Delta-1 variant of SARS-COV-2 acquires spike V1264L and drives the pandemic in Indonesia, Singapore and Malaysia

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

Rosalind and Morris Goodman Cancer Institute, Department of Medicine, McGill University, Montreal, Quebec, Canada

Research Article

Keywords: Coronavirus, virus entry, B.1.1.7, B.1.351, P.1, B.1.617.2, L452R, E484Q, P681R, endocytosis, AP2 receptor, clathrin, cytoplasmic tail

Posted Date: October 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-999390/v1

δ1 variant of SARS-COV-2 acquires spike V1264L and drives the pandemic in Indonesia, Singapore and Malaysia

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

Since April 2021, δ variant of SARS-COV-2 has gradually overtaken all other variants and become a dominant pandemic driver around the world. It has evolved and yielded four subvariants: δ1, δ2, δ3 and δ4. While trying to understand how these subvariants drive the pandemic in Southeast Asia, I noticed that many δ1 genomes from Indonesia encode an extra spike substitution, V1264L. Coincidentally, this confers an acidic dileucine motif because residues 1157-1262 are acidic and residue 1265 is leucine. Such a motif may affect subcellular trafficking of the resulting spike protein. Alarmingly, this V1264L-encoding δ1 subvariant (referred to as δ1L) has become the dominant pandemic driver in Indonesia, Singapore, Malaysia and East Timor. Moreover, it has acquired additional spike substitutions: L1234L in Singapore and D215Y/N in Malaysia. On the average, the resulting sublineages carry 46-48 mutations per genome, making them some of the most mutated variants identified so far. Moreover, a δ1 sublineage from the United Kingdom has acquired V1264L along with spike Y145H and A222V, a signature substitution of a SARS-COV-2 clade that was a major pandemic driver in Europe during the summer of 2020. A222V improves an extensive hydrophobic interaction network at the N-terminal domain of spike protein and may make this sublineage more virulent than δ1 and δ1L. Some δ2 subvariant genomes identified in the United States of America and other countries also encode V1264L. Thus, V1264L is a recurrent spike substitution frequently acquired by δ subvariants during convergent evolution. This recurrence also suggests that V1264L is one key mechanism by which δ variant adopts to expand its 'evolutionary cage.'

Running Title: Evolutionary cage of SARS-COV-2
Keywords: Coronavirus, virus entry, B.1.1.7, B.1.351, P.1, B.1.617.2, L452R, E484Q, P681R, endocytosis, AP2 receptor, clathrin, cytoplasmic tail

INTRODUCTION

Coronavirus disease 2019 (COVID-19) has caused the pandemic for almost two years, which has not only led to tragic loss of life but also crippled the economy around the world. Although successful development of multiple vaccines has brought great hope, we still do not know very little about where this pandemic is heading and when it will end. Thus, it is important to fully understand severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the culprit virus that causes this contagious disease [1]. Genomic surveillance has revealed that its evolution is key to driving waves of cases around the world. For genomic surveillance, the global initiative on sharing avian influenza data (GISAID) has played a critical role and as of October 10, 2021, over 4.25 million SARS-COV-2 genomes have been deposited into the database [2]. From the genomes, it is clear that the virus has yielded many variants, including four variants of concern as designated by the WHO: α (B.1.1.7) [3], β (B.1.351) [4], γ (P.1) [5] and δ (B.1.617.2) [6]. Astonishingly, these variants all emerged from complete obscurity and then rapidly rose to the status of major pandemic drivers. This raises two important questions: 1) how they have been evolving and 2) what future potential variants of concern will look like.

Related to these two questions, I wondered whether there are any simple rules that govern evolution of this virus. If so, such rules will help us predict its evolutionary trajectory and identify potential future variants of concern. Obviously, as the genetic material, the ~29,700-nucleotide viral genome is the key determinant. In theory, every nucleotide of this genome could be altered during SARS-COV-2 evolution. In reality, many such changes, however, are non-functional or even detrimental and thus offer no evolutionary advantage. As a result, just like what occurs with cancer mutations, only favorable or gain-of-function ones are selected during SARS-COV-2 evolution. In cancer biology, mutations are grouped into two categories, known as ‘driver’ and ‘facilitator’ mutations. A third category is ‘passenger’ mutations because they are apparently selected due to association with the first two categories. One thumb of rule is that ‘driver’ and ‘facilitator’ mutations are recurrent whereas ‘passenger’ mutations are often not. I reasoned that such concepts can be used to understand and annotate SARS-COV-2 mutations. I also envisaged that there are genomic elements limiting SARS-COV-2 evolution and such elements form an ‘evolutionary cage’ demarcating the space within or around which SARS-COV-2 evolves and generates new variants. I believe that the recurrent nature of ‘driver’ and ‘facilitator’ mutations will help identify the genomic elements and map out the shape of this invisible cage.

With these considerations as the simple guide, I have tracked genomes in the GISAID database closely. As a result, 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 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] and

the USA [8], with δ1 becoming almost the sole pandemic driver in the United Kingdom and Spain. Thus, a relevant question is how the four subvariants have contributed to the pandemic in other continents and countries around the world.

Related to this question, I tried to understand how the subvariants have driven the pandemic in Southeast Asia. I noticed that many δ1 genomes from Indonesia encode an extra spike substitution, V1264L, along with an extra NSP3 substitution (T678I). Alarmingly, this δ1 sublineage, δ1L, has become the predominant pandemic driver in Indonesia, Singapore, Malaysia and East Timor. Mechanistically, V1264L confers an acidic dileucine motif. This is significant as such motifs are known to regulate endocytosis and subcellular trafficking of membrane-associated receptors and viral envelope proteins [9-12]. Here, I describe how δ1L has evolved and driven the pandemic in Southeast Asia, and demonstrate that V1264L has also been acquired by δ1 and δ2 subvariants identified in the United Kingdom and USA, respectovely.

RESULTS AND DISCUSSION

δ1 and δ2 subvariants drive the pandemic in Indonesia

The epidemiological curve from Indonesia indicates that the country managed relatively well in 2020 (Fig. 1A). The first wave started at the end of 2020. The variant B.1.466.2 accounts for a third of COVID-19 cases from the first wave (Fig. 1A). This variant carries one small deletion (H69V70del) and four spike substitutions, N439K, G614G, P681R and P809S. After April 2021, δ variant has gradually become the predominant driver (Fig. 1A-B). In support for high virulence of δ variant, the second wave is much more powerful than the first one. Vaccination was initiated at the beginning of 2021 and reached a full vaccination rate of ~17% in September 2021 (Fig. 1A). Temporal distribution of δ variant and its subvariants, δ1 and δ2, indicated that δ1 is the dominant driver whereas δ2 has played a less important role (Fig. 1B). This is similar to what has occurred in India, Europe and the USA [7,8]. Then I utilized Coronapp [13,14] for mutation profiling δ genomes identified in Indonesia after July 16, 2021. The results showed that a majority of the genomes encode two extra substitutions, spike V1264L (Fig. 1C) and NSP3 T678I (Fig. 1D). Importantly, these two substitutions are associated with each other and belong to a δ1 sublineage, which is referred to δ1L (with the letter L denoting L1264). Thus, δ1 has evolved and yielded a new sublineage encoding two extra substitutions. Consistent with this, the average mutation load in the δ-genomes analyzed is 42 per genome (Fig. 2A), which is 1-2 mutations more than δ1 genomes identified in India after August 01, 2021 [7].

As for the origin of δ1L, the first such genome from Indonesia was identified on January 19, 2021 (EPI_ISL_3315002, GISAID). It is quite puzzling is that the major δ1L wave only took off in the country 4-5 months later (Fig. 1A). In India, the first δ1L genome was only reported on April 06, 2021, so it is unlikely that δ1 gave rise to δ1L in India. Multiple genomes

with the collection dates of January 08 and 09, 2021 were deposited into the GISAID database from East Timor and Slovakia in October and September 2021, respectively. Out of 2,257 δ genomes from Slovakia, there are only 12 δ1L genomes (GISAID, accessed on October 06, 2021). One complication is that some early genomes in the database appear to have incorrect sample collection date information, perhaps due to typos and book-keeping errors. The geological proximity of East Timor to Indonesia suggests that δ1 may have yielded δ1L in one of these two countries. But this possibility will need to be verified.

