δ1 variant of SARS-COV-2 acquires spike V1176F and yields a highly mutated subvariant in Europe

Xiang-Jiao Yang[1,2,3,4,*]

[1]Rosalind & Morris Goodman Cancer Institute, [2]Department of Medicine and [3]Department of Biochemistry, McGill University, Montreal, Quebec H3A 1A3, Canada
[4]Department of Medicine, McGill University Health Center, Montreal, Quebec H4A 3J1, Canada

[*]Corresponding contact: xiang-jiao.yang@mcgill.ca; Tel: 514-398-5883

ABSTRACT

Genomic surveillance of SARS-COV-2 has revealed that in addition to many variants of interests, this virus has yielded four variants of concern, α, β, γ and δ, as designated by the World Health Organization. δ variant has recently become the predominant pandemic driver around the world and yielded four different subvariants (δ1, δ2, δ3 and δ4). Among them, δ1 has emerged as the key subvariant that drives the pandemic in India, Europe and the USA. A relevant question is whether δ1 subvariant continues to evolve and acquires additional mutations. Related to this, this subvariant has acquired spike V1176F, a signature substitution of γ variant, and yielded a new sublineage, δ1F. The substitution alters heptad repeat 2 of spike protein and is expected to improve interaction with heptad repeat 1 and enhance virus entry. Moreover, there are δ1F sublineages encoding spike N501Y, A783S, Q836E and V1264L. While N501Y is a signature substitution shared by α, β and γ variants, V1264L is a key substitution in a δ1 sublineage that is a major pandemic driver in Southeast Asia. The Q836E-encoding lineage carries an average of 50 mutations per genome, making it the most mutated variant identified so far. Similar to δ1 subvariant, δ2 subvariant has also acquired spike V1176F and yielded new sublineages. Together, these results suggest that V1176F is a recurrent spike substitution that is frequently acquired by SARS-COV-2 variants to improve viral fitness. It is thus important to track the evolutionary trajectory of related variants for considering and instituting the most effective public health measures.

Running Title: Evolution of SARS-COV-2 via continuous branching

Keywords: Severe acute respiratory syndrome coronavirus, cell-cell fusion, B.1.1.7, B.1.351, P.1, B.1.617.2, A222V, R203M, L452R, T478K, phylogenetic analysis, mutation profiling

INTRODUCTION

Coronavirus disease 2019 (COVID-19) has caused the pandemic, led to tragic loss of life, overloaded health care systems and crippled the economy around the world. Just as this pandemic is unprecedented in the recent human history, global genomic surveillance of severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) is also unparalleled. As of October 16, 2021, almost 4.4 million SARS-COV-2 genomes have been sequenced and deposited into the GISAID (global initiative on sharing avian influenza data) genome sequence database [1]. This invaluable resource provides invaluable information and golden opportunities to track the evolutionary trajectory of the virus. Genomic surveillance of SARS-CoV-2 has firmly established that its evolution is the driving force of this pandemic. In addition to many variants of interests, the virus has yielded four variants of concern as designated by the WHO: α (B.1.1.7) [2], β (B.1.351) [3], γ (P.1) [4] and δ (B.1.617.2) [5]. One important question is what we can learn from known variants about the evolutionary trajectory of SARS-COV-2. Answers to this question will provide valuable scientific evidence for making informed decisions on the most effective public health measures, and help us identify potential future variants of concern.

To address this important question, I have tracked genomes in the GISAID database very closely. I have very recently found that δ variant has evolved further and yielded four subvariants (δ1, δ2, δ3 and δ4) in India [6]. Among them, δ1 has emerged as the major pandemic driver and δ2 plays a much less important role, whereas δ3 and δ4 have gradually faded away [6]. δ1 subvariant is also the major pandemic driver in Europe [6] and the USA [7]. These new findings raise a relevant question about whether δ1 subvariant continues to evolve and acquires additional mutations. Related to this question, I tried to determine whether SARS-COV-2 mutations show any conceptual analogy with germline mutations identified in genetic diseases and somatic mutations associated with cancer, because I have been working on these two types of mutations [8-10].

As a result, I realized and postulated that very similar to cancer mutations, SARS-COV-2 mutations are grouped into three large categories: driver, facilitator and passenger mutations. Different from passenger mutations, driver and facilitator mutations are recurrent. Recurrent frequency can sometimes help us distinguish between facilitator and passenger mutations. Notably, driver or facilitator mutations during SARS-COV-2 evolution are not so dominant as those identified in cancer. Still, some SARS-COV-2 mutations clearly act as drivers. One such example is D614G, a key substitution that has changed the entire course of the pandemic [11-13]. This substitution alters conformational state of the spike protein [14]. Another example is spike N501Y, shared by α [2], β [3] and γ [4] variants. This substitution enhances interaction with the cell-try receptor ACE2 [15]. A third example is spike P681R encoded by δ and κ variants [5]. It improves the furin cleavage site and confers evolutionary advantage [16]. With this hypothesis as a guide, I mapped the evolutionary trajectory of SARS-COV-2 evolution.

Among the four variants of concern, only γ variant encodes V1176F [4]. This substitution is located within heptad repeat 2 (HR2) of the S2 domain of spike protein and expected to improve interaction with HR1, which may facilitate virus entry to target cells by enhancing cell fusion potential [17]. This raises the question whether α [2], β [3] and δ [5] variants acquire V1176F during evolution. Here I describe that δ1 and δ2 subvariants have acquired spike V1176F and formed the new sublineages δ1F and δ2F, respectively (where the letter F denotes F1176). Moreover, δ1F has evolved further and acquired additional mutations, including spike N501Y, A783S, Q836E and V1264L. Alarmingly, the Q836E-encoding lineage, referred to as carries δ1F1, carries an average of 50 mutations per genome, thereby making it the most mutated SARS-COV-2 variant identified so far. Therefore, V1176F is recurrent and belongs to the facilitator or driver mutation category.

RESULTS AND DISCUSSION

Spike V1176F is a recurrent substitution that may improve HR2 function

As shown in Fig. 1A, spike protein is a multidomain protein composed of an N-terminal domain, a receptor-binding domain, a fusion peptide, a fusion peptide proximal region, heptad-repeat regions 1 and 2, a transmembrane motif and a C-terminal cytoplasmic tail. A recent study showed that the fusion peptide C-terminal proximal region is important for spike protein function (Fig. 1A) [18]. V1176 is located within HR2 (Fig. 1A). HR2 is composed four heptad repeats and V1176 occupies a key position of repeat 1 (Fig. 1B). At the 3D structural level, V1176 is part of an unstructured loop. As four heptad repeats utilize key hydrophobic residues, such as V1176, for interaction with HR2 [17]. Because phenylalanine is much more hydrophobic than alanine, V1176F is expected to facilitate such interaction. This is also consistent that V1176F is a signature substitution in γ variant and its related variants, P.2 and P.3 [4,19]. As shown in Fig. 1B, phenylalanine is also present at the equivalent position in the spike proteins from HKU1 and OC43. By contrast, a hydrophilic residue is present at the equivalent position in the spike proteins from NL63 and 229E, which is expected to make HR2 of these two coronaviruses less optimal. Thus, HR2 function is one potential avenue, on which SARS-COV-2 and related coronaviruses improves viral fitness.

Hundreds of V1176F-encoding δ genomes form distinct groups

Among the four **v**ariants of **c**oncern, only γ variant encodes V1176F [4]. In light of the potential importance of this substitution in improving viral fitness, I asked whether δ variant has also acquired this substitution. As a result, it was used it as a search criterion to identify δ genomes in the GISAID database. Indeed, this searched uncovered hundreds of δ genomes that encode V1176F as an extra substitution. Then I utilized Coronapp, an efficient web-based SARS-COV-2 mutation profiling application [20,21], to produce reports on mutations in 817 such V1176F-encoding δ genomes that were identified and sequenced in different countries around the world. The profiling results revealed that these genomes are mainly derivatives of δ1 and δ2 subvariants (with the corresponding sublineages referred to δ1F and δ2F, respectively, where the letter F denotes F1176; Fig. 2), indicating that V1176F has been independently acquired. Intriguingly, a subset of genomes encode spike V1264L (Fig. 2B), located at the cytoplasmic tail (Fig. 1A). This substitution is recurrent and may make δ1 and δ2 sublineages more virulent [22]. Another subset encodes spike Q836E (Fig. 2B), located at the fusion peptide proximal region (Fig. 1A). Consistent with this heterogeneity, the distribution mutation load curve forms two distinct peaks, with the major peak carrying an average of 41 mutations per genome and the smaller peak centering at an average of 49-50 mutations per genome (Fig. 3A). The mutation load from the major peak is expected from typical δ genomes identified in August and September 2021 [6,7,22], but the mutation load of the small peak is totally unexpected.

This raises an intriguing possibility that the small peak corresponds to V1264L- or Q836E-encoding δ genomes. Thus, these two types of genomes were identified and analyzed via Coronapp [20,21]. As shown in Fig. 3B-C, V1264L-encoding δ genomes possess 42 mutations per genome, whereas Q836E-encoding δ genomes carry 50 mutations per genome. Thus, the small peak in Fig. 3A is due to Q836E-encoding δ genomes. Manual inspection of the mutations revealed that these Q836E-encoding δ genomes are derivatives of δ1F subvariant, so the resulting sublineage is designated as δ1F1. Its mutation load at the average of ~50 mutations per genome makes δ1F1 the most mutated SARS-COV-2 lineage identified so far. In comparison, the mutation load is an average of 47 mutations per genome for C.1.2 variant, which has been reported to be the most mutated variant [23,24]. There are several δ1 sublineages carrying 46-49 mutations per genome [6,22].

Manual inspection of the mutations from V1264L-encoding δ genomes revealed that they are δ2F derivatives, supporting that V1176F has been acquired by two different δ subvariants. In agreement with this, phylogenetic analysis V1176F-encoding genomes identified two large groups corresponding to δ1F and δ2F (Figs 3D & S1). δ3 and δ4 subvariants have also acquired V1176F even though in terms of population size, the resulting sublineages are minor compared to those from δ1 and δ2 subvariants (Figs 3D & S1). Further phylogenetic analysis of V1264L-encoding δ genomes identified a major group encoding an extra spike substitution, P809S (Figs 3E & S2). These genomes were mainly identified in the USA (Fig. S1). Therefore, V1176F is recurrent and has been acquired by different δ subvariants.

Mutational analysis of δ1F1 as the most mutated SARS-COV-2 variant identified so far

Q836E-encoding δ-genomes were mainly identified in Europe (Figs 3D & S1). Mutational analysis of these genomes uncovered four silent mutations, F1107F of NSP3, S477S and Y707Y of spike protein, and S318S of nucleocapsid protein. Over 50% of the genomes also encode a fifth silent mutation, spike V991V. These five silent mutations contribute to the high mutation load of δ1F1 genomes and may have been subject to positive selection. Alternatively, they are just passenger mutations. However, silent mutations have been found to improve fitness of HIV-1 variants [25,26], so further investigation of the potential impact of these five silent mutations is warranted. In addition to the five silent mutations, spike A783S is encoded in over 20% of the genomes (Fig. 4B). This substitution and V991V are mutually exclusive, so there are at least two subgroups within the group of Q836E-encoding δ-genomes. In support of this, phylogenetic analysis identified three subgroups, with two corresponding to the A783S- and V991V-encoding ones (Fig. 5). Troublingly, inspection of the third subgroup identified one genome that encodes spike N501Y (Figs 5 & S3). This signature substitution is shared by α [2], β [3] and γ [4] variants of concern, and known to improve ACE2 interaction [15]. Thus, δ1F1 has evolved further and acquired additional substitutions, including spike N501Y and A783S.

Mechanistic impact of different spike substitutions and a hypothetical model

As shown in Fig. 1, V1176F is located within HR2 of spike protein and expected to improve interaction with HR1, which may in turn facilitate virus entry by enhancing cell fusion potential [17]. As illustrated in Fig. 1 & 6A, P809 is only 6 residues away from the S2' cleavage site. Structurally, this proline may hinder the cleavage (Fig. 6B), so P809S may improve cleavage at this site. At the 3D structural level, A783 and Q836 are not far away from the cleavage site (Fig. 6B-C). The side chain of Q836 is 3.5 Å away from D839 and may thus form a hydrogen bond (Fig. 6B). A873 is in proximity to K786 (~5.1 Å), and both residues are on the same surface as the S2' cleavage loop (Fig. 6B). so A783S and Q836E may modulate the cleavage. Moreover, Q836 is located within a fusion peptide proximal region, which was recently found to be very important for spike protein function [18]. Thus, the S2' cleavage site (Figs 1A & 6A-B) is a hot spot for SARS-COV-2 mutations to improve its viral fitness.

