work. TK RT acquisition and analysis of data. ALL interpretation of data. TK, RT drafted the work. DW, JS substantial revision. ALL reviewed the manuscript.

Completing interests

The authors declare no competing interests.

Acknowledgements

We gratefully acknowledge the authors, and the originating and submitting laboratories providing sequences from GISAID's EpiFlu™ Database and NCBI on which this research is based. The list of genomes is provided in Supplemental Table 1.

Supplemental Materials

Supplemental Table 1: Genomes Used in Analysis

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

References

1. "COVID-19 Weekly Epidemiological Update," *WHO Coronavirus disease (COVID-2019) situation reports.* (World Health Organization, Geneva, 2021).
2. J. Shigemura, M. Kurosawa, Mental health impact of the COVID-19 pandemic in Japan. *Psychol Trauma* **12**, 478-479 (2020).
3. . (World Health Organization, Geneva, 2020), vol. 2021.
4. E. Nakazawa, H. Ino, A. Akabayashi, Chronology of COVID-19 Cases on the Diamond Princess Cruise Ship and Ethical Considerations: A Report From Japan. *Disaster Med Public Health Prep* **14**, 506-513 (2020).
5. *COVID-19 Situation Report (Jan. 31, 2021),* (2021 https://www.mhlw.go.jp/stf/newpage_16463.html).
6. P. D. Koyama T., Parida L., Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ.*, (2020).
7. T. Koyama, D. Weeraratne, J. L. Snowdon, L. Parida, Emergence of Drift Variants That May Affect COVID-19 Vaccine Development and Antibody Treatment. *Pathogens* **9**, 324 (2020).
8. A. S. Lauring, E. B. Hodcroft, Genetic Variants of SARS-CoV-2—What Do They Mean? *JAMA*, (2021).
9. T. N. Starr *et al.*, Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *bioRxiv : the preprint server for biology*, 2020.2006.2017.157982 (2020).
10. K. Leung, M. H. Shum, G. M. Leung, T. T. Lam, J. T. Wu, Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill* **26**, (2021).
11. Z. Wang *et al.*, mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *bioRxiv*, 2021.2001.2015.426911 (2021).
12. X. Xie *et al.*, Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nature Medicine*, (2021).
13. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **22**, (2017).
14. S. B. Needleman, C. D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**, 443-453 (1970).
15. H. R. Max Roser, Esteban Ortiz-Ospina and Joe Hasell. (OurWorldInData.org, 2020).
16. M. D. Hall, M. E. J. Woolhouse, A. Rambaut, Using genomics data to reconstruct transmission trees during disease outbreaks. *Rev Sci Tech* **35**, 287-296 (2016).
17. *COVID-19 Situation Report (Feb. 8, 2021)* (2021 https://www.mhlw.go.jp/stf/newpage_16605.html).
18. e-Stat. (Immigration Services Agency of Japan, 2020).
19. T. Sekizuka *et al.*, A Genome Epidemiological Study of SARS-CoV-2 Introduction into Japan. *mSphere* **5**, e00786-00720 (2020).
20. Y. Hirotsu *et al.*, Comparison of automated SARS-CoV-2 antigen test for COVID-19 infection with quantitative RT-PCR using 313 nasopharyngeal swabs, including from seven serially followed patients. *Int J Infect Dis* **99**, 397-402 (2020).
21. 20% of cross border traverls are out of contacts according to Health Mnistory (in Japanese), in *Nikkei*. (Nikkei, 2021).
22. C. R. Wells *et al.*, Optimal COVID-19 quarantine and testing strategies. *Nature Communications* **12**, 356 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

23. A. Anzai, H. Nishiura, "Go To Travel" Campaign and Travel-Associated Coronavirus Disease 2019 Cases: A Descriptive Analysis, July–August 2020. *Journal of Clinical Medicine* **10**, 398 (2021).
24. M. Imai *et al.*, Syrian hamsters as a small animal model for SARS-CoV-2 infection and countermeasure development. *Proceedings of the National Academy of Sciences* **117**, 16587 (2020).
25. C. Lutz, L. Maher, C. Lee, W. Kang, COVID-19 preclinical models: human angiotensin-converting enzyme 2 transgenic mice. *Human Genomics* **14**, 20 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figures