δ1L is a major pandemic driver and acquires spike L1234I in Singapore

During the pandemic, Singapore has taken strict public health measures. Such measures have been highly effective in blocking α [3], β [4], γ [5] and many other variants from causing outbreaks in the country from September 2020 through June 2021 (Fig. 1B). In the summer of 2021, the four subvariants (δ1, δ2, δ3 and δ4) led to some small outbreaks in the country, but it managed to control them successfully. By contrast, these subvariants have led to major waves in the rest of the world. The situation took a sharp turn in Singapore at the beginning of August 2021, followed by a surging wave that is still ongoing (Fig. 1B). This epidemiological prognosis raises an important question about the SARS-COV-2 variants that have been driving this surging wave. Troublingly, this surging wave occurred when the country was reaching a high level of vaccination (Fig. 2B). Since April 2021, the vaccination rate has increased rapidly to an impressive level of ~80% (Fig. 1B). Other public control measures have been similar during the pandemic in the country. Thus, this raises an intriguing possibility that the variant(s) behind this surging wave may be much more virulent than many other variants that had spread to the country before the summer of 2020. If so, it would be important to identify the variant(s).

As described above (Fig. 1), during the summer of 2021, δ1L was a key driver of a surging wave in Indonesia, a neighboring country of Singapore. The geological proximity suggests that δ1L may also be one major driver of the surging wave in Singapore. Indeed, as shown in Fig. 2B, δ1L was a major variant in July 2021 and became almost the sole pandemic driver one month later. In support of this, δ1L genomes were identified in almost all 828 COVID-19 cases analyzed (Fig. 2C). Unexpectedly, almost 40% genomes from September belong to a sublineage (referred to as δ1L1) that encodes an extra spike substitution (L1234I) and a silent mutation (P217P, Fig. 2C). Consistent with this, the average mutation load of the genomes from Singapore is 46 per genome, which is 4 mutations more than what was identified in the genomes from Indonesia (Fig. 2A). Alarmingly, since July 2021, δ1L1 genomes have rapidly risen to the current level of almost 40% (Fig. 2C). Thus, spike L1234I may confer evolutionary advantage to δ1L1. Related to this, L1234 is in at the border of the transmembrane domain and cytoplasmic tail. Coincidentally, this is close to V1264, so L1234I may synergize

with V1264L. Although L1234I is a rather conserved substitution at the physicochemical level, it may finetune spike protein function and regulation.

In Singapore, the first δ1L1 case was identified on August 22, 2021. Over one month earlier, a related case (EPI_ISL_4254882, GISAID) was identified in Indonesia on July 03, 2021, suggesting that δ1L1 might have arisen in Indonesia and then spread to Singapore. Now there are 884 genomes in the GISAID database, with 881 from Singapore and the remaining three from Indonesia, Australia and Hong Kong. Thus, δ1L1 remains largely localized to Singapore. This offers a narrow opportunity to control the spread of this new variant.

δ1L drives the pandemic and acquires spike D215Y/N in Malaysia

Having established that δ1L is a major pandemic driver in Indonesia and Singapore, I next investigated whether this variant has spread to neighboring countries and driven the pandemic there. To address this, I analyzed the situation in Malaysia. As shown in Fig. 3A, the country has experienced three waves of COVID-19 cases, with the latest one starting in July 2021 and being the most powerful. Epidemiological curve and immunization progress in Malaysia. Vaccination progress has been reached an impressive level compared to many other countries in Asia (Fig. 3A). Despite this vaccination progress, the third wave is much more powerful than the first two. As occurred in Singapore, this epidemiological prognosis suggests that the variants behind the third wave may be highly virulent. The geological proximity to Indonesia and Singapore suggests that δ1L may be one key driver of the third wave in Malaysia. Indeed, δ1L has been a major driver since July 2021 (Fig. 3B). It was detected in ~50% cases in August 2021 (Fig. 3B). Thus, δ1L has contributed significantly to the latest surging wave of COVID-19 cases in Malaysia.

To understand how δ1L has evolved in the country, I performed mutation profiling via Coronapp [13,14]. As shown in Fig. 3C-D, multiple new spike and NSP3 mutations have been detected in 5-20% of the 446 δ1L-genomes analyzed, suggesting that δ1L have given rise to different sublineages in Malaysia. Consistent with this, the mutation load is 45-46 per genome (Fig. 4A), which is similar to what observed with the genomes from Singapore (Fig. 2D) but 3-4 mutations more than that observed with the genomes from Indonesia (Fig. 2A). This suggests that there are extra mutations in the δ1L genomes from Malaysia. Mutation profiling uncovered two notable substitutions are D215Y and D215N, which alters the same residue, D215. It should be noted, however, that due to its low abundance (~10% each, Fig. 2C), these two substitutions do not account fully for the higher mutation load (Fig. 4A) than those genomes identified in Indonesia (Fig. 2A). These two substitutions could be used to identify two subgroups of genomes. Analysis of these two subgroups revealed average mutations loads of 48 and 46-47 per genomes for the D215Y- and D215N-encoding subgroups, respectively (Fig. 4B-C). Such mutation loads make both sublineages two of the most mutated SARS-COV-

2 variants identified so far. Related to this, the most mutated variant is 49 mutations per genome [7]. While D215Y is associated with spike A845S in some genomes, D215N-encoding genomes carry a small deletion spanning the codons of H69 and V70 (Fig. 4D-E). Therefore, δ1L has yielded two highly mutated sublineages in Malaysia.

V1264L-encoding δ genomes identified around the world form different groups

Having established that δ1L has been a major pandemic driver in Indonesia, Singapore and Malaysia, I asked how it has spread to other countries. Related to this, δ1L appears to be a major driver in East Timor: out of 33 δ-genomes identified there, 17 encode δ1L (GISAID, accessed on October 05, 2021). Thus, δ1L is a key pandemic driver in East Timor.

For the rest of the world, I searched δ1L genomes the GISAID database (accessed on October 07, 2021). For this, T678I of NSP3 was also used to identify δ1L variant originated from Southeast Asia. This is because in the United Kingdom, there is a δ1L variant that encodes A1711V instead of T678I of NSP3 (see below). This research identified 1,554 δ1L-T678I genomes from Europe. There are 129 such genomes identified in North America, with 19 in Canada (16 of them from British Columbia). In Asia, 200, 118 and 2 such genomes have been identified in Japan, South Korea and China, respectively. None have been identified in South America and only one has been found in Africa. Therefore, the spread of δ1L-T678I around the world is still at the very early stage. But this may change soon, so it will be necessary to watch out the spread of this variant due to its higher virulence than δ1 itself.

Then I carried put phylogenetic analysis of δ1L genomes identified in the world by May 31, 2021. As shown in Fig. 5, δ1L genomes from Indonesia form one large group. Within this group are only one genome identified in Singapore and several from the U.K. Strikingly, the genomes from the U.K. encode A1711V instead of T678I of NSP3, suggesting that they are from a distinct root. A majority of δ1 genomes identified in country encode A1711V [7], so δ1 subvariant has acquired V1264L in the U.K., independently from δ1L identified in Indonesia, Singapore and Malaysia. Thus, there are two distinct groups of δ1L genomes, encoding T678I or A1711V of NSP3. Intriguingly, phylogenetic analysis revealed that both δ2 and δ3 subvariants have also acquired spike V1264L, yielding new δ2L and δ3L sublineages, respectively (Fig. 5). Thus, V1264L is a recurrent spike substitution that has been acquired by different δ subvariants as they evolve to yield new sublineages.

A spike A222V-encoding δ1 sublineage acquires spike V1264L in the U.K.

Then I utilized Coronapp [13,14] to analyze the mutation profile of δ1L genomes identified in the U.K. Shown in Fig. 6A are spike substitutions encoded by 963 V1264L-encoding δ genomes identified in the country during the first weeks of September 2021. It is surprising to notice that in addition to typical δ1L substitutions, about 80% of the genomes carry two extra

spike substitutions, Y145H and A222V (Fig. 6A). In the country, there is a δ1 sublineage, δ1V, that encodes these two substitutions, but not V1264L [15]. Thus, it is likely that δ1V has acquired V1264L and given rise to a new sublineage, which has been referred to as δ1V1 [15].

Moreover, almost all genomes carry A1711V of NSP3 (Fig. 6B), with an average mutation load of 43-44 per genome (Fig. 6C). They do not encode T678I (Fig. 6C) as those δ1L genomes identified in Southeast Asia (Fig. 1D). A1711V is a signature substitution of δ1 subvariant identified in the U.K. [15], so it is conceivable that δ1-NSP3_A1711V has acquired V1264L in the U.K. in a manner completely independent of δ1L identified in Southeast Asia.

Shown in Fig. 6D are the monthly genomes corresponding to δ1V1 and the total V1264L-encoding δ genomes (i.e. the sum of δ1L and δ1V1). Since July 2021, the δ1V1 genome number has increased exponentially, suggesting that Y145H, A222V and V1264L confer evolutionary advantage to δ1V1. A222V is a signature substitution of the SARS-COV-2 GV clade that was a major pandemic driver in Europe during 2020. A222V improves an extensive hydrophobic interaction network [15]. Thus, δ1V1 may be even more virulent than δ1L.