Four δ subvariants (δ1, δ2, δ3 and δ4) were initially identified in India [6] and then spread to the rest of the world [6,7,22]. Through convergent evolution, they have then acquired V1176F and yielded four sublineages (δ1F, δ2F, δ3F and δ4F; Figs 3D & 6D). Among them, δ1F is the fastest evolving and has acquired additional mutations. As for δ1F1 (Fig. 5), its mutation load at the average of ~50 mutations per genome makes it the most mutated SARS-COV-2 variant identified so far. Alarmingly, one δ1F genome identified in September 2021 encodes N501Y (Figs 5 & S3), a key spike substitution shared by α, β and γ variants of concern. Compared to δ1F, δ2F carries fewer mutations, but one subgroup encodes P809S (Figs 3E & 5), which is close to the S2' cleavage site and may improve its cleavage efficiency. Thus, it will be important to track evolution of δ1F and δ2F in the coming few months. In summary, different δ subvariants have acquired spike V1176F and formed the new sublineages. Its importance and recurrent nature call for the need to annotate other substitutions located at the HR1 and HR2 regions of spike protein (Fig. 1A).

ACKNOWLEDGEMENT

I gratefully acknowledge the GISAID for diligent and tireless maintenance SARS-COV-2 genomes and numerous investigators for the valuable genome sequences used in this work (see the supplementary section for details). I am also grateful to Professor Federico M. Giorgi at University of Bologna, Italy, for developing Coronapp and generously allowing me timely access to the Coronapp server. This work was supported by funds from Canadian Institutes of Health Research (CIHR), Natural Sciences and Engineering Research Council of Canada (NSERC) and Compute Canada (to X.J.Y.).

DECLARATION OF INTERESTS

The author declares no competing interests.

MATERIALS AND METHODS

SARS-COV-2 genome sequences, mutational profiling and phylogenetic analysis

The genomes were downloaded the GISAID database on the dates specified in the figure legends. CoVsurver (https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/) was used to analyze mutations on representative SARS-COV-2 genomes. Fasta files containing specific groups of genomes were downloaded from the GISAID database. During downloading, each empty space in the Fasta file headers was replaced by an underscore because such a space makes the files incompatible for subsequent mutational profiling, sequence alignment and phylogenetic analysis, as described with details in two other studies [6,24]. The Fasta headers were shortened and modified further as described [6,24]. The cleaned Fasta file was used for mutational profiling via Coronapp (http://giorgilab.unibo.it/coronannotator/), a web-based mutation annotation application [20,21]. The cleaned Fasta file was also uploaded onto SnapGene (version 5.3.2) for multisequence alignment via the MAFFT tool. RAxML-NG version 0.9.0 [27] was used for phylogenetic as described [6].

Defining different variant genomes using various markers

α, β, γ, δ and other variant genomes were downloaded from the GISAID database as defined by the server. δ subvariant genomes were defined as described [6]. Briefly, nucleocapsid substitutions G215C and R385K (Table 1) were used as markers for δ1 or δ3 genomes, respectively. Spike substitutions A222V and K77T were used as markers for δ2 or δ4 genomes, respectively. In Europe, there are many δ1V genomes that also encode spike A222V, so the NSP3 substitution P822L was used together with spike A222V to identify δ2 genomes. As discussed previously [6], there are several limitations with these markers. But they should not affect the overall conclusions.

PyMol structural modeling

The PyMol molecular graphics system (version 2.4.2, https://pymol.org/2/) from Schrödinger, Inc. was used for downloading structure files from the PDB database for further analysis and image generation. Structural images were cropped via Adobe Photoshop for further presentation through Illustrator.

REFERENCES

1. Elbe, S. and G. Buckland-Merrett. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1**, 33-46.
2. Volz, E., S. Mishra, M. Chand, J.C. Barrett, R. Johnson, L. Geidelberg, W.R. Hinsley, D.J. Laydon, G. Dabrera, A. O'Toole, R. Amato, M. Ragonnet-Cronin, I. Harrison, B. Jackson, C.V. Ariani, O. Boyd, N.J. Loman, J.T. McCrone, S. Goncalves, D. Jorgensen, R. Myers, V. Hill, D.K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D.P. Kwiatkowski, C.-G.U. consortium, S. Flaxman, O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, and N.M. Ferguson. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, epub.
3. Tegally, H., E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E.J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A.J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, N.Y. Hsiao, S. Korsman, M.A. Davies, L. Tyers, I. Mudau, D. York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A. Alisoltani-Dehkordi, A. Godzik, C.K. Wibmer, B.T. Sewell, J. Lourenco, L.C.J. Alcantara, S.L. Kosakovsky Pond, S. Weaver, D. Martin, R.J. Lessells, J.N. Bhiman, C. Williamson, and T. de Oliveira. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438-443.
4. Faria, N.R., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815-821.
5. Mlcochova, P., et al. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*, Epub.
6. Yang, X.J. (2021). SARS-COV-2 delta variant drives the pandemic in India and Europe through two subvariants. *medRxiv (submitted)*.
7. Yang, X.J. (2021). SARS-COV-2 delta variant drives the pandemic in the USA through two subvariants. *Research Square*, https://www.researchsquare.com/article/rs-986605/v1.
8. Yan, K., J. Rousseau, R.O. Littlejohn, C. Kiss, A. Lehman, J.R. Rosenfeld, C.T.R.

Stumpel, A.P. Stegmann, L. Robak, F. Scaglia, T.T. Nguyen, H. Fu, N.F. Ajeawung, M.V. Camurri, L. Li, A. Gardham, B. Panis, M. Almannai, M.J. Sacoto, B. Baskin, C. Ruivenkamp, D. Study, C. Study, M.T. Cho, T. Potjer, G.W. Santen, M.J. Parker, N. Canham, M. McKinnon, L. Potocki, J. MacKenzie, E.R. Roeder, P.M. Campeau, and X.J. Yang. (2017). Mutations in the chromatin regulator gene BRPF1 causes syndromic intellectual disability and deficient histone acetylation. *Am. J. Hum. Genet.* **100**, 91-104.

9. Yan, K., J. Rousseau, K. Machol, L.A. Cross, K.E. Agre, C.F. Gibson, A. Goverde, K.L. Engleman, H. Verdin, E. Baere, L. Potocki, D. Zhou, M. Cadieux-Dion, G.A. Bellus, M.D. Wagner, R.J. Hale, N. Esber, A.F. Riley, B.D. Solomon, M.T. Cho, H. Porter, B.C. Lanpher, A.M. Lewis, J. Savatt, I. Thiffault, B. Callewaert, P.M. Campeau, and X.J. Yang. (2020). Deficient histone H3 propionylation by BRPF1-KAT6 complexes in neurodevelopmental disorders and cancer. *Sci. Adv.* **6**, eaax0021.

10. Li, L., M. Ghorbani, M.W. Hubshman, J. Rousseau, I. Thiffault, R.E. Schnur, C. Breen, R. Oegema, J.M. Weiss, Q. Waisfisz, S. Welner, H. Kingston, E.M. Boon, L. Basel-Salmon, O. Konen, H. Goldberg, L. Bazaq, S. Tzur, X. Bi, M. Bruccoleri, K. McWalter, M. Cho, M. Scarano, S.S. Brooks, S.S. Hughes, K.L.I. van Gassen, J.M. van Hagen, T.K. Pandita, P.B. Agrawal, P.M. Campeau, and X.J. Yang. (2020). Lysine acetyltransferase 8 is involved in cerebral development and syndromic intellectual disability. *J. Clin. Investig.* **130**, 1431-1445.

11. Plante, J.A., Y. Liu, J. Liu, H. Xia, B.A. Johnson, K.G. Lokugamage, X. Zhang, A.E. Muruato, J. Zou, C.R. Fontes-Garfias, D. Mirchandani, D. Scharton, J.P. Bilello, Z. Ku, Z. An, B. Kalveram, A.N. Freiberg, V.D. Menachery, X. Xie, K.S. Plante, S.C. Weaver, and P.Y. Shi. (2020). Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. *bioRxiv*.

12. Shi, P.Y., J. Plante, Y. Liu, J. Liu, H. Xia, B. Johnson, K. Lokugamage, X. Zhang, A. Muruato, J. Zou, C. Fontes-Garfias, D. Mirchandani, D. Scharton, B. Kalveram, J. Bilello, Z. Ku, Z. An, A. Freiberg, V. Menachery, X. Xie, K. Plante, and S. Weaver. (2020). Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. *Res Sq*.

13. Plante, J.A., Y. Liu, J. Liu, H. Xia, B.A. Johnson, K.G. Lokugamage, X. Zhang, A.E. Muruato, J. Zou, C.R. Fontes-Garfias, D. Mirchandani, D. Scharton, J.P. Bilello, Z. Ku, Z. An, B. Kalveram, A.N. Freiberg, V.D. Menachery, X. Xie, K.S. Plante, S.C. Weaver, and P.Y. Shi. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116-121.

14. Cai, Y., J. Zhang, T. Xiao, H. Peng, S.M. Sterling, R.M. Walsh, Jr., S. Rawson, S. Rits-Volloch, and B. Chen. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592.

15. Starr, T.N., A.J. Greaney, S.K. Hilton, D. Ellis, K.H.D. Crawford, A.S. Dingens, M.J. Navarro, J.E. Bowen, M.A. Tortorici, A.C. Walls, N.P. King, D. Veesler, and J.D. Bloom. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310 e20.
16. Yang, X.J. (2021). SARS-COV-2 gamma variant acquires spike P681H or P681R for improved viral fitness *bioRxiv*, https://www.biorxiv.org/content/10.1101/2021.10.16.464641v1.
17. Liu, S., G. Xiao, Y. Chen, Y. He, J. Niu, C.R. Escalante, H. Xiong, J. Farmar, A.K. Debnath, P. Tien, and S. Jiang. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* **363**, 938-47.
18. Cai, Y., J. Zhang, T. Xiao, C.L. Lavine, S. Rawson, H. Peng, H. Zhu, K. Anand, P. Tong, A. Gautam, S. Lu, S.M. Sterling, R.M. Walsh, Jr., S. Rits-Volloch, J. Lu, D.R. Wesemann, W. Yang, M.S. Seaman, and B. Chen. (2021). Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. *Science* **373**, 642-648.
19. Tablizo, F.A., K.M. Kim, C.M. Lapid, M.J. Castro, M.S. Yangzon, B.A. Maralit, M.E. Ayes, E.M. Cutiongco-de la Paz, A.R. De Guzman, J.M. Yap, E.M. Salvaña, J.Q. Wong, J.C. Montoya, M.R. Singh-Vergeire, and C.P. Saloma. (2021). Genome sequencing and analysis of an emergent SARS-CoV-2 variant characterized by multiple spike protein mutations detected from the Central Visayas Region of the Philippines. *bioRxiv*, https://doi.org/10.1101/2021.03.03.21252812.
20. Mercatelli, D., L. Triboli, E. Fornasari, F. Ray, and F.M. Giorgi. (2021). Coronapp: A web application to annotate and monitor SARS-CoV-2 mutations. *J Med Virol* **93**, 3238-3245.
21. Mercatelli, D. and F.M. Giorgi. (2020). Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front Microbiol* **11**, 1800.
22. Yang, X.J. (2021). Delta-1 variant of SARS-COV-2 acquires spike V1264L and drives the pandemic in Indonesia, Singapore and Malaysia. *Research Square (submitted)*.
23. Scheepers, C., J. Everatt, D.G. Amoako, A. Mnguni, A. Ismail, B. Mahlangu, C.K. Wibmer, E. Wilkinson, H. Tegally, J.E. Emmanuel San, J. Giandhari, N. Ntuli, S. Pillay, T. Mohale, Y. Naidoo, Z.T. Khumalo, Z. Makatini, A. Sigal, C. Williamson, F. Treurnicht, K. Mlisana, M. Venter, N.Y. Hsiao, N. Wolter, N. Msomi, R. Lessells, T. Maponga, W. Preiser, P.L. Moore, A. von Gottberg, T. de Oliveira, and J.N. Bhiman. (2021). The continuous evolution of SARS-CoV-2 in South Africa: a new lineage with rapid accumulation of mutations of concern and global detection. *medRxiv*, https://www.medrxiv.org/content/10.1101/2021.08.20.21262342v1.full.
24. Yang, X.J. (2021). SARS-COV-2 C.1.2 variant is highly mutated but may exhibit

reduced affinity for ACE2 receptor. *bioRxiv*, https://www.biorxiv.org/content/10.1101/2021.10.16.464644v1.

25. Harrigan, P.R., C.W. Sheen, V.S. Gill, B. Wynhoven, E. Hudson, V.D. Lima, P. Lecocq, R. Aguirre, A.F. Poon, and N. Sluis-Cremer. (2008). Silent mutations are selected in HIV-1 reverse transcriptase and affect enzymatic efficiency. *AIDS* **22**, 2501-8.
26. Telwatte, S., A.C. Hearps, A. Johnson, C.F. Latham, K. Moore, P. Agius, M. Tachedjian, S. Sonza, N. Sluis-Cremer, P.R. Harrigan, and G. Tachedjian. (2015). Silent mutations at codons 65 and 66 in reverse transcriptase alleviate indel formation and restore fitness in subtype B HIV-1 containing D67N and K70R drug resistance mutations. *Nucleic Acids Res* **43**, 3256-71.
27. Kozlov, A.M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453-4455.
28. Yang, X.J. (2021). Delta-1 variant of SARS-COV-2 acquires spike V1176F and yields a highly mutated subvariant in Europe. *bioRxiv*, https://www.biorxiv.org/content/10.1101/2021.10.16.463825v1.

