

1 Genomic Surveillance in Japan of AY.29—A New Sub-lineage of SARS-CoV-2 Delta
2 Variant with C5239T and T5514C Mutations

3

4 Takashi Abe^{1,2,*}, Masanori Arita²

5 1. Faculty of Engineering, Niigata University, Niigata-ken 950-2181, Japan

6 2. Bioinformation and DDBJ Center, National Institute of Genetics, Shizuoka 411-
7 8540, Japan

8

9 * CORRESPONDENCE:

10 Takashi Abe (takaabe@ie.niigata-u.ac.jp)

11

12 Running title: AY.29: new sub-lineage of SARS-CoV-2 Delta variant

13

14 Keywords: SARS-CoV-2, Delta variant (B.1.617.2), AY.29 sub-lineage

15

16 Abstract

17 In the present study, we report a new sub-lineage of the SARS-CoV-2 Delta variant
18 called AY.29, which has C5239T and T5514C mutations. We investigated the monthly
19 trend of AY.29 in Japan within 11,954 Delta variants downloaded on September 3, 2021.
20 Among the total Japanese Delta variants, the AY.29 sub-lineage accounted for 88.4%. In
21 terms of monthly changes, the sequences became predominant in June, and accounted
22 for 93.2% and 94.2% of the reported sequences in July and August, respectively.
23 Furthermore, the number of Delta variants imported from abroad during the Tokyo 2020
24 Olympics and Paralympics (held in August 2021) was extremely low. Therefore, the
25 epidemic of the new Delta variant is attributable to a newly occurring mutation in Japan.

26

27 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread
28 rapidly across the globe since it was first reported in December 2019, and the
29 momentum of its spread has not diminished (1). To address the trend of SARS-CoV-2
30 pandemic, genome sequencing has been performed on a global scale, and the results
31 have been publicized at the global initiative on sharing avian influenza data (GISAID)
32 (2). Currently, Delta variants (Phylogenetic Assignment of Named Global Outbreak
33 (PANGO) (3) lineage: B.1.617.2) are spread worldwide and predominant in many
34 countries including Japan after its first report in India in December 2020 (4, 5).

35 In the present study, genomic surveillance was performed on the genomic
36 sequence data registered at GISAID to investigate the characteristics of the Delta
37 variants isolated in Japan. First, to investigate the diversity, haplotype networks from
38 genomic Delta single nucleotide variants (SNVs) were constructed using median joining
39 network analysis (6) with PopART version 1.7 (7) against 1,348 strains downloaded on

40 August 12, 2021 (Figure 1). The sequences were aligned using MAFFT version 7.475
41 (8), and sequences with gaps were extracted using trimAl version 1.4.rev22 (9). SNVs
42 were detected using snp-sites version 2.5.1 (10), and clustered using CD-HIT version
43 4.8.1 (11). The resulting haplotype network was divided into five groups, described as
44 N1–N5 in Figure 1. An examination of the hub nodes revealed that N1, N2, N3, and N5
45 originated from airport quarantine strains in Japan, all of which were obtained from
46 returning travelers from India. The remaining N4 was of unknown origin. Most strains
47 collected after June belonged to the nodes C1 to C3, generated from N5 containing
48 Japan/IC-1040/2021. There were two mutations (C5239T and T5514C; ORF1ab:
49 V1750A) from N5 to C1, one mutation (C28170T; ORF8: P93S) from C1 to C2, and
50 one mutation (C5365T) from C2 to C3. Therefore, strains in C1 and C3 differed by two
51 mutations only—i.e., C5239T and T5514C (ORF1ab: V1750A). The relationships
52 between protein functions and mutations T5514C (ORF1ab: V1750A) and C28170T
53 (ORF8: P93S)—which are non-synonymous substitutions—are unknown. No other
54 mutations, except for those characterizing the Delta variant, were detected.

