Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

Haplotype distribution of SARS-CoV-2 variants in low and high vaccination rate countries during ongoing global COVID-19 *pandemic in early 2021*

Ngoc-Niem Bui [a,b], Yu-Tzu Lin [a,1], Su-Hua Huang [c], Cheng-Wen Lin [a,c,*]

[a] *Department of Medical Laboratory Science and Biotechnology, China Medical University, Taichung 40402, Taiwan*
[b] *Faculty of Medicine, Can Tho University of Medicine and Pharmacy, Can Tho 94117, Viet Nam*
[c] *Department of Medical Laboratory Science and Biotechnology, Asia University, Taichung 41354, Taiwan*

ARTICLE INFO

Keywords:
Severe acute respiratory syndrome coronavirus 2
Ongoing pandemic
Variant
Haplotype
Partly vaccinated rate

ABSTRACT

The widespread severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) continuously impacts our economic and public health. The potential of emerging variants to increase transmissibility and evade vaccine-induced immunity lets us put more effort to research on viral mutations and explore the pathogenic haplotypes. In this study, we characterized the haplotype and sub-haplotype diversity of SARS-CoV-2 global variants in January–March and the areas with low and high COVID19 vaccination rates in May 2021 by analyzing viral proteome of complete genome sequences published. Phylogenetic tree analysis of the proteomes of SARS-CoV-2 variants with Neighbor-Joining and Maximum Parsimony methods indicated that haplotype 2 variant with nsp12 P323L and Spike D614G was dominant (98.81%), including new sub-haplotypes 2A_1 to 2A_3, 2B_1 to 2B_3, and 2C_1 to 2C_2 emerged post-one-year COVID-19 outbreak. In addition, the profiling of sub-haplotypes indicated that sub-haplotype 2A_1 with the mutations at N501Y, A570D, D614G, P681H, T716I, S982A, and D118H in Spike was over 58% in May 2021 in the high partly vaccinated rate group (US, Canada, and Germany). Meanwhile, the new haplotype 2C_3 bearing the mutations at EFR156-158del, T19R, A222V, L452R, T478K, and D614G in Spike occupied over 54.8% in May 2021 in the low partly vaccinated rate group (India, Malaysia, Taiwan, and Vietnam). Sub-haplotypes 2A_1 and 2C_3 had a meaningful alternation of ACE2-specific recognition site, neutralization epitopes, and furin cleavage site in SARS-CoV-2 Spike protein. The results discovered the haplotype diversity and new sub-haplotypes of SARS-CoV-2 variants post one-year pandemic in January–March 2021, showing the profiles of sub-haplotypes in the groups with low and high partly vaccinated rates in May 2021. The study reports the emergence of new SARS-CoV-2 sub-haplotypes during ongoing pandemic and vaccination in early 2021, which might help inform the response to vaccination strategies.

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), firstly reported in China in December 2019 and defined as the coronavirus disease in 2019 (COVID-19) by WHO on March 11, 2020 (Zhou et al., 2020). This novel virus was declared a highly transmissible and pathogenic virus that caused over 3 million deaths and over 183 million cases worldwide as of July 6, 2021. The ongoing outbreak of COVID-19 has had devastating medical, economic, and social consequences. There is an urgent need to understand the novel virus evolution and devoted essential measures to keep the pandemic from being under control.

It is noted that there has been a lack of effective therapies to treat COVID-19. It is clear that the SARS-CoV-2 shares 81.9% nucleotide similarity to subgenus Sarbecovirus in the Betacoronavirus genus, previously found in China (Wu et al., 2020). Similar to other coronaviruses, the novel SARS-CoV-2 virus is an enveloped, positive-sense-single-stranded RNA virus with a genome of approximately 30 kb encoding open reading frames ORF1a/b, spike (S), envelope (E), membrane (M), nucleocapsid (N), and several accessory proteins (Cao et al., 2021; Sola et al., 2015). The papain-like (within nonstructural protein 3, nsp3) and 3C-like (nsp5) proteases auto-cleave ORF1a and ORF1b polyproteins divided into 16 nsps which are essential for viral replication and

[*] Corresponding author at: Department of Medical Laboratory Science and Biotechnology, China Medical University, No. 91, Hsueh-Shih Road, Taichung 404, Taiwan.
E-mail address: cwlin@mail.cmu.edu.tw (C.-W. Lin).
[1] Co-first author.

https://doi.org/10.1016/j.meegid.2021.105164
Received 3 September 2021; Received in revised form 23 November 2021; Accepted 25 November 2021
Available online 27 November 2021

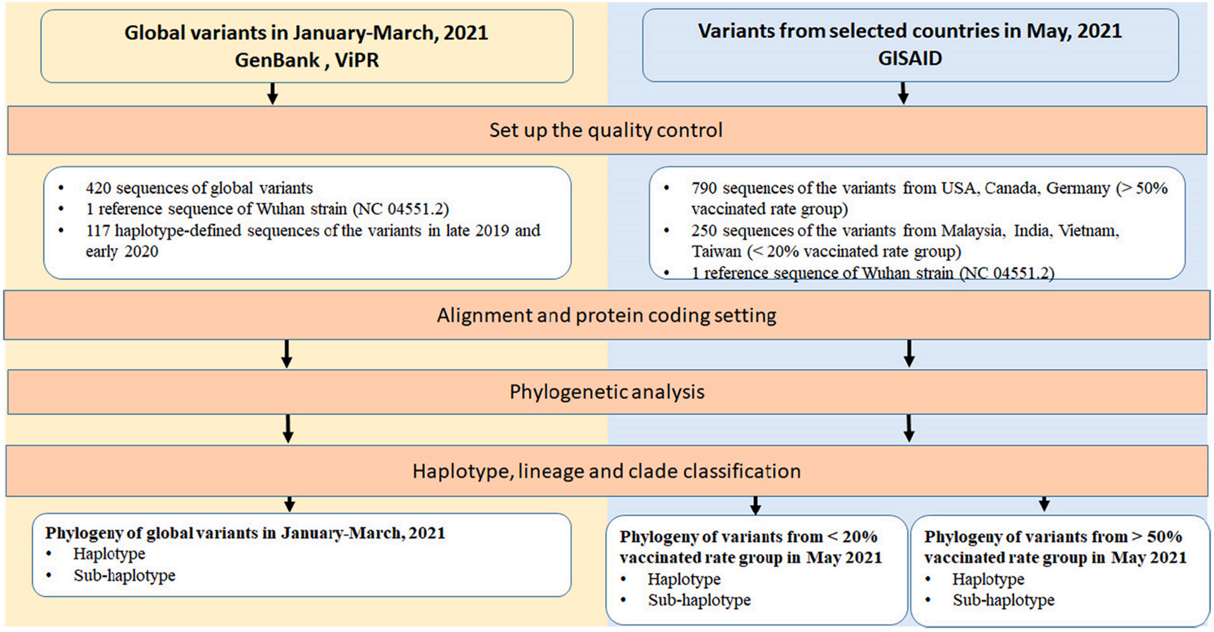
Global variants in January-March, 2021
GenBank , ViPR
Variants from selected countries in May, 2021
GISAID
Set up the quality control
• 420 sequences of global variants
• 1 reference sequence of Wuhan strain (NC 04551.2)
• 117 haplotype-defined sequences of the variants in late 2019 and early 2020
• 790 sequences of the variants from USA, Canada, Germany (> 50% vaccinated rate group)
• 250 sequences of the variants from Malaysia, India, Vietnam, Taiwan (< 20% vaccinated rate group)
• 1 reference sequence of Wuhan strain (NC 04551.2)
Alignment and protein coding setting
Phylogenetic analysis
Haplotype, lineage and clade classification
Phylogeny of global variants in January-March, 2021
• Haplotype
• Sub-haplotype
Phylogeny of variants from < 20% vaccinated rate group in May 2021
• Haplotype
• Sub-haplotype
Phylogeny of variants from > 50% vaccinated rate group in May 2021
• Haplotype
• Sub-haplotype

Fig. 1. The workflow for the haplotype and sub-haplotype profiles of SARS-CoV-2 variants isolated post-one-year outbreak in January–March and the countries with low and high partly vaccinated rates in May 2021. Four hundred twenty complete proteomic sequences of global SARS-CoV-2 variants isolated from January to March 2021 were collected and combined with the reference of the complete proteomic sequence of Wuhan strain (NC 045512.2), and 117 complete proteomic sequences of the variants isolated in late 2019 and early 2020 reported (Infect Genet Evol. 2021 Jul;91: 104800), further followed the steps of alignment, protein-coding setting, phylogenetic analysis and classified the haplotypes and sub-haplotypes of global SARS-CoV-2 variants isolated from January to March 2021. In addition, the complete proteomic sequences of the variants isolated in low- and high-vaccinated rate countries in May 2021 were collected and followed the steps mentioned above to construct a phylogeny tree and classify the haplotypes and sub-haplotypes of the variants in low and high **partly vaccinated** rate groups in May 2021.

transcription. For instance, the nsp12 (RNA-dependent RNA polymerase, RdRp), nsp14 (exoribosome, ExoN), and nsp16 (ribose 2’-*O*-methyltransferase, 2-O-MTase), showing the enzymatic activity in RNA synthesis and processing, have been considered as the targets for the development of antiviral agents (Müller et al., 2018; Ahmad et al., 2020; Hillen et al., 2020). The homotrimer S protein is a class I viral fusion protein. It is the structural protein at the surface of the virion which binds to the host cell angiotensin-converting enzyme 2 (ACE2) receptor and is known as the primary target for vaccine development (Korber et al., 2020; Walls et al., 2020). E protein is a small protein composed of 75 amino acid residues and plays a significant role in viral morphogenesis and viral assembly in cooperation with the viral membrane. It has been shown to participate in activating the host inflammasome (Naqvi et al., 2020). The viral membrane (M) protein is the most abundant viral protein with higher conservation among SARS-CoV-2. The N protein is the only structural protein inside the virion that induces severe immune response during infection. Hence it has potentially an essential target for vaccine development (Yang and Rao, 2021).