FIGURE LEGENDS

Figure 1. Spike V1176F alters heptad repeat 2. (**A**) Domain organization of spike protein. Some key substitutions are shown in black and new substitutions to be described herein are in red. NTD, N-terminal domain; RBD, receptor-binding domain; S1/S2, boundary of S1 and S2 domains after furin cleavage between residues R685 and S686; FP, fusion peptide; FPR, fusion peptide proximal region; HR1 and HR2, heptad-repeat regions 1 and 2, respectively; S2', cleavage site between residues R815 and S816 in S2 domain; TM, transmembrane motif; CT, C-terminal cytoplasmic tail. The domain organization was adapted from a published study [18]. (**B**) Sequence alignment of spike proteins from SARS-COV-2 and five other coronaviruses. For simplicity, only heptad repeat 2 and the transmembrane domain are shown. Positions of key hydrophobic residues defining the four heptad repeats are indicated at the top. V1176 is a key residue in HR2 and V1176F may make the domain more optimal for interaction with HR1. Phenylalanine is also present at the equivalent position in the spike proteins from HKU1 and OC43. By contrast, a hydrophilic residue is present at the equivalent position in the spike proteins from NL63 and 229E. (**C**) Structure around V1176 of spike protein. The S2 domain of the protein form a trimer. For simplicity, only the hexa-helix bundle formed by HR1 and HR2 is shown here. HR1 occupies the core of the bundle whereas HR2 is at the out-skirt. The region from A1174 and I1179 of HR2 could not be defined structurally (depicted with dashed lines) and may be unstructured, so only an approximate position of V1176 is indicated. It is close to V951, V952 and A956 of HR1, so V1176F may enhance hydrophobic interaction with these three residues. Based on PyMol presentation of 6XRA from the PDB database.

Figure 2. Mutation profile of 817 V1176F-encoding δ genomes from different countries around the world. The genomes were downloaded from the GISAID SARS-COV-2 genome sequence database on September 24, 2021 for mutation profiling via Coronapp [20,21]. Shown in (A), (B) and (C) are substitutions in NSP3, spike and nucleocapsid proteins, respectively.

Figure 3. Mutation load and phylogenetic relationship of different δ subvariants. (**A-C**) Mutation load of three different groups of δ genomes. The distribution was generated via Coronapp. The genomes were downloaded from the GISAID database on September 24, 2021 for mutation profiling via Coronapp [20,21]. **(B)** Phylogenetic analysis of the 182 V1176-encoding δ genomes identified by June 15, 2021. Only those with high sequence coverage and complete collection date information were used for phylogenetic analysis via the package RAxML-NG, to generate 20 maximum likelihood trees and one bestTree for presentation via FigTree. The strain names and GISAID accession numbers of the genomes are provided in Figure S1. **(E)** Phylogenetic analysis of the 97 V1176F- and V1264L-encoding δ-genomes. The analysis was done as in (D). The strain names and GISAID accession numbers of the genomes

are provided in Figure S2. The genomes used in panels D-E were downloaded from the GISAID database on September 24, 2021

Figure 4. Mutation profile of 145 V1176F- and Q836E-encoding δ genomes identified and sequenced in different countries around the world. The genomes were downloaded from the GISAID SARS-COV-2 genome sequence database on September 24, 2021 for mutation profiling via Coronapp. Shown in (A), (B) and (C) are substitutions in NSP3, spike and nucleocapsid proteins, respectively.

Figure 5. Phylogenetic analysis of the 143 Q836E-encoding δ genomes. The genomes are the same those used in Fig. 4 except two low-coverage genomes were excluded. The phylogenetic analysis was carried out as in Fig. 3D. The strain names and GISAID accession numbers of the genomes are provided in Figure S3.

Figure 6. Mechanistic impact of spike substitutions and an evolutionary model. (**A**) Sequence comparison of the S2' cleavage site of SARS-COV-2 spike protein with the corresponding regions of spike proteins from five other beta coronaviruses, SARS-COV-1, HKU1, NL63, 229E and OC43. The P809S substitution and the S2' cleavage site are indicated above the SARS-COV-2 spike protein sequence. (**B-C**) Structural details showing spike P809, Q836E and their neighboring residues. All of them are on the surface containing the S1/S2 and S2' cleavage sites. P809 is close the S2' cleavage site between R815 and S816 (B). This site is only about 40 Å away from T676, which is N-terminal to the unstructured S1/S2 cleave loop. The side chain of Q836 is 3.5 Å away from D839 and may thus form a hydrogen bond (B). A873 is proximity to K786 (~5.1 Å), and both residues are in the same surface as the S2' cleavage loop (C). Adapted from PyMol presentation of the spike protein-ACE2 complex structure 6XR8 from the PDB database. (**D**) Hypothetical model explaining how different substitutions are acquired during generation of different variants and subvariants. There are four δ subvariants (δ1, δ2, δ3 and δ4) initially identified in India. Through convergent evolution, they have then acquired V1176F and yielded four δ subvariants (δ1F, δ2F, δ3F and δ4F). Alarmingly, one δ1F genome recently identified in Czech Republic encodes N501Y, a key spike substitution shared by α, β and γ variants of concern. This continuous branching model is similar to what has been presented in several related studies [6,7,16,22,28].

SUPPLEMENTAL INFORMATION

This section includes three supplementary figures with detailed information on three phylogenetic trees and one acknowledgement table for the GISAID genomes used in this work.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Phylogenetic analysis of the 182 V1176-encoding δ genomes identified prior to June 15, 2021. Only those with high sequence coverage and complete collection date information were used for phylogenetic analysis via the package RAxML-NG, to generate 20 maximum likelihood trees and one bestTree for presentation via Figtree. The strain names and GISAID accession numbers of the genomes are detailed here, with an abbreviated version of three shown in Fig. 3D. Note that during RAxML-NG analysis, identical genomes were automatically removed, so not all 182 are shown on the tree presented here.

Figure S2. Phylogenetic analysis of the 97 V1176F- and V1264L-encoding δ-genomes. The analysis was done as in Fig. 3E, with the strain names and GISAID accession numbers of the genomes detailed here, with an abbreviated version of three shown in Fig. 3E. Note that during RAxML-NG analysis, identical genomes were automatically removed, so not all 182 are shown on the tree presented here.

Figure S3. Phylogenetic analysis of the 143 Q836E-encoding δ-genomes. The genomes are the same those used in Fig. 5. The phylogenetic analysis package Raxml-ng was used to generate 20 maximum likelihood trees and the bestTree for presentation via Figtree. The strain names and GISAID accession numbers of the genomes are detailed here, with an abbreviated version of three shown in Fig. 5. Note that during RAxML-NG analysis, identical genomes were automatically removed, so not all 182 are shown on the tree presented here.

Figure 1

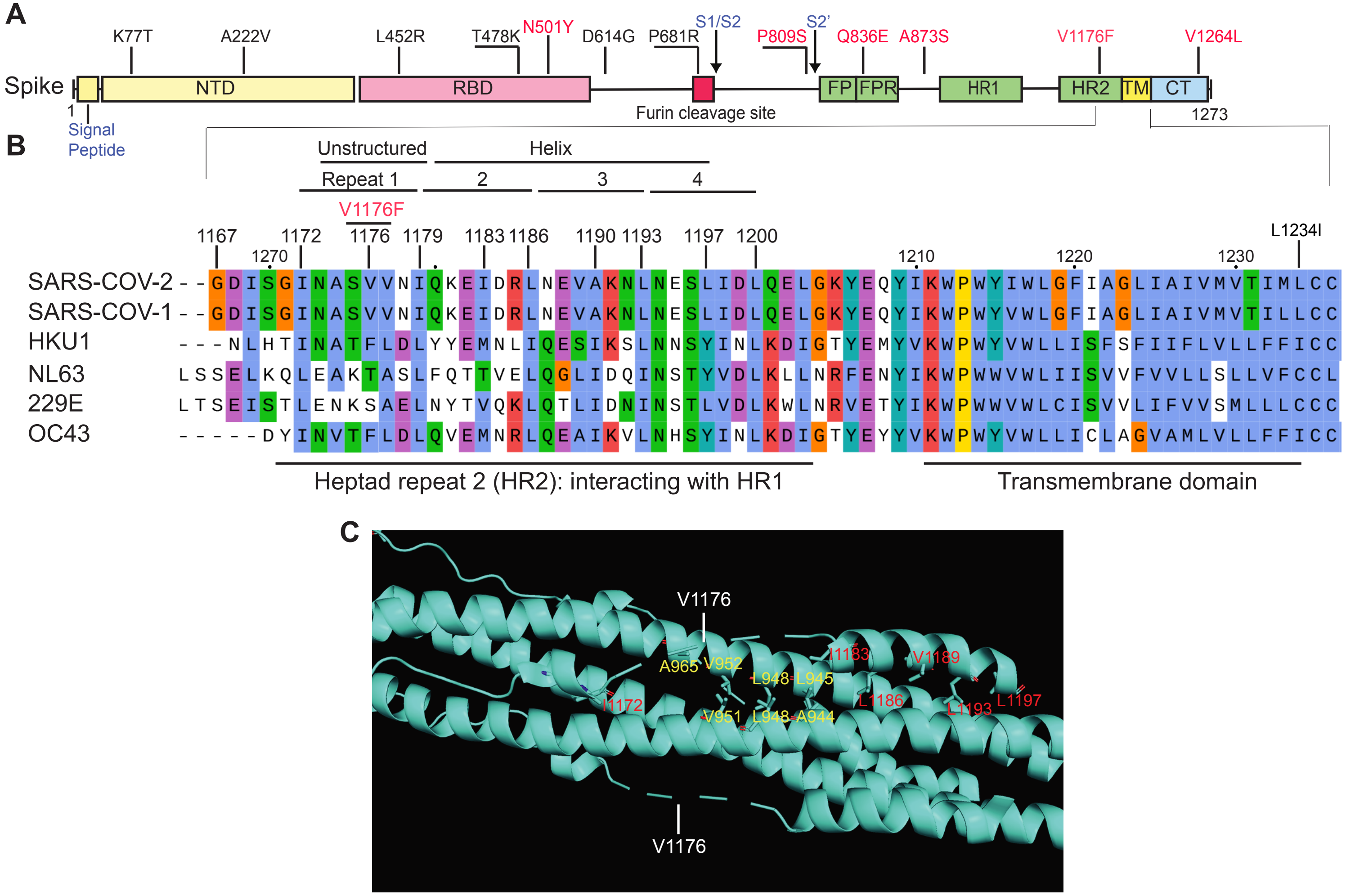

A
Spike
K77T
A222V
L452R
T478K
N501Y
D614G
P681R
S1/S2
S2'
P809S
Q836E
A873S
V1176F
V1264L
NTD
RBD
FP
FPR
HR1
HR2
TM
CT
1
Signal Peptide
Furin cleavage site
1273
B
Unstructured
Helix
Repeat 1
2
3
4
V1176F
1167
1270
1172
1176
1179
1183
1186
1190
1193
1197
1200
1210
1220
1230
L1234I
SARS-COV-2 --GDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCC
SARS-COV-1 --GDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCC
HKU1 ---NLHTINATFLDLYYEMNLIQESIKSLNNSYINLKDIGTYEMYVKWPWYVWLLISFSFIIFLVLLFFICC
NL63 LSSELKQLEAKTASLFQTTVELQGLIDQINSTYVDLKLLNRFENYIKWPWWVWLIISVVFVVLLSLLVFCCL
229E LTSEISTLENKSAELNYTVQKLQTLIDNINSTLVDLKWLNRVETYIKWPWWVWLCISVVLIFVVSMLLLCCC
OC43 -----DYINVTFLDLQVEMNRLQEAIKVLNHSYINLKDIGTYEYYVKWPWYVWLLICLAGVAMLVLLFFICC
Heptad repeat 2 (HR2): interacting with HR1
Transmembrane domain
C
V1176
A965
V952
L948
L945
I1183
V1189
I1172
L1186
V951
L948
A944
L1193
L1197
V1176