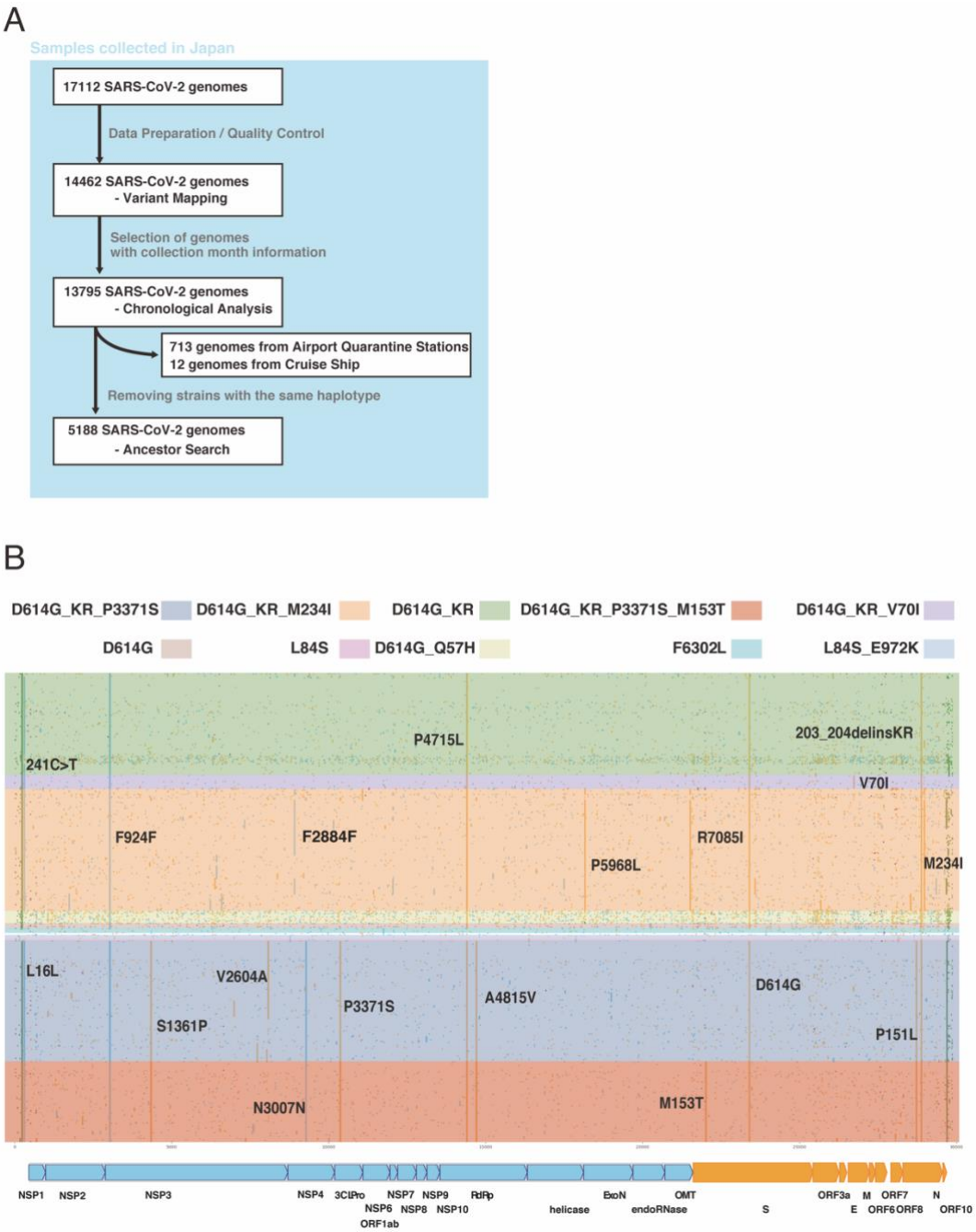

Figure 1. *A Breakdown of Japanese SARS-CoV-2 genomes used in the study. B. Variant plot of SARS-CoV-2 strains in Japan. Missense variants are in orange, synonymous variants are in cyan, non-coding variants are in green and indels are in purple.*

