SARS-COV-2 δ variant drives the pandemic in the USA through two subvariants

Xiang-Jiao Yang (xiang-jiao.yang@mcgill.ca)
McGill University

Research Article

Keywords: B.1.1.7, B.1.351, P.1, B.1.617.2, variant of concern, variant of interest, S202N, R203M, mutation profiling, phylogenetic analysis, nucleocytoplasmic trafficking

Posted Date: October 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-986605/v1

SARS-COV-2 δ variant drives the pandemic in the USA through two subvariants

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

[1]The 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

δ variant of SARS-COV-2 has overtaken all other variants and become a dominant pandemic driver aggressively. In India, it has evolved and yielded δ1, δ2, δ3 and δ4 subvariants. δ1 has also gradually become the dominant pandemic driver there and across Europe, raising the question whether this is true in other regions around the world. Here I demonstrate that δ1 has also become the dominant pandemic driver in the USA. In April and May 2021, α variant was the major pandemic driver, with Ida and γ variants playing minor roles. δ variant only started to emerge in April and May, but it rose exponentially and became a major driver one month later. By September, it was detected in ~99% COVID-19 cases and emerged as almost the sole pandemic driver. In the country, ~50% of its population was fully vaccinated in the summer of 2021; vaccination may have selected against all other variants and thereby helped δ variant achieve such an alarming status. One puzzling question is what genomic features make δ variant so highly competitive. Related to this, δ1, but not δ2, δ3 and δ4, has risen exponentially after May 2021, suggesting that unique NSP3 and nucleocapsid mutations that δ1 carries make it so competitive as a predominant pandemic driver. These results indicate that it is not δ variant *per se*, but its offspring, δ1, that makes δ variant a predominant pandemic driver. Alarmingly, δ1 subvariant has evolved further and gained additional mutations to finetune functions of spike, nucleocapsid and NSP3 proteins. Compared to δ1, δ2 subvariant is less important in driving the ongoing pandemic in the USA, but this subvariant has also evolved further and gained extra mutations. These results suggest a continuously branching model about δ variant evolution and reiterate the urgent need to track and block the evolution so that we will control and end this devastating pandemic as effectively and swiftly as possible.

Running Title: SARS-COV-2 evolution via continuous branching

Keywords: Mutation profiling, phylogenetic analysis, B.1.1.7, B.1.351, P.1, B.1.617.2, L452R, E484Q, P681R, V1176F, V1264L, D63G, D63S

INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic has resulted tragic loss of life, affected health care and crippled the economy around the world. Related to the impact on the economy, a recent estimate by International Monetary Fund indicates that failure to bring this pandemic under control will cost $5.3 trillions of US dollars in lost global growth over the next five year. Thus, it is highly important to control and end this pandemic as swiftly as possible. For this, we need to understand the culprit fully for considering and taking the most effective control measures against it.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the culprit behind this devastating disease [1]. The virus has yielded many variants, including 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 amazing feature about these variants of concern share is that they emerged from complete obscurity and then rapidly rose to the status of major pandemic drivers. One important question is what genetic features make such variants become pandemic drivers. To understand this question, I have tracked SARS-COV-2 genomes in the GISAID database [6]. I have recently found that δ variant has evolved actively and yielded four subvariants (δ1, δ2, δ3 and δ4) in India [7]. Among them, δ1 has emerged as the major pandemic driver and δ2 has played a less important role, whereas δ3 and δ4 have gradually faded away [7]. This emerging theme about δ variant is also true across Europe [7], with δ1 becoming almost the sole pandemic driver in the United Kingdom and Spain. A relevant question is how these δ subvariants have contributed to the pandemic in many other countries around the world.

Related to this, I tried to understand how the subvariants have driven the pandemic in the USA. Along with numerous variants of interest, all four variants of concern have spread to the country and contributed to the pandemic there. While extremely devastating to the country, this also provides a unique opportunity to understand relative virulence of different variants under the same social, political and geological contexts. Moreover, the fully vaccinated population has reached 55.7% in the country, and 13.3% of its entire population has been infected by the virus (according to Our World in Data, https://ourworldindata.org/; accessed on October 10, 2021). Thus, the herd immunity level is in the range from 61.4% (if the recovered population has received vaccination as the general population) to 69% (if no recovered individuals have received vaccination). This thus offers a unique opportunity to examine how different variants may proliferate under such an immunity pressure.

Here I demonstrate that δ has also become the predominant pandemic driver in the USA. It started to emerge in May 2021. It is present in ~99% cases identified in September and has become almost the sole pandemic driver. This may be in part because the herd immunity level in the country has exerted negative selection against other variants. Notably, as has occurred in India and Europe [7], δ1, but not δ2, δ3 and δ4, has risen exponentially in the USA. Moreover,

both δ1 and δ2 subvariants have evolved further and gained additional mutations through a continuously branching model. Therefore, necessary measures are urgently needed to track and block evolution of δ subvariants so that this devastating pandemic will subside and end as swiftly as possible.

RESULTS AND DISCUSSION

δ variant of SARS-COV-2 is the most powerful pandemic driver identified in the USA

As shown in Fig. 1A, there have been multiple waves of COVID-19 cases in the country, with the strongest one occurring from September 2020 through March 2021. The current wave started in June 2021 and formed a peak at the beginning of September. An important question is how SARS-COV-2 variants contributed to the different waves in the country. To answer this question, it is necessary to deconvolve the epidemiological curve at the genomic level. Of relevance, the country has contributed to ~28% of all genome sequences in the GISAID database (Fig. 1B), and genomic surveillance depth reached almost 10% in the summer of 2021 (Fig. 1C). Thus, SARS-COV-2 genomes from the country provide a rich resource for tracking how the variants have driven the epidemiological curve there. Moreover, the vaccination rate has reached over 50% (Fig. 1A; Our World in Data, https://ourworldindata.org/). The herd immunity level in the country should be much higher than this as there is natural immunity in the infected population (13.3% of the entire population, Our World in Data). Thus, it is also possible to extract insights into how different variants fare with a population at such an immunity level.

As shown in Fig. 1D-F, α variant was a major driver behind the pandemic wave that occurring in the spring and early summer for 2021. In May, it was responsible for 65% cases (Fig. 1D). Both Ida and γ variants also made significant contributions, with each detected in 10% genomes sequenced (Fig. 1D). In comparison, other variants, such as β, η, λ and μ, only played very minor roles, with each corresponding to close to or less than 1% genomes sequenced in the country (Fig. 1F). Fortunately, perhaps due to progress of vaccination, these variants all faded away (Fig. 1E-F). One only exception is δ variant, which was still obscure in March and April 2021 but rose exponentially afterwards. In August and September, it became almost the sole pandemic driver in the country (Fig. 1D-E). In comparison, no other variants, including α, reached such a predominant status. Strikingly, κ variant shares multiple mutations with δ variant (e.g. spike T19R, G142D, L452R, P681R and D950N, as well as nucleocapsid D63G, R203M and D377Y), but it is only responsible for a very minor portion of cases (Fig. 1F). Amazingly, few κ genomes were reported from the country after July 2021 (Fig. 1F). Such a drastic difference between these two variants in terms of their potential to become major pandemic drivers is quite puzzling. Together, these results raise an intriguing question about

what makes δ variant so competitive and powerful in driving the pandemic in the USA and other countries around the world.

Related to this question, δ variant has evolved and yielded four subvariants (δ1, δ2, δ3 and δ4) in India [7]. 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 [7]. This emerging theme about δ variant is also true in Europe [7]. By analogy, I thus postulated that δ1 and δ2 subvariants have driven the pandemic in the USA.

δl subvariant rapidly emerged as a dominant pandemic driver in the USA

To test this hypothesis, I downloaded δ genomes from the GISAID database for mutation profiling via Coronapp [8,9]. There are over 0.5 million δ genomes sequenced in the USA, so it would need lots of computing power to process all of them. To simplify the analysis, I started with those δ genomes identified in the country by April 2021 because the corresponding initial COVID-19 cases served as seeds for subsequent cases detected in the country. As shown in Fig. 2A, the initial genomes mainly two large groups. One of them encodes P822L of NSP3 and the other possesses the three mutations for A488S, P122L and P1469S of NSP3. As described in the previous study [7], the latter group corresponds to δ1 subvariant, whereas the former group is composed of δ2, δ3 and δ4 subvariants belongs to the former group. As shown in Fig. 2B, spike K77T and A222V, which are markers of δ4 and δ2 subvariants, respectively [7], are encoded by two subsets of the initial genomes. Moreover, G215C and R385K, which are markers of δ4 and δ2 subvariants, respectively [7], are also encodes by subsets of the genomes (Fig. 2C). Thus, the initial genomes encode all four δ subvariants. In terms of abundance, ~30% of the genomes are for δ1 subvariant and ~20% of them are for δ2 or δ3 subvariant, but ~5% of the genomes are for δ4 subvariant (Fig. 2B-C). These results also indicate that at this initial stage, the abundance difference among the four subvariants was not so dramatic.