Figure 2

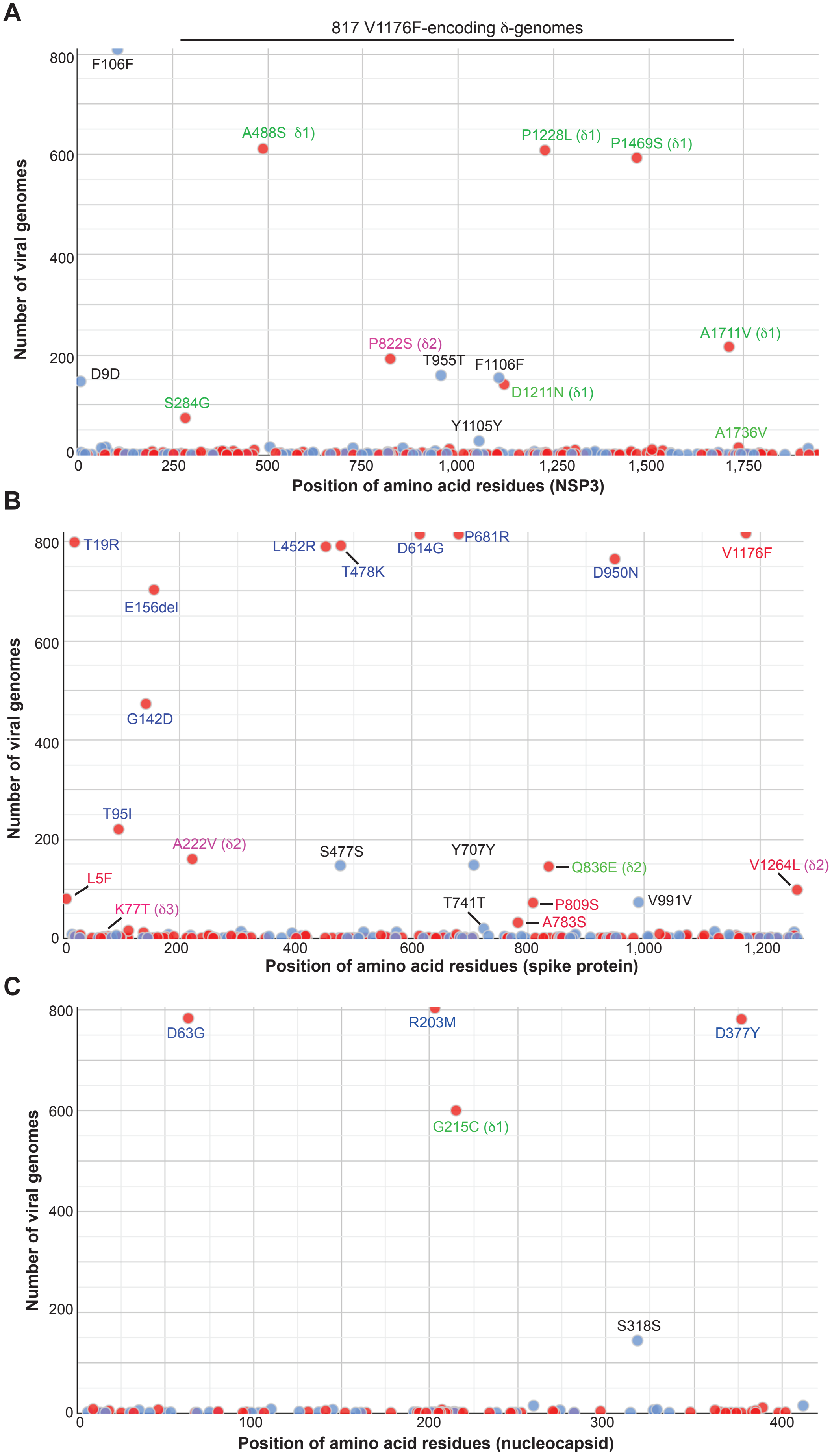
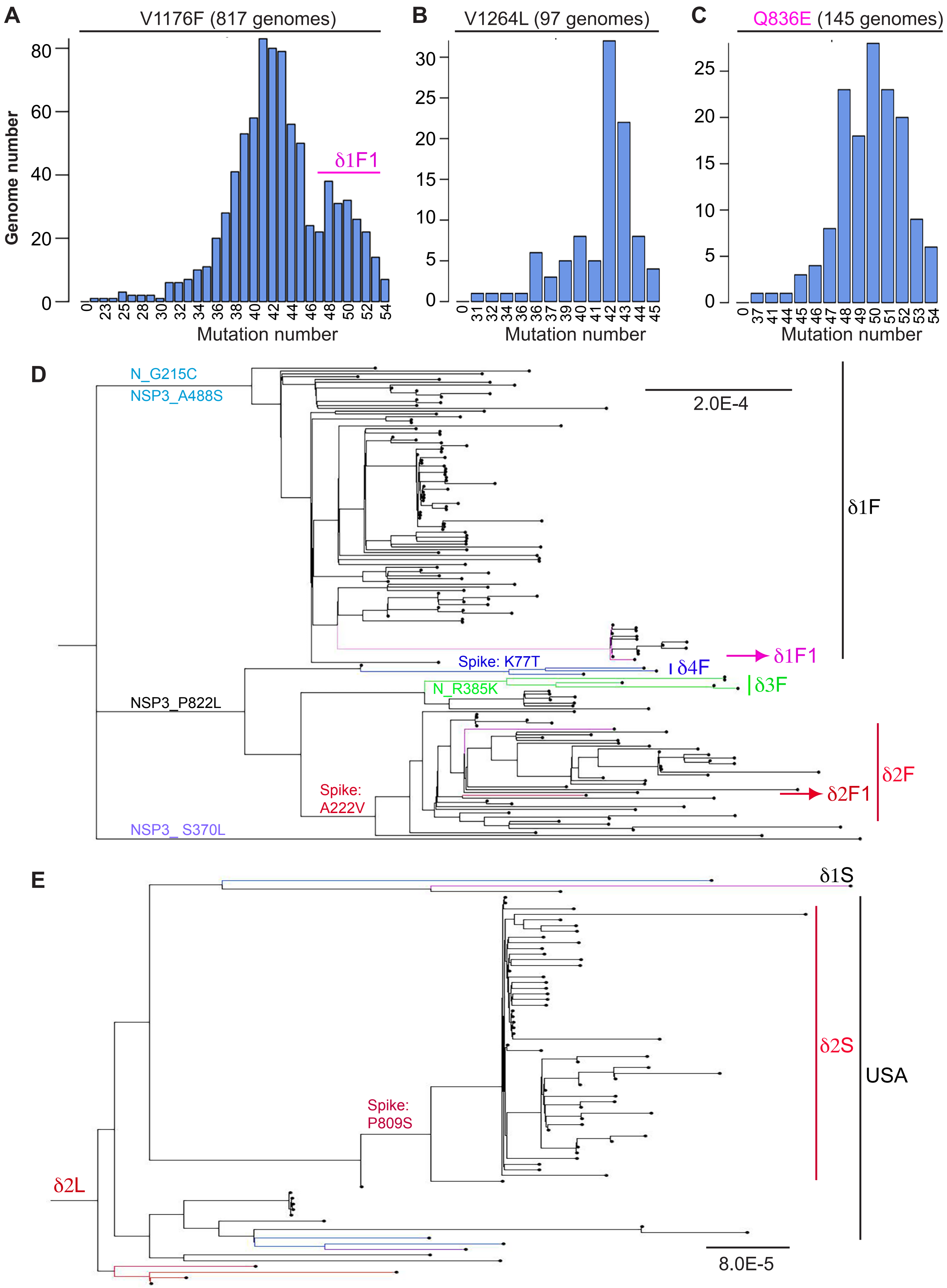
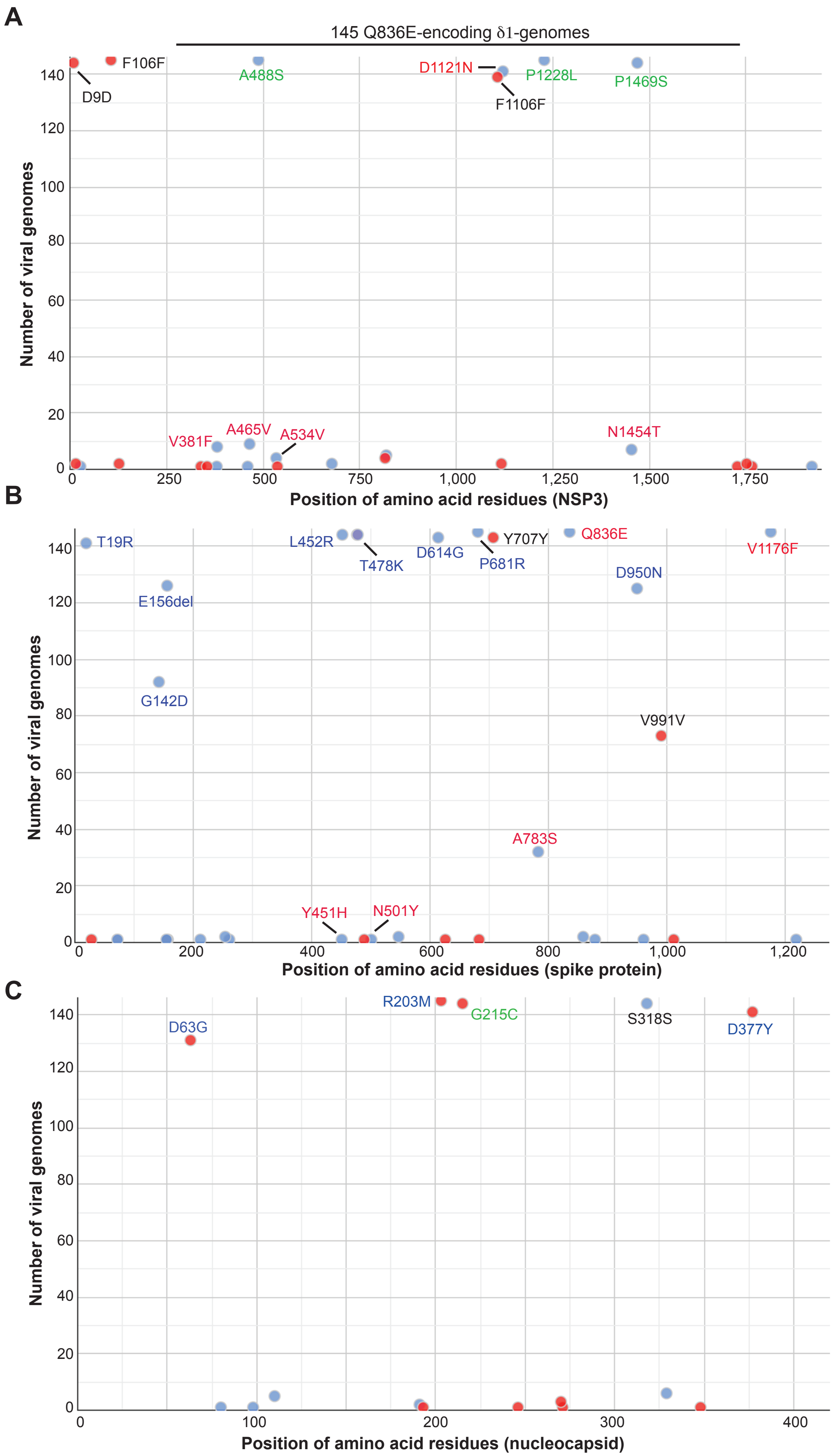

A
817 V1176F-encoding δ-genomes
Number of viral genomes
800
600
400
200
0
F106F
A488S δ1)
P1228L (δ1)
P1469S (δ1)
P822S (δ2)
T955T
F1106F
D1211N (δ1)
D9D
S284G
Y1105Y
A1711V (δ1)
A1736V
0
250
500
750
1,000
1,250
1,500
1,750
Position of amino acid residues (NSP3)
B
Number of viral genomes
800
600
400
200
0
T19R
L452R
T478K
D614G
P681R
D950N
V1176F
E156del
G142D
T95I
A222V (δ2)
S477S
Y707Y
Q836E (δ2)
V1264L (δ2)
L5F
K77T (δ3)
T741T
P809S
A783S
V991V
0
200
400
600
800
1,000
1,200
Position of amino acid residues (spike protein)
C
Number of viral genomes
800
600
400
200
0
D63G
R203M
D377Y
G215C (δ1)
S318S
0
100
200
300
400
Position of amino acid residues (nucleocapsid)

Figure 3



A
V1176F (817 genomes)
Genome number
80
60
40
20
0
δ1F1
0
23
25
28
30
32
34
36
38
40
42
44
46
48
50
52
54
Mutation number
B
V1264L (97 genomes)
30
25
20
15
10
5
0
0
31
32
34
36
36
37
39
40
41
42
43
44
45
Mutation number
C
Q836E (145 genomes)
25
20
15
10
5
0
0
37
41
44
45
46
47
48
49
50
51
52
53
54
Mutation number
D
N_G215C
NSP3_A488S
2.0E-4
δ1F
δ1F1
Spike: K77T
δ4F
δ3F
N_R385K
NSP3_P822L
Spike:
A222V
δ2F
δ2F1
NSP3_ S370L
E
δ1S
δ2S
USA
Spike:
P809S
δ2L
8.0E-5