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

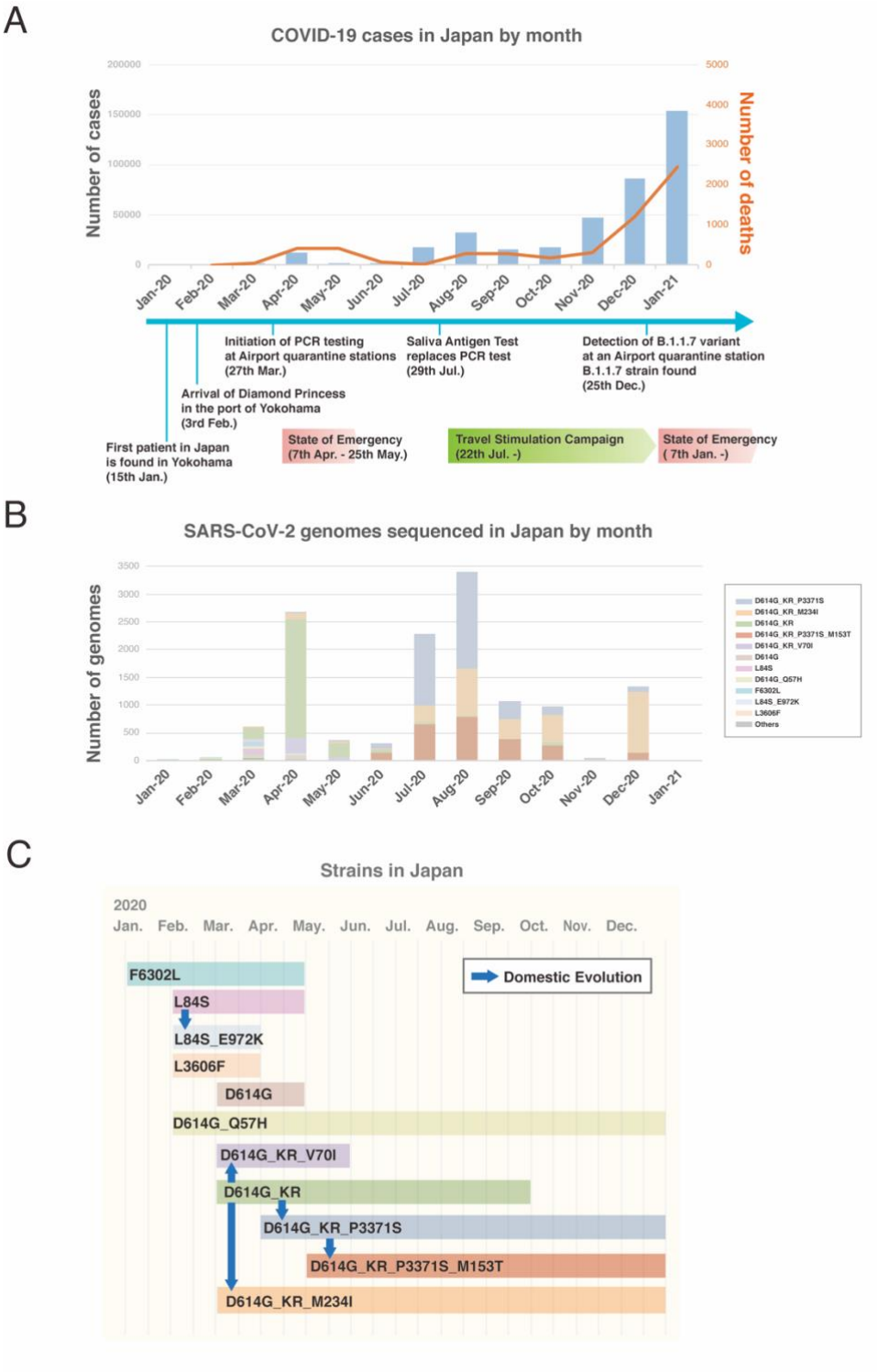

Figure 2. A. Statistics of cases and deaths in Japan from January 2020 to January 2021 with the event timeline. B. Monthly strains with classifications are plotted. C. Strains with the emergence and the last seen month are shown. Evolutionary relationships are indicated with arrows.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