The average mutation load of these initial δ-genomes is 37 per genome (Fig. 3A). In comparison, κ genomes identified around the same time carry 31 mutations per genome (Fig. 3B). Thus, on the average, δ possesses ~6 more mutations than κ variant. This may partly explain why κ variant has not been successful in spreading from India to the USA and causing major outbreaks there. Indeed, despite being a major driver in India, κ variant has not been so successful in causing outbreaks in Europe [7] and many other countries. Analysis of α-genomes revealed a mutation load of 37 per genome (Fig. 3C), which is similar to what δ carries (Fig. 3A). Thus, the mutation load may be one criterion to judge whether a variant has the potential to be a pandemic driver. Related to this, one hallmark of the four Variants of Concern is their high mutation loads. But this should not be used as the sole criterion because the quality of each mutation and the combination of different mutations are two other important criteria.

To substantiate the results from mutation profiling (Fig. 2), I carried out phylogenetic analysis of 356 δ genomes identified in the USA by April 20, 2021. For this, only high-coverage genomes with complete date information on sample collection were used. The results indicate that these genomes form five distinct groups corresponding to δ1, δ2, pre-δ3, δ3 and δ4 (Fig. 3D). This is similar to what was observed with δ-genomes identified in India and Europe by April 2021 [7]. Thus, the initial δ genomes identified in the country by April 2021 form 5 distinct groups. Four of them corresponding to δ1, δ2, δ3 and δ4, whereas the fifth one is related to δ3.

Interestingly, also as has occurred in India and Europe [7], cases with δ2, δ3 and δ4 genomes peaked in June or July 2021 but then declined (Fig. 3E-F). By contrast, the number of δ1 genomes increased exponentially after May 2021 and became responsible for close to or over 80% cases in August and September (Fig. 3E-F). The remaining portion was mainly δ2, where δ3 and δ4 genomes were hardly detectable in August and September 2021 (Fig. 3E-F). This dramatic rise of δ1 in the USA from complete obscurity in such a short period of a few months is astonishing. However, it is very similar to what has occurred in India and across Europe [7]. These findings also support that unique mutations in δ1 variant make it much more competitive than the other three subvariants. One difference is that δ1 variant possesses three NSP3 substitutions whereas δ2, δ3 and δ4 subvariants carry only one NSP3 substitution (Fig. 2A). The other difference is that δ1 variant harbors nucleocapsid G215C, which is absent in δ2, δ3 and δ4 subvariants. Thus, in addition to altering spike protein, SARS-COV-2 may finetune functions of NSP3 and nucleocapsid proteins as important mechanisms to improve its fitness.

Both δ1 and δ2 subvariants have evolved further and yielded sublineages in the USA

To investigate how δ1 subvariant has evolved, I carried out mutation profiling of 1,086 δ1 genomes identified in the USA from September 09-14, 2021. As shown in Fig. 4A, ~20% of the genomes carry A1537S and/or A1736V of NSP3. This is intriguing and raises the question whether these two substitutions are present in all of the 20% genomes. In the GISAID database (accessed on October 09, 2021), there are 17,961 and 80,262 δ1 genomes encoding A1537S and A1736V, respectively, but there are only 3,841 δ1 genomes carring both. These observations suggest that δ1 acquired A1537S and A1736V independently before obtaining the second substitution to yield the sublineage with both. Thus, δ1 has evolved and yielded three sublineages encoding A1537S and/or A1736V (Fig. 4B). Due to its potential virulence, the sublineage with both A1537S and A1736V is referred to as δ1S (the letter S is to denote S1537) and will be analyzed below. In addition, δ1 has also acquired spike S112L, Q613H and V1104L (Fig. 4B), or nucleocapsid R208S (Fig. 4C), in very small subsets of genomes.

To investigate how δ2 subvariant has evolved, I carried out mutation profiling of 1,231 δ2 genomes identified in the USA from September 09-14, 2021. As shown in Fig. 5A, one

sublineage stood out, corresponding to a third of the genomes analyzed. It encodes E815D of NSP3. A1537S of NSP3 is present in a much smaller subset of genomes (Fig. 5A). As shown in Fig. 5B, spike V289I and V1264L are present in about a third of the genomes, with N1047S in about 15% genomes. Intriguingly, G18V is present in about a third of genomes (Fig. 4C). As E815D of NSP3, spike V289I and V1264L, and G18V of nucleocapsid are present in a similar number of δ2 genomes, some of these substitutions may be encoded by the same subset of genomes. To verify this, I inspected mutations in different subsets of δ2 genomes and found that E815D, V289I and G18V are present in the same subset of δ2-genomes whereas V1264L-encoding genomes forms a distinct subset. These two subsets encode two new δ2 sublineages, referred to as δ2D and δ2L, where the letters D and L denote D815 of NSP3 and L1264 of spike protein, respectively. G18V-encoding δ2 genomes were identified before V289I-encoding ones appeared. Afterwards, δ2D genomes encoding all three substitutions were identified. Thus, δ2 subvariant has acquired G18V, V289I and E815D in a sequential manner.

A V1264L-encoding sublinage, δ1L, has been a major pandemic driver in Southeast Asia [10], so emergence of the V1264L-encoding sublinage, δ2L, is alarming. A subset of δ2L genomes encode N1047S (Fig. 5B). As described in another study [10], this substitution confers evolutionary advantage to δ2L. Thus, δ2 subvariant has evolved further and produced two new sublineages in the USA.

As shown in Fig. 6A-B, the average mutation loads in δ1- and δ2-genomes identified in the USA during the second week of September 2021 are 44 and 39, respectively. Alarmingly, the mutation load in δ2S genomes (encoding both A1537S and A1736V, Fig. 4A) load of 46-47 per genome makes this lineage one of the most mutated SARS-COV-2 variant identified so far. A V1176F-encoding sublineage, δ1F, is the most actively evolving SARS-COV variant identified so far, with an average mutation load of ~50 mutations per genome [11]. In comparison, the mutation load is 47 per genome for C.1.2 variant, reported to be the most mutated variant [12,13]. There are several δ1 sublineages carrying 46-49 mutations per genome [7,10]. Thus, δ2S is one of the most mutated SARS-COV-2 lineages.

In comparison, δ2D carries 41 mutations per genome, which low compared to the sublineages mentioned above. However, monthly distribution of δ2 subvariant and δ2D revealed that the extra substitutions make this sublineage more virulent than δ2 subvariant itself. The first three δ2D genomes were identified in North Carolina, Texas and Florida in May 2021. As of October 09, 2021, there are 18,343 such genomes in the GISAID database, with 5,625 from Florida (30.7%). In the state, 503 out of 5,048 genomes (~10%) identified in September are due to δ2D. Thus, this new lineage has played a significant role in driving the pandemic there. As shown in Fig. 6E, δ2D has evolved further and gained additional substitutions in small subsets of its genomes. More alarmingly, eight δ2D genomes identified in Texas in August and September 2021 encode spike V1176F. All of them also carries T141I of

nucleocapsid protein. Moreover, four δ2D genomes identified in Illinois, Ohio and New York in August and September 2021 encode V1264L. Both V1176F and V1264L confer evolutionary advantage [10,11], so it will be important to track how δ2D and its sublineages will evolve.

δ1 and δ2 alter nucleocapsid and NSP3 as important mechanisms to improve viral fitness

Compared to parental δ variant, both δ1 and δ2 subvariants have gained additional substitutions. G215 of nucleocapsid is located at an α-helix and G215C may improve the structure of this helix (Fig. 7A). This is a signature substitution of δ1 subvariant and absent in δ2 subvariant (Figs 2 & 4-5) [7]. Thus, G215C may be one of different reasons that δ1 is much more virulent than δ2. If so, nucleocapsid alteration is an important mechanism by which SARS-COV-2 improves its fitness. Related to this, δ variant carries three other nucleocapsid substitutions, D63G, R203M and D377Y (Fig. 2C), located at three different domains of the protein (Fig. 7A). Moreover, R385K is a signature substitution of δ3 subvariant (Fig. 2C) [7]. δ1 subvariant has evolved further and yielded new sublineages in the USA (Figs 4 & S2). In a subset of δ1S genomes, D63G is replaced by D63S (Fig. S2C). This is significant as D63 is located in the N-terminal domain and directly involved in RNA binding [14]. D63G may interfere with RNA binding [14], so it is quite puzzling why δ1 variant has acquired such a substitution. One possibility is that D63G is due to a passenger mutation and thus serves an 'evolutionary trap' that the variant enters accidentally. In this regard, D63S is expected to be more favorable for RNA interaction than D63G and may thus improve viral fitness of δ1S.

Like δ1 subvariant, δ2 subvariant has also evolved further and yielded sublineages in the USA (Figs 5 & 6D-E). G18V of nucleocapsid is encoded in a third of δ2D genomes analyzed (Fig. 5C). G18 of nucleocapsid is close to the N-terminal tail, a known immune epitope, so G18V may confer immune evasion. Thus, this reiterates that during evolution, δ variant often acquires nucleocapsid substitutions as an important mechanism to improve viral fitness. In support for the importance of nucleocapsid alteration during SARS-COV-2 evolution, genetic screens with an artificially weakened β-coronavirus revealed that its nucleocapsid gene is frequently mutated during evolution in cultured cells [15,16]. Strikingly, some of the mutations from the screens are reminiscent of R203M in δ and k variants and similar to R203K in α and γ variants of concern.