Figure 4


A
145 Q836E-encoding δ1-genomes
Number of viral genomes
140
120
100
80
60
40
20
0
F106F
D9D
A488S
D1121N
P1228L
F1106F
P1469S
V381F
A465V
A534V
N1454T
0
250
500
750
1,000
1,250
1,500
1,750
Position of amino acid residues (NSP3)
B
Number of viral genomes
140
120
100
80
60
40
20
0
T19R
L452R
T478K
D614G
P681R
Y707Y
Q836E
V1176F
D950N
E156del
G142D
V991V
A783S
Y451H
N501Y
0
200
400
600
800
1,000
1,200
Position of amino acid residues (spike protein)
C
Number of viral genomes
140
120
100
80
60
40
20
0
D63G
R203M
G215C
S318S
D377Y
0
100
200
300
400
Position of amino acid residues (nucleocapsid)

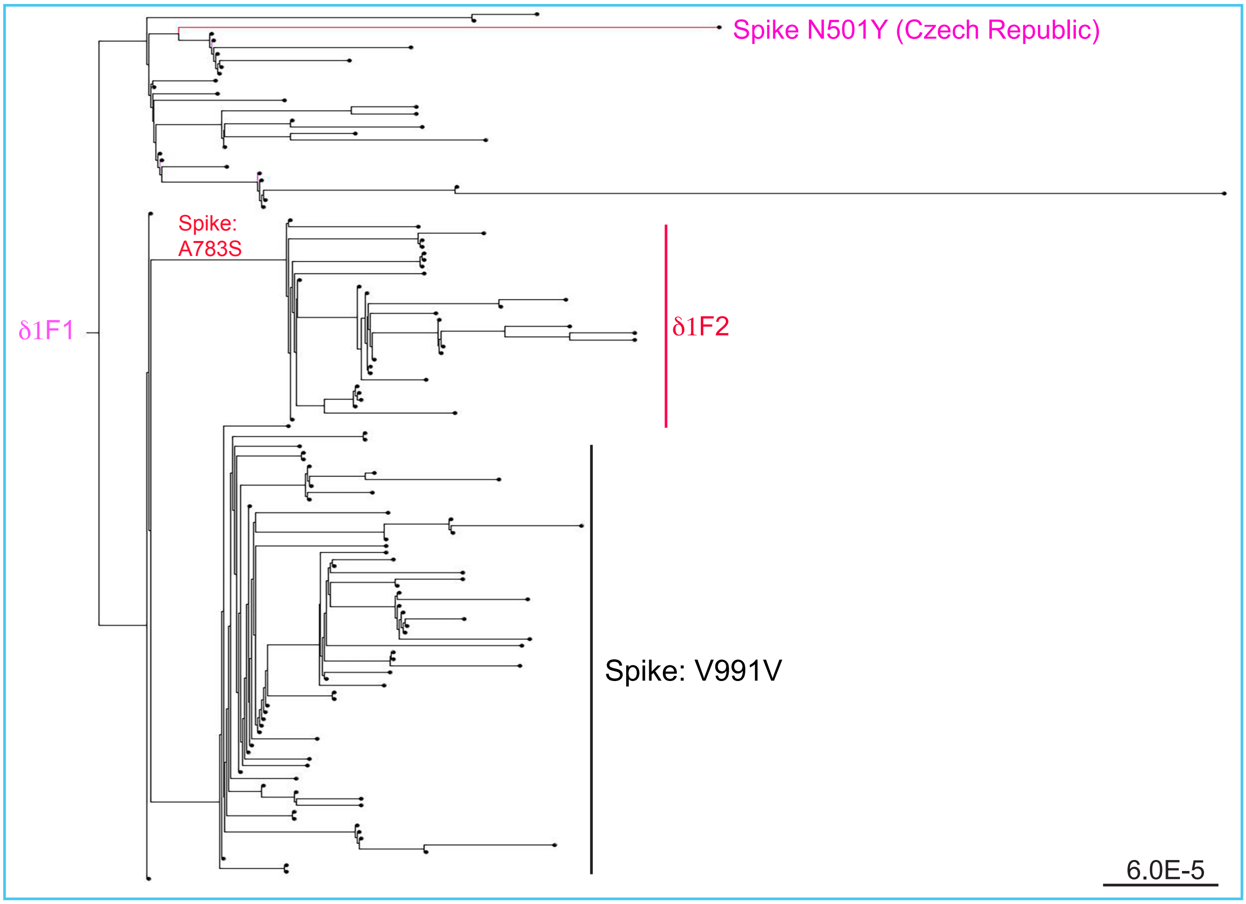

Spike N501Y (Czech Republic)
Spike:
A783S
δ1F1
δ1F2
Spike: V991V
6.0E-5

Figure 6

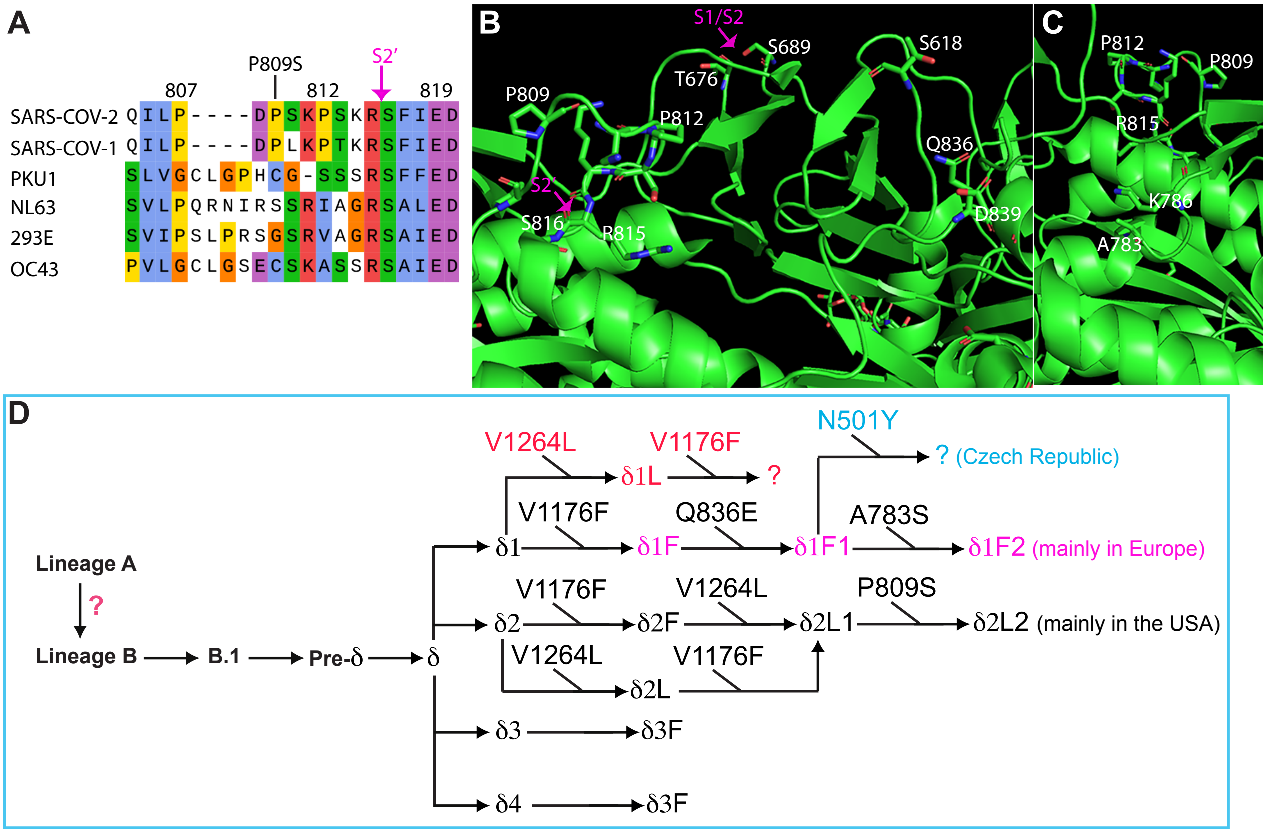

A
P809S
S2′
807
812
819
SARS-COV-2 Q I L P - - - - D P S K P S K R S F I E D
SARS-COV-1 Q I L P - - - - D P L K P T K R S F I E D
PKU1 S L V G C L G P H C G - S S S R S F F E D
NL63 S V L P Q R N I R S S R I A G R S A L E D
293E S V I P S L P R S G S R V A G R S A I E D
OC43 P V L G C L G S E C S K A S S R S A I E D
B
S1/S2
S689
S618
T676
P809
P812
Q836
S2′
S816
R815
D839
C
P812
P809
R815
K786
A783
D
Lineage A
?
Lineage B
B.1
Pre-δ
δ
δ1
V1264L
δ1L
V1176F
?
V1176F
δ1F
Q836E
δ1F1
N501Y
? (Czech Republic)
A783S
δ1F2 (mainly in Europe)
δ2
V1176F
δ2F
V1264L
δ2L1
P809S
δ2L2 (mainly in the USA)
V1264L
δ2L
V1176F
δ3
δ3F
δ4
δ3F

