Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion ΔH69/ΔV70

Kemp SA[1], Datir RP[1], Collier DA[1], Ferreira IATM[2,3], Carabelli A[3], Harvey W[5], Robertson DL[4], Gupta RK[2,3]

[1]Division of Infection and Immunity, University College London, London, UK.

[2]Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Cambridge, UK.

[3]Department of Medicine, University of Cambridge, Cambridge, UK.

[4]MRC - University of Glasgow Centre for Virus Research, Glasgow, UK.

[5]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK

Address for correspondence:
Ravindra K. Gupta
Cambridge Institute for Therapeutic Immunology and Infectious Diseases
Jeffrey Cheah Biomedical Centre
Puddicombe Way
Cambridge CB2 0AW, UK
Tel: +44 1223 331491
rkg20@cam.ac.uk

Key words: SARS-CoV-2; COVID-19; antibody escape; neutralising antibodies; mutation; evasion; resistance; fitness; evolution

Abstract

SARS-CoV-2 Spike amino acid replacements in the receptor binding domain (RBD) occur relatively frequently and some have a consequence for immune recognition. Here we report recurrent emergence and significant onward transmission of a six nucleotide deletion in the Spike gene, which results in loss of two amino acids: **Δ**H69/**Δ**V70. Of particular note this deletion often co-occurs with the receptor binding motif amino acid replacements N501Y, N439K and Y453F. In addition, we report a sub-lineage of over 350 sequences bearing seven spike mutations across the RBD (N501Y, A570D), S1 (**Δ**H69/V70) and S2 (P681H, T716I, S982A and D1118H) in England. Some of these mutations have possibly arisen as a result of the virus evolving from immune selection pressure in infected individuals. Enhanced surveillance for the **Δ**H69/**Δ**V70 deletion with and without RBD mutations should be considered as a priority.

Background

SARS-CoV-2's Spike surface glycoprotein engagement of ACE2 is essential for virus entry and infection[1], and the receptor is found in respiratory and gastrointestinal tracts[2]. Despite this critical interaction and related mutational constraints, it appears the RBD can tolerate mutations in this region[3,4], raising the real possibility of virus escape from vaccines and monoclonal antibodies. Spike mutants exhibiting reduced susceptibility to monoclonal antibodies have been identified in *in vitro* screens[5,6]. Some of these have been found in clinical isolates[7]. The unprecedented scale of whole genome SARS-CoV-2 sequencing has enabled identification and epidemiological analysis of transmission. As of December 11th, there were 246,534 SARS-CoV-2 sequences available in the GISAID initiative (https:gisaid.org/).

We recently documented *de novo* emergence of antibody escape mediated by Spike in an individual treated with convalescent plasma (CP), on the background of D614G[8]. Similarly, deletions in the NTD have been reported to provide escape for N-Terminal Domain-specific neutralising antibodies[11]. Dynamic changes in prevalence of Spike variants **Δ**H69/**Δ**V70 (an out of frame deletion) and D796H variant followed repeated use of CP, and in vitro the mutant displayed reduced susceptibility to the CP and multiple other sera, whilst retaining infectivity comparable to wild type[8]. We hypothesised that Spike **Δ**H69/**Δ**V70 arises either

as a compensatory change and/or antibody evasion mechanism as suggested for other NTD deletions[1], and therefore aimed to characterise specific circumstances around emergence of **Δ**H69/**Δ**V70 globally. Here we analysed the publicly available GISAID data for circulating SARS-CoV-2 sequences containing **Δ**H69/**Δ**V70.

Results

A deletion H69/V70 was present in over 3,000 sequences worldwide (2.5% of the available data) (Figure 1), and largely in Europe from where most of the sequences in GISAID are derived (Table 1). Many are from the UK and Denmark where sequencing rates are high compared to other countries. **Δ**H69/**Δ**V70 is observed in multiple different lineages, representing at least five independent acquisitions of the SARS-CoV-2 Spike **Δ**H69/**Δ**V70 deletion (Figure 1). The earliest samples that include the **Δ**H69/**Δ**V70 were detected in Thailand and Germany in January and February 2020, respectively, and are independent deletion events. The prevalence has since increased in other countries since August 2020 (Table 1). Further analysis of sequences revealed firstly that single deletions of either 69 or 70 were uncommon and secondly that some lineages of **Δ**H69/**Δ**V70 alone were present, as well as **Δ**H69/**Δ**V70 in the context of other mutations in Spike, specifically those in the RBD (Figure 1).

The structural impact of the double deletion was predicted by homology modelling of the spike NTD possessing **Δ**H69/**Δ**V70 using SWISS-MODEL. The **Δ**H69/**Δ**V70 deletion was predicted to alter the conformation of a protruding loop comprising residues 69-76, with the loop being predicted to be pulled in towards the NTD (Fig 2A). In the pre- and post-deletion conformations, the positions of the alpha carbons of residues 67 and 68 are roughly equivalent whereas the position of Ser71 in the post-deletion structure is estimated to have moved by approximately 6.7Å to approximately occupy the position of His69 in the pre-deletion structure. Concurrently, the positions of Gly72, Thr73, Asn74 and Gly75 are predicted to have changed by 6.5Å, 6.5Å, 4.7Å and 1.9Å respectively, with the overall effect of drawing these residues inwards, resulting in a less dramatically protruding loop; the position of Thr76 in the post-deletion model is roughly equivalent to its position in the pre-deletion structure.

We next examined the lineages where Spike mutations in the RBD were identified at high frequency, in particular co-occurring with N439K, an amino acid replacement recently reported to be expanding in Europe and detected across the world[3] (Figure 3, Supplementary figure 1). N439K appears to have reduced susceptibility to a small subset of monoclonals targeting the RBD, whilst retaining affinity for ACE2 *in vitro*[3]. The proportion of viruses with **Δ**H69/**Δ**V70 only increased from August 2020 when it co-occurred with the second N439K lineage[3] (Figure 3). As of November 26th, remarkably there were twice as many cumulative sequences with the deletion as compared to the single N439K (Figure 3). Due to their high sampling rates the country with the highest proportion of N439K+ **Δ**H69/**Δ**V70 versus N439K alone is England. The low levels of sequencing in most countries indicate N439K's prevalence could be relatively high[3]. In Scotland, where early growth of N439K was high (forming N439K lineage i that subsequently went extinct with other lineages after the lockdown[3]), there is now an inverse relationship with 546 versus 177 sequences for N439K and N439K+**Δ**H69/**Δ**V70respectively (as of November 26th). These differences therefore likely reflect differing epidemic growth characteristics and timings of the introductions the N439K variants with or without the deletion.