Following the rapid expansion of this novel coronavirus, natural selection mutations may occur and cause changes in its infectivity, pathogenicity. The mutation of the virus also results in the emergence of highly infectious and lethal lineages that lead to failures of existing antibodies and vaccines (Parlikar et al., 2020; Phan, 2020). In early 2020, ORF1a/b became one of the mutation hot spots with a mutation rate of up to 29.47%, particularly in nsp2, nsp3, and nsp12 (Pachetti et al., 2020; Ren et al., 2020)· The variations within the S protein–cell receptor interface is more vulnerable to viral infectivity, in which the variant with S D614G emerged is the most prevalent clade at multiple geographic areas (Korber et al., 2020; Ogawa et al., 2020). SARS-CoV-2 variants have been defined as the variants of interest (VOI), variants of concern (VOC), and variants of high consequence. Emerging SARS-CoV-2 variants in early 2021, including B.1.1.7/B.1.1.7 + E484K lineage (United Kingdom), B.1.351 lineage (South Africa), P.1 lineage (Brazil), and B.1.427/B.1.429 lineage (California) has rapidly become dominant in domestic and arousing global concerns (Wibmer et al., 2021). The B.1.1.7 lineage harbors three amino acid deletions and seven missense mutations in spike protein, including D614G and N501Y in the ACE2 receptor-binding domain (Volz et al., 2021). This variant showed a significant increase in the adequate reproduction number but has a low impact on neutralization by monoclonal antibody therapy and convalescent and post-vaccination sera (Collier et al., 2021; Shen et al., 2021; Wang et al., 2021). The P.1 and the B.1.351 variants have been shown a moderate impact on neutralization by BNT162b2- and mRNA-1273 fully-vaccinated sera. Like B.1.1.7 lineage, the P.1 and the B.1.351 have undergone an unusually large number of mutations (Collier et al., 2021; Wang et al., 2021; Garcia-Beltran et al., 2021).

At the moment of the ongoing COVID-19 outbreak and increasing partly and fully vaccinated rates, exploring haplotypes of current global SARS-CoV-2 variants is of great importance, particularly monitoring the prevalence of SARS-CoV-2 variants in the countries with low and high vaccinated rates. Moreover, the surveillance of the novel variants can be overcome by further evaluating the potential reduction in neutralization by monoclonal antibody treatments or convalescent and post-vaccination sera. The present study characterized the haplotype diversity and sub-haplotype distribution of SARS-CoV-2 variants post one-year spread in January–March 2021 and after increasing the COVID-19 vaccination rate in May 2021 by analyzing the proteome of the variants sequenced in the open data. The study provided the difference of amino acid substitutions between proteomes of main haplotypes and sub-haplotypes in May 2021, understanding the impact on the efficacy of vaccines, therapeutics, and even diagnosis.

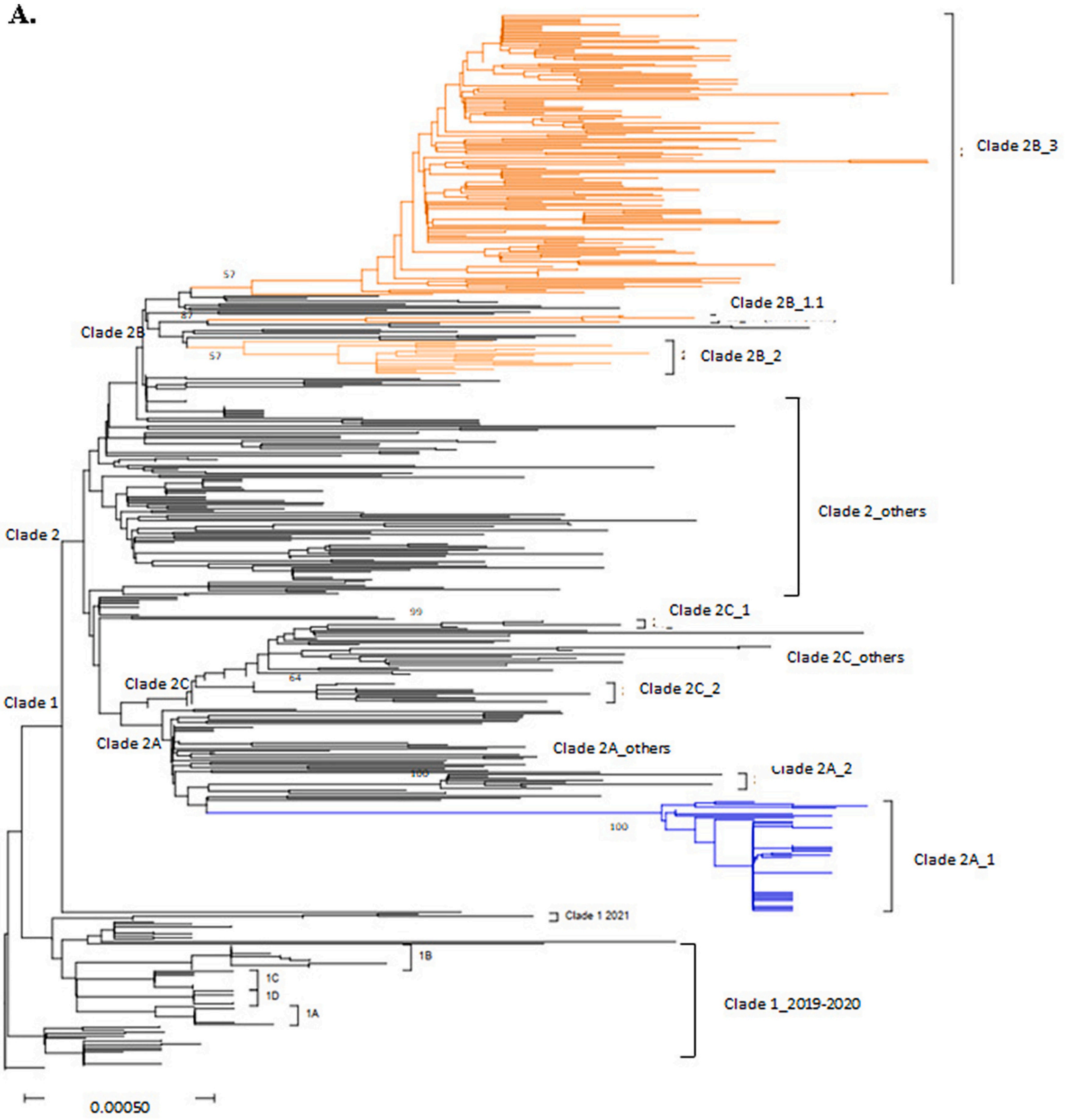

Fig. 2. The haplotype and sub-haplotype profiling of global SARS-CoV-2 variants post 1-year pandemic. NJ phylogenetic tree analysis of complete proteomic sequences from 420 SARS-CoV-2 global variants, sequenced in January–March 2021 and 117 haplotype-defined variants, collected post-6-month spread. The evolutionary distances were computed to scale, with the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test. The data of bootstrap of specific haplotype and sub-haplotype were marked next to each branch (A). The circle version of the phylogenetic tree analyzed was shown (B).

2. Materials and methods

2.1. Data collection, phylogenetic tree construction, and (sub)haplotype classification of SARS-CoV-2 variants post one-year spread

In order to generate the dataset of SARS-CoV-2 variants post one-year spread, the complete genome sequences of the variants were downloaded from the Virus Pathogen Resource database (www.viprbrc.org) follow the criteria: complete full-length genome (at least 29,000 bp) of SARS-CoV-2 in humans during the period from January to March 2021. By the data collection time of March 14, the number of sequences was loaded up to 10,477 complete and non-duplicate genomes sequences from the USA. In order to collect the sequences, we selected and downloaded the 1st and the 26th sequence on each page displaying 50 sequences per page by upload date. Hence, two sequences per page from 210 pages of USA data (total 420 sequences) were retrieved to generate the equilibrium data. Together with 420 sequences from the USA, 212 sequences from other global countries were recruited. Initially, 632 sequences were in the raw dataset, and then removed the duplicate sequences and the sequences with a high variant with gaps or high numbers of mismatched over 20 nucleotides with 'N' or other ambiguous IUPAC code (Korber et al., 2020). Finally, 420 complete genomes of the global SARS-CoV-2 variant were now ready for the next step. The cleaned dataset (420 sequences) was analyzed using the MEGA X software, and the mutation data was exported into a .csv file. Later, these files were observed mutation contained in variant clades and sub-clades defined from the phylogenetic tree and bootstrap data. The data was later aligned against the Wuhan-Hu-1 strain NC_045512.2 as a reference by using the alignment program online MAFFT (Multiple Alignment using Fast Fourier Transform) (Katoh and Standley, 2013). To keep the coding regions, the upstream of nsp1, the non-coding regions, and the downstream of ORF10 were deleted using the MEGA X software (Kumar