Analysis of V1264L-encoding δ2-genomes

Like δ1V1, δ2 genomes encode spike A222V [15]. Phylogenetic analysis identified two large groups of δ2 genomes encoding V1264L (Fig. 5). One such group contains genomes mainly from the USA, and the other possesses genomes from Europe and Africa. By analogy to δ1L, δ2 V1264L-encoding δ2 subvariant is referred to as δ2L. Mutation profiling of 1,047 δ2L-genomes identified in the USA during September 2021 revealed an extra spike substitution N1074S (Fig. 7A). As shown in Fig. 7B, N1074S-encoding δ2L genomes carry an average of 40-41 mutations per genome. This is much lower than what is observed with different δ1L sublineages (e.g. Figs 2B & 4B-C). As shown in Fig. 7C, N1074S does confer viral fitness to δ2L. As described previously, δ2 is typically less virulent than δ1 [15]. Indeed, the growth of δ2L or its N1074S-encoding sublineage is less dramatic than δ1L and its sublineages. Thus, δ2L or its N1074S-encoding sublineage may be less dangerous than δ1L and its sublineages described above.

Mechanistic impact of V1264L and other spike substitutions

As shown in Fig. 8A-B, V1264 is located at the C-terminal cytoplasmic tail, which is important for cytoplasmic trafficking [16,17]. A recent study showed that the cytoplasmic tail of spike protein facilitates its expression on cell surface and promotes syncytia formation between infected cells [18]. Coincidentally, V1264L confers an acidic dileucine motif to spike protein because residues 1157-1262 are acidic and residue 1265 is leucine. Such a motif may affect endocytosis of many membrane-associated receptors of the resulting protein [9,10,12].

Moreover, a conserved dileucine motif mediates clathrin- and AP2-dependent endocytosis of the HIV-1 envelope protein [11]. Thus, V1264L may finetune the function of spike protein.

Notably, L1234I encoded in δ1L is located at the border between the transmembrane domain and the cytoplasmic tail (Fig. 8A). L1234I may synergize with V1264L in finetuning the function and regulation of spike protein. H69 and Y145 are within a super antigenic site in the N-terminal domain (Fig. 8A), so their alterations (H69del and Y145H, respectively; Fig. 4) may confer immune evasions. A222 is part of an extensive hydrophobic network (Fig. 8D) [7], so A222V should improve this network. Spike A845 is in proximity to Q836 (Fig. 8E). Both residues are part of a region C-terminal from the fusion peptide (Fig. 8A). This region is important for spike function in promoting cell-cell fusion [19].

The cartoon in Fig. 8F summarizes evolutionary relationship between δ variant and its subvariants described herein. Two other studies of δ variant genomes from India, Europe and the U.S.A. supports a model in which this variant yields mainly four subvariants [7,8]. Alarmingly, δ1 sublineages such as δ1L, δ1L1, δ1V and δ1V1 appear to be even more virulent than δ1 itself. Moreover, δ1L1 and δ1V1 may be more virulent than δ1L. This is alarming. Therefore, we will need to track evolution of these sublineages very closely and map their evolutionary trajectory in a precise and timely fashion, which will help us adopt the best public health measures and develop the new generation of vaccines.

Identification of δ1L and its sublineages herein also provides support for a continuously branching model on SARS-COV-2 evolution [7,8]. Moreover, according to the 'evolutionary cage' hypothesis, there are genomic elements limiting the evolution and such elements form an 'evolutionary cage' that demarcate the space within which SARS-COV-2 evolves and generates new variants. Thus, the recurrent nature of spike V1264L suggests that V1264 is one such element and not optimal for SARS-COV-2. If so, V1264L is a driver or facilitator mutation that optimizes the function of the cytoplasmic tail of spike protein. Of relevance, spike P681R, P681H and V1176F are also such types of mutations [20,21]. Thus, acquisition of V1264L is an important mechanism by which δ1 subvariant expands its evolutionary cage. Because V1264L is expected to finetune the cytoplasmic tail of spike protein and modulate its role in endocytosis, subcellular trafficking and cell-cell fusion, these results highlight the need to characterize other substitutions that also alter the cytoplasmic tail of spike protein.

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 has completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: He have received grants from Canadian Institutes of Health Research (CIHR) and Natural Sciences and Engineering Research Council of Canada (NSERC) for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

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. The 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 [13,14]. 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 [22] was used for phylogenetic as described [7].

Defining different variant genomes using various mutation markers

Genomes of α, β, γ, δ and other SARS-COV-2 variants were downloaded from the GISAID database as defined by the server. δ subvariant genomes were defined as described [7]. Briefly, the nucleocapsid substitutions G215C and R385K were used as mutation 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. Elbe, S. and G. Buckland-Merrett. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1**, 33-46.
3. 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.
4. 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.
5. Faria, N.R., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in

Manaus, Brazil. *Science* **372**, 815-821.
6. Mlcochova, P., et al. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*, Epub.
7. Yang, X.J. (2021). SARS-COV-2 delta variant drives the pandemic in India and Europe through two subvariants. *medRxiv (submitted)*.
8. 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.
9. Doray, B., J.M. Knisely, L. Wartman, G. Bu, and S. Kornfeld. (2008). Identification of acidic dileucine signals in LRP9 that interact with both GGAs and AP-1/AP-2. *Traffic* **9**, 1551-62.
10. Kelly, B.T., A.J. McCoy, K. Spate, S.E. Miller, P.R. Evans, S. Honing, and D.J. Owen. (2008). A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. *Nature* **456**, 976-979.
11. Byland, R., P.J. Vance, J.A. Hoxie, and M. Marsh. (2007). A conserved dileucine motif mediates clathrin and AP-2-dependent endocytosis of the HIV-1 envelope protein. *Mol Biol Cell* **18**, 414-25.
12. Tikkanen, R., S. Obermuller, K. Denzer, R. Pungitore, H.J. Geuze, K. von Figura, and S. Honing. (2000). The dileucine motif within the tail of MPR46 is required for sorting of the receptor in endosomes. *Traffic* **1**, 631-40.
13. 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.
14. Mercatelli, D. and F.M. Giorgi. (2020). Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front Microbiol* **11**, 1800.
15. Zhang, J., T. Xiao, Y. Cai, C.L. Lavine, H. Peng, H. Zhu, K. Anand, P. Tong, A. Gautam, M.L. Mayer, R.M. Walsh, S. Rits-Volloch, D.R. Wesemann, W. Yang, M.S. Seaman, J. Lu, and B. Chen. (2021). Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *bioRxiv*, https://www.biorxiv.org/content/10.1101/2021.08.17.456689v1.
16. Sadasivan, J., M. Singh, and J.D. Sarma. (2017). Cytoplasmic tail of coronavirus spike protein has intracellular targeting signals. *J Biosci* **42**, 231-244.
17. Jennings, B.C., S. Kornfeld, and B. Doray. (2021). A weak COPI binding motif in the cytoplasmic tail of SARS-CoV-2 spike glycoprotein is necessary for its cleavage, glycosylation, and localization. *FEBS Lett* **595**, 1758-1767.
18. Cattin-Ortola, J., L.G. Welch, S.L. Maslen, G. Papa, L.C. James, and S. Munro. (2021). Sequences in the cytoplasmic tail of SARS-CoV-2 Spike facilitate expression at the cell surface and syncytia formation. *Nat Commun* **12**, 5333.

19. 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.
20. 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.
21. 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.
22. 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. SARS-COV-2 evolution drives the pandemic in Indonesia. (**A**) Epidemiological curve and immunization progress in Indonesia. While B.1.466.2 is a variant that accounted for a third of cases during the first wave in the country, δ variant is the predominant driver of the second wave. Despite up to ~17% full vaccination, the second wave is much more powerful than the first one, supporting high virulence of δ variant. For preparation of this panel, the Our World in Data website (https://ourworldindata.org/) was accessed on September 26, 2021. (**B**) Monthly distribution of δ variant and its subvariants. δ1 and δ2 subvariants are defined as in two other studies [7,8]. δ1L is a δ1 sublineage that encodes an extra spike substitution, V1264L. For the analysis, the GISAID SARS-COV-2 genome sequence database was accessed on September 27, 2021. (**C-D**) Mutation profile of 904 SARS-COV-2 genomes identified in Indonesia after July 15, 2021. The genomes were downloaded from the GISAID database on September 27, 2021 for mutation profiling via Coronapp [13,14]. Shown here are spike (C) and NSP3 substitutions (D).