The second significant cluster with **Δ**H69/**Δ**V70 and RBD mutants involves Y453F, another RBD mutation that increases binding affinity to ACE2, along with F486L and N501T related to human-mink transmissions in Denmark[9] (Figure 4). This sub-lineage, termed 'Cluster 5' was part of a wider lineage in which the same deleted region (**Δ**H69/**Δ**V70) was observed. In Y453F lineages, the mutant virus demonstrates reduced susceptibility to sera from recovered COVID-19 patients (https://files.ssi.dk/Mink-cluster-5-short-report_AFO2). The **Δ**H69/**Δ**V70 was first detected in the Y453F background on August 24th and thus far appears limited to Danish sequences.

A third lineage containing the same out of frame deletion **Δ**H69/**Δ**V70 has arisen with another RBD mutation N501Y (Figure 5, supplementary figure 2). Based on its location it might be expected to escape antibodies similar to COV2-2499[5]. In addition, when SARS-CoV-2 was passaged in mice for adaptation purposes, N501Y emerged and increased pathogenicity[10]. Early sequences with N501Y alone were isolated both in Brazil and USA in April 2020. N501Y + **Δ**H69/**Δ**V70sequences appear to have been detected first in the UK in

September 2020, with the crude cumulative number of N501Y + ΔH69/ΔV70 mutated sequences now exceeding the single mutant (Figure 5). Of particular concern is a sub-lineage of around 350 sequences (Figure 6) bearing six spike mutations across the RBD (N501Y, A570D) and S2 (P681H, T716I, S982A and D1118H) as well as the ΔH69/ΔV70 in England (Figure 7). This cluster has a very long branch (Figure 6).

Discussion

We have presented data demonstrating multiple, independent, and circulating lineages of SARS-CoV-2 variants bearing a Spike ΔH69/ΔV70. This deletion spanning six nucleotides, is mostly due to an out of frame deletion of six nucleotides, has frequently followed receptor binding amino acid replacements (N501Y, N439K and Y453F that have been shown to reduce binding with monoclonal antibodies) and its prevalence is rising in parts of Europe, with the greatest increases since August 2020.

A recent analysis highlighted the potential for enhanced transmissibility of viruses with deletions in the N terminal domain, including ΔH69/ΔV70[11]. The potential for SARS-CoV-2 mutations to rapidly emerge and fix is exemplified by D614G, an amino acid replacement in S2 that alters linkages between S1 and S2 subunits on adjacent protomers as well as RBD orientation, infectivity, and transmission[12-14]. The example of D614G also demonstrates that mechanisms directly impacting important biological processes can be indirect. Similarly, a number of possible mechanistic explanations may underlie ΔH69/ΔV70. For example, the fact that it sits on an exposed surface might be suggestive of immune interactions and escape, although allosteric interactions could alternatively lead to higher infectivity.

The finding of a sub-lineage of over 350 sequences bearing seven spike mutations across the RBD (N501Y, A570D), S1 (ΔH69/ΔV70) and S2 (P681H, T716I, S982A and D1118H) in England requires careful monitoring. The detection of a high number of novel mutations suggests this lineage has either been introduced from a geographic region with very poor sampling or viral evolution may have occurred in a single individual in the context of a chronic infection[8].

Given the emergence of multiple clusters of variants carrying RBD mutations and the ΔH69/ΔV70 deletion, limitation of transmission takes on a renewed urgency. Concerted

global vaccination efforts with wide coverage should be accelerated. Continued emphasis on testing/tracing, social distancing and mask wearing are essential, with investment in other novel methods to limit transmission[15]. Detection of the deletion by rapid diagnostics should be a research priority as such tests could be used as a proxy for antibody escape mutations to inform surveillance at global scale.

Acknowledgements

RKG is supported by a Wellcome Trust Senior Fellowship in Clinical Science (WT108082AIA). DLR is funded by the MRC (MC UU 1201412). WH is funded by the MRC (MR/R024758/1). We thank Dr James Voss for the kind gift of HeLa cells stably expressing ACE2.

Conflicts of interest

RKG has received consulting fees from UMOVIS lab, Gilead Sciences and ViiV Healthcare, and a research grant from InvisiSmart Technologies.

Methods

Phylogenetic Analysis

All available full-genome SARS-CoV-2 sequences were downloaded from the GISAID database (http://gisaid.org/) [16] on 26th November. Duplicate and low-quality sequences (>5% N regions) were removed, leaving a dataset of 194,265 sequences with a length of >29,000bp. All sequences were realigned to the SARS-CoV-2 reference strain MN908947.3, using MAFFT v7.473 with automatic flavour selection and the --keeplength --addfragments options[17]. Major SARS-CoV-2 clade memberships were assigned to all sequences using the Nextclade server v0.9 (https://clades.nextstrain.org/).

Maximum likelihood phylogenetic trees were produced using the above curated dataset using IQ-TREE v2.1.2 [18]. Evolutionary model selection for trees were inferred using ModelFinder [19] and trees were estimated using the GTR+F+I model with 1000 ultrafast bootstrap replicates[20]. All trees were visualised with Figtree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), rooted on the SARS-CoV-2 reference sequence and nodes arranged in descending order. Nodes with bootstraps values of <50 were collapsed using an in-house script.

Pseudotype virus preparation

Viral vectors were prepared by transfection of 293T cells by using Fugene HD transfection reagent (Promega). 293T cells were transfected with a mixture of 11ul of Fugene HD, 1µg of pCDNAΔ19Spike-HA, 1ug of p8.91 HIV-1 gag-pol expression vector[22,23], and 1.5µg of pCSFLW (expressing the firefly luciferase reporter gene with the HIV-1 packaging signal). Viral supernatant was collected at 48 and 72h after transfection, filtered through 0.45um filter and stored at -80˚C as previously described[24]. Infectivity was measured by luciferase detection in target TZMBL transduced to express TMPRSS2 and ACE2.

Normalisation of virus titre by SG-PERT to measure RT activity in lentivirus preparation

Supernatant was subjected to SG-PERT as previously described.[25]

Homology modelling

Prediction of conformational change in the spike N-terminal assessed by homology modelling of the NTD (residues 14-306) predicted by homology modelling using SWISS-MODEL[26] with template chain A of PDB 7C2L[27] and aligned with 7C2L using PyMOL. Figures prepared with PyMOL (Schrödinger) using PDBs 7C2L, 6ZGE28 and 6ZGG[28].