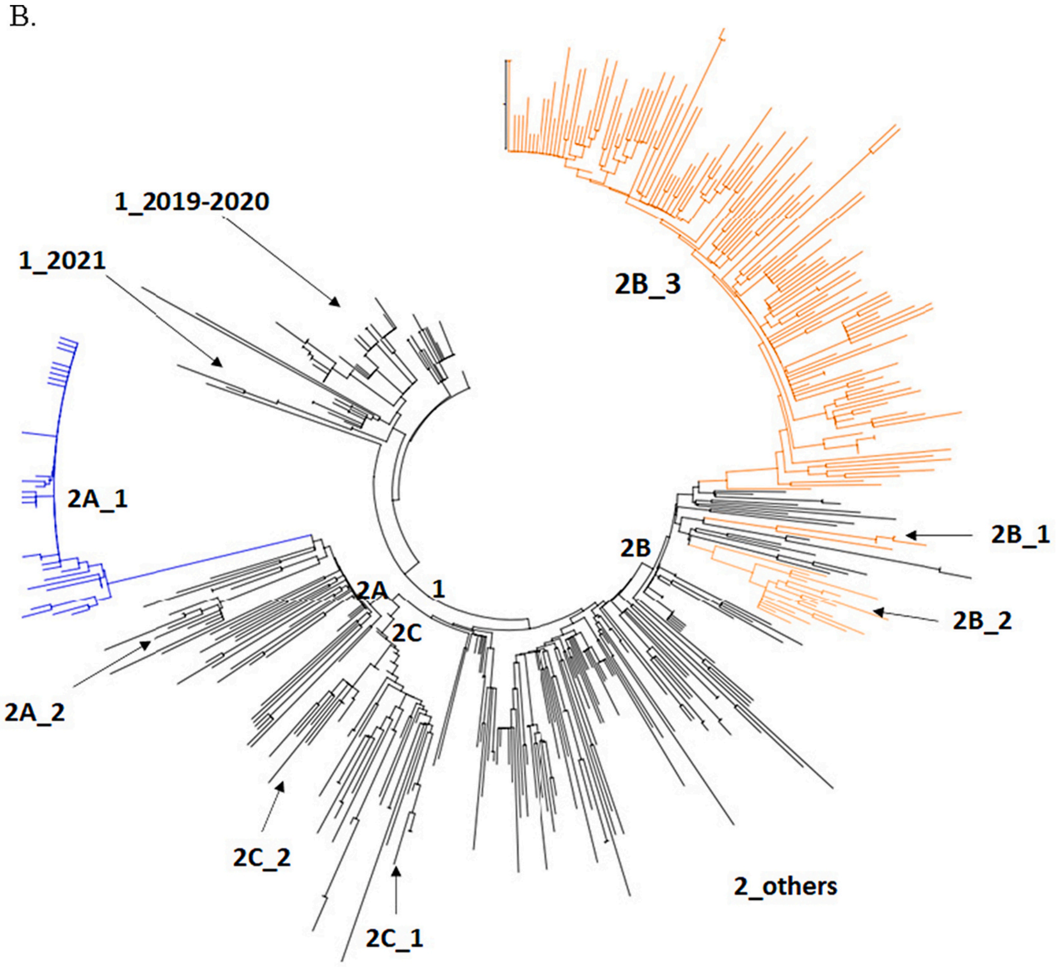

Fig. 2. (*continued*).

et al., 2018). The final clean and coding regions only dataset were used to construct the phylogenetic tree of the proteomes of SARS-CoV-2 variants with both two methods, including NJ (Neighbor-Joining) (Felsenstein, 1985; Saitou and Nei, 1987) and MCL (Maximum composite Likelihood)(Kumar et al., 2018). Since the 100 replicates of phylogenetic tree analysis produced stable bootstrap values (Pattengale et al., 2010), the bootstrap consensus tree of 100 replicates was inferred using the NJ for analyzing 117 and 420 complete proteomes of variants that emerged in 2020 and 2021, respectively. Finally, the phylogenetic tree was validated by comparing with our prior report in that the bootstrap consensus tree of 1000 replicates of NJ and MCL method for analyzing the variants collected post-six-month spread (Bui et al., 2021). The Poisson correction method was computed to see the evolutionary distance. The SPR algorithm was proposed to compare the MP and MCL distance. The other files (bootstrap consensus tree, phylogenetic criteria construct, and sequence data excel files) were also generated for analysis (Fig. 1).

2.2. *Haplotype and sub-haplotype distribution of SARS-CoV-2 variants in low- and high-vaccination rate countries during the periods of January to March and May 2021*

Since some countries have favored vaccinating as many people as possible, while others have tried to prioritize vaccinating specific vulnerable groups of the population, the vaccine rollout strategy varies from country to country. Therefore, it is crucial to evaluate the prevalence of SARS-CoV-2 variants in different vaccination rate countries. To track the SARS-CoV-2 variant extend and prevalent in low- and high-vaccination rate countries, we based on the report from OWID (Our World in Data) (Ritchie et al., 2021) to categorize the countries with the partly vaccinated rates (at least one-dose of COVID-19 vaccination) of lower than 20% and higher than 50%, respectively, on May 28, 2021. India, Malaysia, Taiwan, and Vietnam were enrolled in the Group with a lower partly vaccinated rate of 20% in May 2021. The US, Canada, and Germany had a higher than 50% partly vaccinated rate on May 28, 2021, placed in one group. Finally, 250 sequences of SARS-CoV-2 variants in the low-vaccinated rate group and 790 sequences of variants from the high partly vaccinated rate group in May 2021 were selected and run to construct the phylogenetic tree after cleaning and alignment by MAFF and followed to define the haplotypes and sub-haplotypes (Fig. 1), described as above.

2.3. *Mapping amino acid substitutions in the spike trimer of main sub-haplotypes*

Since three-dimensional structures of the mutated spike proteins of the main sub-haplotypes are not yet solved, the location of the amino acid substitutions within the spike trimer was generated using the published cryo-electron microscopy (cryo-EM) structure of the spike protein (RCSB Protein Data Bank ID 7DK3) as the template (Xu et al., 2021), and displayed by the PyMOL software (Molecular Graphics System, Version 2.0 Schrödinger, LLC). In addition, the ACE2-specific recognition site (T470-T478 loop and Y505) (Xu et al., 2021) and four regions of predicted B- and T-cell epitopes (VRQIAPGQT, YQAGSTPCN, FQPTNGVGF, ILPDPSKPS) (Bhattacharya et al., 2020) were also marked in the spike protein to elucidating the relationship among the ACE2-

Table 1
Haplotype and sub-haplotype profiles of global SARS-CoV-2 variants post 1-year spread from January to March 2021.

Haplotype	Sub-haplotype	Amino acid substitutions	Pangolin lineage	WHO label	Percentage of variants from Dec 2019 to Jun 2020[a]	Percentage of variants from Jan 2021 to Mar 2021
1		ORF8 L84S	A		25.34% (1072/4230)	1.19% (5/420)
2		nsp12 P323L, Spike D614G	B		63.54% (2688/4230)	98.81% (415/420)
2A		N R203K, N G204R			9.1% (385/4230)	15.23% (64/420)
	2A_1	nsp3 T183I, nsp3 A890D, nsp3 I1412T, spike V143del, spike N501Y, spike A570D, spike P681H, spike T716I, spike S982A, spike D1118H, ORF8 R52I, ORF8 Y73C, N S235F	B.1.1.7	Alpha		13.33% (56/420)
	2A_2	nsp3 P141S, nsp4 T492I, nsp6 I49V, nsp9 T35I, spike T478K, spike P681H, spike T732A	B.1.1.519			1.90% (8/420)
	2A_3[b]	nsp3 S370L, nsp3 K977Q, nsp13 E341D, spike L18F, Spike T20N, spike P26S, spike D138Y, spike K417T, spike E484K, spike N501Y, spike H655Y, spike T1027I, spike V1176F, ORF3a S253P, ORF8 E92K, N P80R	P.1	Gamma		
2B		ORF3a Q57H			36.21% (1532/4230)	38.81% (163/420)
	2B_1	nsp2 T85I, nsp3 K837N, spike D80A, spike D215G	B.1.351	Beta		0.95% (4/420)
	2B_2	nsp2 T85I, spike W152C, N T205I	B.1.429	Epsilon		4.05% (17/420)
	2B_3	nsp2 T85I, nsp5 L89F, nsp14 N129D, nsp16 R216C, ORF8 S24L, N P67S, N P199L	B.1.596			33.81% (142/420)
	2B_4[b]	nsp2 T85I, nsp4 L438P, spike L5F, spike T95I, spike D253G, ORF3a P42L, ORF8 T11I	B.1.526	Iota		
	2B_5[b]	nsp3 S126L, nsp3 T350I, nsp3 P822L, nsp6 L75F, nsp13 S259L, spike N439K, spike P681R spike G1251V, N T205I	B.1.1.466.2 AU.2 B.1.459			
2C		spike P681R				3.33% (14/420)
	2C_1	nsp3 T428I, nsp5 G15S, snp8 T148I, spike Q677H, M V70L, N R203K, N G204del, N G121V	C.36			0.95% (4/420)
	2C_2	nsp3 T428I, nsp5 G15S, ORF3a S171L, N R203K, N G204del, N G212V	C.36			2.38% (10/420)
	2C_3[b]	nsp3 P822L, nsp4 A446V, nsp6 V149A, nsp6 T181I, nsp12 G671S, nsp13 P77L, spike T19R, spike EFR156-158del, spike A222V, spike L452R, spike T478K, ORF3a S26L, M I82T, ORF8 DF118,119del, N D63G, N R203M, N D377Y	B.1.617.2	Delta		
	2C_4[b]	nsp3 T749I, nsp6 T77A, **nsp12 P323L**, nsp13 M429I, nsp15 K259R, nsp15 S261A, spike L452R, spike E484, **spike D614G, spike P681R**, ORF3a S26L, ORF7a V82A, N R203M, N D377Y				
2D[b]		spike H69del, E L21F	B.1.525	Eta		

[a] The data were adapted from our prior report (Bui et al., 2021).
[b] The haplotypes and sub-haplotypes were identified in Table 2, which SARS-CoV-2 variants were isolated from low and high partly vaccinated rate groups in May 2021.

specific recognition site, the B-cell epitopes, and the amino acid substitutions within the spike protein trimer of sub-haplotypes.