The genetic screens also uncovered NSP3 gene mutations [15,16], thereby supporting that δ variant may acquire NSP3 substitutions as an important mechanism to improve viral fitness. Related to this, δ1 subvariant carries A488S, P822L and P1469S, whereas δ2 subvariant encodes P822L (Fig. 2A). Different δ1 sublineages have acquired H1307Y, A1537S, W1545L and/or A1736V (Figs 4 & S2). One such sublineage encodes both A1537S and A1736V (Fig. S2). Strikingly, this sublineage carries a mutation load of 46-47 per genome (Fig. 6C), making

it one of the most mutated SARS-COV-2 variant identified so far. Moreover, a subset of the genomes encode D63S, rather than D63G, of nucleocapsid. Similar to δ1, δ2 subvariant has also produced new sublineages with additional NSP3 substitutions (Fig. 5). The most prominent one is δ2D, which carries E815D of NSP3. E815 forms a salt bridge with K610 and may clash with E812. Thus, E815D may improve the local structure. This region is close to P822 (Fig. 6B), is replaced by leucine in δ2 subvariant. Thus, E815D may synergize with P822L to improve the fitness of δ2D. This δ2 sublineage also encodes G18V of nucleocapsid and V289I of spike protein. As shown in Fig. 6C, V289 is close to L276 and F306, so V289I may improve hydrophobic interaction among these three residues. New genomes identified in September 2021 revealed that δ2D has evolved further and acquired virulent substitutions such as V1176F and V1264L. Thus, it will be important to watch out this sublineage.

δ variant evolves and generates sublineages through a continuously branching model

δ variant has evolved dynamically and yielded four subvariants (δ1, δ2, δ3 and δ4) in India [7]. All of them have been detected in the USA (Fig. 3E), but δ1 has emerged as the dominant pandemic driver there. It is now responsible for >80% cases, whereas δ2 is present in ~10% cases (Fig. 3F). This emerging theme about δ subvariants is also true in Europe [7], with δ1 becoming almost the sole pandemic driver in the United Kingdom and Spain. Based on their temporal abundance from March to September 2021, it is tempting to propose the relative virulence of the subvariants as compared to δ variant itself and α variant (Fig. 7D).

δ1 subvariant has evolved further, gained additional mutations (Fig. 4) and yielded new sublineages (Fig. 7E). Among these sublineages are δ1-A1537S and δ1-A1736V, which account for a significant portion of δ1 genomes (Fig. 4). Moreover, δ1S encodes both A1537S and A1736V (Fig. 7E). δ1S has evolved further and gained new substitutions, such as H1307Y of NSP3 and D63S of nucleocapsid (Fig. S2), further attesting to the continuously evolving nature of SARS-COV-2. Like δ1 subvariant, δ2 subvariant has evolved further and gained additional mutations (Fig. 5). One resulting sublineage is δ2D, encoding E815D of NSP3, spike V289I and nucleocapsid G18V (Figs 5 & 6E). Alarmingly, this lineage is more virulent than δ2 subvariant itself (Fig. 6D). Furthermore, δ2D has recently gained additional substitutions, such as spike D253G and D1259Y (Fig. 6E). Moreover, V1176F and V1264L, two virulent spike substitutions [10,11], are present in 12 δ2D genomes identified in the USA from August to September 2021. These observations suggest that δ2D is on the way to become even more virulent. Thus, this is perhaps the most dangerous δ2 sublineage identified so far.

δ1L is a major pandemic driver in Southeast Asia [10] and has been detected in a small portion of cases identified in the USA at the beginning of September 2021 (Fig. 4B). This lineage has been a key pandemic driver in Indonesia, Singapore, Malaysia and East Timor [10], so it is wise to watch out this emerging δ1-sublineage while following those that have recently

emerged inside the USA (e.g., δ1S and δ2D). Emergence of different δ1 and δ2 sublineages in the USA and other countries supports that SARS-COV-2 generates variants by evolving and branching out continuously (Fig. 7E). Some of the variants are selected and emerge as major pandemic drivers. Deep genomic surveillance and systematic genomic annotation should help identify such drivers at the initial stages when they are about to cause major outbreaks. Overall, to help end this tragic pandemic swiftly, it is important to track and block the continual evolutionary process of SARS-COV-2 (Fig. 7E).

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 another study [7]. The Fasta headers were shortened and modified further [7]. The cleaned Fasta file was used for mutational profiling via Coronapp (http://giorgilab.unibo.it/coronannotator/), a web-based mutation annotation application [8,9]. 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 [17] was used for phylogenetic analysis as described [7].

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 [7]. 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 [7], 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.

Pandemic and vaccination data

Pandemic and vaccination data were downloaded from the Our World in Data website as described [7].

REFERENCES

1. Hu, B., H. Guo, P. Zhou, and Z.L. Shi. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* **19**, 141-154.
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., I.M. Claro, D. Candido, L.A. Franco, P.S. Andrade, T.M. Coletti, C.A. Silva, F.C. Sales, E.R. Manuli, R.S. Aguiar, N. Gaburo, C. da C. Camilo, N.A. Fraiji, M. Esashika Crispim, M. Carvalho, A. Rambaut, N. Loman, O.G. Pybus, E.C. Sabino, and C.G. Network. (2021). Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *https://virological.org*.
5. Mlcochova, P., et al. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*.
6. Elbe, S. and G. Buckland-Merrett. (2017). Data, disease and diplomacy: GISAID's

innovative contribution to global health. *Glob Chall* **1**, 33-46.
7. Yang, X.J. (2021). SARS-COV-2 delta variant drives the pandemic in India and Europe through two subvariants. *medRxiv*.
8. 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.
9. Mercatelli, D. and F.M. Giorgi. (2020). Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front Microbiol* **11**, 1800.
10. Yang, X.J. (2021). Delta-1 variant of SARS-COV-2 acquires spike V1264L and drives the pandemic in Indonesia, Singapore and Malaysia. *bioRxiv*.
11. Yang, X.J. (2021). Delta-1 variant of SARS-COV-2 acquires spike V1176F and yields a highly mutated subvariant in Europe. *bioRxiv*.
12. 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.
13. Yang, X.J. (2021). SARS-COV-2 C.1.2 variant is highly mutated but may exhibit reduced affinity for ACE2 receptor. *bioRxiv*.
14. Dinesh, D.C., D. Chalupska, J. Silhan, E. Koutna, R. Nencka, V. Veverka, and E. Boura. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog* **16**, e1009100.
15. Hurst, K.R., R. Ye, S.J. Goebel, P. Jayaraman, and P.S. Masters. (2010). An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. *J Virol* **84**, 10276-88.
16. Hurst, K.R., C.A. Koetzner, and P.S. Masters. (2013). Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *J Virol* **87**, 9159-72.
17. 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.

FIGURE LEGENDS

Figure 1. α and δ variants of SARS-COV-2 as key pandemic drivers in the USA during 2021. (**A**) Epidemiological curve and vaccination progress in the country. While different variants, including α, β, γ and Ida, drove the pandemic in the spring and early summer for 2021, δ variant is the predominant driver of the latest wave. Despite up to ~50% full vaccination, the current wave is much more powerful than the previous one, supporting high virulence of δ variant. The entire pandemic is divided into pre-VOC (Variant of Concern) and VOC phases. The peak dates of the waves are indicated, with the latest peak occurred at the beginning of September 2021. (**B**) Genomic surveillance depth in the USA compared to that in the world, United Kingdom, Europe and Asia. (**C**) Genomic surveillance depth in the USA steadily increased from ~0.5% at the end of 2020 to ~10% in the summer of 2021. (**D-F**) Temporal distribution of different variants detected in the USA. Except for δ, all other variants have gradually faded away. In August and September 2021, δ became almost the sole pandemic driver. No other variants have been so predominant during the pandemic, attesting to the uniqueness of this variant. For preparation of panel A, the Our World in Data website (https://ourworldindata.org/) was accessed on September 18, 2021, and for preparation of panels B-F, the GISAID database was accessed on the same day. Note that the values for August 2021 in panel C and for September 2021 in panels C, D & F were not complete yet when the website and database were accessed.

Figure 2 Mutation profile of 902 δ-genomes identified in the USA by April 2021**.** The genomes were downloaded from the GISAID SARS-COV-2 genome sequence database on September 15, 2021 for mutation profiling via Coronapp [8,9]. Shown in (A), (B) and (C) are substitutions in NSP3, spike and nucleocapsid proteins, respectively.

Figure 3 Analysis of SARS-COV-2 genomes identified in the USA. (**A**) Mutation load in 902 δ genomes sequenced in the country by April 2021. The distribution was generated via Coronapp as in Fig. 2, but only the mutation load is shown here. (**B**) Mutation load in 270 κ-genomes identified in the country by April 2021. The distribution was generated via Coronapp, but only the mutation load is shown here. (**C**) Mutation load in 544 α genomes identified in the country April 30, 2021. The distribution was generated via Coronapp, but only the mutation load is shown here. (**D**) Phylogenetic analysis of 356 δ-genomes identified in the USA by April 20, 2021. The genomes were downloaded from the GISAID database on September 18, 2021. Only high-coverage genomes with complete date information on sample collection were used for phylogenetic analysis. The package RAxML-NG was used to generate this bestTree and 20 maximum likelihood trees for presentation via FigTree. The strain names and GISAID accession numbers of the genomes are provided in Figure S1. (**E-F**) Monthly distribution of δ

variant and its subvariants. For the analyses, the GISAID database was accessed on September 15, 2021 (panels A & D); September 18, 2021 (E-F) and September 19, 2021 (B-C). δ subvariants are defined as in another study [7]. The value for September 2021 was not complete (indicated by an asterisk in panel E) when the database was accessed.