1 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

2 Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature medicine* 26, 681-687, doi:10.1038/s41591-020-0868-6 (2020).

3 Thomson, E. C. *et al.* The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. *bioRxiv*, 2020.2011.2004.355842, doi:10.1101/2020.11.04.355842 (2020).

4 Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* 182, 1295-1310 e1220, doi:10.1016/j.cell.2020.08.012 (2020).

223 5 Greaney, A. J. *et al.* Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-
224 Binding Domain that Escape Antibody Recognition. *Cell host & microbe*,
225 doi:10.1016/j.chom.2020.11.007 (2020).
226 6 Starr, T. N. *et al.* Prospective mapping of viral mutations that escape antibodies used to
227 treat COVID-19. *bioRxiv*, doi:10.1101/2020.11.30.405472 (2020).
228 7 Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host.
229 *N Engl J Med*, doi:10.1056/NEJMc2031364 (2020).
230 8 Kemp, S. A. *et al.* Neutralising antibodies drive Spike mediated SARS-CoV-2 evasion.
231 *medRxiv*, 2020.2012.2005.20241927, doi:10.1101/2020.12.05.20241927 (2020).
232 9 Oude Munnink, B. B. *et al.* Transmission of SARS-CoV-2 on mink farms between humans
233 and mink and back to humans. *Science*, doi:10.1126/science.abe5901 (2020).
234 10 Gu, H. *et al.* Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science*
235 369, 1603-1607, doi:10.1126/science.abc4730 (2020).
236 11 McCarthy, K. R. *et al.* Natural deletions in the SARS-CoV-2 spike glycoprotein drive
237 antibody escape. *bioRxiv*, 2020.2011.2019.389916, doi:10.1101/2020.11.19.389916
238 (2020).
239 12 Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases
240 Infectivity of the COVID-19 Virus. *Cell*, doi:10.1016/j.cell.2020.06.043 (2020).
241 13 Yurkovetskiy, L. *et al.* Structural and Functional Analysis of the D614G SARS-CoV-2 Spike
242 Protein Variant. *Cell* 183, 739-751 e738, doi:10.1016/j.cell.2020.09.032 (2020).
243 14 Hou, Y. J. *et al.* SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and
244 transmission in vivo. *Science*, doi:10.1126/science.abe8499 (2020).
245 15 Mlcochova, P. *et al.* Extended in vitro inactivation of SARS-CoV-2 by titanium dioxide
246 surface coating. *bioRxiv*, 2020.2012.2008.415018, doi:10.1101/2020.12.08.415018
247 (2020).
248 16 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision
249 to reality. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European*
250 *communicable disease bulletin* 22, 30494, doi:10.2807/1560-7917.ES.2017.22.13.30494
251 (2017).
252 17 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
253 Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772-
254 780, doi:10.1093/molbev/mst010 (2013).

18 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37, 1530-1534, doi:10.1093/molbev/msaa015 (2020).

19 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14, 587-589, doi:10.1038/nmeth.4285 (2017).

20 Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution* 30, 1188-1195, doi:10.1093/molbev/mst024 (2013).

21 Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus evolution* 4, vex042-vex042, doi:10.1093/ve/vex042 (2018).

22 Naldini, L., Blomer, U., Gage, F. H., Trono, D. & Verma, I. M. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. *Proceedings of the National Academy of Sciences of the United States of America* 93, 11382-11388 (1996).

23 Gupta, R. K. *et al.* Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. *Aids* 24, 1651-1655, doi:10.1097/QAD.0b013e3283398216 (2010).

24 MIcochova, P. *et al*. Combined point of care nucleic acid and antibody testing for SARS-CoV-2 following emergence of D614G Spike Variant. *Cell Rep Med*, 100099, doi:10.1016/j.xcrm.2020.100099 (2020).

25 Gregson, J. *et al.* HIV-1 viral load is elevated in individuals with reverse transcriptase mutation M184V/I during virological failure of first line antiretroviral therapy and is associated with compensatory mutation L74I. *The Journal of infectious diseases*, doi:10.1093/infdis/jiz631 (2019).

26 Waterhouse *et al*. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46, 296-303, doi:https://doi.org/10.1093/nar/gky427 (2018).

27 Chi *et al*. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* 369, 650-655, doi:10.1126/science.abc6952 (2020).

28 Wrobel *et al*. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol* 27, 763-767, doi:10.1038/s41594-020-0468-7 (2020).

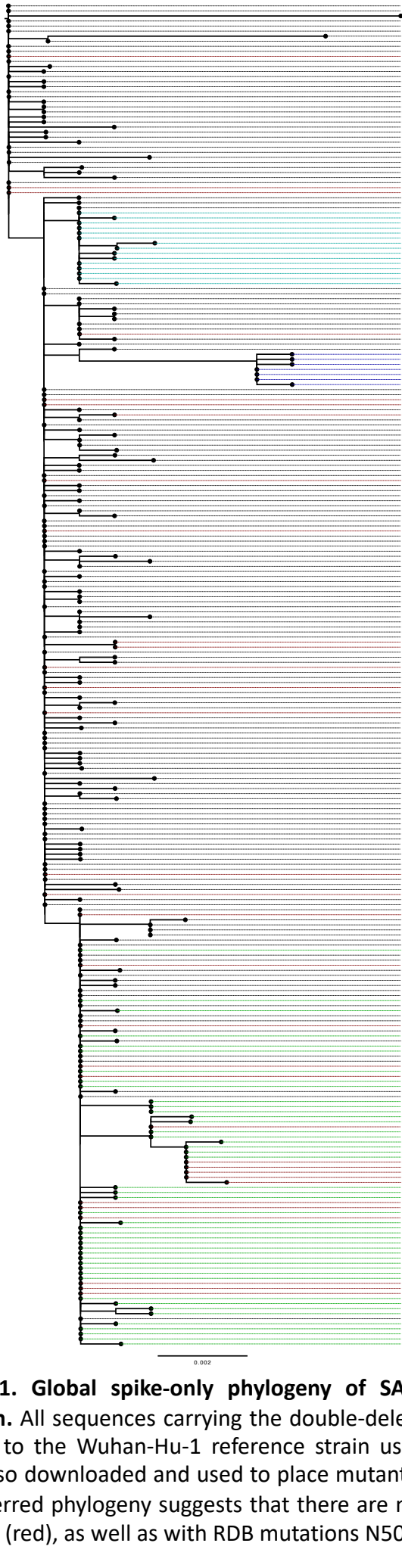


Figure 1. Global spike-only phylogeny of SARS-CoV-2 sequences carrying the Spike ΔH69/V70 deletion. All sequences carrying the double-deletion were downloaded from the GISAID database and aligned to the Wuhan-Hu-1 reference strain using MAFFT. A series of global background sequences were also downloaded and used to place mutants into context. All duplicate sequences were removed. The inferred phylogeny suggests that there are multiple lineages of sequences carrying the ΔH69/V70, by itself (red), as well as with RDB mutations N501Y (dark blue), N453F (cyan) and Y439K (green).