3. Results

3.1. *Haplotype trend of SARS-CoV-2 variants post 1-year spread*

After a 1-year spread of ongoing COVID-19 pandemic, the haplotype profile of 420 SARS-CoV-2 variants from global areas in January–March 2021 was characterized by phylogenetic analysis based on amino acid substitution rate using Maximum Composite Likelihood distance by Neighbor-Joining tree from MEGA CC (Fig. 2, Table 1). The haplotype and sub-haplotype distribution of the variants from global areas post 1-year spread from January 2021 to March 2021 was an observable change compared to the previous data post-6-month spread from December 2019 to June 2020 (Bui et al., 2021). Haplotype 1 with the mutation at ORF8 L84S was prevalent post-6-month spread (25.34%) but rarely identified post-1-year pandemic (1.19%). Meanwhile, haplotype 2 with the mutations at nsp12 P323L and spike G614D was almost total post-1-year spread (98.81%) (Fig. 2, Table 1). Among haplotype 2, haplotype 2A with the mutations at nsp12 P323L, Spike D614G, and N R203K, G204R, markedly increased from 9.1% post-6-month spread to 15.23% post-1-year spread, in which sub-haplotypes 2A _1 (13.33%) and 2A_2 (1.90%) belonged to Pangolin lineages B.1.1.7 (WHO label Alpha) and B1.1.519 variants, respectively. Haplotype 2B containing the mutations at nsp12 P323L, spike D614G, ORF3a Q57H, nsp2 I85I, slightly increased from 36.21% post 6-month spread to 38.81% post 1-year spread, including 0.95% for sub-haplotype 2B_1 (Pangolin lineage B1.351, WHO label Beta), 4.05% for 2B_2 (Pangolin lineage B.1.429, WHO label Epsilon), 33.81% for 2B_3 (Pangolin lineage B.1.596), respectively. In addition, new sub-haplotype 2C (3.33%) appeared post-1-year spread. The results demonstrated the difference in haplotype and sub-haplotype profiling of global SARS-CoV-2 variants circulated after the first six months and post one year of the COVID-19 pandemic.

3.2. *Sub-haplotype distribution of SARS-CoV-2 variants in the areas with low and high COVID-19 vaccination rates*

Since Our World in Data (OWID, https://ourworldindata.org/) reported on May 6, 2021, the US, Canada, and Germany reached up 50% partly vaccinated rate (at least one-dose of COVID-19 vaccination), which were categorized as the high vaccination rate group. Meanwhile, the partly vaccinated rate of India, Malaysia, Taiwan, and Vietnam were lower than 20%, defined as the low partly vaccinated rate group. The sub-haplotype profiles of SARS-CoV-2 variants sequenced in May 2021were further examined in these two groups (Fig. 3, Table 2). In the haplotype profiling of the variants collected in May 2021, the haplotype 2A occupied 10% in the low partly vaccinated rate group but up to 63.54% in the high partly vaccinated rate group. Notably, the sub-

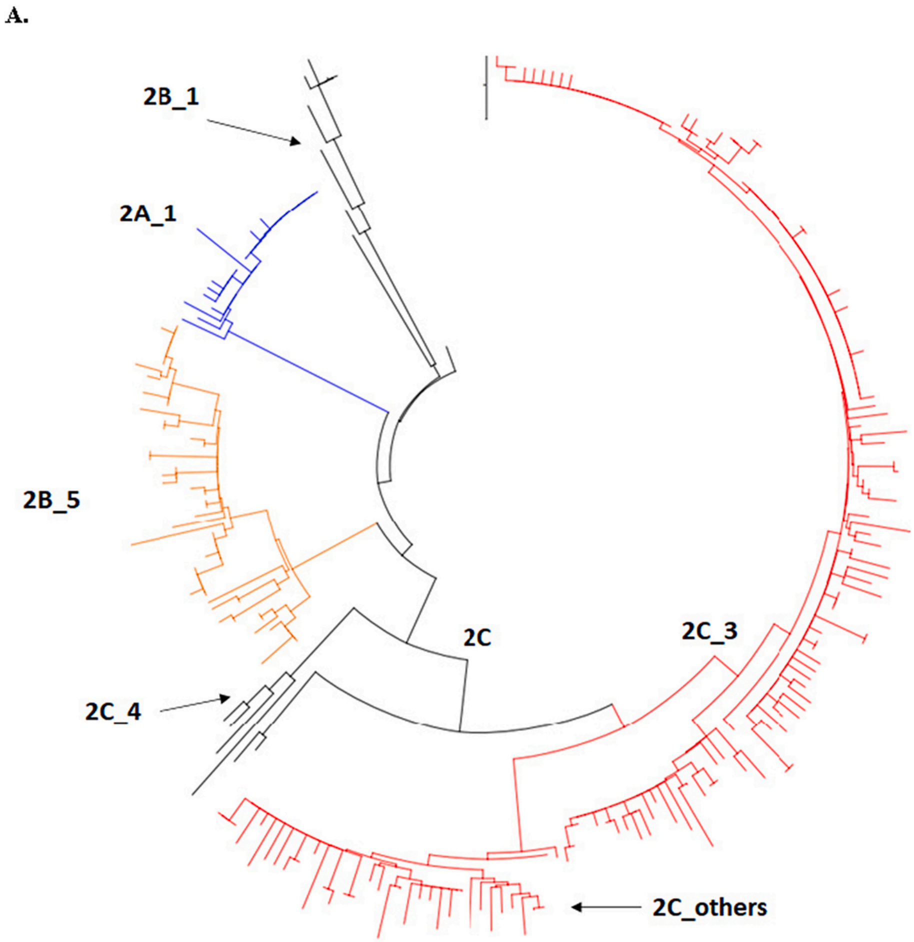

Fig. 3. Sub-haplotype distribution of SARS-CoV-2 variants collected from low and high partly vaccinated rate groups using NJ phylogenetic analysis. The evolutionary relationships among SARS-CoV-2 variants were inferred using the NJ method. The haplotype and sub-haplotype profiles of complete proteomic sequences of the variants in the low- (A) and high- (B) partly vaccinated rate groups in May 2021 were performed by following the steps of alignment protein-coding setting, phylogenetic analysis, and clade classification. The main haplotypes and sub-haplotypes in both groups were marked in different colors.

haplotype 2A_1 variant was dominant in both groups, but the sub-haplotype 2A_3 (P1, Gamma) variant showed a small proportion (5.19%) in the high partly vaccinated rate group. While, haplotype 2C variant had the highest proportion (66.8%) in the low partly vaccinated rate group, but the lowest proportion (11.4%) in the high partly vaccinated rate group. Notably, the haplotype 2C_3 belonged to Pangolin lineage B.1.617.2 (WHO label Delta) made up approximately 54.8% and 11.14% in low and high partly vaccinated rate groups, respectively (Fig. 3A and B, Table 2). The comparison among the sub-haplotype profiles, sub-haplotype 2A_1, increased from around 15.23% in the global variants in January–March to over 58% in the high partly vaccinated rate group in May 2021 (Figs. 2 and 3, Tables 1 and 2). Moreover, haplotype 2B was reduced from just over 38.81% in the global variants collected in January–March to around 14.05% in the high partly vaccinated rate group in May 2021. However, new sub-haplotypes 2B_4 (B.1.526, Iota) and 2D (B.1.525, Eta) emerged in the high v partly vaccinated rate group; the novel sub-haplotype 2B_5 (B.1.1.466.2) appeared in the low partly vaccinated rate group in May 2021. The results indicated that amino acid substitutions in SARS-CoV-2 variants caused the noticeable change of haplotypes and the emergence of new sub-haplotypes in the areas with different vaccination rates.

3.3. *Unique features in new sub-haplotype variants in May 2021*

The haplotype 2A variant contained four mutations at nsp12 P323L, Spike D614G, and N R203K, G204R merged post-6-month outbreak (Fig. 4, Table 1). Among the sub-haplotypes of haplotype 2A variants, sub-haplotype 2A_1 was the most prevalent sub-haplotype in the partly vaccinated rate group, containing three mutations in nsp3 (T183I, A890D, I1412T), seven new mutations in Spike protein (V143del, N501Y, A570, P681H, T716I, D614G, S982A, D1118H), two mutations in ORF8 (R52I, Y73C), and one new mutation in N (S235F) (Figs. 4 and 5C, Table 1). Haplotype 2B containing the mutations at nsp12 P323L, Spike D614G, ORF3a Q57H, and nsp2 I85I emerged post 6-month spread, evolved into five sub-haplotypes, such as additional mutations at nsp3 K837N, spike protein D80A and D215G in sub-haplotype 2B_1, spike L5F, T95I, D253G in sub-haplotype 2B_4, and spike N439K, P681R in sub-haplotype 2B_5 (Figs. 4 and 5D, Table 1). Significantly, haplotype 2C variant (66.8%) was leading in the partly vaccinated rate group, which sub-haplotype 2C_3 contained two unique mutations in spike protein (EFR156-158del, L452R, P681R), one mutation in ORF3a (S26L), and one mutation in N (R203M) (Figs. 4 and 5E, Table 1). Significantly, new sub-haplotypes 2A_1 and 2C_3 had meaningful amino acid substitutions like T478K, N501Y, and P681H in SARS-COV-2 spike protein, locating ACE2-specific recognition site and the neutralization epitopes, and near the protease cleavage site (Fig. 5). The results revealed that new sub-haplotypes had an alteration in the properties of the spike protein, including receptor binding, neutralization by immunized sera, and even virus entry.