Figure 4 Mutation profile of 1,086 δ1 genomes identified in the USA from September 9-14, 2021. The genomes were downloaded from the GISAID database on September 19, 2021 for mutation profiling via Coronapp [8,9]. Shown in (A), (B) and (C) are substitutions in NSP3, spike and nucleocapsid proteins, respectively. δ1 subvariant is defined as in another study [7].

Figure 5 Mutation profiles of 1,231 δ2 genomes identified in the USA. δ2 genomes identified from September 09-14, 2021 and submitted by September 30, 2021 were downloaded from the GISAID database on October 09, 2021 for mutation profiling via Coronapp [8,9]. Shown in (A), (B) and (C) are substitutions in NSP3, spike and nucleocapsid proteins, respectively. δ2 subvariant is defined as in another study [7].

Figure 6**.** Mutation load in δ1/2 subvariants and analysis of their sublineages. (**A**) Mutation load in 1,086 δ1-genomes identified in the country. The distribution was generated via Coronapp as Fig. 4, but the mutation load is shown here. (**B**) Mutation load in 1,231 δ2 genomes identified in the USA. The distribution was generated via Coronapp as Fig. 5, but the mutation load is shown here. The mutation load in δ1 genomes (A) is 4-5 mutations more than that in δ2 genomes (B). Compared to parental δ variant, all δ1 and δ2 genomes carry 4 and 3 extra NSP3 gene mutations, respectively (Figs 4A & 5A). While δ2 genomes encode spike A222V as an extra substitution (Figs 4B & 5B), δ1 subvariant possesses G215C as an extra nucleocapsid substitution (Figs 4C & 5C). Thus, mutations in genes for NSP3, spike and nucleocapsid proteins do not explain fully the mutation load difference between δ1- and δ2-genomes. (**C**) Mutation load in 1,041 δ2S-genomes identified in the USA. δ2S-genomes encode both A1537S and A1736V of NSP3. The genomes were downloaded from the GISAID SARS-COV-2 genome sequence database on October 09, 2021 for mutation profiling via Coronapp [8,9]. The average mutation load of 46-47 per genome makes this lineage one of the most mutated SARS-COV-2 variant identified so far. NSP3, spike and nucleocapsid substitutions are provided in Fig. S2. (**D**) Monthly distribution of δ2 subvariant and a sublineage (δ2D) encoding E815D of NSP3. For the analyses, the GISAID website was accessed on October 09, 2021. δ2 subvariant is defined as in another study [7]. The sublineage encodes an extra NSP3 substitution, E815D. (**E**) Mutation profile of 663 V289I-encoding δ2-genomes identified in the USA after September 16, 2021. The genomes were downloaded from

the GISAID SARS-COV-2 genome sequence database on October 09, 2021 for mutation profiling via Coronapp [8,9]. Shown here are substitutions in spike protein.

Figure 7 Mechanistic impact of substitutions in δ variant and a continuously branching model on its evolution. (**A-B**) Structural details of nucleocapsid and NSP3 residues altered in δ variant and its subvariants. Panels A-B are based on PyMol presentation of structural models, QHD43423.pdb and QHD43415.pdb (https://zhanglab.ccmb.med.umich.edu/COVID-19/), respectively. Both were built by Dr. Yang Zhang's group at University of Michigan. The former was rom crystal structures of the N- and C-terminal RNA-binding domains of nucleocapsid protein (6M3M and 6YUN from the PDP database, respectively). (**C**) Structural details of spike V289 and its neighboring residues. Adapted from PyMol presentation of the spike protein structure 6XR8 from the PDB database. (**D**) Relative virulence of δ variant and its subvariants compared to α variant. (**E**) A continuously branching model about how δ variant evolves and generates subvariants. It has evolved actively and yielded four subvariants (δ1, δ2, δ3 and δ4) in India [7]. All of them have been detected in the USA (Fig. 3D), but only δ1 has emerged as the major pandemic driver there. This emerging theme about δ subvariants is also true in Europe [7], with δ1 becoming almost the sole pandemic driver in the United Kingdom and Spain. δ1 and δ2 subvariants have evolved and gained additional mutations, but the sublineages are still minor (Figs 4-5). δ1L is a major pandemic driver in Southeast Asia [10].

SUPPLEMENTAL INFORMATION

This section includes two supplementary figures and six acknowledgement tables for the GISAID genomes used in this work.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Phylogenetic analysis of 356 δ-genomes identified in the USA by April 20, 2021. The genomes were downloaded from the GISAID database for phylogenetic analysis as in Fig. 3D. Only high-coverage genomes with complete date information on sample collection were used. The package RAxML-NG was used to generate 20 maximum likelihood trees and the bestTree for presentation via FigTree.

Figure S2 Mutation profiling of 1,041 δ1S-genomes identified in the USA after September 01, 2021. δ1S genomes encode two additional substitutions of NSP3, A1537S and A1736V. The genomes were downloaded from the GISAID sequence database on October 09, 2021 for mutation profiling via Coronapp [8,9]. The average mutation load of 46-47 per genome (Fig. 6C) makes δ2S one of the most mutated SARS-COV-2 variants identified so far. Note that in a subset of the genomes, D63G of nucleocapsid is replaced by D63S. D63G is a signature substitution of δ1 variant. D63 is located within the N-terminal RNA-bidning domain (Fig. 7A) and directly involved in RNA binding. D63G may affect this binding [14], whereas D63S may be more favorable for RNA interaction than D63G and thus improve viral fitness of δ1S.

Figure 1

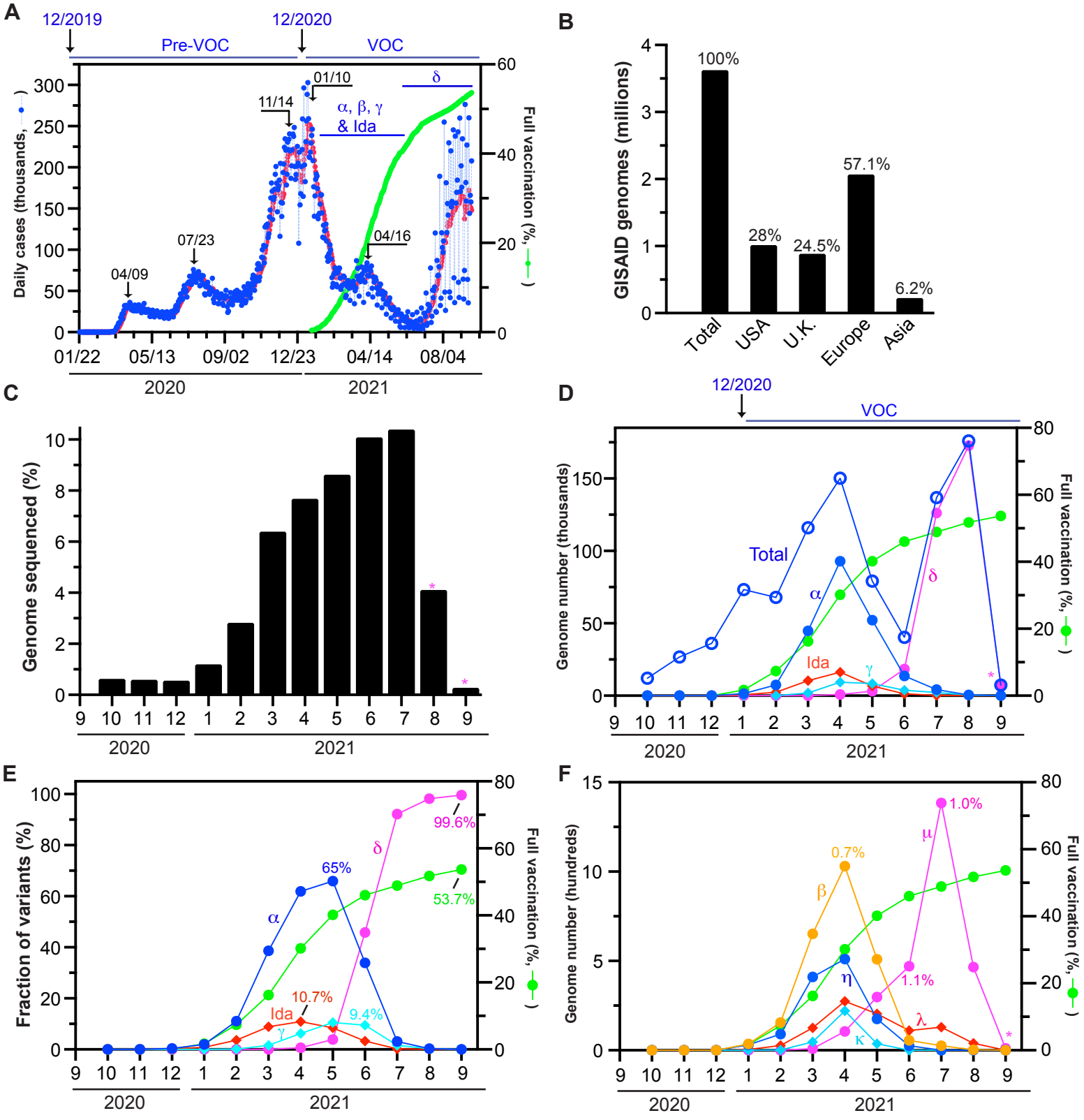

A
12/2019
12/2020
Pre-VOC
VOC
01/10
δ
11/14
α, β, γ
& Ida
07/23
04/16
04/09
Daily cases (thousands,)
Full vaccination (%,)
01/22
05/13
09/02
12/23
04/14
08/04
2020
2021
B
GISAID genomes (millions)
100%
28%
24.5%
57.1%
6.2%
Total
USA
U.K.
Europe
Asia
C
Genome sequenced (%)
2020
2021
D
12/2020
VOC
Genome number (thousands)
Full vaccination (%,)
Total
α
δ
Ida
γ
2020
2021
E
Fraction of variants (%)
Full vaccination (%,)
99.6%
65%
53.7%
10.7%
9.4%
α
δ
Ida
γ
2020
2021
F
Genome number (hundreds)
Full vaccination (%,)
1.0%
0.7%
1.1%
μ
β
η
λ
κ
2020
2021