Figure S1

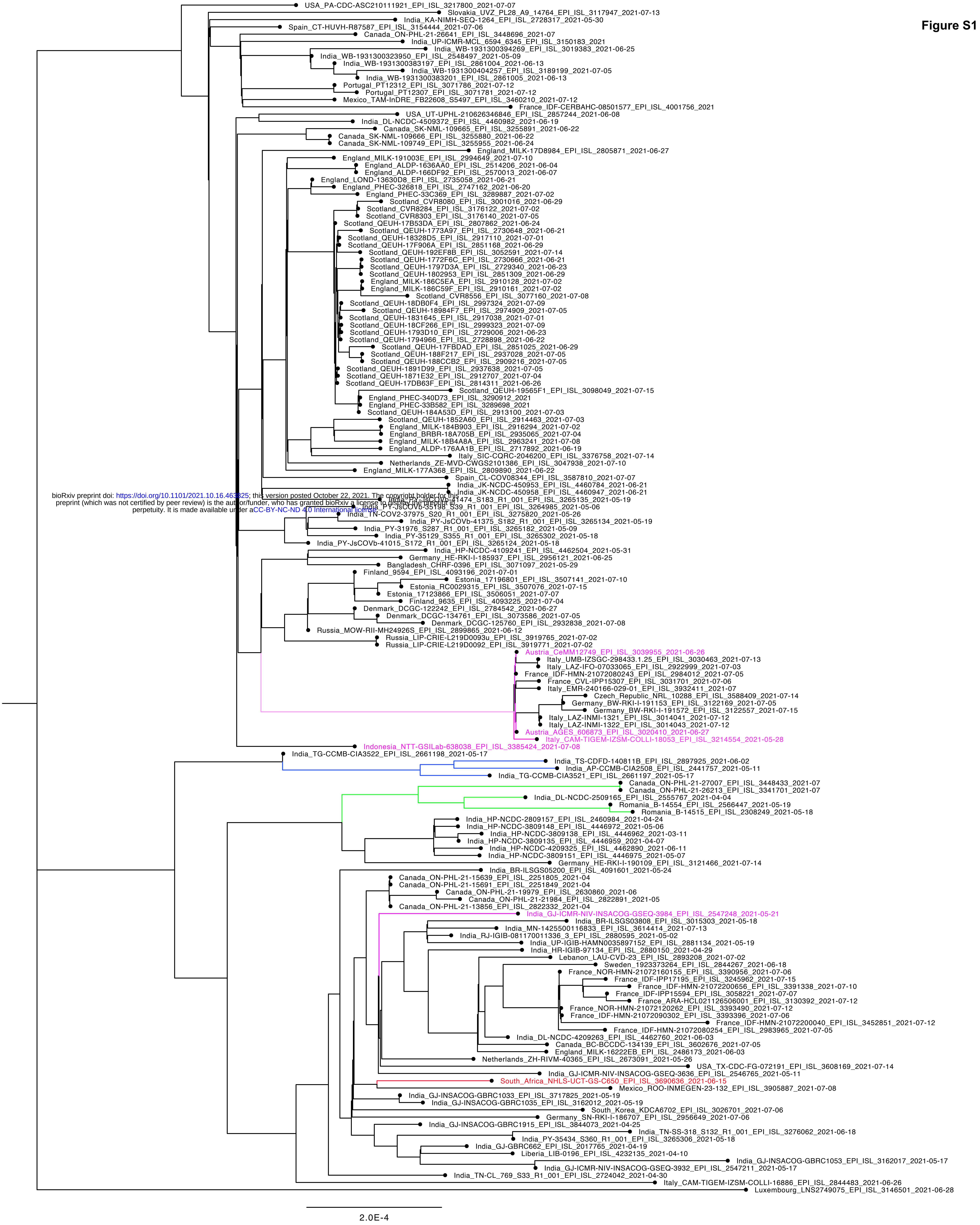

USA_PA-CDC-ASC210111921_EPI_ISL_3217800_2021-07-07
Slovakia_UVZ_PL28_A9_14764_EPI_ISL_3117947_2021-07-13
India_KA-NIMH-SEQ-1264_EPI_ISL_2728317_2021-05-30
Spain_CT-HUVH-R87587_EPI_ISL_3154444_2021-07-06
Canada_ON-PHL-21-26641_EPI_ISL_3448696_2021-07
India_UP-ICMR-MCL_6594_6345_EPI_ISL_3150183_2021
India_WB-1931300394269_EPI_ISL_3019383_2021-06-25
India_WB-1931300323950_EPI_ISL_2548497_2021-05-09
India_WB-1931300383197_EPI_ISL_2861004_2021-06-13
India_WB-1931300404257_EPI_ISL_3189199_2021-07-05
India_WB-1931300383201_EPI_ISL_2861005_2021-06-13
Portugal_PT12312_EPI_ISL_3071786_2021-07-12
Portugal_PT12307_EPI_ISL_3071781_2021-07-12
Mexico_TAM-InDRE_FB22608_S5497_EPI_ISL_3460210_2021-07-12
France_IDF-CERBAHC-08501577_EPI_ISL_4001756_2021
USA_UT-UPHL-210626346846_EPI_ISL_2857244_2021-06-08
India_DL-NCDC-4509372_EPI_ISL_4460982_2021-06-19
Canada_SK-NML-109665_EPI_ISL_3255891_2021-06-22
Canada_SK-NML-109666_EPI_ISL_3255880_2021-06-22
Canada_SK-NML-109749_EPI_ISL_3255955_2021-06-24
England_MILK-17D8984_EPI_ISL_2805871_2021-06-27
England_MILK-191003E_EPI_ISL_2994649_2021-07-10
England_ALDP-1636AA0_EPI_ISL_2514206_2021-06-04
England_ALDP-166DF92_EPI_ISL_2570013_2021-06-07
England_LOND-13630D8_EPI_ISL_2735058_2021-06-21
England_PHEC-326818_EPI_ISL_2747162_2021-06-20
England_PHEC-33C369_EPI_ISL_3289887_2021-07-02
Scotland_CVR8080_EPI_ISL_3001016_2021-06-29
Scotland_CVR8284_EPI_ISL_3176122_2021-07-02
Scotland_CVR8303_EPI_ISL_3176140_2021-07-05
Scotland_QEUH-17B53DA_EPI_ISL_2807862_2021-06-24
Scotland_QEUH-1773A97_EPI_ISL_2730648_2021-06-21
Scotland_QEUH-18328D5_EPI_ISL_2917110_2021-07-01
Scotland_QEUH-17F906A_EPI_ISL_2851168_2021-06-29
Scotland_QEUH-192EF8B_EPI_ISL_3052591_2021-07-14
Scotland_QEUH-1772F6C_EPI_ISL_2730666_2021-06-21
Scotland_QEUH-1797D3A_EPI_ISL_2729340_2021-06-23
Scotland_QEUH-1802953_EPI_ISL_2851309_2021-06-29
England_MILK-186C5EA_EPI_ISL_2910128_2021-07-02
England_MILK-186C59F_EPI_ISL_2910161_2021-07-02
Scotland_CVR8556_EPI_ISL_3077160_2021-07-08
Scotland_QEUH-18DB0F4_EPI_ISL_2997324_2021-07-09
Scotland_QEUH-18984F7_EPI_ISL_2974909_2021-07-05
Scotland_QEUH-1831645_EPI_ISL_2917038_2021-07-01
Scotland_QEUH-18CF266_EPI_ISL_2999323_2021-07-09
Scotland_QEUH-1793D10_EPI_ISL_2729006_2021-06-23
Scotland_QEUH-1794966_EPI_ISL_2728898_2021-06-22
Scotland_QEUH-17FBDAD_EPI_ISL_2851025_2021-06-29
Scotland_QEUH-188F217_EPI_ISL_2937028_2021-07-05
Scotland_QEUH-188CCB2_EPI_ISL_2909216_2021-07-05
Scotland_QEUH-1891D99_EPI_ISL_2937638_2021-07-05
Scotland_QEUH-1871E32_EPI_ISL_2912707_2021-07-04
Scotland_QEUH-17DB63F_EPI_ISL_2814311_2021-06-26
Scotland_QEUH-19565F1_EPI_ISL_3098049_2021-07-15
England_PHEC-340D73_EPI_ISL_3290912_2021
England_PHEC-33B582_EPI_ISL_3289698_2021
Scotland_QEUH-184A53D_EPI_ISL_2913100_2021-07-03
Scotland_QEUH-1852A60_EPI_ISL_2914463_2021-07-03
England_MILK-184B903_EPI_ISL_2916294_2021-07-02
England_BRBR-18A705B_EPI_ISL_2935065_2021-07-04
England_MILK-18B4A8A_EPI_ISL_2963241_2021-07-08
England_ALDP-176AA1B_EPI_ISL_2717892_2021-06-19
Italy_SIC-CQRC-2046200_EPI_ISL_3376758_2021-07-14
Netherlands_ZE-MVD-CWGS2101386_EPI_ISL_3047938_2021-07-10
England_MILK-177A368_EPI_ISL_2809890_2021-06-22
Spain_CL-COV08344_EPI_ISL_3587810_2021-07-07
India_JK-NCDC-450953_EPI_ISL_4460784_2021-06-21
India_JK-NCDC-450958_EPI_ISL_4460947_2021-06-21
India_PY-JsCOVb-41474_S183_R1_001_EPI_ISL_3265135_2021-05-19
India_PY-JsCOVb-35198_S39_R1_001_EPI_ISL_3264985_2021-05-06
India_TN-COV2-37975_S20_R1_001_EPI_ISL_3275820_2021-05-26
India_PY-JsCOVb-41375_S182_R1_001_EPI_ISL_3265134_2021-05-19
India_PY-31976_S287_R1_001_EPI_ISL_3265182_2021-05-09
India_PY-35129_S355_R1_001_EPI_ISL_3265302_2021-05-18
India_PY-JsCOVb-41015_S172_R1_001_EPI_ISL_3265124_2021-05-18
India_HP-NCDC-4109241_EPI_ISL_4462504_2021-05-31
Germany_HE-RKI-I-185937_EPI_ISL_2956121_2021-06-25
Bangladesh_CHRF-0396_EPI_ISL_3071097_2021-05-29
Finland_9594_EPI_ISL_4093196_2021-07-01
Estonia_17196801_EPI_ISL_3507141_2021-07-10
Estonia_RC0029315_EPI_ISL_3507076_2021-07-15
Estonia_17123866_EPI_ISL_3506051_2021-07-07
Finland_9635_EPI_ISL_4093225_2021-07-04
Denmark_DCGC-122242_EPI_ISL_2784542_2021-06-27
Denmark_DCGC-134761_EPI_ISL_3073586_2021-07-05
Denmark_DCGC-125760_EPI_ISL_2932838_2021-07-08
Russia_MOW-RII-MH24926S_EPI_ISL_2899865_2021-06-12
Russia_LIP-CRIE-L219D0093u_EPI_ISL_3919765_2021-07-02
Russia_LIP-CRIE-L219D0092_EPI_ISL_3919771_2021-07-02
Austria_CeMM12749_EPI_ISL_3039955_2021-06-26
Italy_UMB-IZSGC-298433.1.25_EPI_ISL_3030463_2021-07-13
Italy_LAZ-IFO-07033065_EPI_ISL_2922999_2021-07-03
France_IDF-HMN-21072080243_EPI_ISL_2984012_2021-07-05
France_CVL-IPP15307_EPI_ISL_3031701_2021-07-06
Italy_EMR-240166-029-01_EPI_ISL_3932411_2021-07
Czech_Republic_NRL_10288_EPI_ISL_3588409_2021-07-14
Germany_BW-RKI-I-191153_EPI_ISL_3122169_2021-07-05
Germany_BW-RKI-I-191572_EPI_ISL_3122557_2021-07-15
Italy_LAZ-INMI-1321_EPI_ISL_3014041_2021-07-12
Italy_LAZ-INMI-1322_EPI_ISL_3014043_2021-07-12
Austria_AGES_606873_EPI_ISL_3020410_2021-06-27
Italy_CAM-TIGEM-IZSM-COLLI-18053_EPI_ISL_3214554_2021-05-28
Indonesia_NTT-GSILab-638038_EPI_ISL_3385424_2021-07-08
India_TG-CCMB-CIA3522_EPI_ISL_2661198_2021-05-17
India_TS-CDFD-140811B_EPI_ISL_2897925_2021-06-02
India_AP-CCMB-CIA2508_EPI_ISL_2441757_2021-05-11
India_TG-CCMB-CIA3521_EPI_ISL_2661197_2021-05-17
Canada_ON-PHL-21-27007_EPI_ISL_3448433_2021-07
Canada_ON-PHL-21-26213_EPI_ISL_3341701_2021-07
India_DL-NCDC-2509165_EPI_ISL_2555767_2021-04-04
Romania_B-14554_EPI_ISL_2566447_2021-05-19
Romania_B-14515_EPI_ISL_2308249_2021-05-18
India_HP-NCDC-2809157_EPI_ISL_2460984_2021-04-24
India_HP-NCDC-3809148_EPI_ISL_4446972_2021-05-06
India_HP-NCDC-3809138_EPI_ISL_4446962_2021-03-11
India_HP-NCDC-3809135_EPI_ISL_4446959_2021-04-07
India_HP-NCDC-4209325_EPI_ISL_4462890_2021-06-11
India_HP-NCDC-3809151_EPI_ISL_4446975_2021-05-07
Germany_HE-RKI-I-190109_EPI_ISL_3121466_2021-07-14
India_BR-ILSGS05200_EPI_ISL_4091601_2021-05-24
Canada_ON-PHL-21-15639_EPI_ISL_2251805_2021-04
Canada_ON-PHL-21-15691_EPI_ISL_2251849_2021-04
Canada_ON-PHL-21-19979_EPI_ISL_2630860_2021-06
Canada_ON-PHL-21-21984_EPI_ISL_2822891_2021-05
Canada_ON-PHL-21-13856_EPI_ISL_2822332_2021-04
India_GJ-ICMR-NIV-INSACOG-GSEQ-3984_EPI_ISL_2547248_2021-05-21
India_BR-ILSGS03808_EPI_ISL_3015303_2021-05-18
India_MN-1425500116833_EPI_ISL_3614414_2021-07-13
India_RJ-IGIB-081170011336_3_EPI_ISL_2880595_2021-05-02
India_UP-IGIB-HAMN0035897152_EPI_ISL_2881134_2021-05-19
India_HR-IGIB-97134_EPI_ISL_2880150_2021-04-29
Lebanon_LAU-CVD-23_EPI_ISL_2893208_2021-07-02
Sweden_1923373264_EPI_ISL_2844267_2021-06-18
France_NOR-HMN-21072160155_EPI_ISL_3390956_2021-07-06
France_IDF-IPP17195_EPI_ISL_3245962_2021-07-15
France_IDF-HMN-21072200656_EPI_ISL_3391338_2021-07-10
France_IDF-IPP15594_EPI_ISL_3058221_2021-07-07
France_ARA-HCL021126506001_EPI_ISL_3130392_2021-07-12
France_NOR-HMN-21072120262_EPI_ISL_3393490_2021-07-12
France_IDF-HMN-21072090302_EPI_ISL_3393396_2021-07-06
France_IDF-HMN-21072200040_EPI_ISL_3452851_2021-07-12
France_IDF-HMN-21072080254_EPI_ISL_2983965_2021-07-05
India_DL-NCDC-4209263_EPI_ISL_4462760_2021-06-03
Canada_BC-BCCDC-134139_EPI_ISL_3602676_2021-07-05
England_MILK-16222EB_EPI_ISL_2486173_2021-06-03
Netherlands_ZH-RIVM-40365_EPI_ISL_2673091_2021-05-26
USA_TX-CDC-FG-072191_EPI_ISL_3608169_2021-07-14
India_GJ-ICMR-NIV-INSACOG-GSEQ-3636_EPI_ISL_2546765_2021-05-11
South_Africa_NHLS-UCT-GS-C650_EPI_ISL_3690636_2021-06-15
Mexico_ROO-INMEGEN-23-132_EPI_ISL_3905887_2021-07-08
India_GJ-INSACOG-GBRC1033_EPI_ISL_3717825_2021-05-19
India_GJ-INSACOG-GBRC1035_EPI_ISL_3162012_2021-05-19
South_Korea_KDCA6702_EPI_ISL_3026701_2021-07-06
Germany_SN-RKI-I-186707_EPI_ISL_2956649_2021-07-06
India_GJ-INSACOG-GBRC1915_EPI_ISL_3844073_2021-04-25
India_TN-SS-318_S132_R1_001_EPI_ISL_3276062_2021-06-18
India_PY-35434_S360_R1_001_EPI_ISL_3265306_2021-05-18
India_GJ-GBRC662_EPI_ISL_2017765_2021-04-19
Liberia_LIB-0196_EPI_ISL_4232135_2021-04-10
India_GJ-INSACOG-GBRC1053_EPI_ISL_3162017_2021-05-17
India_GJ-ICMR-NIV-INSACOG-GSEQ-3932_EPI_ISL_2547211_2021-05-17
India_TN-CL_769_S33_R1_001_EPI_ISL_2724042_2021-04-30
Italy_CAM-TIGEM-IZSM-COLLI-16886_EPI_ISL_2844483_2021-06-26
Luxembourg_LNS2749075_EPI_ISL_3146501_2021-06-28
2.0E-4