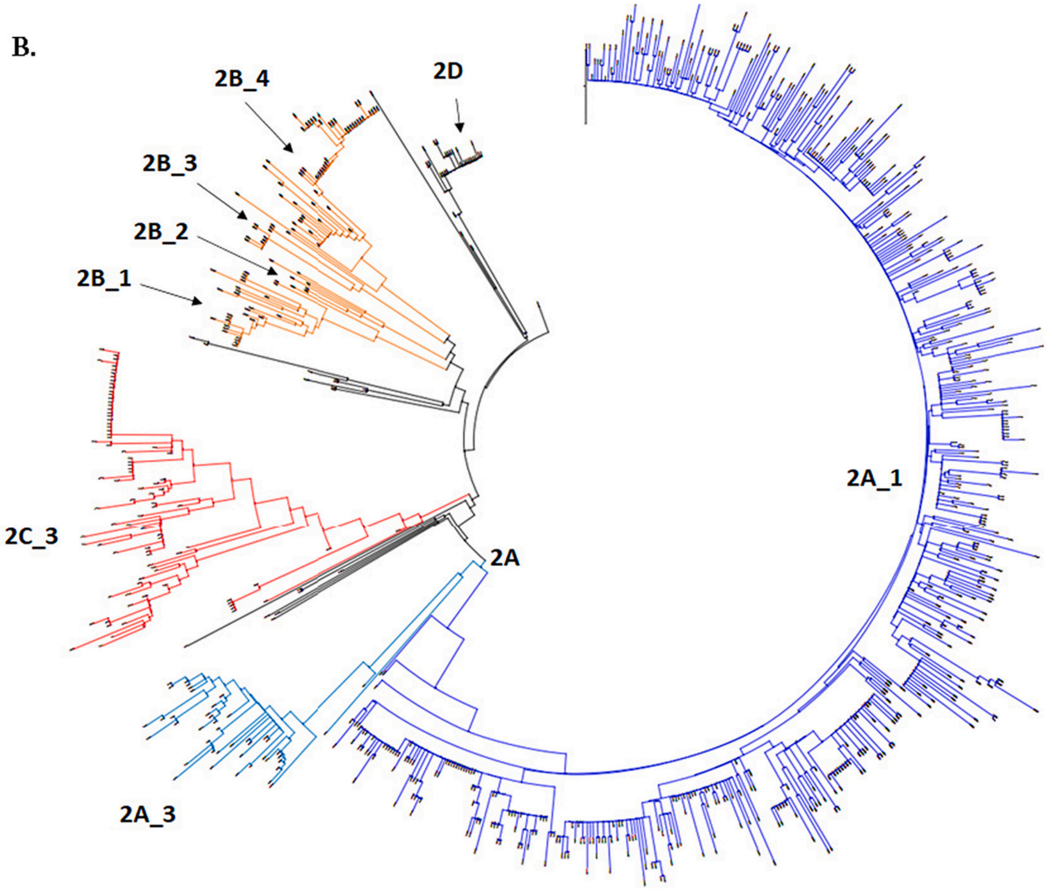

Fig. 3. (*continued*).

Table 2
Haplotypes and sub-haplotype distribution of SARS-CoV-2 variants in low and high partly vaccinated rate countries in May 2021.

Haplotype/Sub-haplotype	Low partly vaccinated rate group[a]	High partly vaccinated rate group[b]
1	0	0
2	100% (250/250)	100% (790/790)
2A	10% (25/250)	63.54% (502/790)
2A_1	10.0% (25/250)	58.35% (461/790)
2A_3	0	5.19% (41/790)
2B	22.80% (57/250)	14.05% (111/790)
2B_1	2.80% (7/250)	3.80% (30/790)
2B_2	0	0.76% (6/790)
2B_3	0	1.39% (11/790)
2B_4	0	6.83% (54/790)
2B_5	20.00% (50/250)	0
2B_others	0	1.27% (10/790)
2C	66.80% (167/250)	11.14% (88/790)
2C_3	54.80% (137/250)	11.14% (88/790)
2C_4	1.6% (4/250)	0
2C_others	10.40% (26/250)	0
2D	0	3.29% (26/790)
2 others	0.40% (1/250)	7.97% (63/790)

[a] SARS-CoV-2 variants from India, Malaysia, Taiwan, and Vietnam on May 2021, in which the partly vaccinated rate was < 20%.

[b] SARS-CoV-2 variants from the US, Canada, and Germany on May 2021, in which the partly vaccinated rate was > 50%.

4. Discussion

Compared with the profiles of haplotypes post-6-month outbreak of COVID-19, SARS-CoV-2 variants accumulated many new amino acid substitutions post one-year pandemic and emerged many distinct sub-haplotypes in low and high partly vaccinated rate countries (Figs. 2 and 3, Tables 1 and 2). For instance, the haplotype 2A variants defined post-6-month outbreak in our prior report (Bui et al., 2021) progressed into three sub-haplotypes, in which sub-haplotype 2A_1 (B.1.1.7, Alpha) variants emerged in January 2021 had 17 new mutations, became highly prevalent in the high partly vaccinated rate countries (US, Canada, and Germany) in May 2021 (Table 2). Especially, sub-haplotype 2C_3 (B.1.617.2, Delta) variants with 22 to 24 amino acid changes unexpectedly became the mainly circulating variant in the low vaccination-rate group (particularly in India) in May 2021 (Figs. 2 and 3, Tables 1 and 2). Many shreds of evidence were found that only the variants bearing key mutations with critical biological functions demonstrated high transmissibility (Zhou and Wang, 2021), suggesting that these critical mutations had a crucial role in viral biology and replication. In terms of biological function effect, the most concern mutations belong to spike protein, which increased transmission and reduced the current vaccine and antibody resistance efficiency. During the early pandemic stage, the first concern mutation in spike protein was D614G that appeared and is emerging as an increasingly common variant in the world. After a year-round spread, there are a lot of new mutations accumulated in spike protein. Our complete genome sequence dataset of SARS-CoV-2 variants in 2021, sub-haplotype 2A_1 variant had two deleted-amino acids (H69del, V143del) and six amino acid changes in spike protein (N501Y, A570D, P681H, T716I, S982A, D1118H), as demonstrated in many studies that the deleted at position 69–70, N501Y, P681H mutation had increased viral transmissibility (Wang et al., 2021). The mutation N501Y happens in one of six essential residues of RBD for the binding capacity of SARS-CoV-2 (L455, F486, Q493, S494, N501, and Y505) (Andersen et al., 2020). Under the accumulating

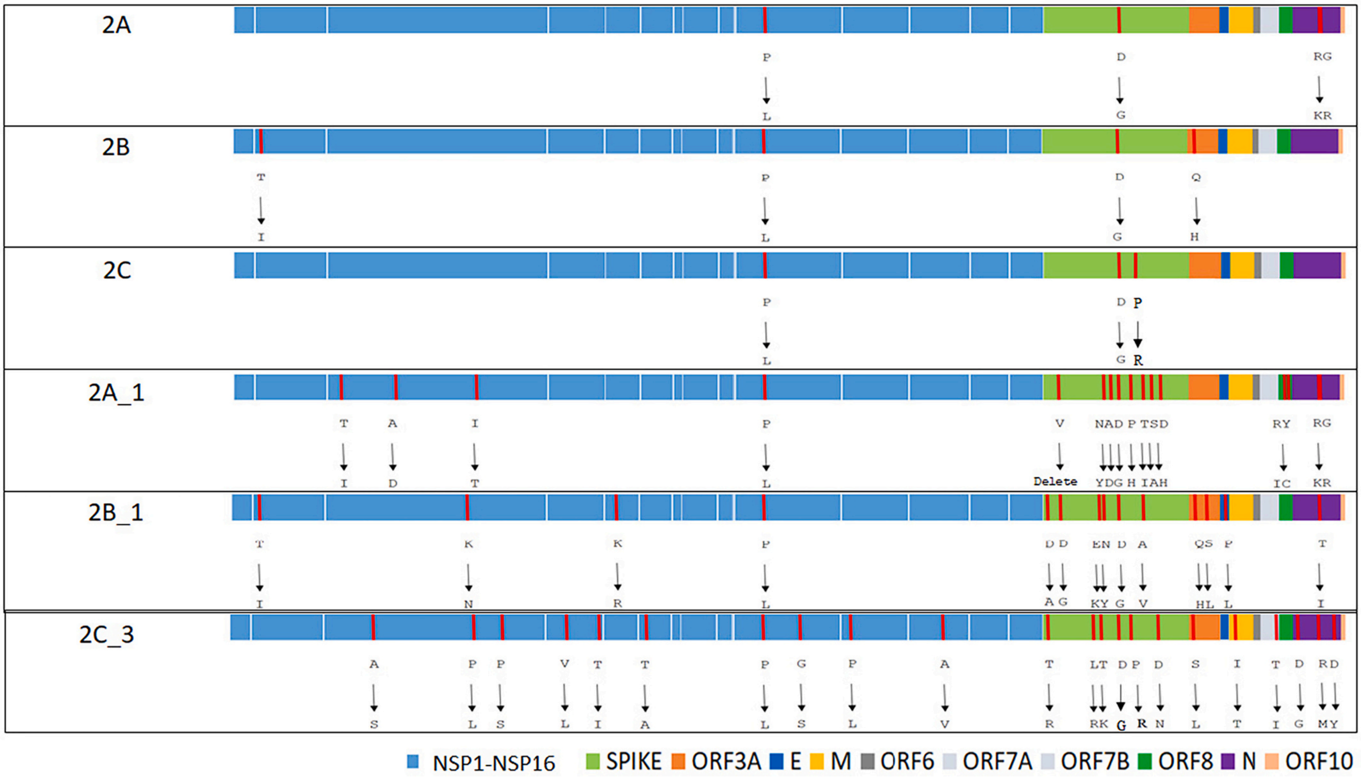

Fig. 4. Amino acid substitution in the proteome of main haplotypes and sub-haplotypes. Non-structure proteins were marked as blue bars and separated by white lines (nsp1 to nsp16). The structural proteins S (green bar), E (light blue bar), M (yellow bar), and N (violet bar), as well as the accessory proteins (ORF3a, ORF7, ORF8, ORF10), were marked as different colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of many mutations affected to the transmission, sub-haplotype 2A_1 variant was found increasing over the survey period from January–March to May 2021. Notably, sub-haplotypes 2A_3 and 2B_1 had the mutations at K417T, and E484K in the spike was a small proportion in the high partly vaccinated rate group in May 2021, which was found to decrease neutralization of human immune serum (Greaney et al., 2021; Liu et al., 2021) and cause severe disease even in patients that have been previously infected (Cele et al., 2021; Wibmer et al., 2021). Although sub-haplotypes 2A_3 and 2B_1 were less easily neutralized than original SARS-CoV-2, there was no good evidence that these variants could evade the vaccination. The new sub-haplotypes appeared in the amino acid substitutions in B- and T-cell epitopes of Spike protein predicted (Bhattacharya et al., 2020), which might alter the immune recognition potential SARS-CoV2 epitopes by HLA alleles among different populations (Bose et al., 2021). Therefore, the new sub-haplotypes could trigger the differential T-cell-based immunological responses to SARS-CoV-2 infection in different ethnic populations and reduce the efficacy of SARS-CoV-2 epitope-based vaccines in the future.