Figure 2

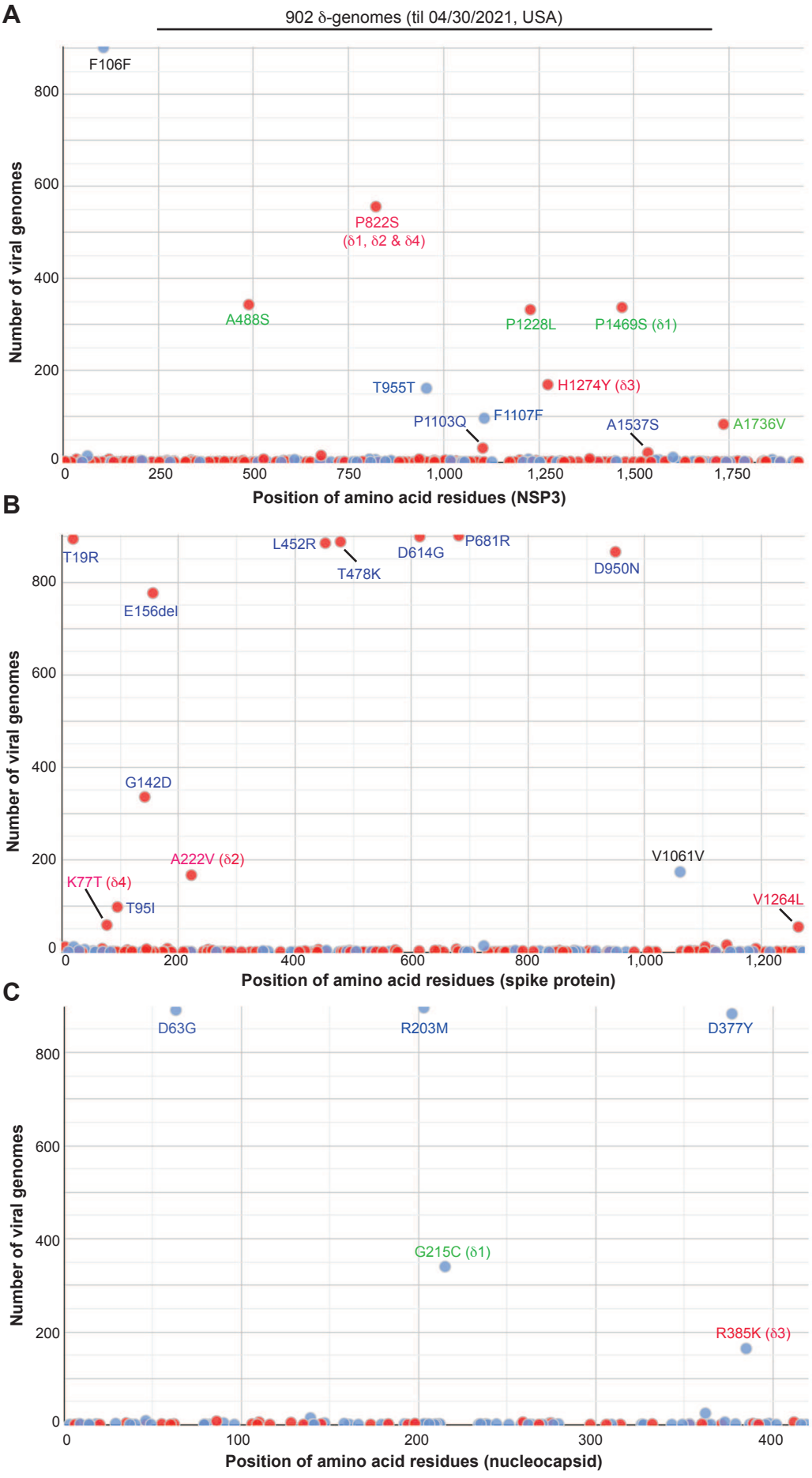

A
902 δ-genomes (til 04/30/2021, USA)
Number of viral genomes
Position of amino acid residues (NSP3)
F106F
P822S
(δ1, δ2 & δ4)
A488S
P1228L
P1469S (δ1)
T955T
H1274Y (δ3)
P1103Q
F1107F
A1537S
A1736V
B
Number of viral genomes
Position of amino acid residues (spike protein)
T19R
L452R
T478K
D614G
P681R
D950N
E156del
G142D
A222V (δ2)
V1061V
K77T (δ4)
T95I
V1264L
C
Number of viral genomes
Position of amino acid residues (nucleocapsid)
D63G
R203M
D377Y
G215C (δ1)
R385K (δ3)

Figure 3

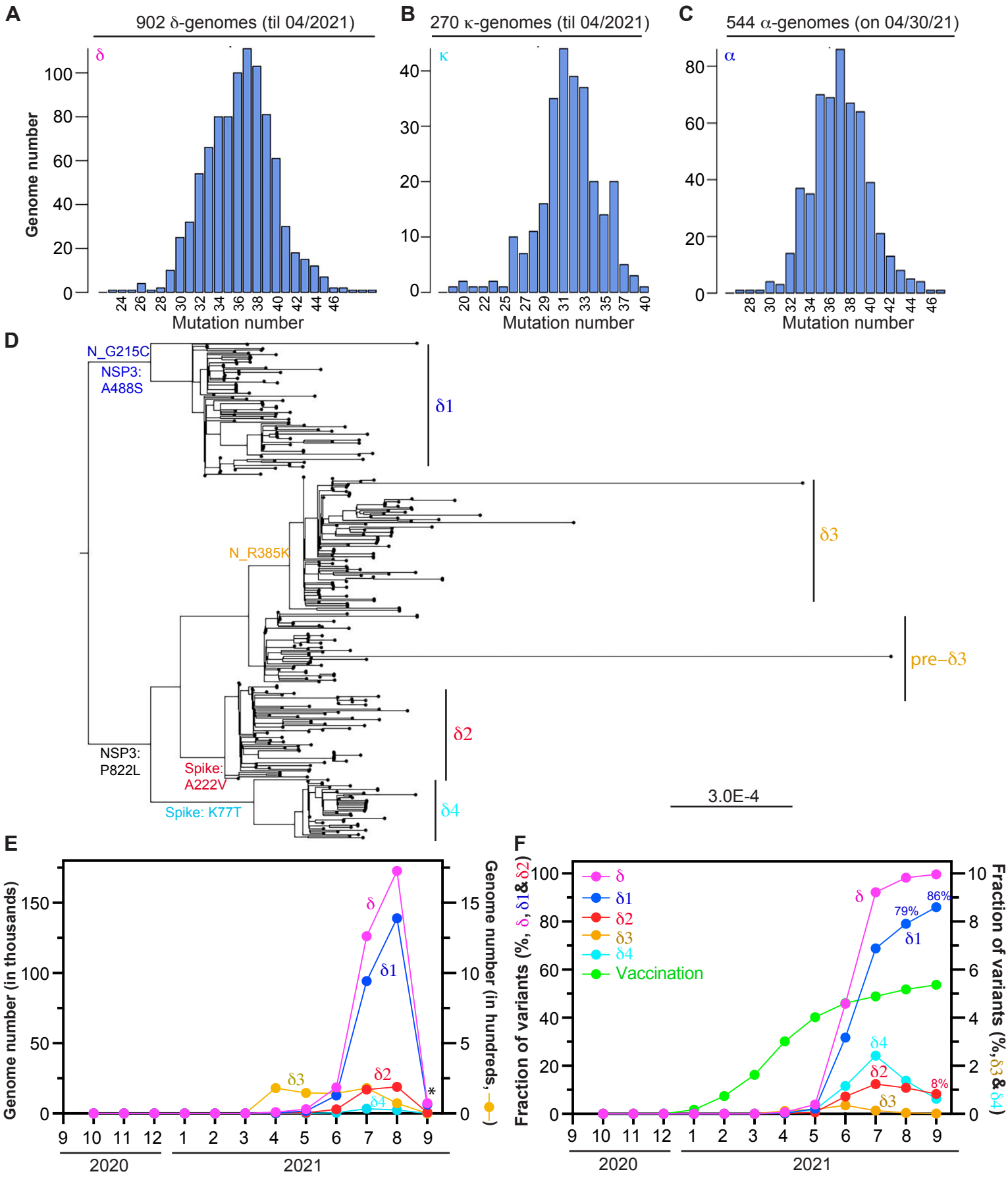

A
902 δ-genomes (til 04/2021)
δ
Genome number
Mutation number
B
270 κ-genomes (til 04/2021)
κ
Mutation number
C
544 α-genomes (on 04/30/21)
α
Mutation number
D
N_G215C
NSP3:
A488S
δ1
δ3
N_R385K
pre–δ3
δ2
NSP3:
P822L
Spike:
A222V
Spike: K77T
δ4
3.0E-4
E
Genome number (in thousands)
Genome number (in hundreds)
δ
δ1
δ2
δ3
δ4
2020
2021
F
Fraction of variants (%, δ, δ1& δ2)
Fraction of variants (%,δ3 & δ4)
δ
δ1
δ2
δ3
δ4
Vaccination
86%
79%
8%
2020
2021