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Total
All ΔH69/V70 and RBD mutations 439K, 453F and 501Y	Australia					1	1			1	2		5
	Burkina Faso								1	1			2
	Canada									1			1
	Cote d'Ivoire						5						5
	Czech Republic									12	53		65
	Denmark				1				**53**	**281**	**611**	**3**	949
	England				3			2	**184**	**253**	**572**	**449**	1465
	Faroe Islands									9			9
	France			1						9			10
	Germany		1						1	3			5
	Italy			3									3
	Malaysia			1	1								2
	Netherlands										9		9
	New Zealand								1				1
	Northern Ireland								20	1	1		22
	Norway									1			1
	Saudi Arabia				1								1
	Scotland								25	68	83	2	178
	Slovenia			4									4
	South Africa							1					1
	Sweden			1						1	1		3
	Switzerland								9				9
	Taiwan			1									1
	Thailand	1											1
	Turkey			2									2
	USA				1		1						2
	Wales								14	101	174	20	309
	Grand Total	1	1	13	7	1	7	3	308	742	1506	474	3065
Only 69/70 (without RBD replacements)	Denmark								28	109	212	4	353
	Faroe Islands									3			3
	Turkey						1						1
	United Kingdom								2	6	6		14
	Grand Total						1		30	118	218	4	371

Table 1. Number of reported SARS-CoV-2 sequences with the ΔH69/V70 deletion. All sequences containing the ΔH69/V70 deletion were extracted from the GISAID database (Accessed 26th Nov 2020) and tabulated according to both reporting country of origin and date in which they were posted online. The lineages carrying ΔH69/V70 began to expand in both Denmark and England in August 2020.