Figure 2 Mutation profiles of SARS-COV-2 genomes identified in Indonesia and Singapore. (**A**) Mutation load in 904 SARS-COV-2 genomes identified in Indonesia after July 15, 2021. The analysis was carried out with Coronapp as Fig. 1C-D, but the mutation load is shown here. (**B**) Epidemiological curve and vaccination progress in Singapore. Since April 2021, different δ subvariants have been detected in imported and local COVID-19 cases, but δ1L has been the main pandemic driver since July 2021 (see panel B). Alarmingly, although Singapore is now one of the top vaccinated countries (reaching almost 80% of the population), the current wave is the worst that the country has experienced since the beginning of the pandemic. This suggests a potential high breakthrough rate by δ1L subvariant. For preparation of this panel, the Our World in Data website was accessed on September 26, 2021. (**B**) Monthly distribution of δ variant and its subvariants. For the analysis, the GISAID website was accessed on September 26, 2021. δ1L1 is a new δ1L sublineage that encodes spike L1234I (see panel E). (**D-E**) Mutational profiling of SARS-COV-2 genomes from 828 COVID-19 cases identified in Singapore in September 2021. The genomes were downloaded from the GISAID database on September 22, 2021 for mutation profiling via Coronapp. Shown in (B) is the mutation load among the genomes, whereas shown in (D) and (E) are substitutions in spike and NSP3 proteins, respectively. The average mutation load of δ1L1 is 46 per genome.

Figure 3. Mutation profile of δ-genomes identified in Malaysia. (**A**) Epidemiological curve and immunization progress in Malaysia. While AU.2 and B.1.524 drove the first wave in the country, δ variant has been the predominant driver of the latest wave that started in the summer of 2021. Despite some vaccination, the third wave is much more powerful than the first two.

For preparation of this panel, the Our World in Data website was accessed on September 27, 2021. (**B**) Monthly distribution of δ variant and its δ1L subvariant. For the analysis, the GISAID website was accessed on September 27, 2021. Note incomplete data (indicated with an asterisk) for September 2021 when the website was accessed (**C-D**) Mutation profile of SARS-COV-2 genomes from 446 δ1L-genomes identified in Malaysia. The genomes were downloaded from the GISAID SARS-COV-2 sequence database on September 27, 2021 for mutation profiling via Coronapp [13,14]. Shown here are spike (C) and nucleocapsid substitutions (D). Note that a small deletion (H69V70del) is only detected as H69del by Coronapp.

Figure 4 Mutation profiles of δ1L-genomes identified in Malaysia. (**A**) Mutation load in 446 δ1L genomes identified in the country. The distribution was generated via Coronapp as Fig. 3C-D, but the mutation load is shown here. (**B-D**) Mutational profiling of δ1L genomes that also encode spike D215Y (B & D) or D215N (C & E). The genomes were downloaded from the GISAID database on September 27, 2021 for mutation profiling via Coronapp. Shown in (B) is the mutation load among the genomes, whereas shown in (B-C) and (D-E) are the mutation loads and spike substitutions, respectively. Note that a small deletion (corresponding to H69V70del) is only detected as H69del by Coronapp.

Figure 5. Phylogenetic analysis of 473 δ1L-genomes identified around the world by May 31, 2021. The genomes were downloaded from the GISAID database on September 26, 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.

Figure 6. Mutation profile of δ1L-genomes identified in the U.K. (**A-B**) Mutation profile of 963 δ1L-genomes identified in the country in September 2021. The genomes were downloaded from the GISAID SARS-COV-2 sequence database on September 27, 2021 for mutation profiling via Coronapp [13,14]. Shown are spike (A) and NSP3 (B) substitutions. δ1L from Indonesia encodes T678I of NSP3 (Fig. 1D), but δ1L originated from the U.K. encodes A1711V of NSP3 (panel B). δ1V1 carries spike Y145H, A222V and V1264L [7]. Its parental strain is δ1V, encoding spike Y145H and A222V [7], indicating that spike V1264L has been independently acquired by different variants. (**C**) Mutation load in δ1L-genomes identified in the U.K. The analysis was performed via Coronapp as panels A-B, but the mutation load is shown here. δ1V1 genomes carry an average of 43-44 mutations per genome. (**D**) Temporal distribution of δ variant and its δ1L subvariant. For the analysis, the GISAID website was

accessed on September 27, 2021. δ1L3 encodes NSP3 and spike V1264L, whereas δ1L4 encodes an additional spike substitution, A222V.

Figure 7. Mutation profile of δ1L-genomes identified in the USA. (**A-B**) Mutation profile of 1,047 V1264L-encoding δ-genomes identified in the country in September 2021. The genomes were downloaded from the GISAID database on September 27, 2021 for mutation profiling via Coronapp [13,14]. Shown in (A) and (B) are the mutation loads and spike substitutions, respectively. (**C**) Monthly distribution of δ2L subvariant and its N1074S-encoding sublineage. For the analysis, the GISAID website was accessed on September 28, 2021.

Figure 8. Mechanistic impact of V1264L and other substitutions. (**A**) Domain organization of spike protein. Some key substitutions of δ variant are shown in black and new substitutions 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 C-terminal 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. Adapted from a published study [19]. (**B**) Sequence alignment of spike proteins from SARS-COV-2 and five other coronaviruses. Only the cytoplasmic tail is shown along with a portion of the transmembrane domain. (**C-D**) Structural details of spike H69, D215, Y145 and A222. H69 and Y145 are part of a super antigenic site in the N-terminal domain (Fig. 8A). A222 is part of a hydrophobic core [7]. (**E**) Structural details of spike A845 and a neighboring residue, Q836. The structural models (C-E) were adapted from PyMol presentation of the spike protein structure 6XR8 from the PDB database. (**E**) Cartoon summarizing evolutionary relationship between δ variant and its subvariants described herein. Two other studies of δ variant genomes from India, Europe and the USA supports a similar model in which this variant branches out and yields mainly four subvariants [7,8].

SUPPLEMENTAL INFORMATION

This section includes one supplementary figure with detailed information on a phylogenetic tree and six acknowledgement tables for the GISAID genomes used in this work.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Phylogenetic analysis of 473 δ1L-genomes identified around the world by May 30, 2021. The genomes were downloaded from the GISAID SARS-COV-2 genome sequence database on September 26, 2021 for phylogenetic analysis. Only high-coverage genomes with complete date information on sample collection were used. The software package RAxML-NG was used to generate 20 maximum likelihood trees and the bestTree for presentation via Figtree as in Fig. 5.