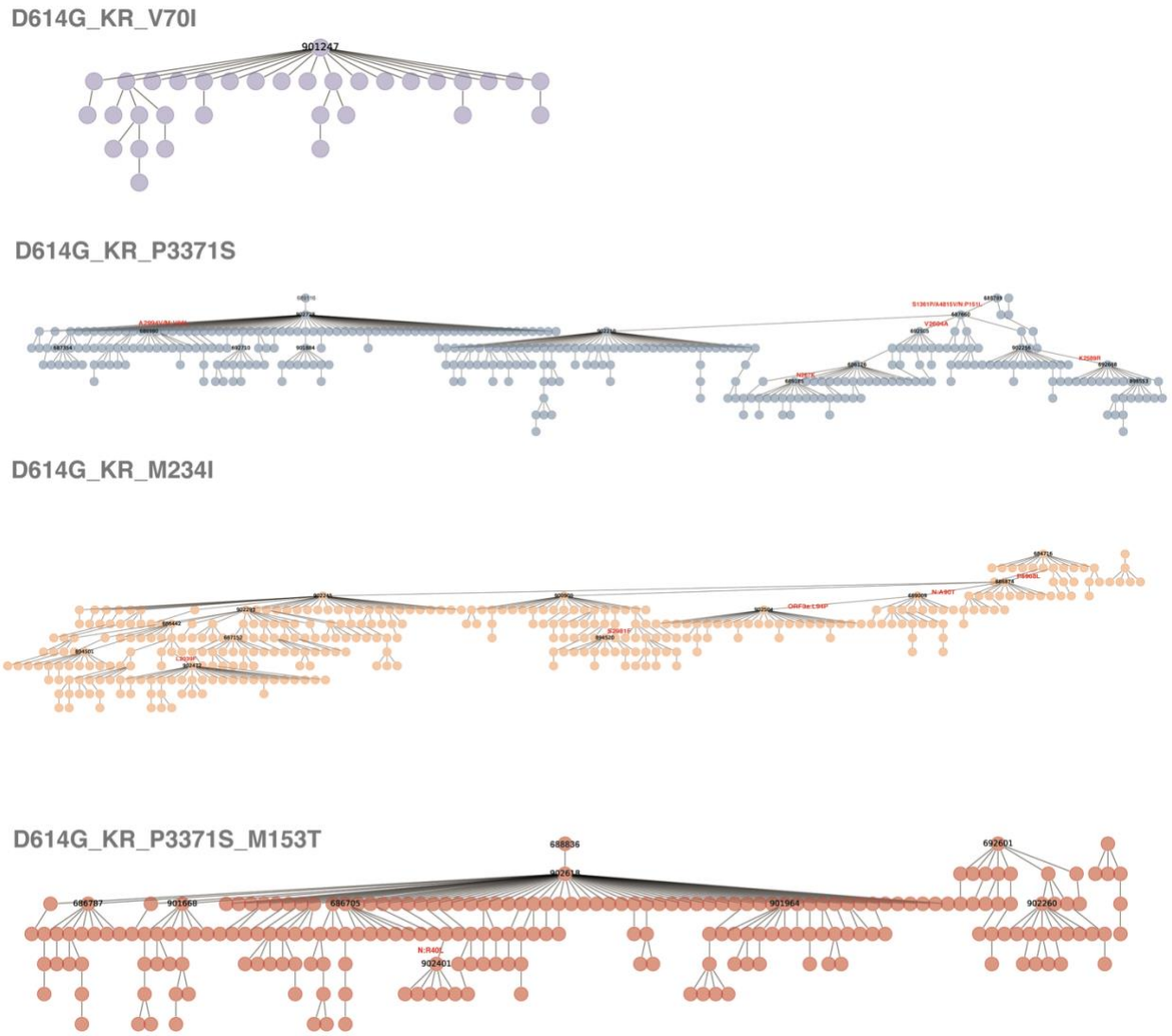

Figure 3. Transmission trees of Japan specific strains, D614G_KR_V70I, D614G_KR_P3371S, D614G_KR_M234I, and D614G_KR_P3371S_M153T are shown. For strains with more than 5 children are labeled with accession number and gained non-synonymous variants.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.26.21252555; this version posted March 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