Figure 4

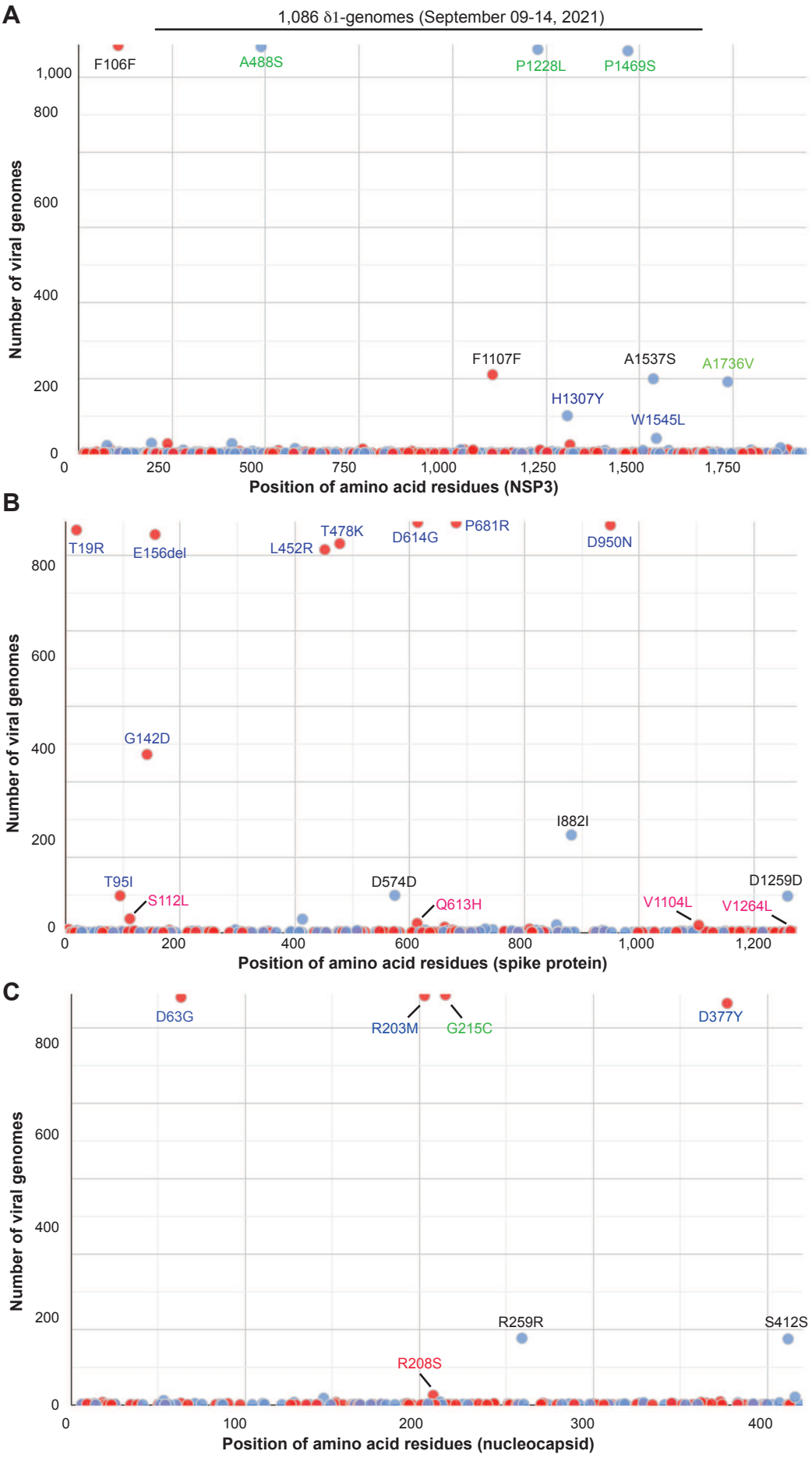

A
1,086 δ1-genomes (September 09-14, 2021)
F106F
A488S
P1228L
P1469S
F1107F
A1537S
A1736V
H1307Y
W1545L
Number of viral genomes
Position of amino acid residues (NSP3)
B
T19R
E156del
L452R
T478K
D614G
P681R
D950N
G142D
I882I
T95I
S112L
D574D
Q613H
V1104L
V1264L
D1259D
Number of viral genomes
Position of amino acid residues (spike protein)
C
D63G
R203M
G215C
D377Y
R259R
S412S
R208S
Number of viral genomes
Position of amino acid residues (nucleocapsid)

Figure 5

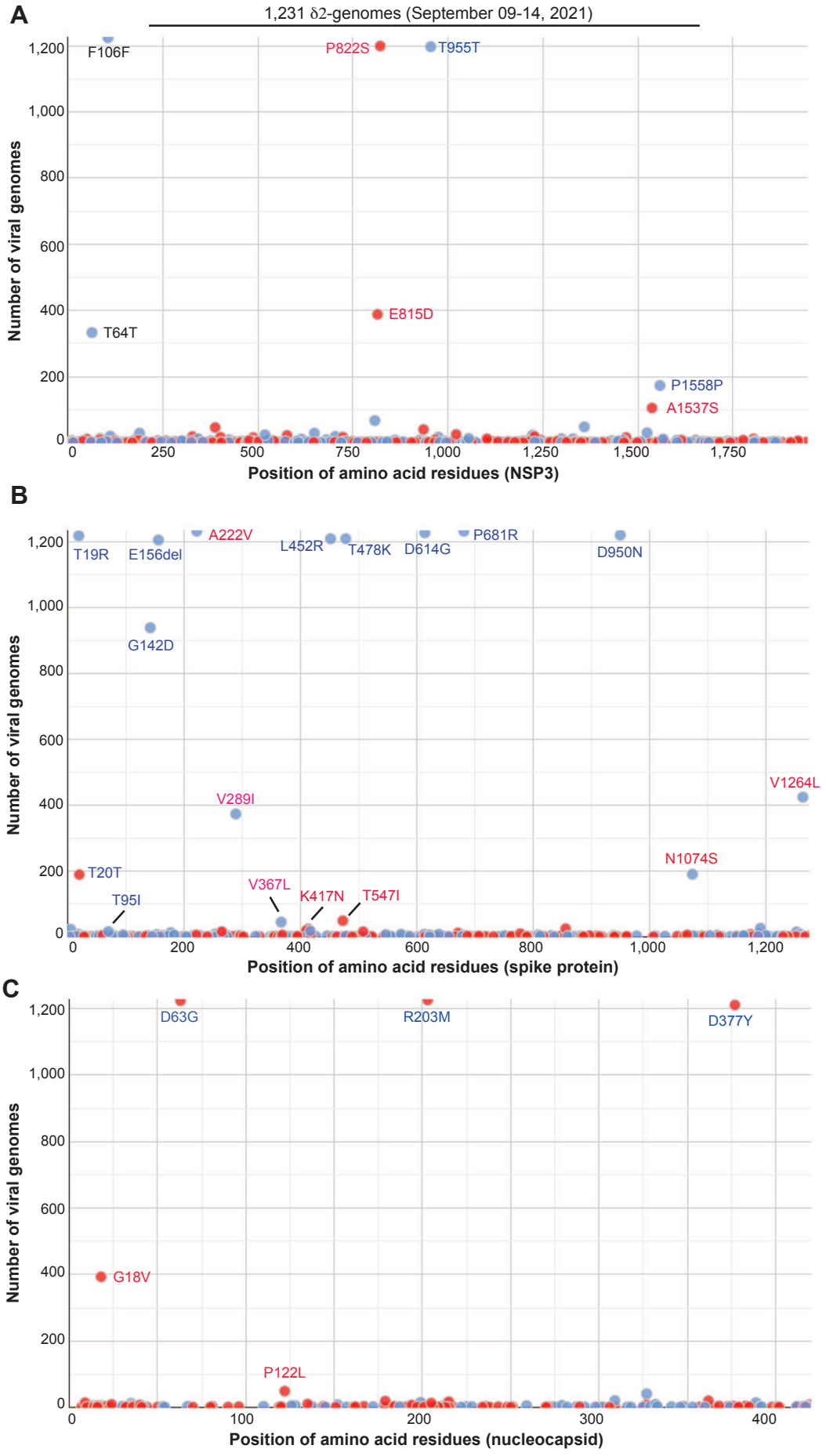

A
1,231 δ2-genomes (September 09-14, 2021)
Number of viral genomes
Position of amino acid residues (NSP3)
F106F
P822S
T955T
E815D
T64T
P1558P
A1537S
B
Number of viral genomes
Position of amino acid residues (spike protein)
T19R
E156del
A222V
L452R
T478K
D614G
P681R
D950N
G142D
V289I
V1264L
T20T
N1074S
V367L
K417N
T547I
T95I
C
Number of viral genomes
Position of amino acid residues (nucleocapsid)
D63G
R203M
D377Y
G18V
P122L