Figure 1

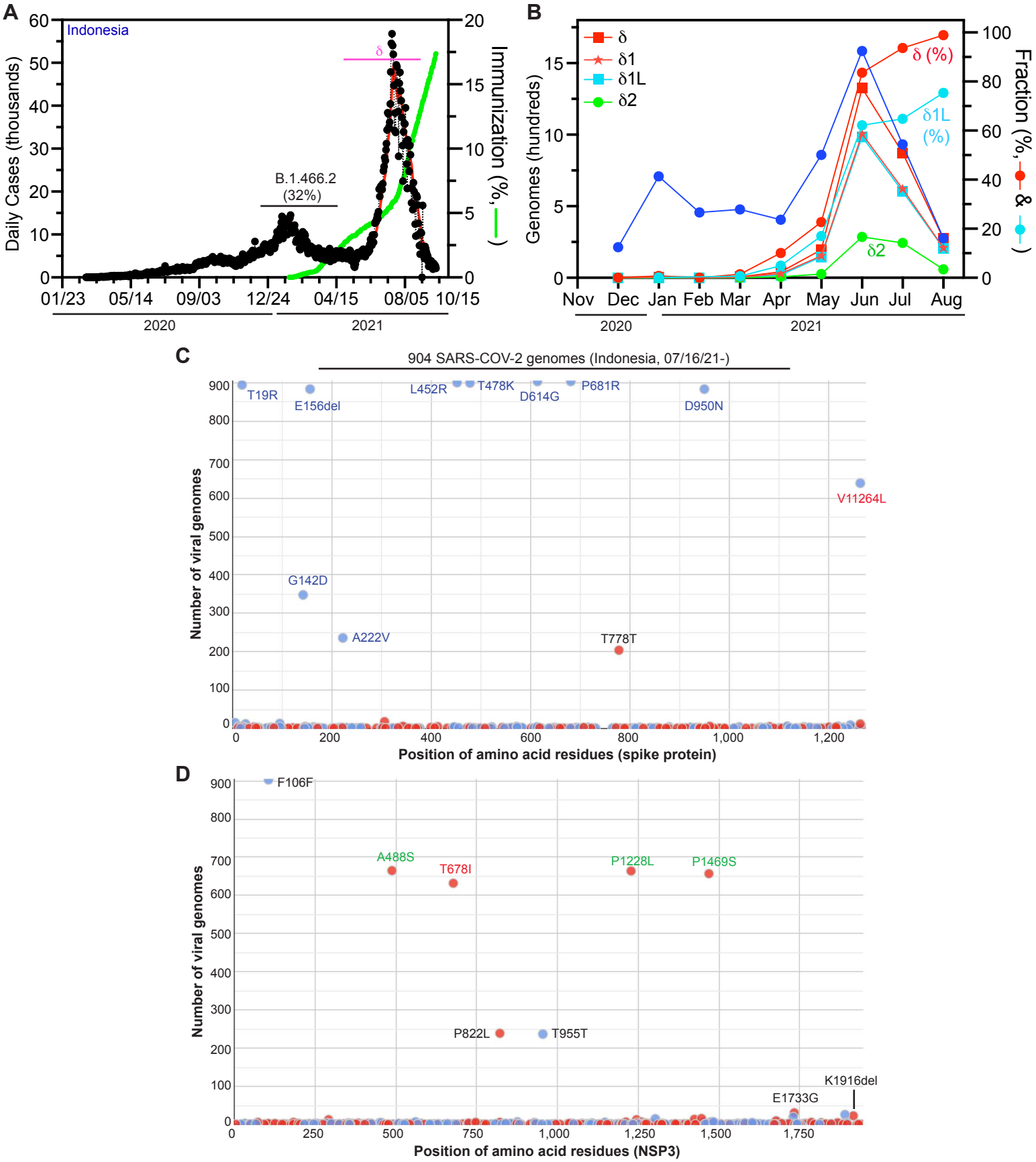

A
Indonesia
δ
B.1.466.2
(32%)
Daily Cases (thousands)
Immunization (%)
01/23
05/14
09/03
12/24
04/15
08/05
10/15
2020
2021
B
δ
δ1
δ1L
δ2
δ (%)
δ1L
(%)
δ2
Genomes (hundreds)
Fraction (%, &,)
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
2020
2021
C
904 SARS-COV-2 genomes (Indonesia, 07/16/21-)
T19R
E156del
L452R
T478K
D614G
P681R
D950N
V11264L
G142D
A222V
T778T
Number of viral genomes
Position of amino acid residues (spike protein)
D
F106F
A488S
T678I
P1228L
P1469S
P822L
T955T
K1916del
E1733G
Number of viral genomes
Position of amino acid residues (NSP3)

Figure 2

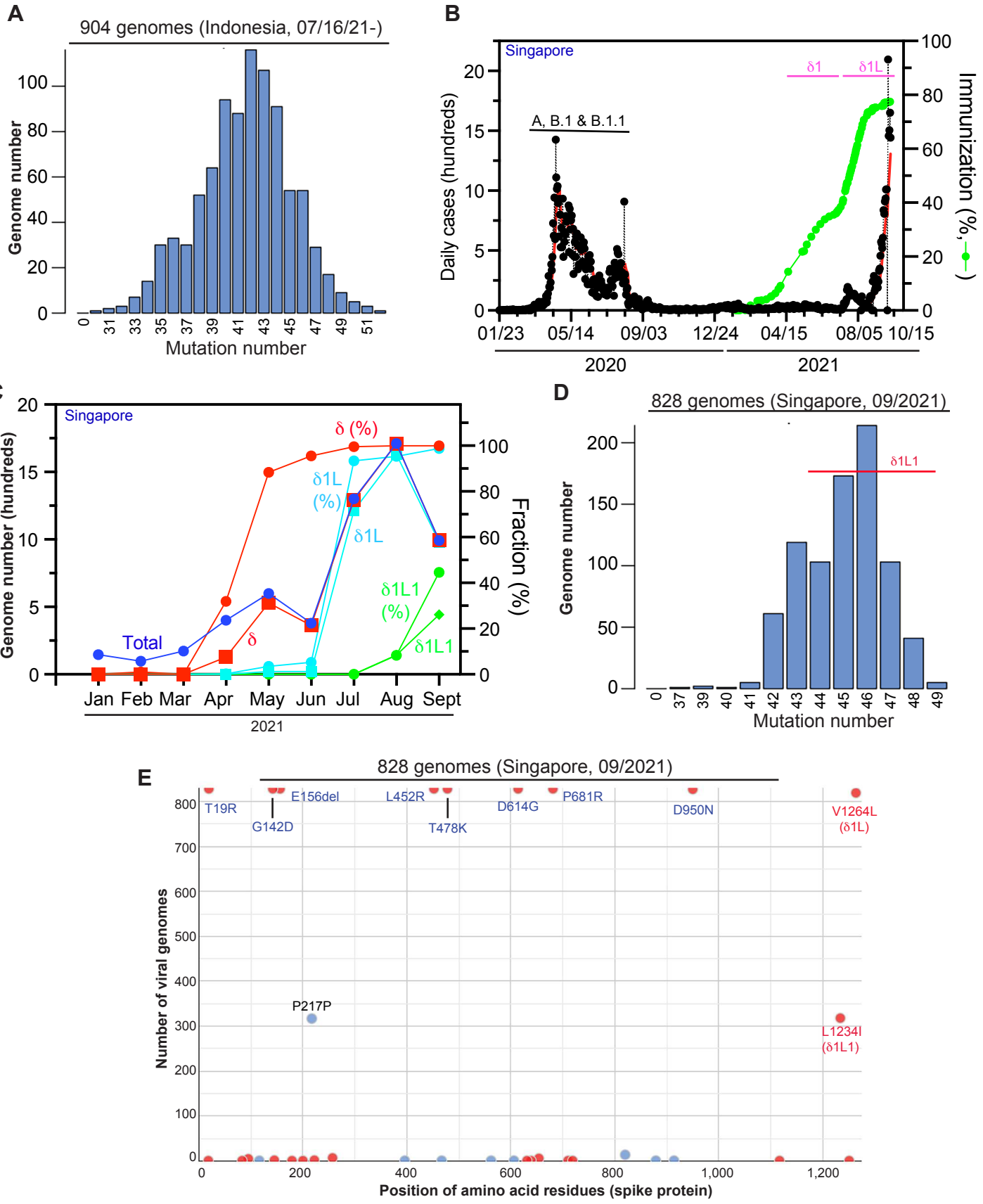

A
904 genomes (Indonesia, 07/16/21-)
Genome number
Mutation number
B
Singapore
A, B.1 & B.1.1
δ1
δ1L
Daily cases (hundreds)
Immunization (%)
01/23
05/14
09/03
12/24
04/15
08/05
10/15
2020
2021
C
Singapore
Genome number (hundreds)
Fraction (%)
δ (%)
δ1L (%)
δ1L
δ1L1 (%)
δ1L1
Total
δ
Jan Feb Mar Apr May Jun Jul Aug Sept
2021
D
828 genomes (Singapore, 09/2021)
Genome number
δ1L1
Mutation number
E
828 genomes (Singapore, 09/2021)
Number of viral genomes
T19R
E156del
G142D
L452R
T478K
D614G
P681R
D950N
V1264L (δ1L)
P217P
L1234I (δ1L1)
Position of amino acid residues (spike protein)