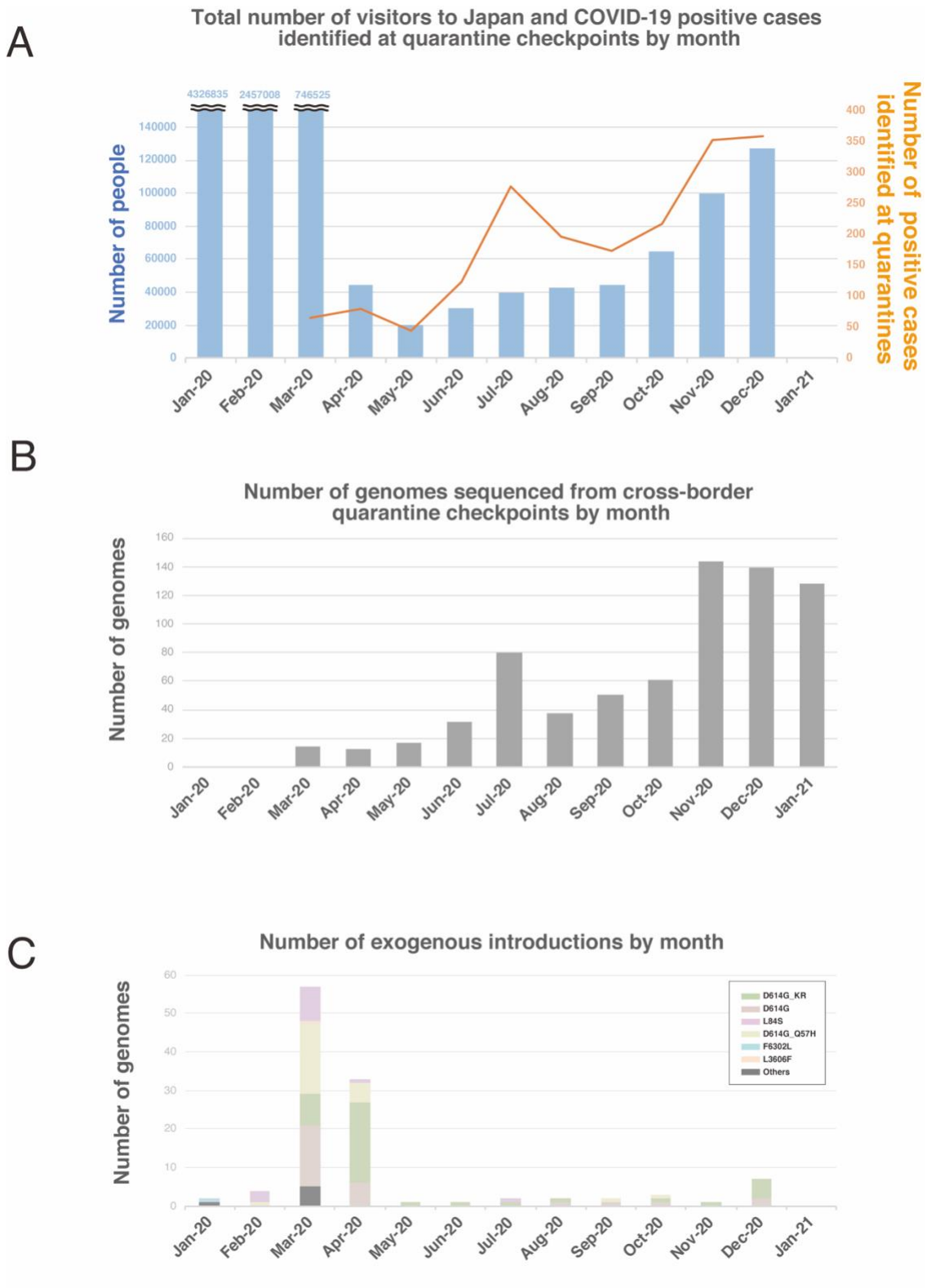

Figure 4. A. Number of inbound travelers are plotted with number of positive cases found at airport quarantine stations. B. Monthly numbers of SARS-CoV-2 genomes from airport quarantines and cruise ships are plotted. C. Number of identified exogenous strains in each month are shown with strain classifications.