Sub-haplotype 2C_3 was still high emerging in the group of countries lower than 20% partly vaccinated rate and occupied a part in the group of countries higher than 50% partly vaccinated rate (Table 2). Sub-haplotype 2C_3, which belonged to the B.1.617.2 in India (Dhar et al., 2021), became the most commonly reported variant in this country from mid-April 2021. Sub-haplotype 2C_3 had three deleted-amino acids (EFR 156,157,158 deleted) six amino acid changes (T19R, L452R, T478K, D614G, P681R, D950N) in spike protein (Figs. 4 and 5). The amino acid change P681R near the protease cleavage site might relate with the spike protein stability (Chakraborty et al., 2021; Wrobel et al., 2020a,b), and the mutations L452RT and 478 K could alter B-cell epitopes of the spike protein to weaken the binding ability of serum or convalescent patients' antibodies and even the transmission (Bhattacharya et al., 2020; Greaney et al., 2021; Li et al., 2020). Notably, sub-haplotype 2C_3 also showed a novel mutation G671S in nsp12 (RNA-dependent RNA polymerase, RdRp) that was the target for the antiviral drug in treatment such as remdesivir in many countries (Kupferschmidt and Cohen, 2020). Moreover, sub-haplotype 2C_3 contained three mutations on N protein (N D63G, N R203M, N D377Y). N protein, an essential role in genome packaging, has been demonstrated to have a role in intracellular protein transport, interference in host translation (Cubuk et al., 2021; Zeng et al., 2020). The mutations RG203,204KR in N protein were detected in haplotype 2A in the early stage of pandemic and continuously fast-spreading in 2021, resulting in the increase of the infectivity in human lung cell line and the enhancement in damaging lung blood vessels in hamster model (Wu et al., 2021). In this case, it is interesting to demonstrate the role of the mutations in the N protein of sub-haplotype 2C_3 variant in the correlation with serious symptoms.

Among haplotype 2B variants, sub-haplotype 2B_1 that belonged to lineage B.1.351 and labeled as Beta by WHO still keep a small proportion during the survey time in 2021. The sub-haplotypes 2B_2 (Pangolin lineage B.1.429, Epsilon label by WHO) and 2B_3 (Pangolin lineage B.1.596) were 0.76% and 1.39%, respectively, in the high partly vaccinated rate group in May 2021. There are two new haplotypes found in the May dataset. The haplotype 2B_4 occupied 6.83% in the high partly vaccinated rate group, belonged to lineage B.1.526 (Iota), and the haplotype 2B_5 occupied 20% in Malaysia in May 2021. Sub-haplotype 2B_4 haplotype contained one mutation in nsp4 L89F, three other mutations in spike protein (L5F, T95I, D253G), one mutation in ORF3a (P42L), one mutation in ORF8 (T11I). Of note, the dataset on May 2021, we found a haplotype 2B_5 contained three mutations in nsp3 (S126L, T350I, P822L), one mutation in nsp6 (L75F), one mutation in nsp13 (S259L), three mutations in spike protein (N439K, P681R, G1251V), and one mutation in N (T205I). The E protein, the smallest protein with only 75 amino acids, has been demonstrated the in vitro and in vivo ability to cause acute respiratory distress syndromes (ARDS)- like damage (Xia

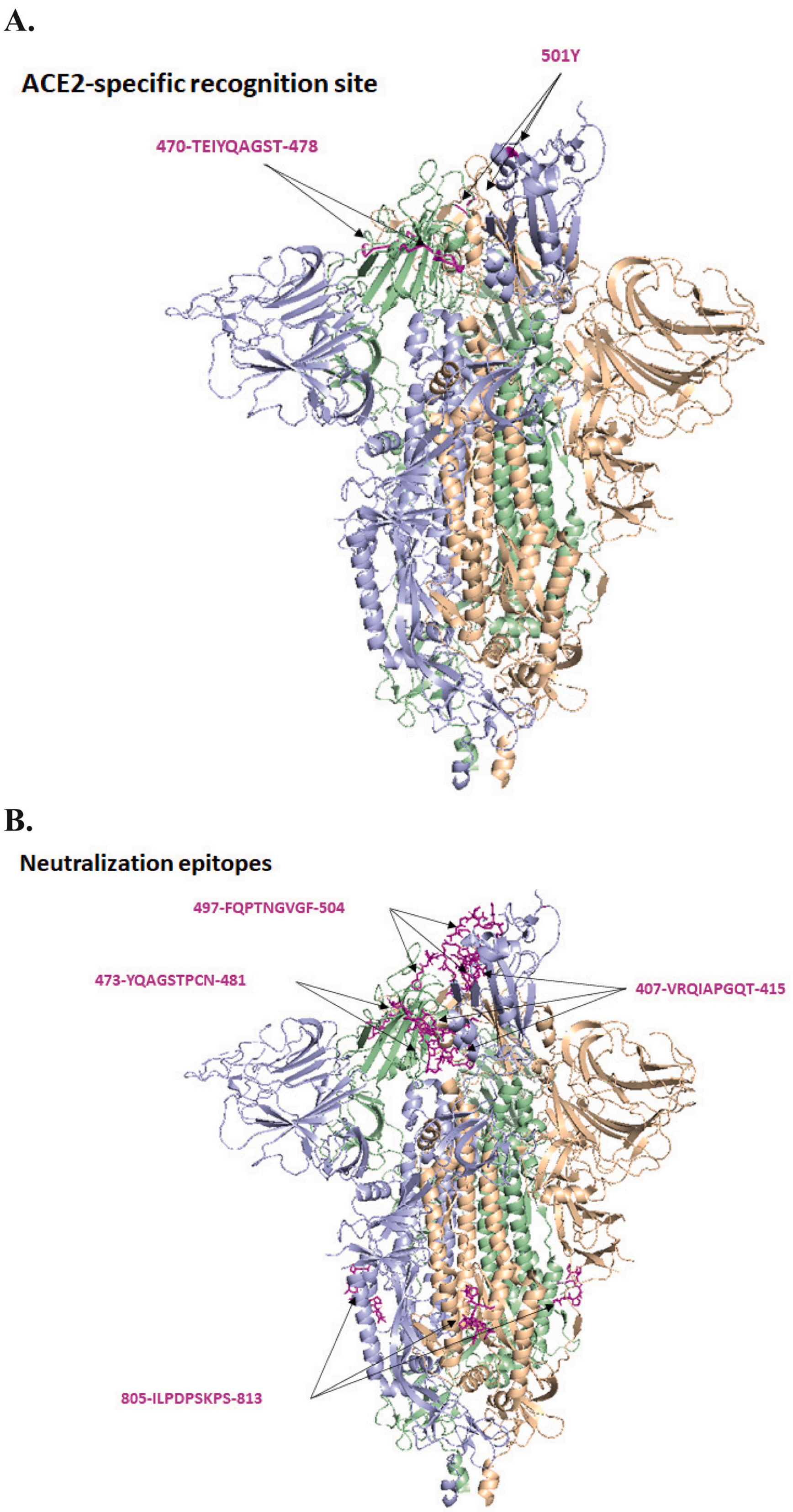

Fig. 5. Amino acid substitutions within spike trimer of main sub-haplotypes from the SARS-CoV-2 variants in May 2021. Cartoons displayed the tightly closed spike protein trimer with packed fusion peptide (RCSB Protein Data Bank ID 7DK3) view by the PyMOL software. The pink spheres are indicated the mutation belongs to the epitope region. The red spheres are indicated mutation in other regions. The ACE2-specific recognition site (A) and four B-cell epitope regions (B) were also marked as pink ribbon. Amino acid substitutions within the spike protein trimer of sub-haplotypes 2A_1 (C), 2B_1 (D), and 2C_3 (E) were displayed by the Spacefill command; the changes in the proposed neutralization epitopes of the spike protein were marked in purple colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