Figure 6

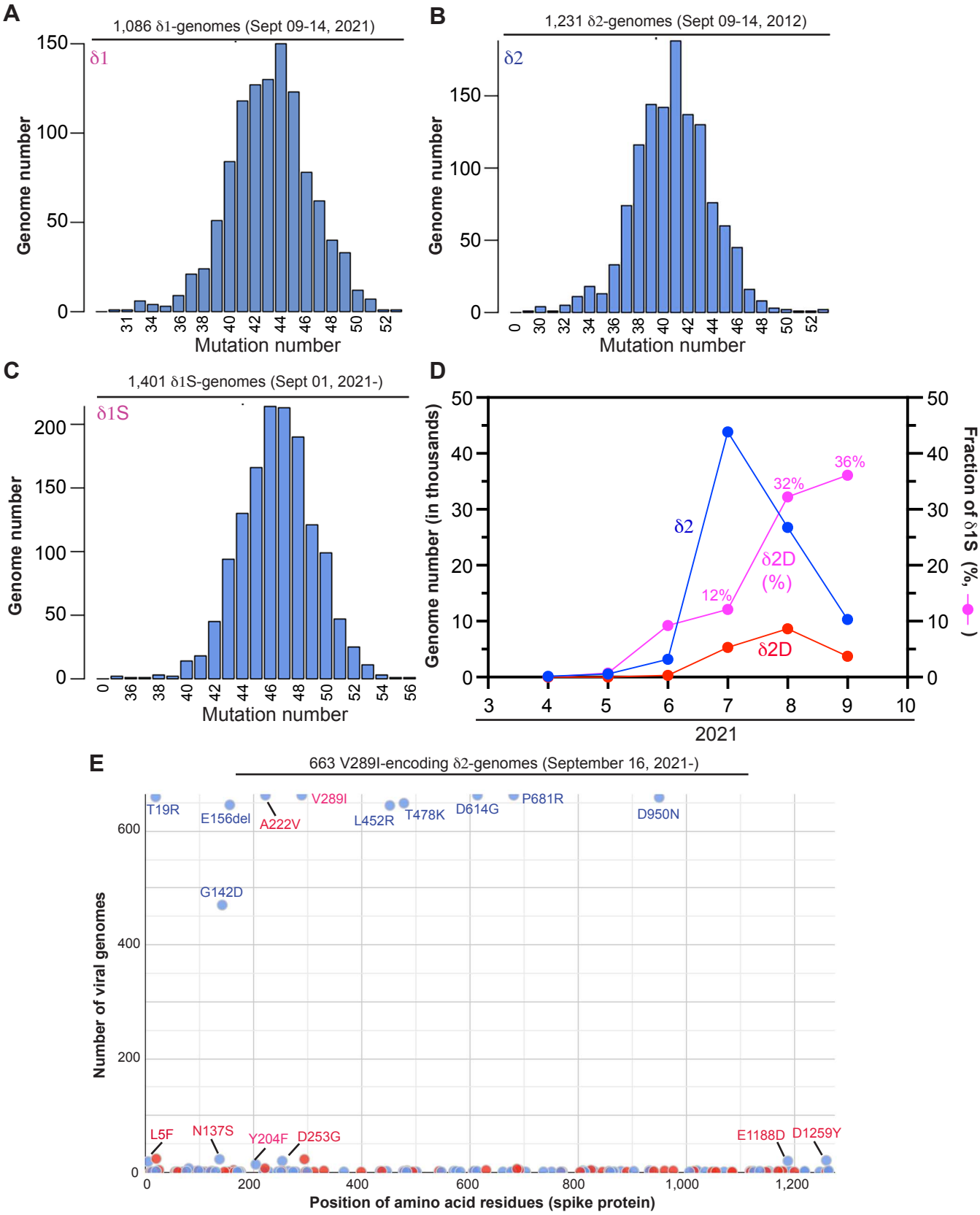

A
1,086 δ1-genomes (Sept 09-14, 2021)
δ1
Genome number
Mutation number
B
1,231 δ2-genomes (Sept 09-14, 2012)
δ2
Genome number
Mutation number
C
1,401 δ1S-genomes (Sept 01, 2021-)
δ1S
Genome number
Mutation number
D
Genome number (in thousands)
Fraction of δ1S (%)
δ2
δ2D
δ2D (%)
12%
32%
36%
2021
E
663 V289I-encoding δ2-genomes (September 16, 2021-)
Number of viral genomes
Position of amino acid residues (spike protein)
T19R
E156del
A222V
V289I
L452R
T478K
D614G
P681R
D950N
G142D
L5F
N137S
Y204F
D253G
E1188D
D1259Y

Figure 7

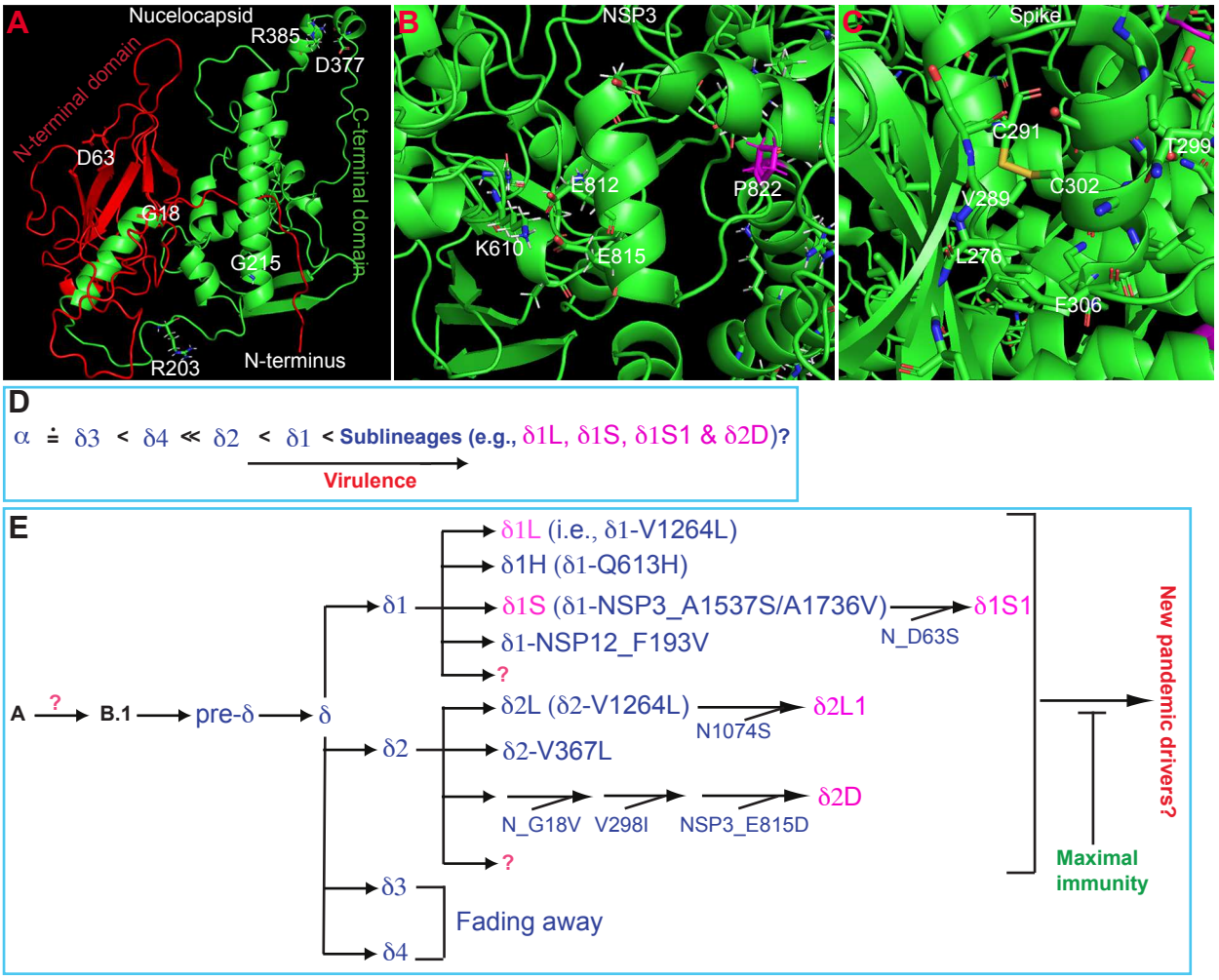

A
Nucelocapsid
R385
D377
N-terminal domain
C-terminal domain
D63
G18
G215
R203
N-terminus
B
NSP3
E812
P822
K610
E815
C
Spike
C291
T299
C302
V289
L276
E306
D
α ≐ δ3 < δ4 ≪ δ2 < δ1 < Sublineages (e.g., δ1L, δ1S, δ1S1 & δ2D)?
Virulence
E
A
?
B.1
pre-δ
δ
δ1
δ2
δ3
δ4
δ1L (i.e., δ1-V1264L)
δ1H (δ1-Q613H)
δ1S (δ1-NSP3_A1537S/A1736V)
δ1-NSP12_F193V
?
δ1S1
N_D63S
δ2L (δ2-V1264L)
δ2L1
N1074S
δ2-V367L
δ2D
N_G18V
V298I
NSP3_E815D
?
Fading away
New pandemic drivers?
Maximal immunity