55 Next, to determine whether mutations C5239T and T5514C (ORF1ab:
56 V1750A) are unique to Japanese strains, we searched all Delta variants downloaded on
57 September 1, 2021. Of the 908 sequences detected with these two mutations, the
58 countries and numbers of strains were as follows: Japan (883 strains), Hong Kong (2),
59 Belgium (1), Italy (1), United Kingdom (13), USA (2), and New Zealand (6). Japanese
60 strains accounted for 97.2% of the total population. The earliest strains were detected on
61 April 2020 in Japan, and the most recent strains were detected on August 23, 2021 in
62 the UK and Belgium. When the relationship between these strains on the phylogenetic
63 tree was investigated, they were assigned to the new sub-lineage into the Ultrafast

64 Sample placement on Existing tRee (UShER) global phylogenetic tree (12) (Figure 2).

65 Therefore, mutations C5239T and T5514C (ORF1ab: V1750A) are unique to Japan.

66 Based on these results, on September 2, 2021 we proposed to the PANGO
67 nomenclature team—which manages the genetic lineage information about SARS-CoV-
68 2—that the strains with these mutations constitute a new sub-lineage of the Delta
69 variant. After confirmation by PANGO, AY.29 was assigned as a new sub-lineage for
70 Delta variants (13).

71 Finally, we investigated the monthly trend of AY.29 in Japan within 11,954
72 Japanese Delta variants downloaded on September 3, 2021 (Figure 3). Sequences
73 derived from AY.29 accounted for 88.4% (10620/11954 strains). The sequences became
74 predominant in June, and accounted for 93.2% (6939/7446 strains) and 94.2%
75 (2959/3140 strains) of the detected sequences in July and August, respectively. In
76 addition to these two mutations, strains with C28170T (ORF8: P93S) and C5365T
77 mutations also accounted for approximately half of the AY.29 cases.

78 In conclusion, the predominant strain of the Japanese Delta variants currently
79 prevalent in Japan is AY.29, especially in July and August, 2021. In the GISAID data,
80 the number of Delta variants imported from abroad during the Tokyo 2020 Olympics
81 and Paralympics (held in August 2021) was extremely low. C5239T and T5514C
82 (ORF1ab: V1750A) are mutation markers that can be used to confirm whether a Delta
83 variant is of Japanese origin. Using these markers, it will be possible to quickly confirm
84 whether new Delta variants originate from AY.29 or from other possibly new variants.

85

86 ACKNOWLEDGMENTS

87 We gratefully acknowledge the authors that have submitted sequences to the GISAID

88 database. We are grateful to Dr. So Nakagawa (Tokai University) for helpful
89 discussions.

90

91 FUNDING INFORMATION

92 This research was supported by JST and CREST (grant number JPMJCR20H1).

93

94 CONFLICTS OF INTEREST

95 The authors have no conflicts of interest to declare.

96

97 REFERENCES

- 98 1. World Health Organization (WHO). 2021. Coronavirus Disease (COVID-2019),
99 Situation Reports. Available at <[https://www.who.int/emergencies/diseases/novel-](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports)
100 [coronavirus-2019/situation-reports](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports)> . Accessed September 8, 2021.
- 101 2. Shu, Y, McCauley, J. GISAID: Global initiative on sharing all influenza data - from
102 vision to reality. Euro Surveill. 2017;22.
- 103 3. Rambaut A, Holmes EC, O'Toole Á et al. A dynamic nomenclature proposal for
104 SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol.
105 2020;5(11):1403-1407.
- 106 4. World Health Organization (WHO). 2021. Tracking SARS-CoV-2 variants.
107 Available at <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>.
108 Accessed September 8, 2021.
- 109 5. Outbreak.info. 2021. COVID-19 Cases & Deaths. Available at
110 <<https://outbreak.info/>>. Accessed September 8, 2021.
- 111 6. Bandelt, HJ, Forster, P, Röhl, A. Median-joining networks for inferring intraspecific

- 112 phylogenies. *Mol. Biol. Evol.* 1999;16:37–48.
- 113 7. Leigh, JW, Bryant, D. popart: Full-feature software for haplotype network
114 construction. *Methods Eco. Evol.* 2015;6:1110–1116.
- 115 8. Katoh, K, Standley, DM. MAFFT multiple sequence alignment software version 7:
116 Improvements in performance and usability. *Mol. Biol. Evol.* 2013;30:772–780.
- 117 9. Capella-Gutiérrez, S, Silla-Martínez, JM, Gabaldón, T. trimAl: A tool for automated
118 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*
119 2009;25:1972–1973.
- 120 10. Page AJ, Taylor B, Delaney AJ et al. SNP-sites: rapid efficient extraction of SNPs
121 from multi-FASTA alignments. *Microb Genom.* 2016;29:2(4):e000056.
- 122 11. Fu L, Niu B, Zhu Z et al. CD-HIT: accelerated for clustering the next-generation
123 sequencing data. *Bioinformatics.* 2012;1:28(23):3150-3152.
- 124 12. Turakhia Y, Thornlow B, Hinrichs AS et al. Ultrafast Sample Placement on Existing
125 Trees (USHER) Empowers Real-Time Phylogenetics for the SARS-CoV-2
126 Pandemic. *Nat Genet.* 2021; 53(6):809-816.
- 127 13. Proposal for new sub-lineage of AY with C5239T and T5514C (ORF1ab:V1750A)
128 in Japan. Available at < [https://github.com/cov-lineages/pango-](https://github.com/cov-lineages/pango-designation/issues/208)
129 [designation/issues/208](https://github.com/cov-lineages/pango-designation/issues/208)>. Accessed September 8, 2021.