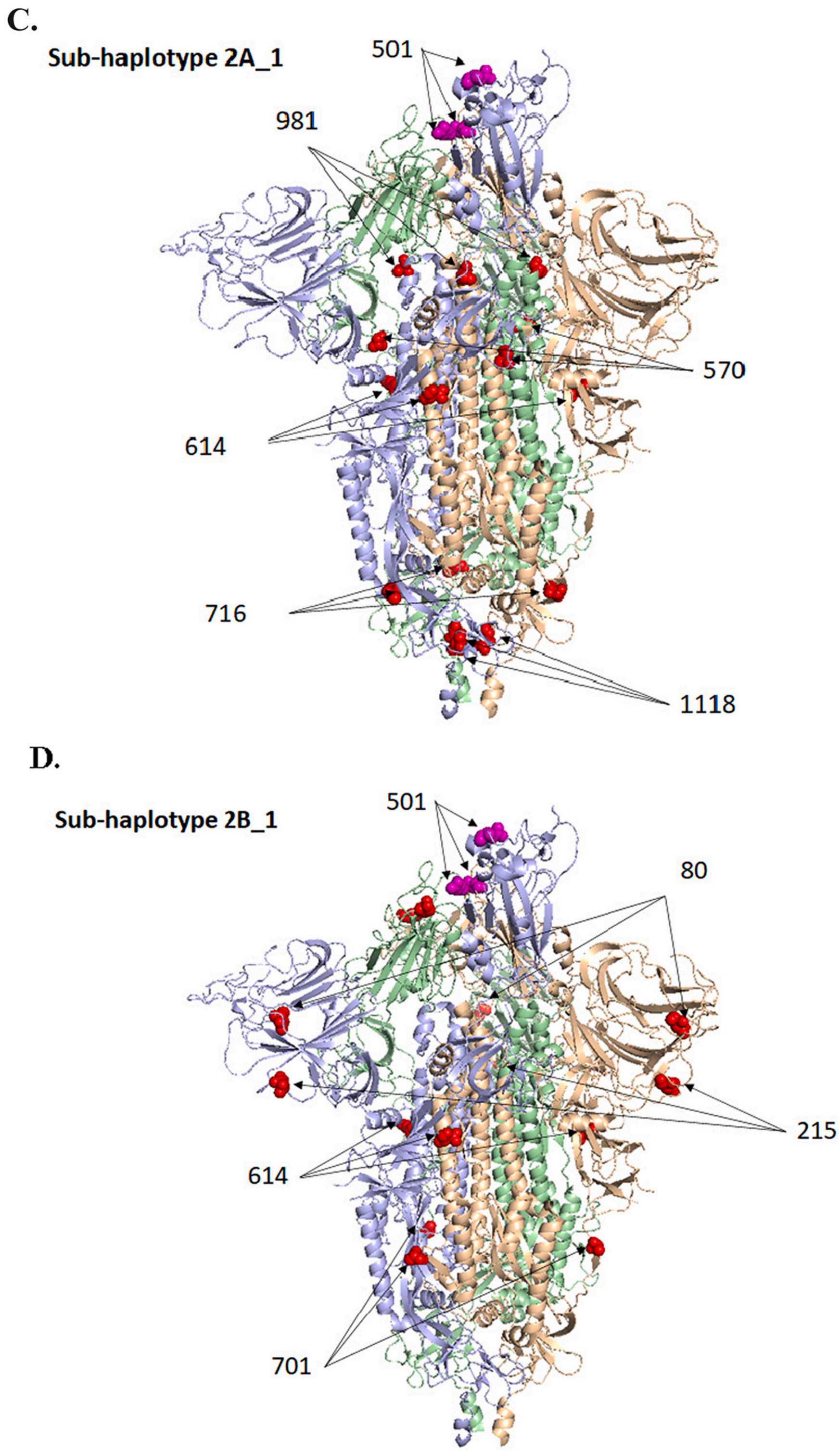

Fig. 5. (*continued*).

et al., 2021), such as the effect of the mutations on the channel activity and the capability of E protein-induced cell death (Xia et al., 2021). Over the rise of vaccine rate worldwide, the haplotype 2B was reduced in May 2021.

The study provided the usefulness of integrating surveillance methods to monitor haplotype and sub-haplotype profiling of SARS-CoV-2 variant during the ongoing COVID-19 pandemic and vaccination. The geographic tracking of new SARS-CoV-2 variants like sub-haplotype 2C_3 is helpful to control and defeat the novel variant in further COVID-19 outbreaks. Our data might be useful in informing and consolidating the investigation of SARS-CoV-2 evolution and the response to vaccination strategies, such as the increase of global vaccination coverage, and the *booster vaccine dose when* human and SARS-CoV-2 coexist.

Declaration of Competing Interest

All authors declare no competing financial interest.

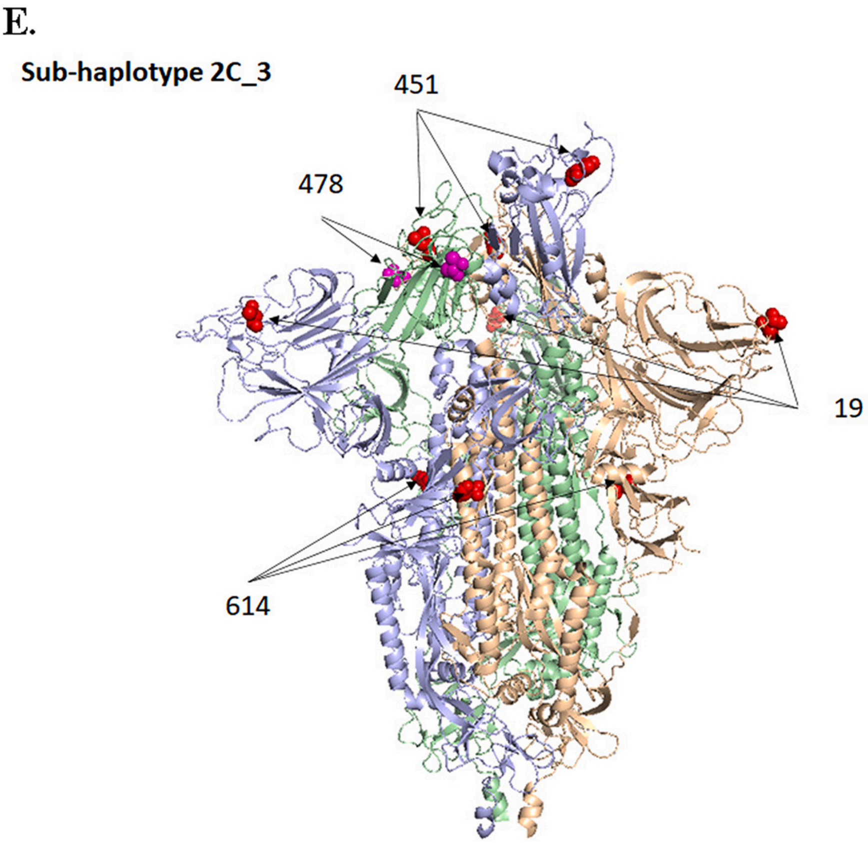

Fig. 5. (*continued*).

Acknowledgments

This work was financially supported by China Medical University, Taiwan (CMU109-ASIA-07, CMU109-MF-57, and CMU109-S-09) and funded by grants from the Ministry of Science and Technology, Taiwan (MOST108-2320-B-039-039-MY3; MOST110-2923-B-039 -001 -MY3). The experiments and data analysis were performed in part using the Medical Research Core Facilities Center, Office of Research & Development at China Medical University, Taiwan.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.105164.

References

Ahmad, J., Ikram, S., Ahmad, F., Rehman, I.U., Mushtaq, M., 2020. SARS-CoV-2 RNA dependent RNA polymerase (RdRp) – A drug repurposing study. Heliyon 6, e04502. https://doi.org/10.1016/j.heliyon.2020.e04502.

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. The proximal origin of SARS-CoV-2. Nat. Med. 264, 450–452. https://doi.org/10.1038/s41591-020-0820-9.

Bhattacharya, M., Sharma, A.R., Mallick, B., Sharma, G., Lee, S.S., Chakraborty, C., 2020. Immunoinformatics approach to understand molecular interaction between multi-epitopic regions of SARS-CoV-2 spike-protein with TLR4/MD-2 complex. Infect. Genet. Evol. 85, 104587 https://doi.org/10.1016/j.meegid.2020.104587.

Bose, T., Pant, N., Pinna, N.K., Bhar, S., Dutta, A., Mande, S.S., 2021. Does immune recognition of SARS-CoV2 epitopes vary between different ethnic groups? Virus Res. 305, 198579 https://doi.org/10.1016/j.virusres.2021.198579.

Bui, N.N., Lin, Y.T., Huang, S.H., Lin, C.W., 2021. The extent of molecular variation in novel SARS-CoV-2 after the six-month global spread. Infect. Genet. Evol. 91 https://doi.org/10.1016/j.meegid.2021.104800.

Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., Yang, M., Xing, X., Wang, Y., Li, M., Zhou, B., Wang, X., Wang, J., Xue, Y., 2021. The architecture of the SARS-CoV-2 RNA genome inside virion. Nat. Commun. 121, 1–14. https://doi.org/10.1038/s41467-021-22785-x.

Cele, S., Gazy, I., Jackson, L., Hwa, S.-H., Tegally, H., Lustig, G., Giandhari, J., Pillay, S., Wilkinson, E., Naidoo, Y., Karim, F., Ganga, Y., Khan, K., Bernstein, M., Balazs, A.B., Gosnell, B.I., Hanekom, W., Moosa, M.-Y.S., Lessells, R.J., de Oliveira, T., Sigal, A., 2021. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. Nature 5937857, 142–146. https://doi.org/10.1038/s41586-021-03471-w.

Chakraborty, C., Sharma, A.R., Bhattacharya, M., Saha, R.P., Ghosh, S., Biswas, S., Samanta, S., Sharma, G., Agoramoorthy, G., Lee, S.S., 2021. SARS-CoV-2 and other human coronaviruses: mapping of protease recognition sites, antigenic variation of spike protein and their grouping through molecular phylogenetics. Infect. Genet. Evol. 89, 104729 https://doi.org/10.1016/J.MEEGID.2021.104729.

Collier, D.A., De Marco, A., Ferreira, I.A.T.M., Meng, B., Datir, R., Walls, A.C., Kemp, S., Bassi, J., Pinto, D., Fregni, C.S., Bianchi, S., Tortorici, M.A., Bowen, J., Culap, K., Jaconi, S., Cameroni, E., Snell, G., Pizzuto, M.S., Pellanda, A.F., Garzoni, C., Riva, A., Elmer, A., Kingston, N., Graves, B., McCoy, L.E., Smith, K.G.C., Bradley, J.R., Temperton, N., Lourdes Ceron-Gutierrez, L., Barcenas-Morales, G., Harvey, W., Virgin, H.W., Lanzavecchia, A., Piccoli, L., Doffinger, R., Wills, M., Veesler, D., Corti, D., Gupta, R.K., 2021. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. Nature 10, 1–10. https://doi.org/10.1038/s41586-021-03412-7.

Cubuk, J., Alston, J.J., Incicco, J.J., Singh, S., Stuchell-Brereton, M.D., Ward, M.D., Zimmerman, M.I., Vithani, N., Griffith, D., Wagoner, J.A., Bowman, G.R., Hall, K.B., Soranno, A., Holehouse, A.S., 2021. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. Nat. Commun. 121, 1–17. https://doi.org/10.1038/s41467-021-21953-3.

Dhar, M.S., Marwal, R., Radhakrishnan, V., Ponnusamy, K., Jolly, B., Bhoyar, R.C., Fatihi, S., Datta, M., Singh, P., Sharma, U., Ujjainia, R., Naushin, S., Bhateja, N., Divakar, M.K., Sardana, V., Singh, M.K., Imran, M., Senthivel, V., Maurya, R., Jha, N., Mehta, P., Rophina, M., Arvinden, V., Chaudhary, U., Thukral, L., Pandey, R., Dash, D., Faruq, M., Lall, H., Gogia, H., Madan, P., Kulkarni, S., Chauhan, H., Sengupta, S., Kabra, S., (INSACOG), T.I.S.-C.-2 G.C, Singh, S.K., Agrawal, A., Rakshit, P., 2021. Genomic characterization and Epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. medRxiv. https://doi.org/10.1101/2021.06.02.21258076, 2021.06.02.21258076.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution (N. Y) 39, 783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.