Figure S2

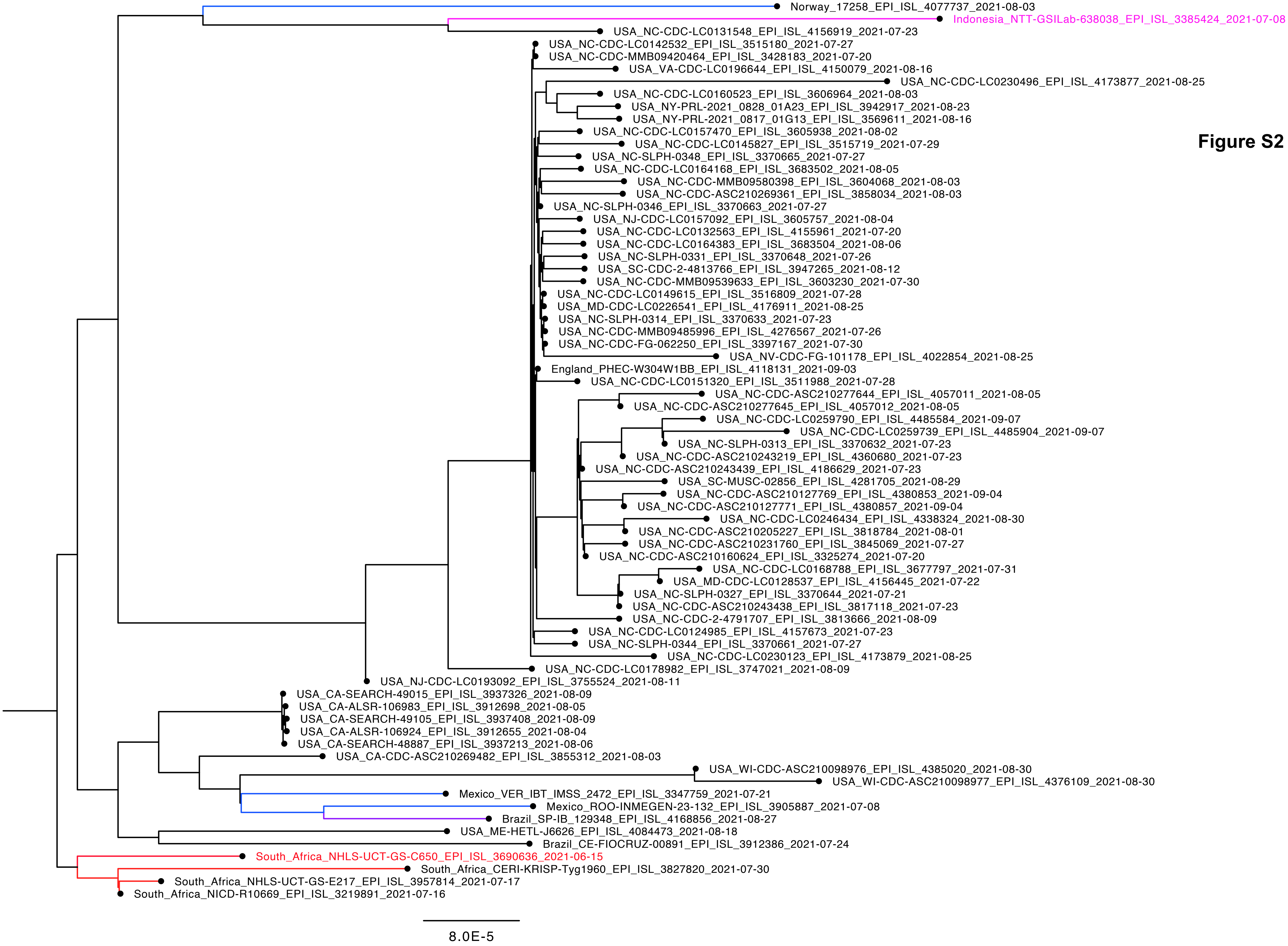
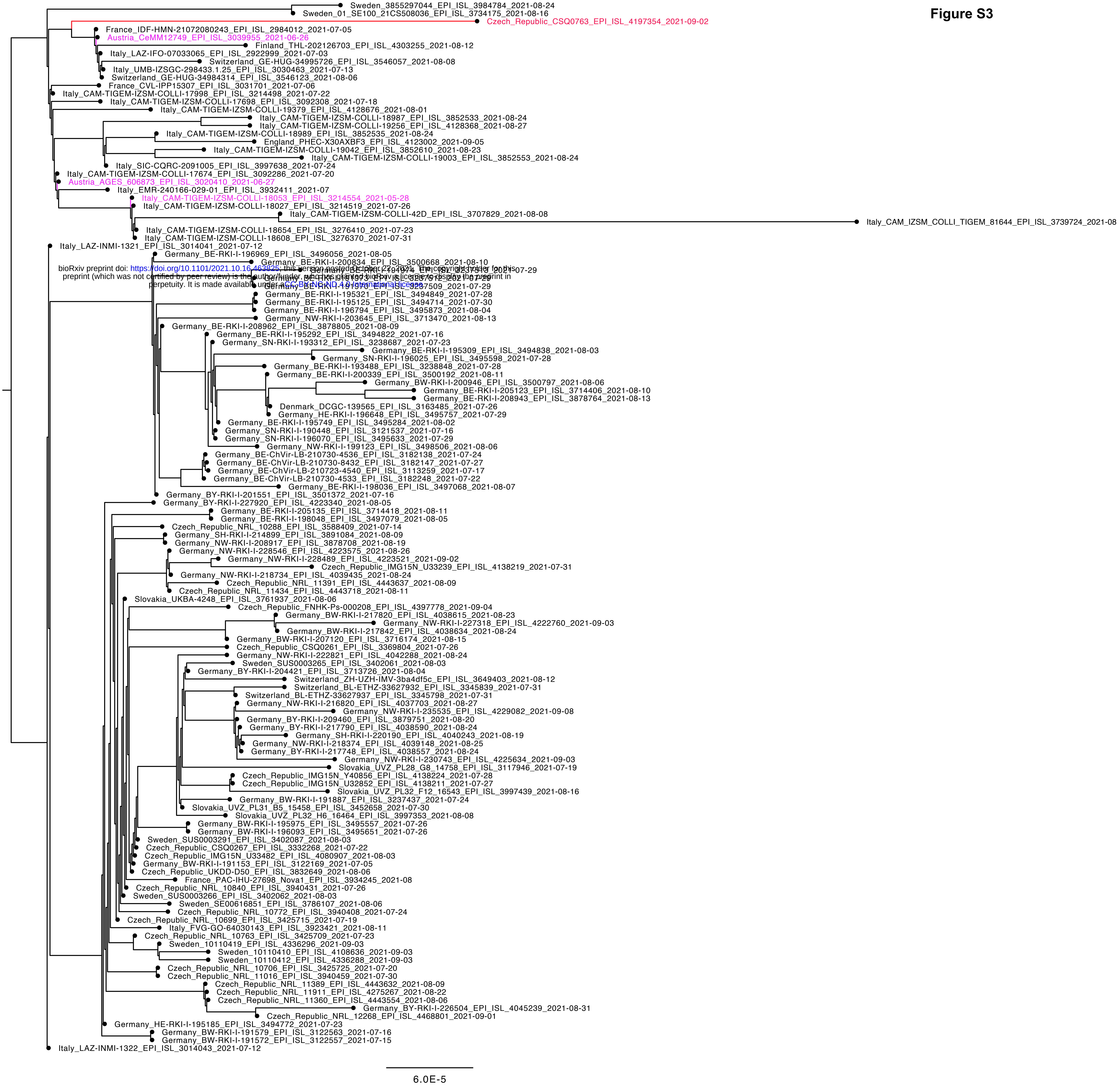

Norway_17258_EPI_ISL_4077737_2021-08-03
Indonesia_NTT-GSILab-638038_EPI_ISL_3385424_2021-07-08
USA_NC-CDC-LC0131548_EPI_ISL_4156919_2021-07-23
USA_NC-CDC-LC0142532_EPI_ISL_3515180_2021-07-27
USA_NC-CDC-MMB09420464_EPI_ISL_3428183_2021-07-20
USA_VA-CDC-LC0196644_EPI_ISL_4150079_2021-08-16
USA_NC-CDC-LC0230496_EPI_ISL_4173877_2021-08-25
USA_NC-CDC-LC0160523_EPI_ISL_3606964_2021-08-03
USA_NY-PRL-2021_0828_01A23_EPI_ISL_3942917_2021-08-23
USA_NY-PRL-2021_0817_01G13_EPI_ISL_3569611_2021-08-16
USA_NC-CDC-LC0157470_EPI_ISL_3605938_2021-08-02
USA_NC-CDC-LC0145827_EPI_ISL_3515719_2021-07-29
USA_NC-SLPH-0348_EPI_ISL_3370665_2021-07-27
USA_NC-CDC-LC0164168_EPI_ISL_3683502_2021-08-05
USA_NC-CDC-MMB09580398_EPI_ISL_3604068_2021-08-03
USA_NC-CDC-ASC210269361_EPI_ISL_3858034_2021-08-03
USA_NC-SLPH-0346_EPI_ISL_3370663_2021-07-27
USA_NJ-CDC-LC0157092_EPI_ISL_3605757_2021-08-04
USA_NC-CDC-LC0132563_EPI_ISL_4155961_2021-07-20
USA_NC-CDC-LC0164383_EPI_ISL_3683504_2021-08-06
USA_NC-SLPH-0331_EPI_ISL_3370648_2021-07-26
USA_SC-CDC-2-4813766_EPI_ISL_3947265_2021-08-12
USA_NC-CDC-MMB09539633_EPI_ISL_3603230_2021-07-30
USA_NC-CDC-LC0149615_EPI_ISL_3516809_2021-07-28
USA_MD-CDC-LC0226541_EPI_ISL_4176911_2021-08-25
USA_NC-SLPH-0314_EPI_ISL_3370633_2021-07-23
USA_NC-CDC-MMB09485996_EPI_ISL_4276567_2021-07-26
USA_NC-CDC-FG-062250_EPI_ISL_3397167_2021-07-30
USA_NV-CDC-FG-101178_EPI_ISL_4022854_2021-08-25
England_PHEC-W304W1BB_EPI_ISL_4118131_2021-09-03
USA_NC-CDC-LC0151320_EPI_ISL_3511988_2021-07-28
USA_NC-CDC-ASC210277644_EPI_ISL_4057011_2021-08-05
USA_NC-CDC-ASC210277645_EPI_ISL_4057012_2021-08-05
USA_NC-CDC-LC0259790_EPI_ISL_4485584_2021-09-07
USA_NC-CDC-LC0259739_EPI_ISL_4485904_2021-09-07
USA_NC-SLPH-0313_EPI_ISL_3370632_2021-07-23
USA_NC-CDC-ASC210243219_EPI_ISL_4360680_2021-07-23
USA_NC-CDC-ASC210243439_EPI_ISL_4186629_2021-07-23
USA_SC-MUSC-02856_EPI_ISL_4281705_2021-08-29
USA_NC-CDC-ASC210127769_EPI_ISL_4380853_2021-09-04
USA_NC-CDC-ASC210127771_EPI_ISL_4380857_2021-09-04
USA_NC-CDC-LC0246434_EPI_ISL_4338324_2021-08-30
USA_NC-CDC-ASC210205227_EPI_ISL_3818784_2021-08-01
USA_NC-CDC-ASC210231760_EPI_ISL_3845069_2021-07-27
USA_NC-CDC-ASC210160624_EPI_ISL_3325274_2021-07-20
USA_NC-CDC-LC0168788_EPI_ISL_3677797_2021-07-31
USA_MD-CDC-LC0128537_EPI_ISL_4156445_2021-07-22
USA_NC-SLPH-0327_EPI_ISL_3370644_2021-07-21
USA_NC-CDC-ASC210243438_EPI_ISL_3817118_2021-07-23
USA_NC-CDC-2-4791707_EPI_ISL_3813666_2021-08-09
USA_NC-CDC-LC0124985_EPI_ISL_4157673_2021-07-23
USA_NC-SLPH-0344_EPI_ISL_3370661_2021-07-27
USA_NC-CDC-LC0230123_EPI_ISL_4173879_2021-08-25
USA_NC-CDC-LC0178982_EPI_ISL_3747021_2021-08-09
USA_NJ-CDC-LC0193092_EPI_ISL_3755524_2021-08-11
USA_CA-SEARCH-49015_EPI_ISL_3937326_2021-08-09
USA_CA-ALSR-106983_EPI_ISL_3912698_2021-08-05
USA_CA-SEARCH-49105_EPI_ISL_3937408_2021-08-09
USA_CA-ALSR-106924_EPI_ISL_3912655_2021-08-04
USA_CA-SEARCH-48887_EPI_ISL_3937213_2021-08-06
USA_CA-CDC-ASC210269482_EPI_ISL_3855312_2021-08-03
USA_WI-CDC-ASC210098976_EPI_ISL_4385020_2021-08-30
USA_WI-CDC-ASC210098977_EPI_ISL_4376109_2021-08-30
Mexico_VER_IBT_IMSS_2472_EPI_ISL_3347759_2021-07-21
Mexico_ROO-INMEGEN-23-132_EPI_ISL_3905887_2021-07-08
Brazil_SP-IB_129348_EPI_ISL_4168856_2021-08-27
USA_ME-HETL-J6626_EPI_ISL_4084473_2021-08-18
Brazil_CE-FIOCRUZ-00891_EPI_ISL_3912386_2021-07-24
South_Africa_NHLS-UCT-GS-C650_EPI_ISL_3690636_2021-06-15
South_Africa_CERI-KRISP-Tyg1960_EPI_ISL_3827820_2021-07-30
South_Africa_NHLS-UCT-GS-E217_EPI_ISL_3957814_2021-07-17
South_Africa_NICD-R10669_EPI_ISL_3219891_2021-07-16
8.0E-5