Figure 3

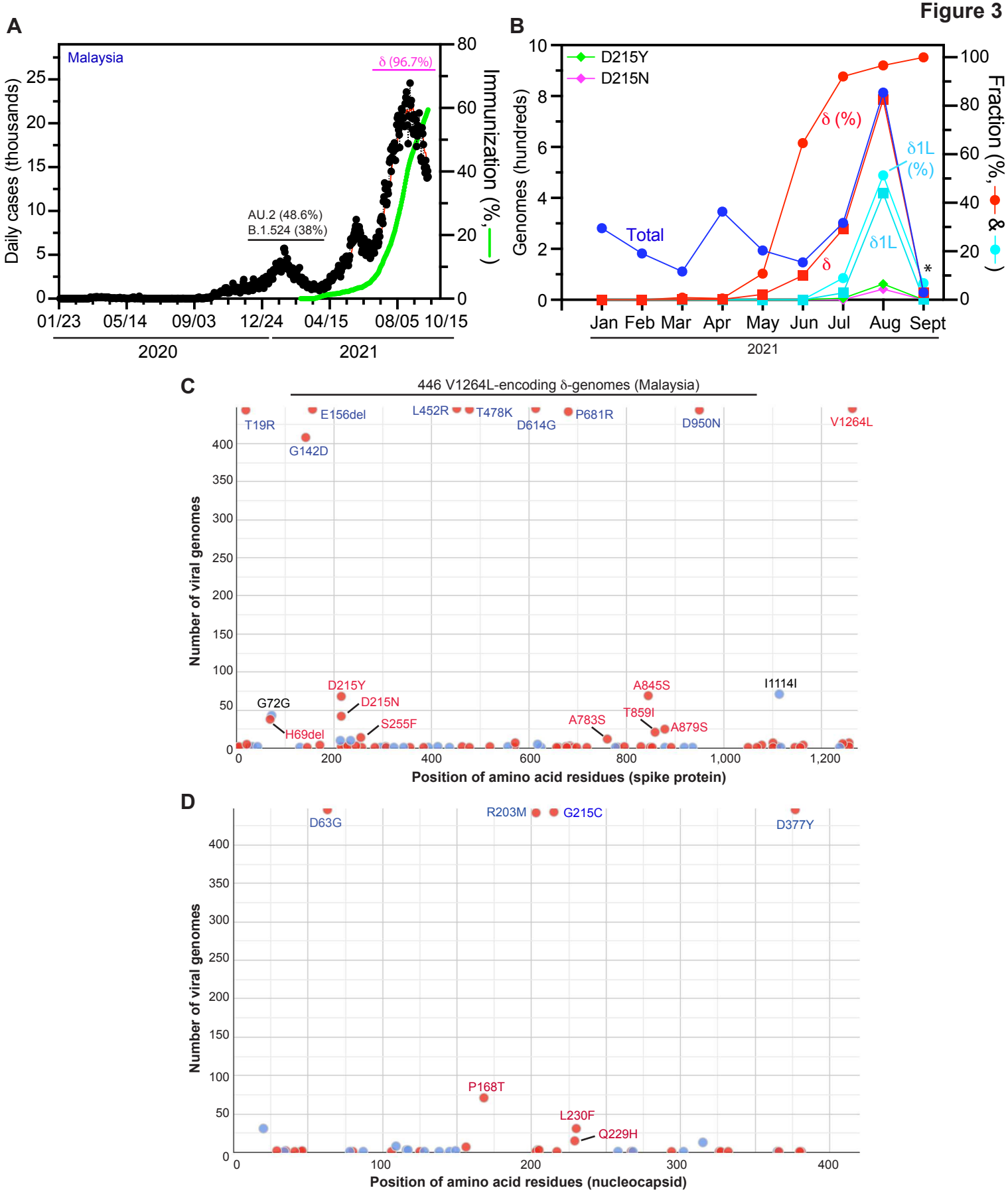

A
Malaysia
δ (96.7%)
AU.2 (48.6%)
B.1.524 (38%)
Daily cases (thousands)
Immunization (%)
01/23
05/14
09/03
12/24
04/15
08/05
10/15
2020
2021
B
D215Y
D215N
Total
δ (%)
δ1L (%)
δ
δ1L
*
Genomes (hundreds)
Fraction (%, δ & δ1L)
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sept
2021
C
446 V1264L-encoding δ-genomes (Malaysia)
T19R
E156del
L452R
T478K
D614G
P681R
D950N
V1264L
G142D
D215Y
D215N
S255F
H69del
G72G
A845S
T859I
A879S
A783S
I1114I
Number of viral genomes
Position of amino acid residues (spike protein)
D
D63G
R203M
G215C
D377Y
P168T
L230F
Q229H
Number of viral genomes
Position of amino acid residues (nucleocapsid)

Figure 4

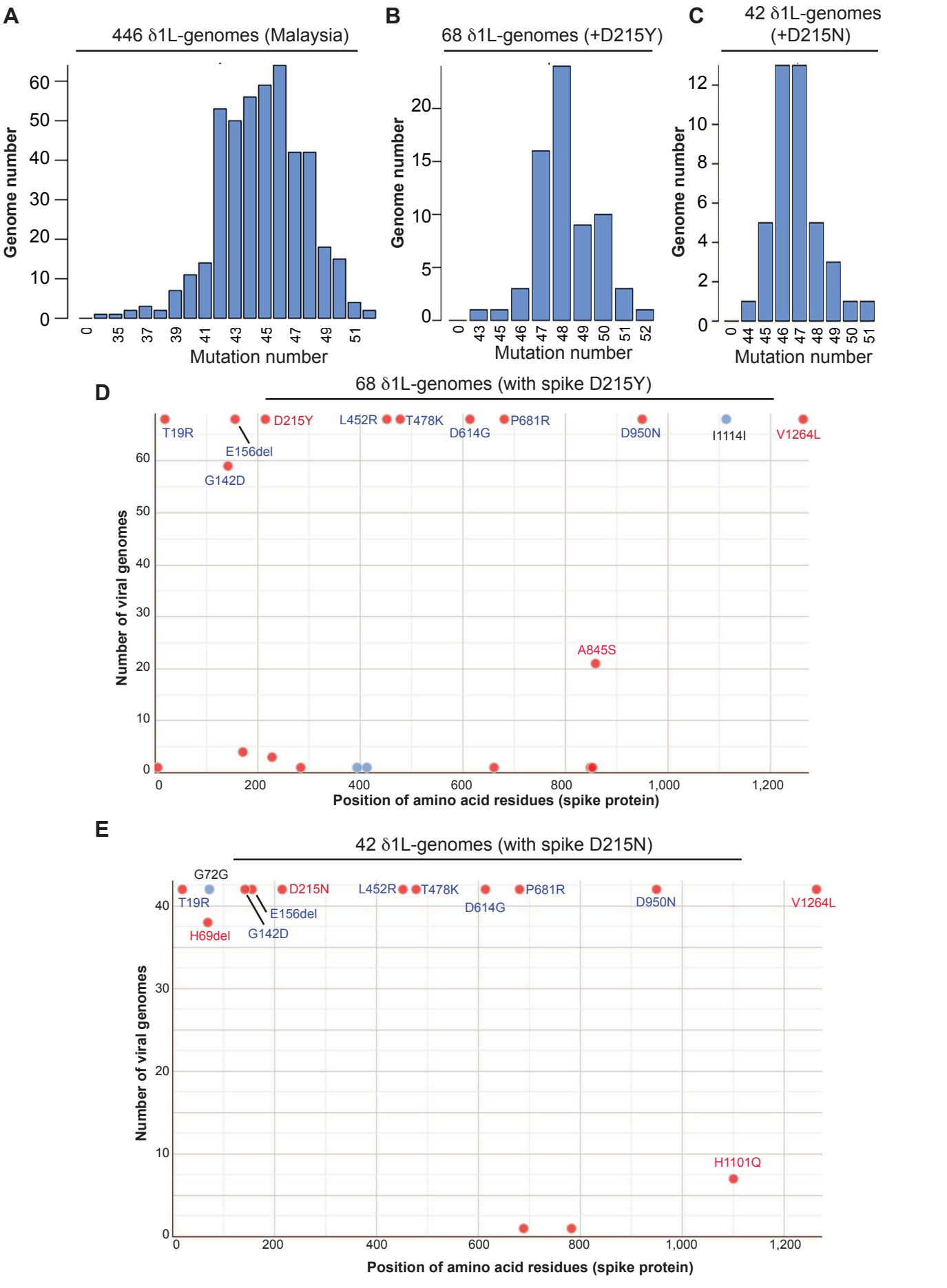

A
446 δ1L-genomes (Malaysia)
Genome number
Mutation number
B
68 δ1L-genomes (+D215Y)
Genome number
Mutation number
C
42 δ1L-genomes (+D215N)
Genome number
Mutation number
D
68 δ1L-genomes (with spike D215Y)
T19R
D215Y
E156del
G142D
L452R
T478K
D614G
P681R
D950N
I1114I
V1264L
A845S
Number of viral genomes
Position of amino acid residues (spike protein)
E
42 δ1L-genomes (with spike D215N)
G72G
T19R
H69del
D215N
E156del
G142D
L452R
T478K
D614G
P681R
D950N
V1264L
H1101Q
Number of viral genomes
Position of amino acid residues (spike protein)