Garcia-Beltran, W.F., Lam, E.C., St Denis, K., Nitido, A.D., Garcia, Z.H., Hauser, B.M., Feldman, J., Pavlovic, M.N., Gregory, D.J., Poznansky, M.C., Sigal, A., Schmidt, A.G., Iafrate, A.J., Naranbhai, V., Balazs, A.B., 2021. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell 184, 2372–2383.e9. https://doi.org/10.1016/j.cell.2021.03.013.

Greaney, A.J., Starr, T.N., Gilchuk, P., Zost, S.J., Binshtein, E., Loes, A.N., Hilton, S.K., Huddleston, J., Eguia, R., Crawford, K.H.D., Dingens, A.S., Nargi, R.S., Sutton, R.E., Suryadevara, N., Rothlauf, P.W., Liu, Z., Whelan, S.P.J., Carnahan, R.H., Crowe, J.E., Bloom, J.D., 2021. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host Microbe 29, 44–57.e9. https://doi.org/10.1016/J.CHOM.2020.11.007.

Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., Cramer, P., 2020. Structure of replicating SARS-CoV-2 polymerase. Nature 584, 154–156. https://doi.org/10.1038/s41586-020-2368-8.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Angyal, A., Brown, R.L., Carrilero, L., Green, L.R., Groves, D.C., Johnson, K.J., Keeley, A.J., Lindsey, B.B., Parsons, P.J., Raza, M., Rowland-Jones, S., Smith, N., Tucker, R.M., Wang, D., Wyles, M.D., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182, 812–827.e19. https://doi.org/10.1016/j.cell.2020.06.043.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Kupferschmidt, K., Cohen, J., 2020. Race to find COVID-19 treatments accelerates. Science 367, 1412–1413. https://doi.org/10.1126/science.367.6485.1412.

Li, Q., Wu, J., Nie, J., Zhang, Li, Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., Qin, H., Wang, M., Lu, Q., Li, Xiaoyu, Sun, Q., Liu, J., Zhang, Linqi, Li, Xuguang, Huang, W., Wang, Y., 2020. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell 182, 1284–1294.e9. https://doi.org/10.1016/J.CELL.2020.07.012.

Liu, Z., VanBlargan, L.A., Bloyet, L.M., Rothlauf, P.W., Chen, R.E., Stumpf, S., Zhao, H., Errico, J.M., Theel, E.S., Liebeskind, M.J., Alford, B., Buchser, W.J., Ellebedy, A.H., Fremont, D.H., Diamond, M.S., Whelan, S.P.J., 2021. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe 29, 477–488.e4. https://doi.org/10.1016/J.CHOM.2021.01.014.

Müller, C., Schulte, F.W., Lange-Grünweller, K., Obermann, W., Madhugiri, R., Pleschka, S., Ziebuhr, J., Hartmann, R.K., Grünweller, A., 2018. Broad-spectrum antiviral activity of the eIF4A inhibitor silvestrol against corona- and picornaviruses. Antivir. Res. 150, 123–129. https://doi.org/10.1016/j.antiviral.2017.12.010.

Naqvi, A.A.T., Fatima, K., Mohammad, T., Fatima, U., Singh, I.K., Singh, A., Atif, S.M., Hariprasad, G., Hasan, G.M., Hassan, M.I., 2020. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. Biochim. Biophys. Acta Mol. basis Dis. 1866, 165878 https://doi.org/10.1016/j.bbadis.2020.165878.

Ogawa, J., Zhu, W., Tonnu, N., Singer, O., Hunter, T., Ryan, A.L., Pao, G.M., 2020. The D614G mutation in the SARS-CoV2 spike protein increases infectivity in an ACE2 receptor dependent manner. bioRxiv Prepr. Serv. Biol. https://doi.org/10.1101/2020.07.21.214932.

Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R.C., Zella, D., Ippodrino, R., 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J. Transl. Med. 18, 179. https://doi.org/10.1186/s12967-020-02344-6.

Parlikar, A., Kalia, K., Sinha, S., Patnaik, S., Sharma, N., Vemuri, S.G., Sharma, G., 2020. Understanding genomic diversity, pan-genome, and evolution of SARS-CoV-2. PeerJ 8, e9576. https://doi.org/10.7717/peerj.9576.

Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R., Moret, B.M., Stamatakis, A., 2010. How many bootstrap replicates are necessary? J. Comput. Biol. 17, 337–354. https://doi.org/10.1089/cmb.2009.0179.

Phan, T., 2020. Genetic diversity and evolution of SARS-CoV-2. Infect. Genet. Evol. 81, 104260 https://doi.org/10.1016/j.meegid.2020.104260.

Ren, Y., Shu, T., Wu, D., Mu, J., Wang, C., Huang, M., Han, Y., Zhang, X.-Y., Zhou, W., Qiu, Y., Zhou, X., 2020. The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. Cell. Mol. Immunol. 17, 881–883. https://doi.org/10.1038/s41423-020-0485-9.

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., MacDonald, B., Beltekian, D., Dattani, S., Roser, M., 2021. Coronavirus Pandemic (COVID-19). Published online at OurWorldInData.org.. https://ourworldindata.org/covid-vaccinations. Online Resource.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Shen, X., Tang, H., McDanal, C., Wagh, K., Fischer, W., Theiler, J., Yoon, H., Li, D., Haynes, B.F., Sanders, K.O., Gnanakaran, S., Hengartner, N., Pajon, R., Smith, G., Glenn, G.M., Korber, B., Montefiori, D.C., 2021. SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. Cell Host Microbe 29, 529–539.e3. https://doi.org/10.1016/j.chom.2021.03.002.

Sola, F.A., S, Z., L, E., 2015. Continuous and discontinuous RNA synthesis in coronaviruses. Annu. Rev. Virol. 2, 265–288. https://doi.org/10.1146/ANNUREV-VIROLOGY-100114-055218.

Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C.V, Boyd, O., Loman, N.J., McCrone, J.T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D.K., Gaythorpe, K., Groves, N., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., Ferguson, N.M., 2021. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature 593, 266–269. https://doi.org/10.1038/s41586-021-03470-x.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 181, 281–292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., Graham, B.S., Mascola, J.R., Chang, J.Y., Yin, M.T., Sobieszczyk, M., Kyratsous, C.A., Shapiro, L., Sheng, Z., Huang, Y., Ho, D.D., 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 593, 130–135. https://doi.org/10.1038/s41586-021-03398-2.

Wibmer, C.K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Oosthuysen, B., Lambson, B.E., de Oliveira, T., Vermeulen, M., van der Berg, K., Rossouw, T., Boswell, M., Ueckermann, V., Meiring, S., von Gottberg, A., Cohen, C., Morris, L., Bhiman, J.N., Moore, P.L., 2021. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. Nat. Med. 274, 622–625. https://doi.org/10.1038/s41591-021-01285-x.

Wrobel, D.J.B., P, X., C, R., SR, M., PB, R., JJ, S., SJ, G., 2020a. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. Nat. Struct. Mol. Biol. 27, 763–767. https://doi.org/10.1038/S41594-020-0468-7.

Wrobel, D.J.B., P, X., C, R., SR, M., PB, R., JJ, S., SJ, G., 2020b. Author correction: SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. Nat. Struct. Mol. Biol. 27, 1001. https://doi.org/10.1038/S41594-020-0509-2.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C., Zhang, Y.-Z., 2020. A new coronavirus associated with human respiratory disease in China. Nat. 579, 265–269. https://doi.org/10.1038/s41586-020-2008-3.

Wu, H., Xing, N., Meng, K., Fu, B., Xue, W., Dong, P., Xiao, Y., Liu, G., Luo, H., Zhu, W., Lin, X., Meng, G., Zhu, Z., 2021. Nucleocapsid mutation R203K/G204R increases the infectivity, fitness and virulence of SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2021.05.24.445386.

Xia, B., Shen, X., He, Y., Pan, X., Liu, F.-L., Wang, Y., Yang, F., Fang, S., Wu, Y., Duan, Z., Zuo, X., Xie, Z., Jiang, X., Xu, L., Chi, H., Li, S., Meng, Q., Zhou, H., Zhou, Y., Cheng, X., Xin, X., Jin, L., Zhang, H.-L., Yu, D.-D., Li, M.-H., Feng, X.-L., Chen, J., Jiang, H., Xiao, G., Zheng, Y.-T., Zhang, L.-K., Shen, J., Li, J., Gao, Z., 2021. SARS-CoV-2 envelope protein causes acute respiratory distress syndrome (ARDS)-like pathological damages and constitutes an antiviral target. Cell Res. 2021 318, 847–860. https://doi.org/10.1038/s41422-021-00519-4.

Xu, C., Wang, Y., Liu, C., Zhang, C., Han, W., Hong, X., Wang, Y., Hong, Q., Wang, S., Zhao, Q., Wang, Y., Yang, Y., Chen, K., Zheng, W., Kong, L., Wang, F., Zuo, Q., Huang, Z., Cong, Y., 2021. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. Sci. Adv. 7, eabe5575. https://doi.org/10.1126/sciadv.abe5575.

Yang, H., Rao, Z., 2021. Structural biology of SARS-CoV-2 and implications for therapeutic development. Nat. Rev. Microbiol. 19, 685–700. https://doi.org/10.1038/s41579-021-00630-8.

Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Yunru, Liu, M., Mohammed, A., Zhao, C., Yang, Yun, Xie, J., Ding, C., Ma, X., Weng, J., Gao, Y., He, H., Jin, T., 2020. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochem. Biophys. Res. Commun. 527, 618–623. https://doi.org/10.1016/j.bbrc.2020.04.136.

Zhou, W., Wang, W., 2021. Fast-spreading SARS-CoV-2 variants: challenges to and new design strategies of COVID-19 vaccines. Signal Transduct. Target. Ther. 61, 1–6. https://doi.org/10.1038/s41392-021-00644-x.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273. https://doi.org/10.1038/s41586-020-2012-7.