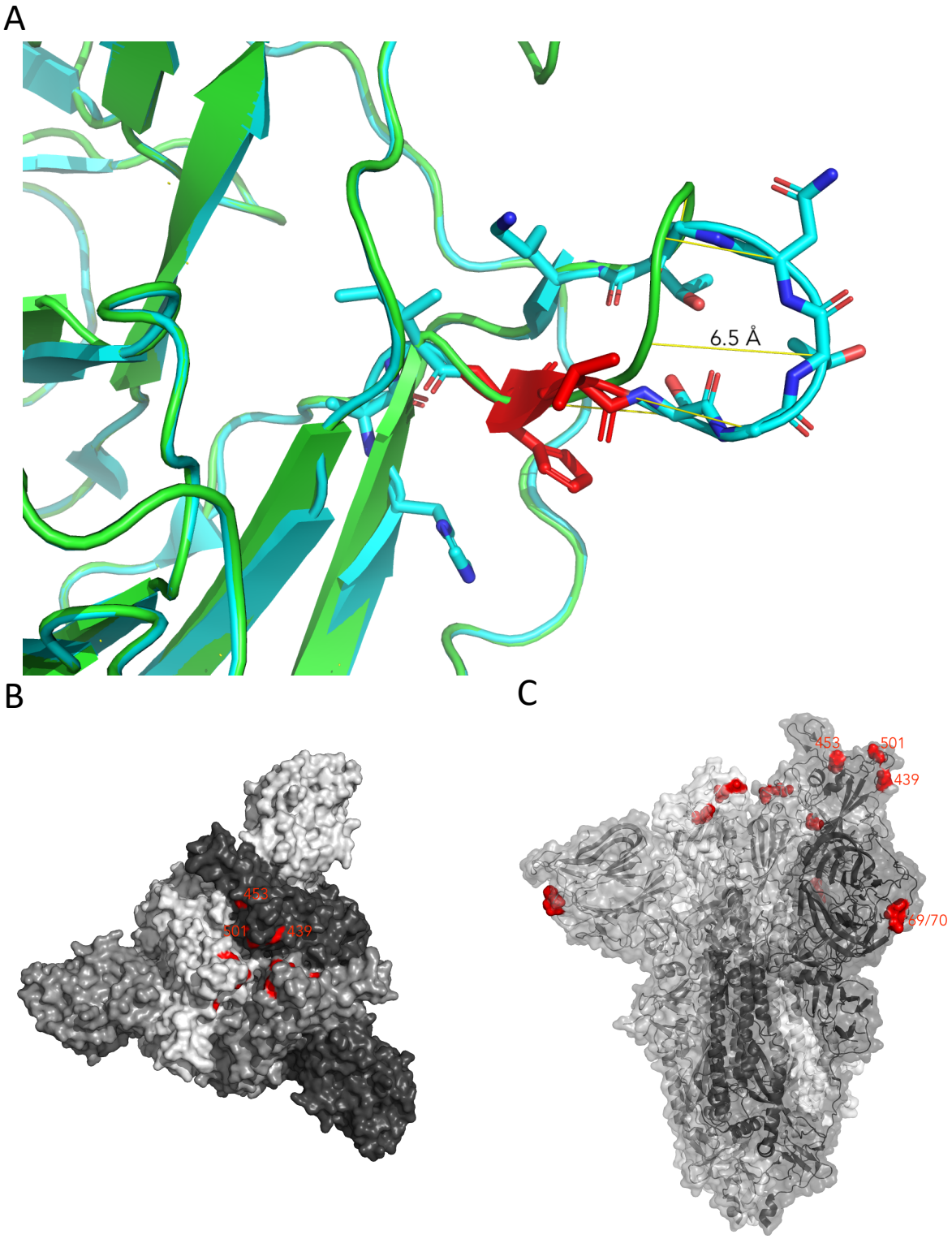


Figure 2. Structural aspects of ΔH69/V70 A) Prediction of conformational change in the spike N-terminal domain due to deletion of residues His69 and Val70. The pre-deletion structure is shown in cyan, except for residues 69 and 70, which are shown in red. The predicted post-deletion structure is shown in green. Residues 66-77 of the pre-deletion structure are shown in stick representation and coloured by atom (carbon in cyan, nitrogen in blue, oxygen in coral). Yellow lines connect aligned residues 66-77 of the pre- and post-deletion structures and the distance of 6.5 Å between aligned alpha carbons of Thr73 in the pre- and post-deletion conformation is labelled. **B**) Surface representation of spike homotrimer in closed conformation (PDB: 6ZGE, Wrobel et al., 2020) homotrimer viewed in a 'top-down' view along the trimer axis with each monomer in shown in different shades of grey and locations of RBD mutations at residues 439, 453 and 501 highlighted in red. **C**) Spike in open conformation with a single erect RBD (PDB: 6ZGG, Wrobel et al. 2020) in trimer axis vertical view with the locations of deleted residues His69 and Val70 in the N-terminal domain and RBD mutations highlighted as red spheres and labelled on the monomer with erect RBD. Residues 71-75, which form the exposed loop undergoing conformational change in **A**, are omitted from this structure.

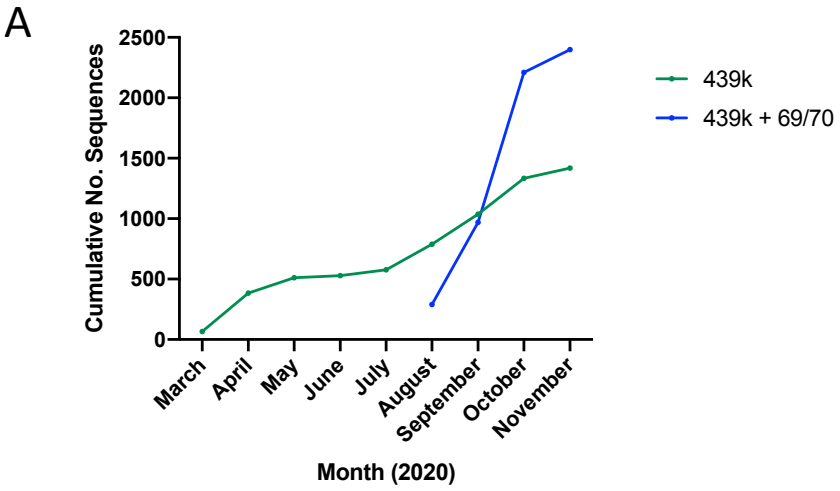


Figure 3. The relative increase in frequency of sequences bearing Spike mutants 439K and 501Y based on sampling dates. Between August and October 2020, an exponential increase in mutants carrying both 439K and **Δ**H69/V70 saw the latter become dominant in terms of cumulative cases.

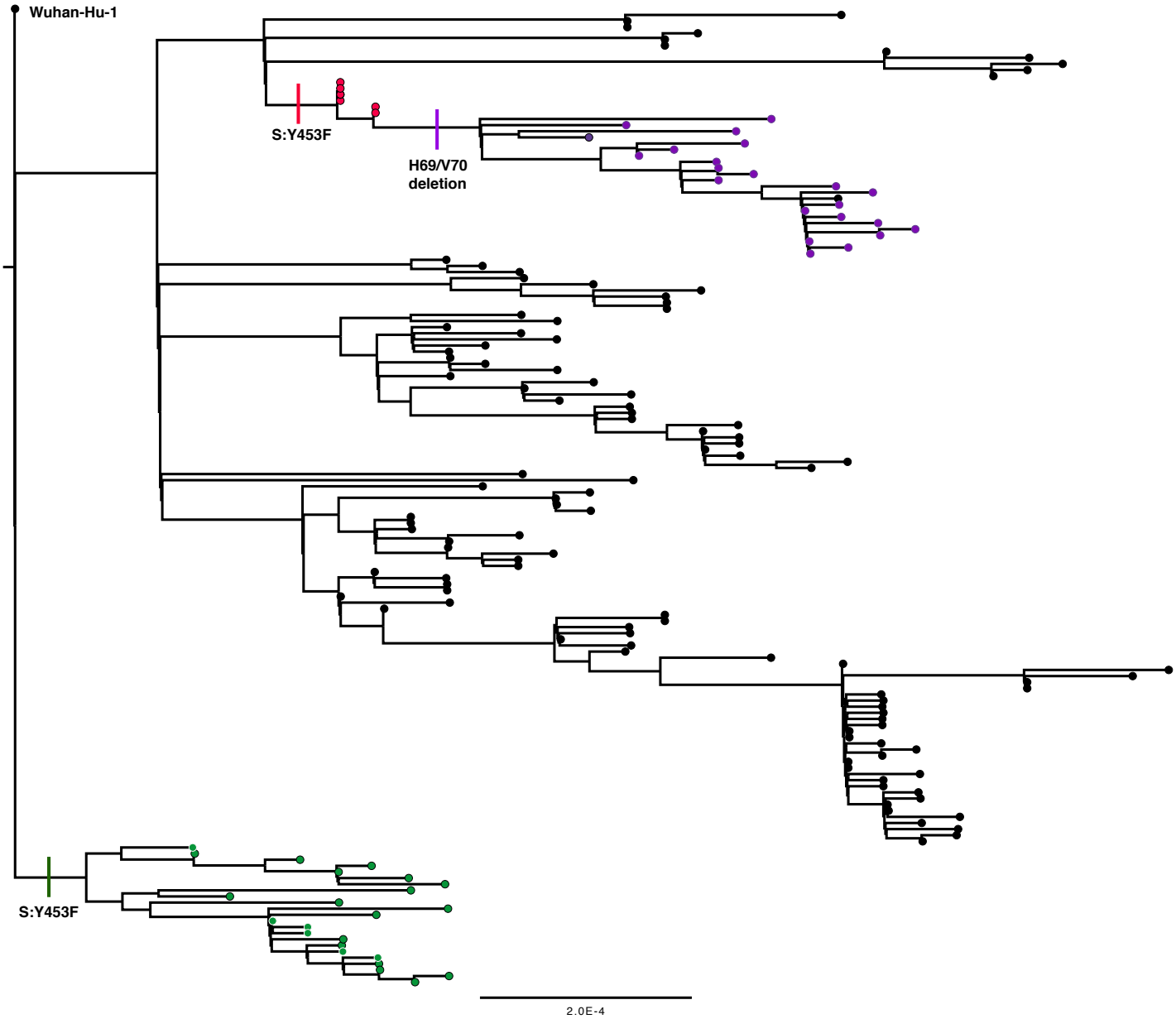


Figure 4. Maximum likelihood phylogeny of all Mink-origin SARS-CoV-2 Spike sequences. All 753 publicly available Mink origin Spike sequences were downloaded from the GISAID database (accessed 12th December) and aligned to the Wuhan-Hu-1 reference sequence using MAFFT. Acquisition of the Spike mutant Y453F, an exposed region in the RBD, occurred early on in Dutch Mink (green circles). Later in infection, 453F was acquired by Danish Mink (red) and subsequently, those with the 453F mutant also acquired the **Δ**H69/V70 double deletion (purple).