130

131 FIGURE LEGENDS

132 Figure 1. Haplotype networks. Representative groups are shown in N1 to N5. Nodes in
133 which the strains isolated after June were mainly classified are indicated in C1 to C3.
134 The colors indicate the months in which the strains were isolated at the bottom right of
135 the figure. The sizes of the circles indicate the number of strains.

136

137 Figure 2. Phylogenetic tree derived from candidate sequences assigned by UShER. The
138 colors indicate the clades. The candidate strains are shown in red, and the other strains
139 are shown in the upper right of the figure.

140

141 Figure 3. Monthly trend of the AY.29 sub-lineage in Japan. C1 to C3 are the group IDs
142 defined in Figure 1. The numbers in parentheses are the total numbers of strains isolated
143 in a particular month. The number of strains belonging to each group is shown at the
144 bottom of the figure.

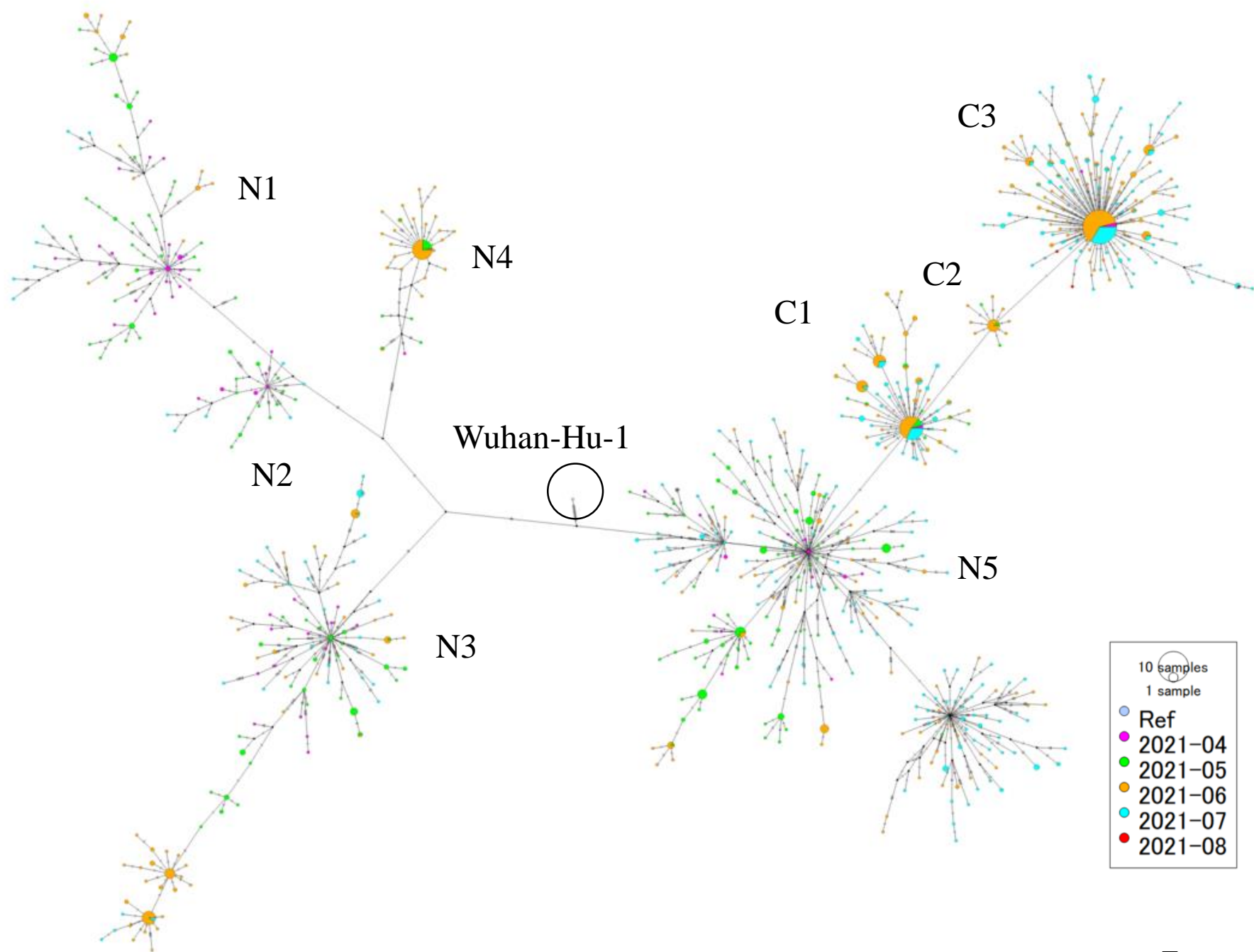


Figure 1

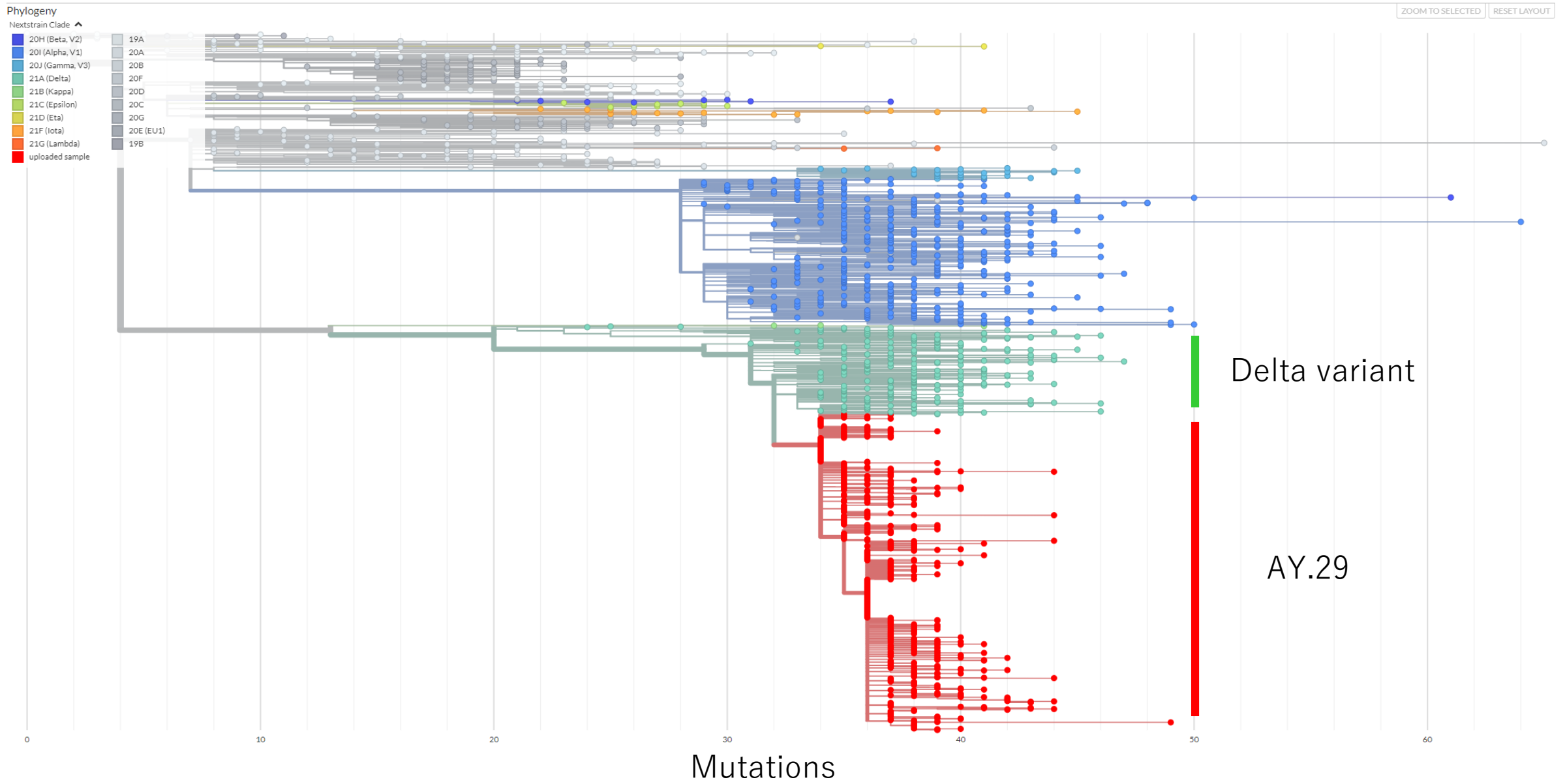
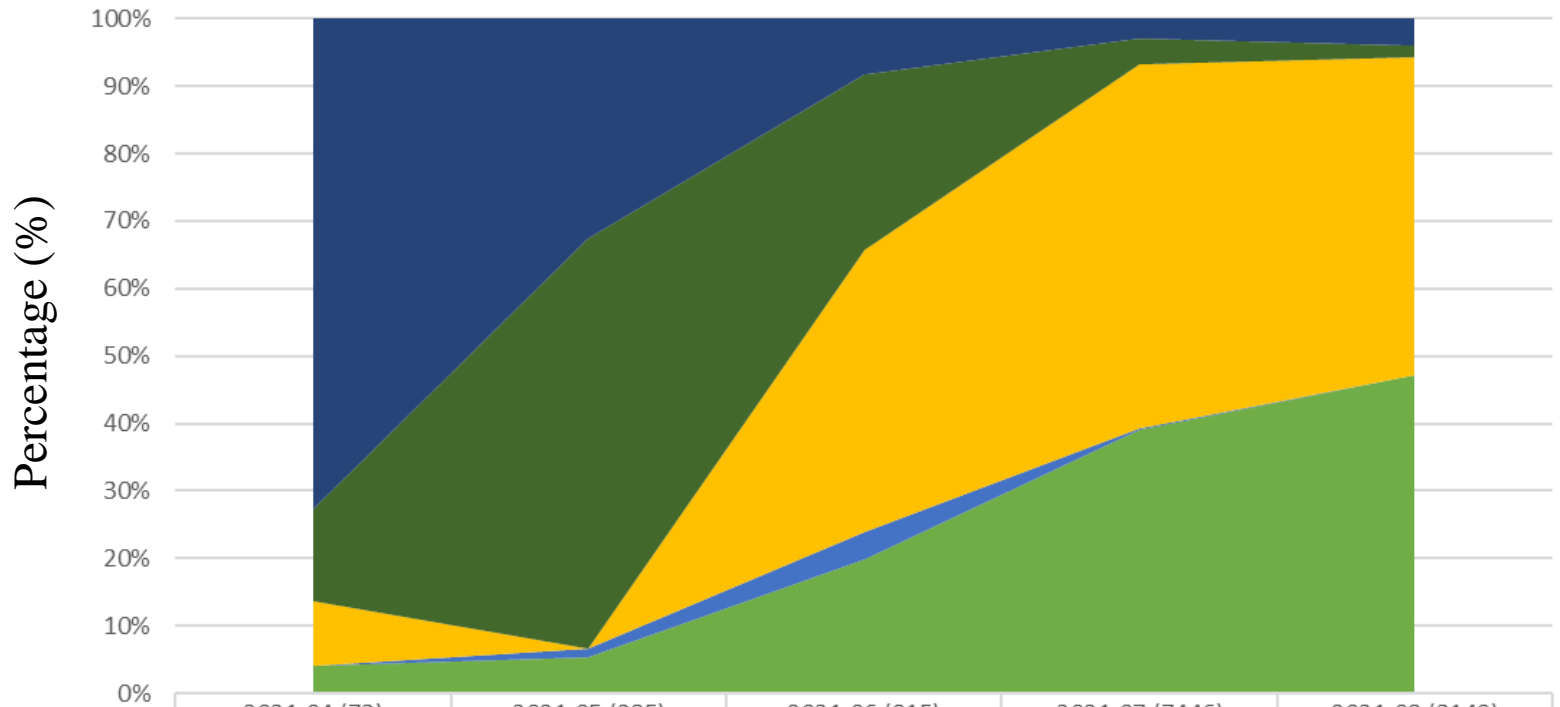


Figure 2



	2021-04 (73)	2021-05 (285)	2021-06 (915)	2021-07 (7446)	2021-08 (3140)
■ Airport quarantine strains	53	93	76	220	124
■ Other Delta variant	10	173	237	287	57
■ C3	7		384	4021	1476
■ C2	0	4	37	12	1
■ C1	3	15	181	2906	1482

■ C1 ■ C2 ■ C3 ■ Other Delta variant ■ Airport quarantine strains

Figure 3