Figure 5

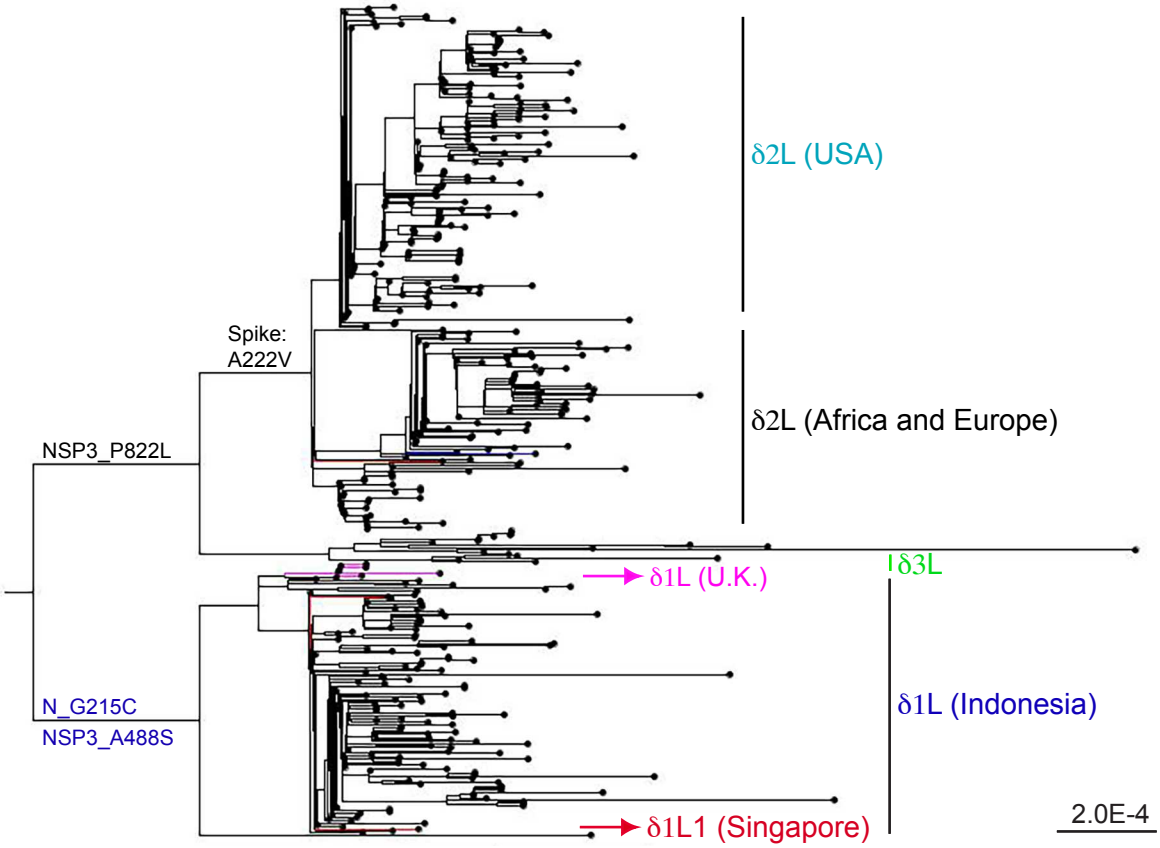

δ2L (USA)
Spike:
A222V
δ2L (Africa and Europe)
NSP3_P822L
δ3L
δ1L (U.K.)
δ1L (Indonesia)
N_G215C
NSP3_A488S
δ1L1 (Singapore)
2.0E-4

Figure 6

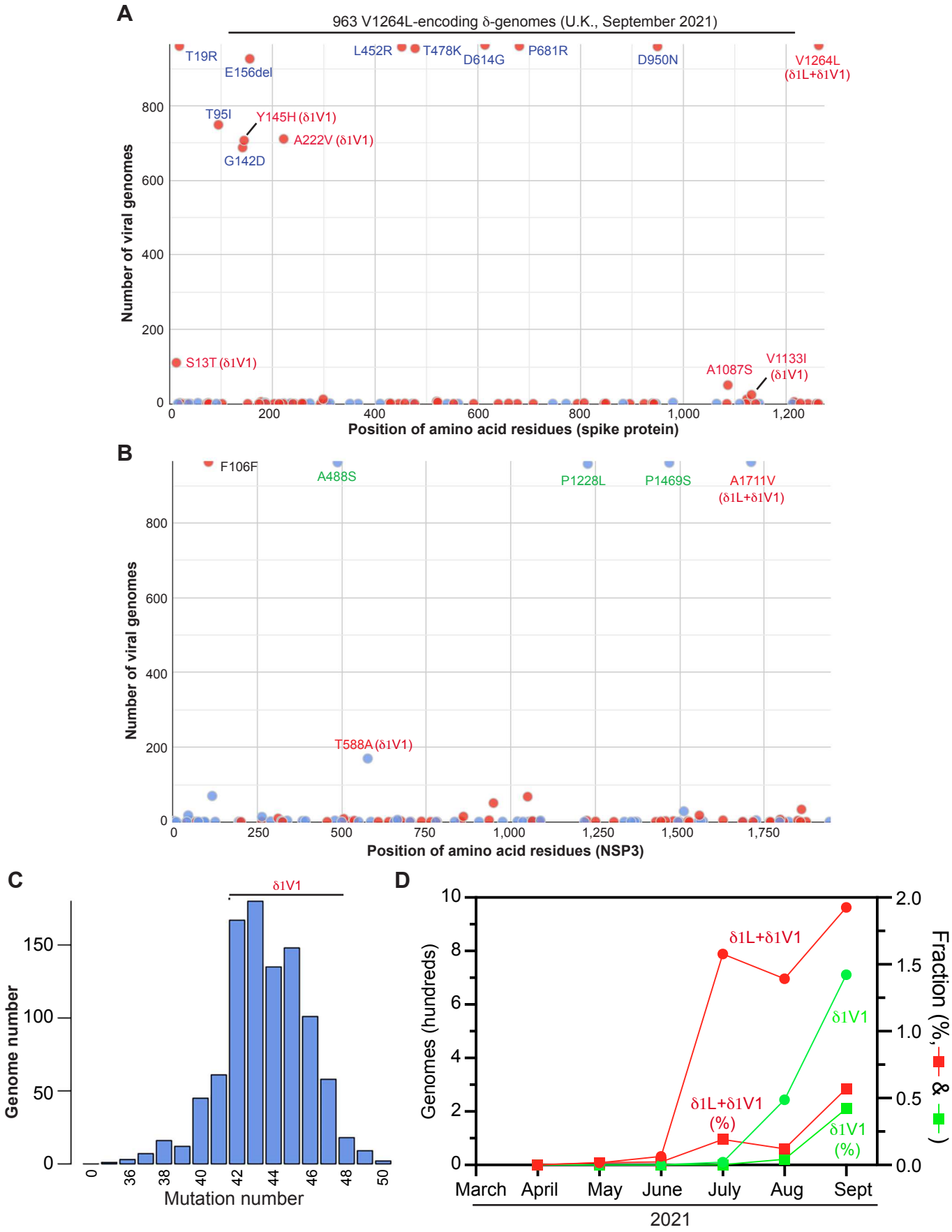

A
963 V1264L-encoding δ-genomes (U.K., September 2021)
T19R
E156del
L452R
T478K
D614G
P681R
D950N
V1264L (δ1L+δ1V1)
T95I
Y145H (δ1V1)
A222V (δ1V1)
G142D
S13T (δ1V1)
A1087S
V1133I (δ1V1)
Number of viral genomes
Position of amino acid residues (spike protein)
B
F106F
A488S
P1228L
P1469S
A1711V (δ1L+δ1V1)
T588A (δ1V1)
Number of viral genomes
Position of amino acid residues (NSP3)
C
δ1V1
Genome number
Mutation number
D
δ1L+δ1V1
δ1V1
δ1L+δ1V1 (%)
δ1V1 (%)
Genomes (hundreds)
Fraction (%)
March
April
May
June
July
Aug
Sept
2021