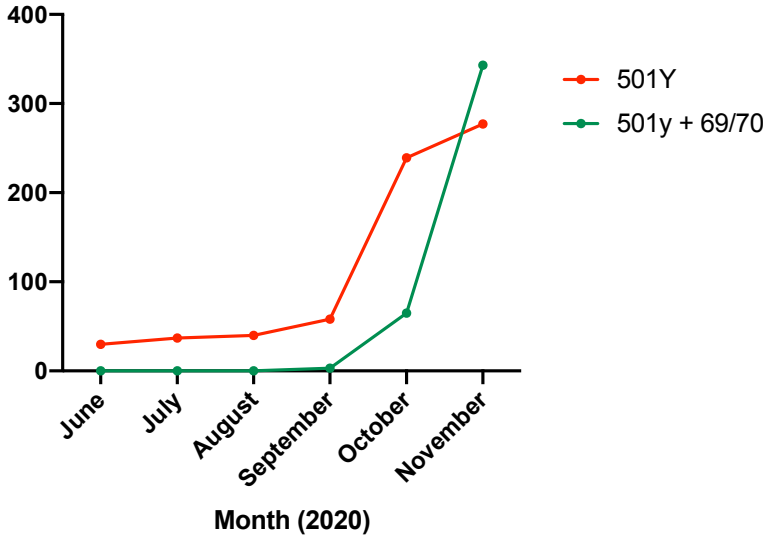


Figure 5. The relative increase in frequency of sequences bearing Spike mutant 501Y based on sampling dates. Between October and November 2020, the sequences carrying both 501Y and the **Δ**H69/V70 also became dominant.

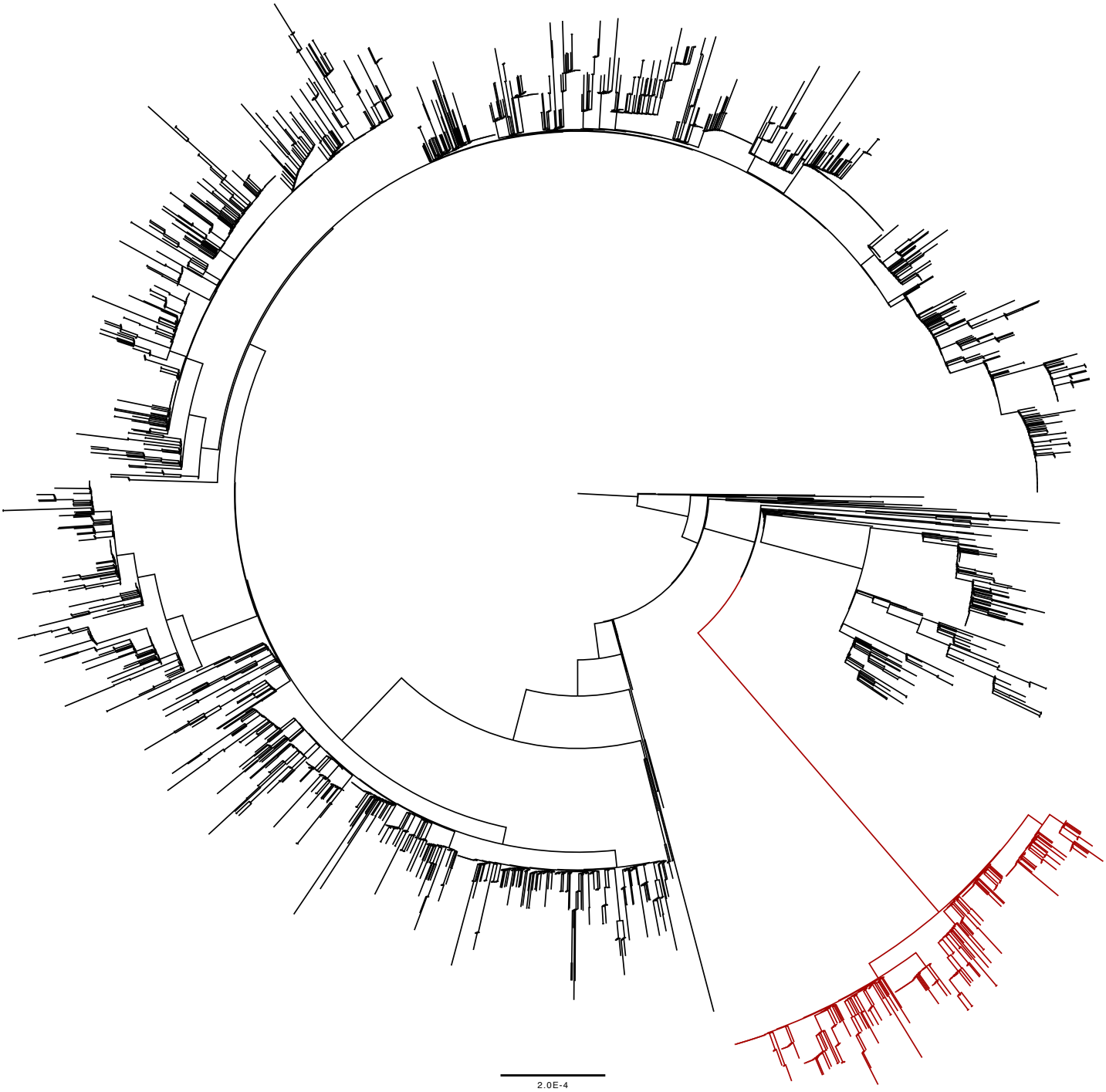


Figure 6: Lineage bearing multiple Spike mutations including ΔH69/V70: Global maximum likelihood phylogeny of all ΔH69/V70 sequences downloaded from the GISAID database (3066 sequences, accessed 26th November 2020). A distinct sub-lineage of ΔH69/V70 sequences (red) developed with six linked S mutations, two of which are in the RBD (N501Y and A570D) and four others in S2 (P681H, T716I, S982A and D1118H). All five mutations occur together in all cases with the deletion.