Figure S1

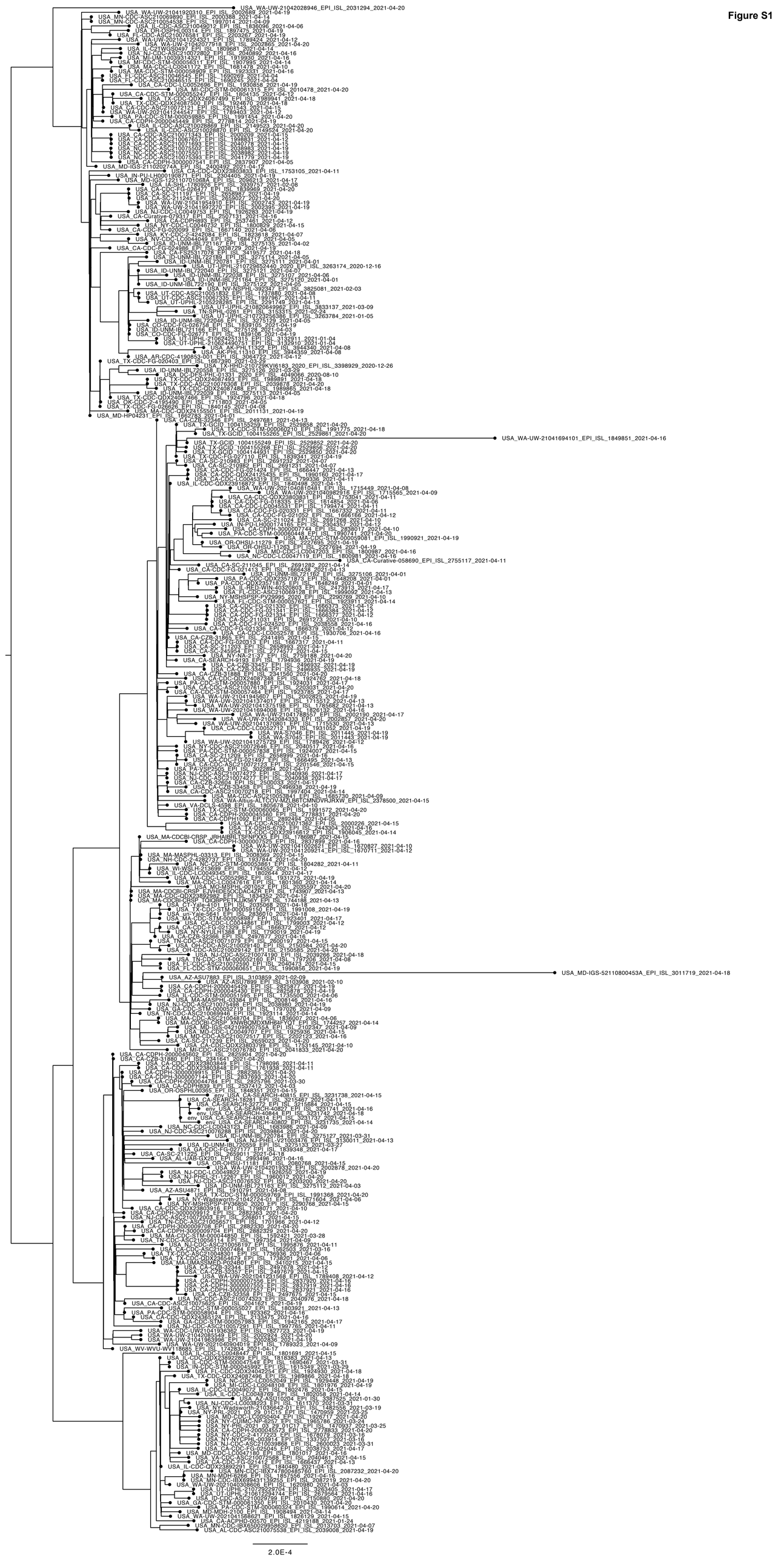
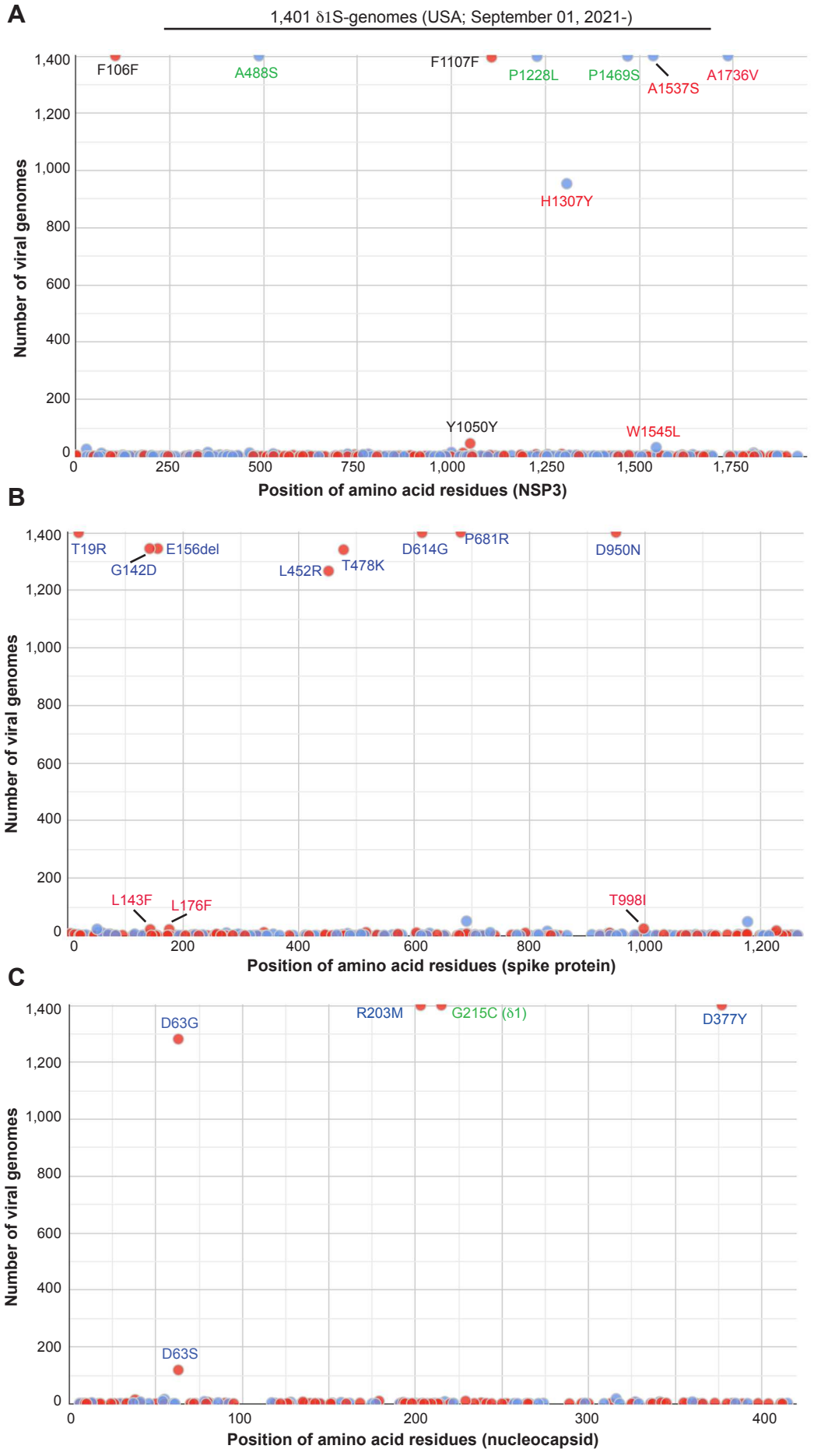

USA_WA-UW-21042028946_EPI_ISL_2031294_2021-04-20
USA_WA-UW-21041694101_EPI_ISL_1849851_2021-04-16
USA_MD-IGS-52110800453A_EPI_ISL_3011719_2021-04-18
2.0E-4

Figure S2


A
1,401 δ1S-genomes (USA; September 01, 2021-)
Number of viral genomes
Position of amino acid residues (NSP3)
F106F
A488S
F1107F
P1228L
P1469S
A1537S
A1736V
H1307Y
Y1050Y
W1545L
B
Number of viral genomes
Position of amino acid residues (spike protein)
T19R
G142D
E156del
L452R
T478K
D614G
P681R
D950N
L143F
L176F
T998I
C
Number of viral genomes
Position of amino acid residues (nucleocapsid)
D63G
R203M
G215C (δ1)
D377Y
D63S

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1USdeltaApril302021902.pdf
- TableS2USdeltaApril202021356.pdf
- TableS3USD1Sept061920211086.pdf
- TableS4USD2Sept01192021657.pdf
- TableS5Sept16Oct092021663.pdf
- TableS6USSept01Oct0920211401.pdf