Figure 7

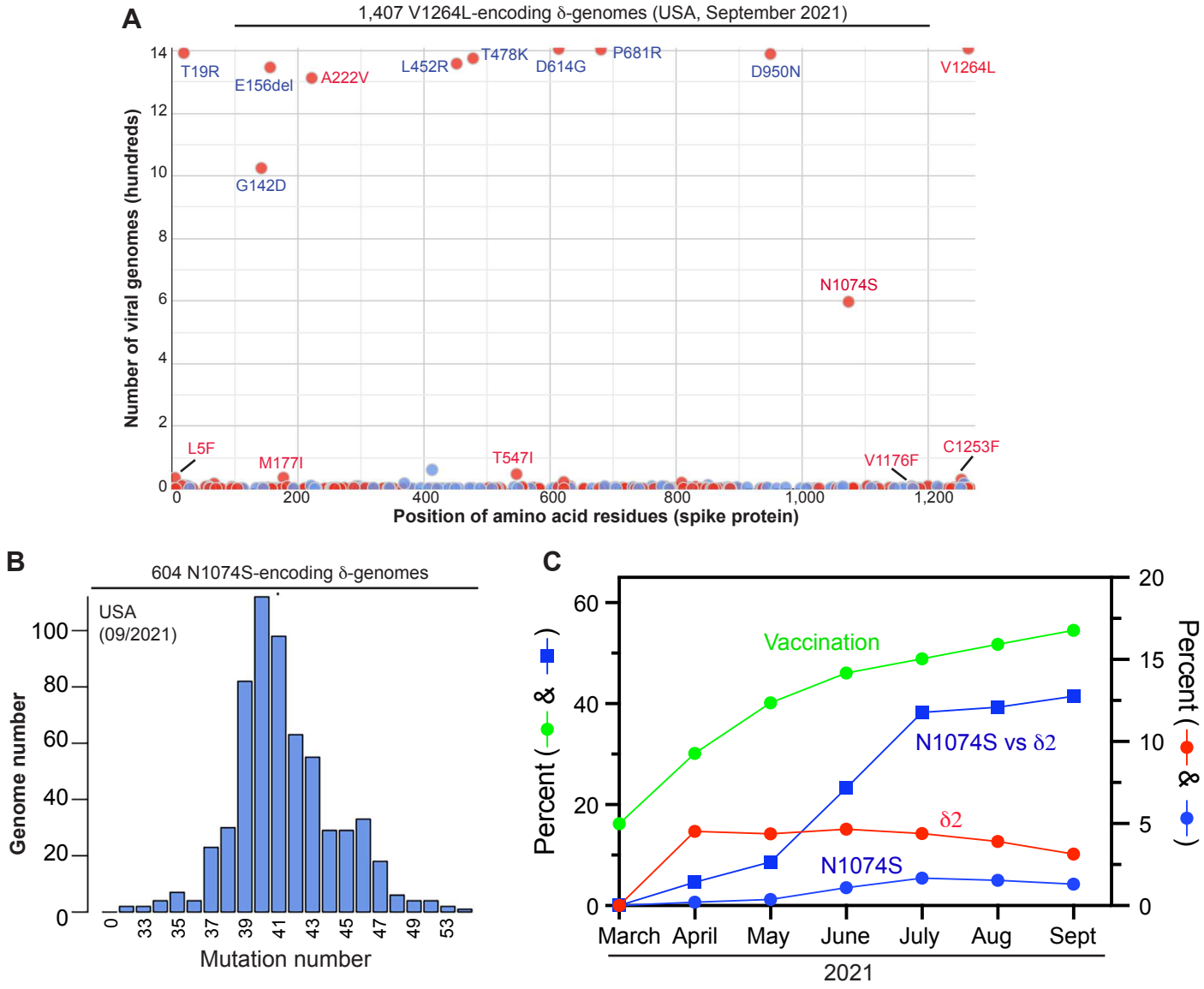
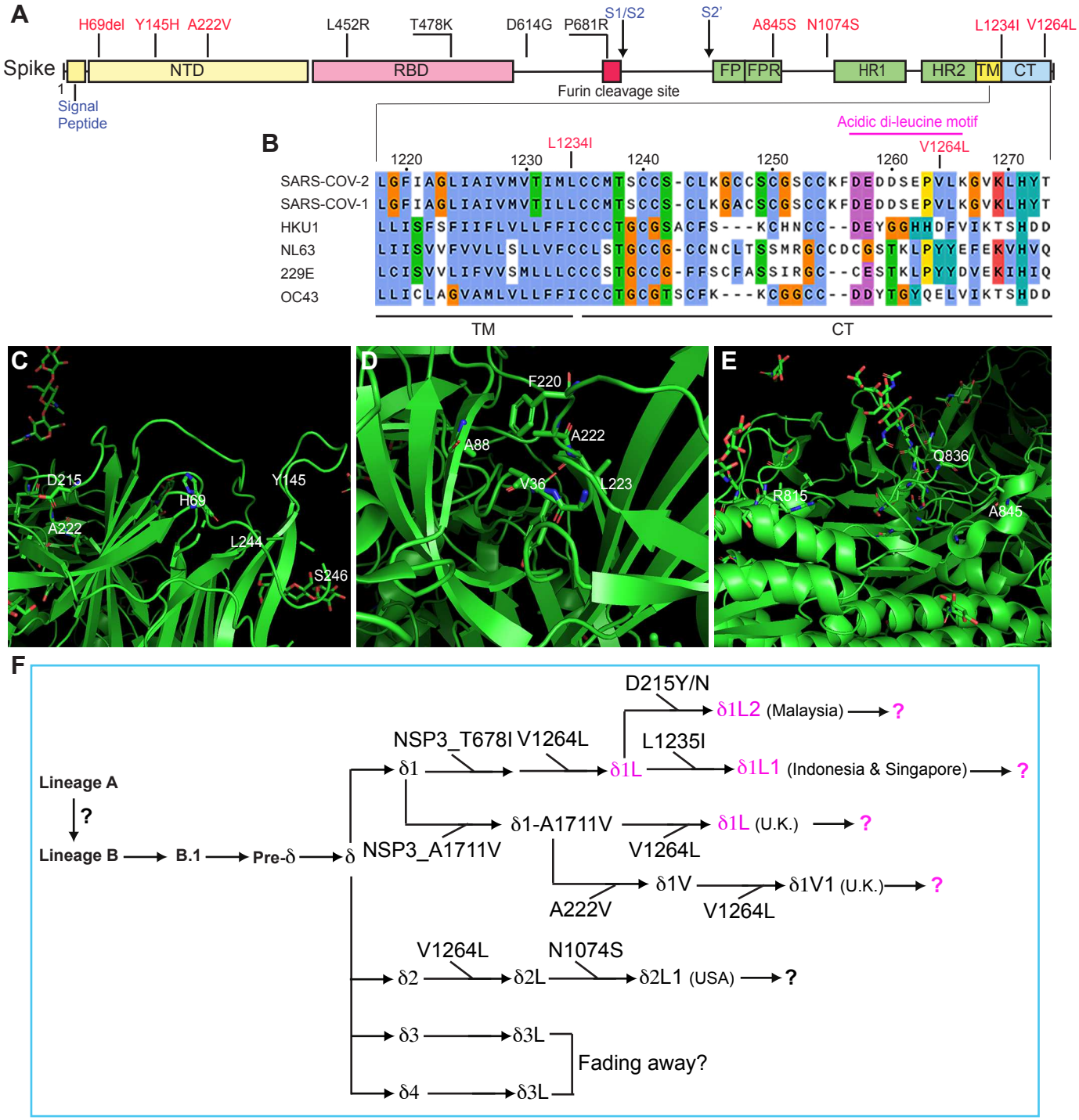

A
1,407 V1264L-encoding δ-genomes (USA, September 2021)
Number of viral genomes (hundreds)
Position of amino acid residues (spike protein)
T19R
E156del
A222V
L452R
T478K
D614G
P681R
D950N
V1264L
G142D
N1074S
L5F
M177I
T547I
V1176F
C1253F
B
604 N1074S-encoding δ-genomes
USA
(09/2021)
Genome number
Mutation number
C
Vaccination
N1074S vs δ2
δ2
N1074S
Percent (δ &)
March
April
May
June
July
Aug
Sept
2021

Figure 8


A
H69del Y145H A222V
L452R
T478K
D614G P681R
S1/S2
S2'
A845S N1074S
L1234I V1264L
Spike
NTD
RBD
FP FPR
HR1
HR2 TM CT
1
Signal Peptide
Furin cleavage site
Acidic di-leucine motif
B
L1234I
V1264L
1220 1230 1240 1250 1260 1270
SARS-COV-2
SARS-COV-1
HKU1
NL63
229E
OC43
TM
CT
C
D215
H69
Y145
A222
L244
S246
D
F220
A88
A222
V36
L223
E
Q836
R815
A845
F
D215Y/N
δ1L2 (Malaysia)
?
NSP3_T678I V1264L
L1235I
Lineage A
?
δ1
δ1L
δ1L1 (Indonesia & Singapore)
?
Lineage B
B.1
Pre-δ
δ
NSP3_A1711V
δ1-A1711V
δ1L (U.K.)
?
V1264L
δ1V
δ1V1 (U.K.)
?
A222V
V1264L
V1264L
N1074S
δ2
δ2L
δ2L1 (USA)
?
δ3
δ3L
Fading away?
δ4
δ3L

Figures S1

Supplementary Files

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

- TableS1IndonesiaV1264LJul16Sept272021904.pdf
- TableS2SingaporeSept01222021828.pdf
- TableS3MalaysiaV1264LSept272021446.pdf
- TableS4DeltaL1tilMay302021473.pdf
- TableS5UKV1264LSept1262021963.pdf
- TableS6USV1264LSept12620211407.pdf