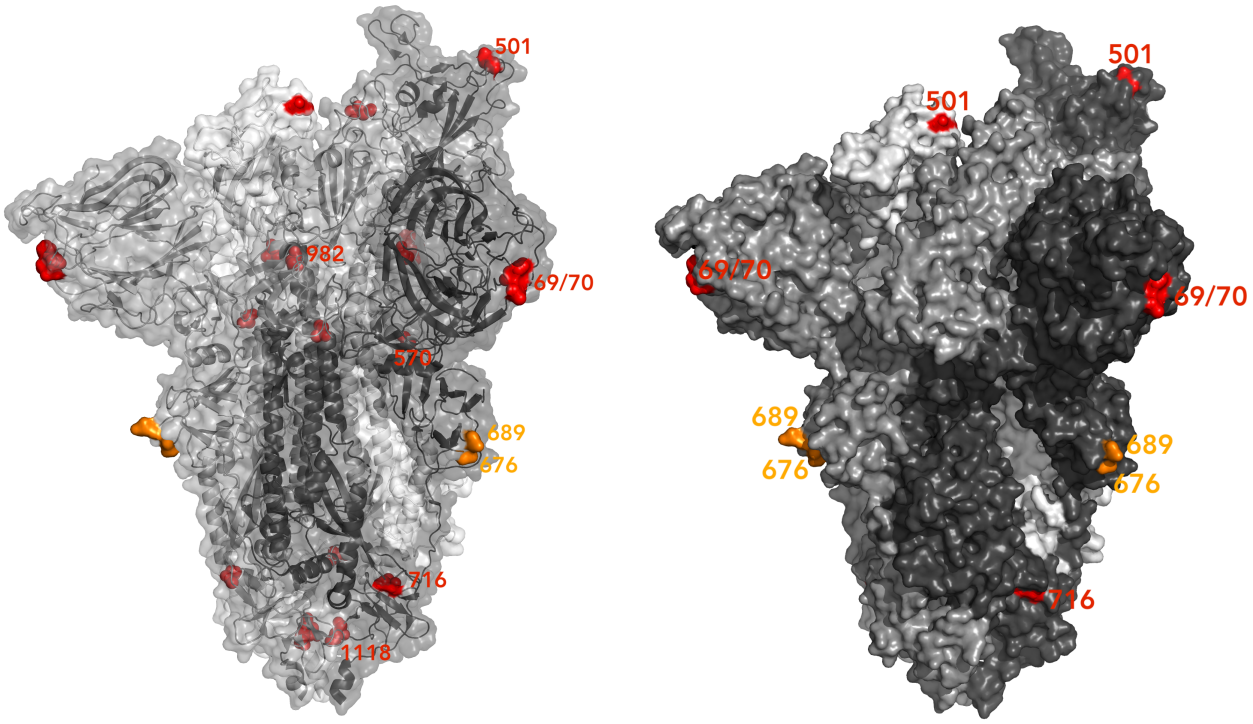
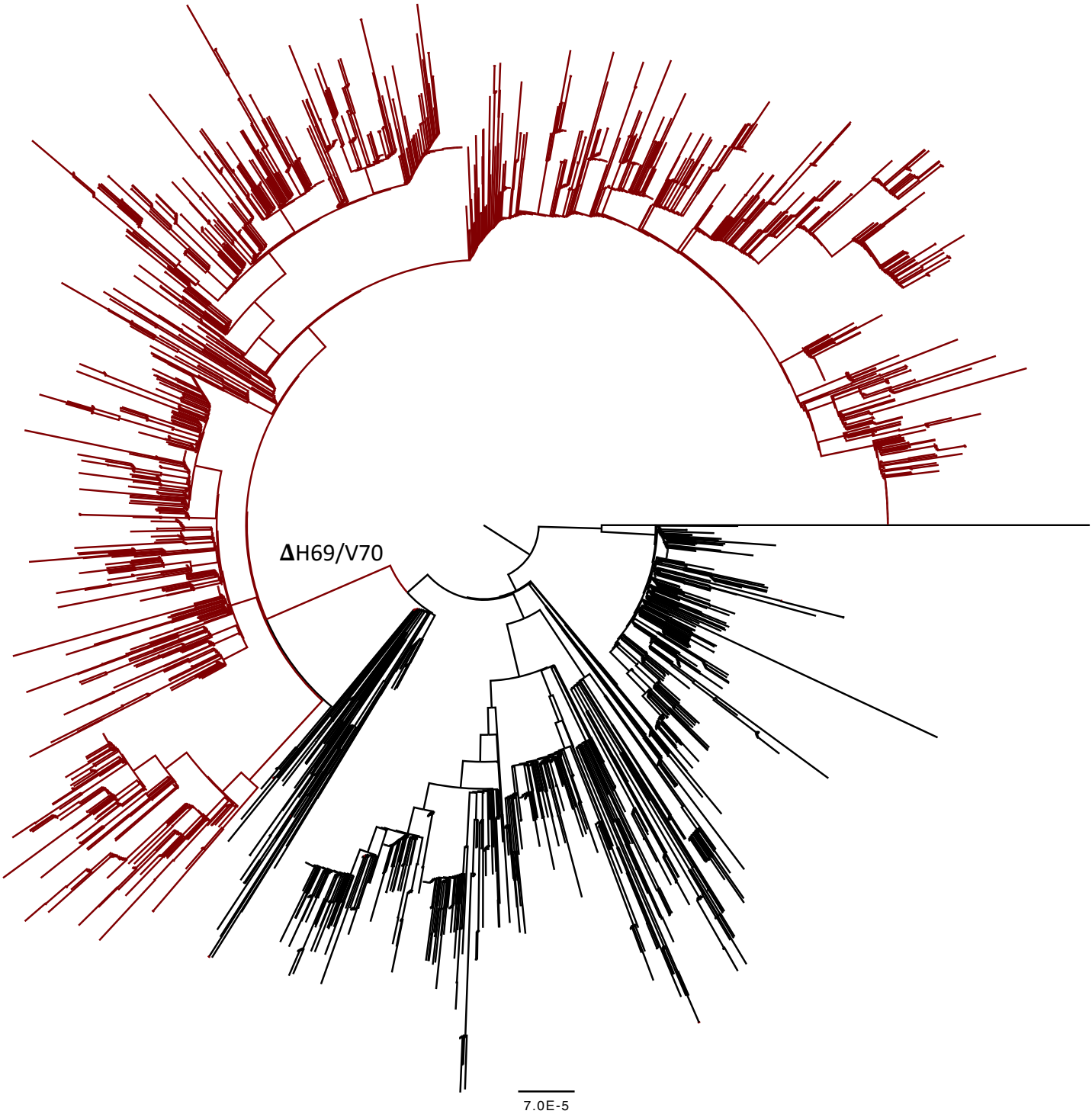
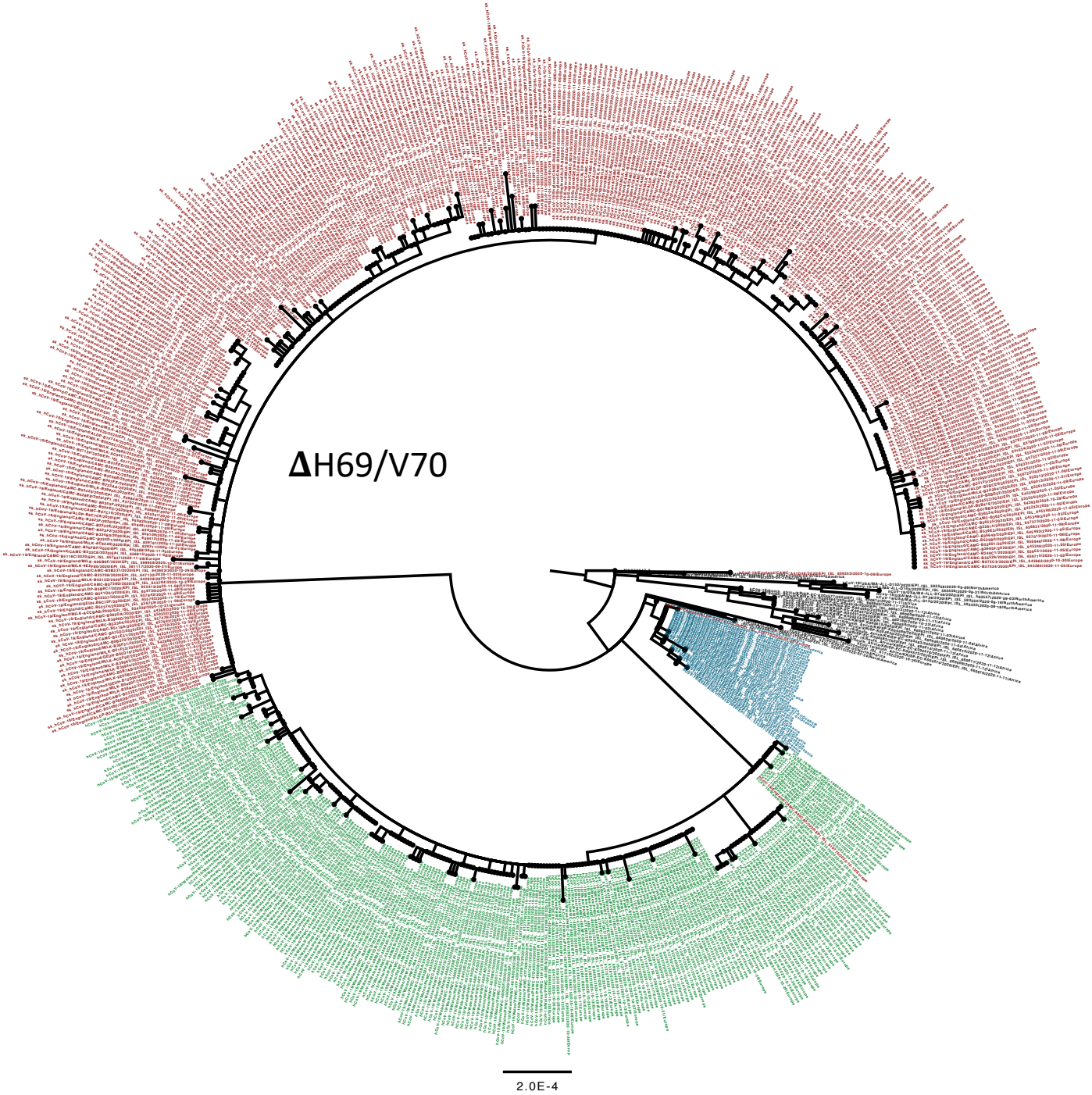