Figure S3


Sweden_3855297044_EPI_ISL_3984784_2021-08-24
Sweden_01_SE100_21CS508036_EPI_ISL_3734175_2021-08-16
Czech_Republic_CSQ0763_EPI_ISL_4197354_2021-09-02
France_IDF-HMN-21072080243_EPI_ISL_2984012_2021-07-05
Austria_CeMM12749_EPI_ISL_3039955_2021-06-26
Finland_THL-202126703_EPI_ISL_4303255_2021-08-12
Italy_LAZ-IFO-07033065_EPI_ISL_2922999_2021-07-03
Switzerland_GE-HUG-34995726_EPI_ISL_3546057_2021-08-08
Italy_UMB-IZSGC-298433.1.25_EPI_ISL_3030463_2021-07-13
Switzerland_GE-HUG-34984314_EPI_ISL_3546123_2021-08-06
France_CVL-IPP15307_EPI_ISL_3031701_2021-07-06
Italy_CAM-TIGEM-IZSM-COLLI-17998_EPI_ISL_3214498_2021-07-22
Italy_CAM-TIGEM-IZSM-COLLI-17698_EPI_ISL_3092308_2021-07-18
Italy_CAM-TIGEM-IZSM-COLLI-19379_EPI_ISL_4128676_2021-08-01
Italy_CAM-TIGEM-IZSM-COLLI-18987_EPI_ISL_3852533_2021-08-24
Italy_CAM-TIGEM-IZSM-COLLI-19256_EPI_ISL_4128368_2021-08-27
Italy_CAM-TIGEM-IZSM-COLLI-18989_EPI_ISL_3852535_2021-08-24
England_PHEC-X30AXBF3_EPI_ISL_4123002_2021-09-05
Italy_CAM-TIGEM-IZSM-COLLI-19042_EPI_ISL_3852610_2021-08-23
Italy_CAM-TIGEM-IZSM-COLLI-19003_EPI_ISL_3852553_2021-08-24
Italy_SIC-CQRC-2091005_EPI_ISL_3997638_2021-07-24
Italy_CAM-TIGEM-IZSM-COLLI-17674_EPI_ISL_3092288_2021-07-20
Austria_AGES_606873_EPI_ISL_3020410_2021-06-27
Italy_EMR-240166-029-01_EPI_ISL_3932411_2021-07
Italy_CAM-TIGEM-IZSM-COLLI-18053_EPI_ISL_3214554_2021-05-28
Italy_CAM-TIGEM-IZSM-COLLI-18027_EPI_ISL_3214519_2021-07-26
Italy_CAM-TIGEM-IZSM-COLLI-42D_EPI_ISL_3707829_2021-08-08
Italy_CAM_IZSM_COLLI_TIGEM_81644_EPI_ISL_3739724_2021-08
Italy_CAM-TIGEM-IZSM-COLLI-18654_EPI_ISL_3276410_2021-07-23
Italy_CAM-TIGEM-IZSM-COLLI-18608_EPI_ISL_3276370_2021-07-31
Italy_LAZ-INMI-1321_EPI_ISL_3014041_2021-07-12
Germany_BE-RKI-I-196969_EPI_ISL_3496056_2021-08-05
Germany_BE-RKI-I-200834_EPI_ISL_3500668_2021-08-10
Germany_BE-RKI-I-194974_EPI_ISL_3237518_2021-07-29
Germany_BE-RKI-I-191973_EPI_ISL_3237512_2021-07-29
Germany_BE-RKI-I-191970_EPI_ISL_3237509_2021-07-29
Germany_BE-RKI-I-195321_EPI_ISL_3494849_2021-07-28
Germany_BE-RKI-I-195125_EPI_ISL_3494714_2021-07-30
Germany_BE-RKI-I-196794_EPI_ISL_3495873_2021-08-04
Germany_NW-RKI-I-203645_EPI_ISL_3713470_2021-08-13
Germany_BE-RKI-I-208962_EPI_ISL_3878805_2021-08-09
Germany_BE-RKI-I-195292_EPI_ISL_3494822_2021-07-16
Germany_SN-RKI-I-193312_EPI_ISL_3238687_2021-07-23
Germany_BE-RKI-I-195309_EPI_ISL_3494838_2021-08-03
Germany_SN-RKI-I-196025_EPI_ISL_3495598_2021-07-28
Germany_BE-RKI-I-193488_EPI_ISL_3238848_2021-07-28
Germany_BE-RKI-I-200339_EPI_ISL_3500192_2021-08-11
Germany_BW-RKI-I-200946_EPI_ISL_3500797_2021-08-06
Germany_BE-RKI-I-205123_EPI_ISL_3714406_2021-08-10
Germany_BE-RKI-I-208943_EPI_ISL_3878764_2021-08-13
Denmark_DCGC-139565_EPI_ISL_3163485_2021-07-26
Germany_HE-RKI-I-196648_EPI_ISL_3495757_2021-07-29
Germany_BE-RKI-I-195749_EPI_ISL_3495284_2021-08-02
Germany_SN-RKI-I-190448_EPI_ISL_3121537_2021-07-16
Germany_SN-RKI-I-196070_EPI_ISL_3495633_2021-07-29
Germany_NW-RKI-I-199123_EPI_ISL_3498506_2021-08-06
Germany_BE-ChVir-LB-210730-4536_EPI_ISL_3182138_2021-07-24
Germany_BE-ChVir-LB-210730-8432_EPI_ISL_3182147_2021-07-27
Germany_BE-ChVir-LB-210723-4540_EPI_ISL_3113259_2021-07-17
Germany_BE-ChVir-LB-210730-4533_EPI_ISL_3182248_2021-07-22
Germany_BE-RKI-I-198036_EPI_ISL_3497068_2021-08-07
Germany_BY-RKI-I-201551_EPI_ISL_3501372_2021-07-16
Germany_BY-RKI-I-227920_EPI_ISL_4223340_2021-08-05
Germany_BE-RKI-I-205135_EPI_ISL_3714418_2021-08-11
Germany_BE-RKI-I-198048_EPI_ISL_3497079_2021-08-05
Czech_Republic_NRL_10288_EPI_ISL_3588409_2021-07-14
Germany_SH-RKI-I-214899_EPI_ISL_3891084_2021-08-09
Germany_NW-RKI-I-208917_EPI_ISL_3878708_2021-08-19
Germany_NW-RKI-I-228546_EPI_ISL_4223575_2021-08-26
Germany_NW-RKI-I-228489_EPI_ISL_4223521_2021-09-02
Czech_Republic_IMG15N_U33239_EPI_ISL_4138219_2021-07-31
Germany_NW-RKI-I-218734_EPI_ISL_4039435_2021-08-24
Czech_Republic_NRL_11391_EPI_ISL_4443637_2021-08-09
Czech_Republic_NRL_11434_EPI_ISL_4443718_2021-08-11
Slovakia_UKBA-4248_EPI_ISL_3761937_2021-08-06
Czech_Republic_FNHK-Ps-000208_EPI_ISL_4397778_2021-09-04
Germany_BW-RKI-I-217820_EPI_ISL_4038615_2021-08-23
Germany_NW-RKI-I-227318_EPI_ISL_4222760_2021-09-03
Germany_BW-RKI-I-217842_EPI_ISL_4038634_2021-08-24
Germany_BW-RKI-I-207120_EPI_ISL_3716174_2021-08-15
Czech_Republic_CSQ0261_EPI_ISL_3369804_2021-07-26
Germany_NW-RKI-I-222821_EPI_ISL_4042288_2021-08-24
Sweden_SUS0003265_EPI_ISL_3402061_2021-08-03
Germany_BY-RKI-I-204421_EPI_ISL_3713726_2021-08-04
Switzerland_ZH-UZH-IMV-3ba4df5c_EPI_ISL_3649403_2021-08-12
Switzerland_BL-ETHZ-33627932_EPI_ISL_3345839_2021-07-31
Switzerland_BL-ETHZ-33627937_EPI_ISL_3345798_2021-07-31
Germany_NW-RKI-I-216820_EPI_ISL_4037703_2021-08-27
Germany_NW-RKI-I-235535_EPI_ISL_4229082_2021-09-08
Germany_BY-RKI-I-209460_EPI_ISL_3879751_2021-08-20
Germany_BY-RKI-I-217790_EPI_ISL_4038590_2021-08-24
Germany_SH-RKI-I-220190_EPI_ISL_4040243_2021-08-19
Germany_NW-RKI-I-218374_EPI_ISL_4039148_2021-08-25
Germany_BY-RKI-I-217748_EPI_ISL_4038557_2021-08-24
Germany_NW-RKI-I-230743_EPI_ISL_4225634_2021-09-03
Slovakia_UVZ_PL28_G8_14758_EPI_ISL_3117946_2021-07-19
Czech_Republic_IMG15N_Y40856_EPI_ISL_4138224_2021-07-28
Czech_Republic_IMG15N_U32852_EPI_ISL_4138211_2021-07-27
Slovakia_UVZ_PL32_F12_16543_EPI_ISL_3997439_2021-08-16
Germany_BW-RKI-I-191887_EPI_ISL_3237437_2021-07-24
Slovakia_UVZ_PL31_B5_15458_EPI_ISL_3452658_2021-07-30
Slovakia_UVZ_PL32_H6_16464_EPI_ISL_3997353_2021-08-08
Germany_BW-RKI-I-195975_EPI_ISL_3495557_2021-07-26
Germany_BW-RKI-I-196093_EPI_ISL_3495651_2021-07-26
Sweden_SUS0003291_EPI_ISL_3402087_2021-08-03
Czech_Republic_CSQ0267_EPI_ISL_3332268_2021-07-22
Czech_Republic_IMG15N_U33482_EPI_ISL_4080907_2021-08-03
Germany_BW-RKI-I-191153_EPI_ISL_3122169_2021-07-05
Czech_Republic_UKDD-D50_EPI_ISL_3832649_2021-08-06
France_PAC-IHU-27698_Nova1_EPI_ISL_3934245_2021-08
Czech_Republic_NRL_10840_EPI_ISL_3940431_2021-07-26
Sweden_SUS0003266_EPI_ISL_3402062_2021-08-03
Sweden_SE00616851_EPI_ISL_3786107_2021-08-06
Czech_Republic_NRL_10772_EPI_ISL_3940408_2021-07-24
Czech_Republic_NRL_10699_EPI_ISL_3425715_2021-07-19
Italy_FVG-GO-64030143_EPI_ISL_3923421_2021-08-11
Czech_Republic_NRL_10763_EPI_ISL_3425709_2021-07-23
Sweden_10110419_EPI_ISL_4336296_2021-09-03
Sweden_10110410_EPI_ISL_4108636_2021-09-03
Sweden_10110412_EPI_ISL_4336288_2021-09-03
Czech_Republic_NRL_10706_EPI_ISL_3425725_2021-07-20
Czech_Republic_NRL_11016_EPI_ISL_3940459_2021-07-30
Czech_Republic_NRL_11389_EPI_ISL_4443632_2021-08-09
Czech_Republic_NRL_11911_EPI_ISL_4275267_2021-08-22
Czech_Republic_NRL_11360_EPI_ISL_4443554_2021-08-06
Germany_BY-RKI-I-226504_EPI_ISL_4045239_2021-08-31
Czech_Republic_NRL_12268_EPI_ISL_4468801_2021-09-01
Germany_HE-RKI-I-195185_EPI_ISL_3494772_2021-07-23
Germany_BW-RKI-I-191579_EPI_ISL_3122563_2021-07-16
Germany_BW-RKI-I-191572_EPI_ISL_3122557_2021-07-15
Italy_LAZ-INMI-1322_EPI_ISL_3014043_2021-07-12
6.0E-5