Figure 7. Spike residues in a highly mutated circulating strain with ΔH69/V70. Spike homotrimer in open conformation with one upright RBD (PDB: 6ZGE, Wrobel et al., 2020) with different monomers shown in shades of grey. To the left, surface representation overlaid with ribbon representation and to the right, opaque surface representation accentuating the locations of surface-exposed residues. The deleted residues 69 and 70 and the residues involved in amino acid substitutions (501, 570, 716, 982 and 1118) are coloured red. The location of an exposed loop including residue 681 is absent from the structure, though the residues either side of the unmodelled residues, 676 and 689, are coloured orange. On the left structure, highlighted residues are labelled on the monomer with an upright RBD; on the right structure, all visible highlighted residues are labelled.



Supplementary Figure 1. Circularised maximum likelihood phylogeny of global sequences carrying Spike mutant 439K. All sequences in the GISAID database containing S:439K (3820 sequences, 26th November 2020) were realigned to Wuhan-Hu-1 using MAFFT. Viruses carrying the Spike double deletion **Δ**H69/V70 (red) emerged and expanded from viruses with S:439K (black).



Supplementary Figure 2. Circularised maximum likelihood phylogeny of global sequences carrying Spike mutant 501Y. All sequences in the GISAID database containing S:501Y were downloaded and realigned to Wuhan-Hu-1 using MAFFT. Sequences were broadly split into four major clades; sequences carrying the Spike double deletion **Δ**H69/V70 (red) formed an entirely separate clade from non-carriers. Sequences carrying 501Y but an absence of **Δ**H69/V70 formed a second lineage and appeared to expand only in Wales (green). Another major clade (blue) was limited entirely to Australia and finally a fourth clade (black) was limited to several African countries and Brazil.