

1 **Rapidly emerging SARS-CoV-2 B.1.1.7 sub-lineage in the United States of America with**
2 **spike protein D178H and membrane protein V70L mutations**

3 Lishuang Shen¹, Jennifer Dien Bard¹, Timothy J. Triche¹, Alexander R. Judkins¹, Jaclyn A.
4 Biegel¹, Xiaowu Gai¹

5

6 ¹Children's Hospital Los Angeles, Department of Pathology and Laboratory Medicine,
7 Keck School of Medicine of University of Southern California, Los Angeles, CA, U.S.

8

9 Correspondence to:

10

11 Xiaowu Gai, PhD

12 Center for Personalized Medicine

13 Division of Genomic Medicine

14 Department of Pathology and Laboratory Medicine

15 4650 Sunset Blvd., Mailstop #173

16 Los Angeles, CA 90027

17 Phone: (323) 644-8506

18 Email: xgai@chla.usc.edu

19

20 Abstract

21 The SARS-CoV-2 B.1.1.7 lineage is highly infectious and as of April 2021 accounted for 92% of
22 COVID-19 cases in Europe and 59% of COVID-19 cases in the U.S. It is defined by the N501Y
23 mutation in the receptor binding domain (RBD) of the Spike (S) protein, and a few other
24 mutations. These include two mutations in the N terminal domain (NTD) of the S protein, HV69-
25 70del and Y144del (also known as Y145del due to the presence of tyrosine at both positions).
26 We recently identified several emerging SARS-CoV-2 variants of concerns, characterized by
27 Membrane (M) protein mutations, including I82T and V70L. We now identify a sub-lineage of
28 B.1.1.7 that emerged through sequential acquisitions of M:V70L in November 2020 followed by
29 a novel S:D178H mutation first observed in early February 2021. The percentage of B.1.1.7
30 isolates in the U.S. that belong to this sub-lineage increased from 0.15% in February 2021 to
31 1.8% in April 2021. To date this sub-lineage appears to be U.S.-specific with reported cases in
32 31 states, including Hawaii. As of April 2021 it constituted 36.8% of all B.1.1.7 isolates in
33 Washington. Phylogenetic analysis and transmission inference with Nextstrain suggests this sub-
34 lineage likely originated in either California or Washington. Structural analysis revealed that the
35 S:D178H mutation is in the NTD of the S protein and close to two other signature mutations of
36 B.1.1.7, HV69-70del and Y144del. It is surface exposed and may alter NTD tertiary
37 configuration or accessibility, and thus has the potential to affect neutralization by NTD directed
38 antibodies.

39 Keywords:

40 SARS-CoV-2, D178H, V70L, B.1.1.7 lineage , COVID-19

41 Introduction

42 B.1.1.7 emerged in the UK and was the first major SARS-CoV-2 variant of concern (VOC) that
43 is both more transmissible and apparently more virulent (1). It now accounts for 50-90% of the
44 COVID-19 cases in U.S. and Europe. The Spike (S) protein N501Y mutation in the receptor-
45 binding domain (RBD) confers higher binding affinity of the S protein for ACE2, while the other
46 two deletions, HV69-70del and Y144del in the N-terminal domain (NTD) may also play a role in
47 ACE2 receptor binding or neutralizing antibody escape (2). With millions of new B.1.1.7 cases
48 in recent months, there is a very high probability of continuous acquisitions of new mutations,
49 some of which may result in the emergence of new and even more infectious sub-lineages of
50 B.1.1.7. While these new mutations may not be significantly deleterious by themselves, but when
51 they appear in the context of other mutations within this VOC the result may be a more
52 transmissible or pathogenic virus. This calls for rigorous genomic surveillance for newly
53 acquired mutations in previously reported VOCs, including but not limited to B.1.1.7 and
54 B.1.351.

55 Using the Children's Hospital Los Angeles (CHLA) COVID-19 Analysis Research Database
56 (CARD) (3), and viral sequences submitted to GISAID and NCBI GenBank, we have routinely
57 performed genomic epidemiology and genomic surveillance studies of local, national and
58 international databases (4-9). This allowed us to identify a new VOC (B.1.575) with a signature
59 mutation I82T in the M gene (7). In the same study, we identified multiple other M mutations
60 including V70L that are currently being encountered with significantly increased frequency. We

61 have identified the M:V70L mutation in multiple SARS-CoV-2 lineages but primarily in the
62 B.1.1.7 lineage. The B.1.1.7-M:V70L sub-lineage has been circulating at consistently moderate
63 prevalence since November 2020. Through continuous genomic surveillance we identified the
64 subsequent acquisition of yet another Spike mutation, D178H, within this lineage. This new
65 B.1.1.7 sub-lineage, carrying both M:V70L and S:D178H mutations, appeared in February 2021
66 but by April 2021 quickly increased to account for 36.8% and 1.8% of all reported B.1.1.7
67 genomes in Washington and the U.S., respectively. This B.1.1.7-M:V70L-S:D178H sub-lineage
68 is currently exclusive to the U.S. as of May 8, 2021, and had been detected in 31 states, with the
69 majority of cases found in Washington, California and Ohio. Here, we report its detection,
70 characterization, transmission, and evolution.

71 **Materials and methods**

72 **Ethics approval.**

73 Study design conducted at Children’s Hospital Los Angeles was approved by the Institutional
74 Review Board under IRB CHLA-16-00429.

75 **SARS-CoV-2 whole genome sequencing**

76 Whole genome sequencing of the 2900 samples previously confirmed at Children’s Hospital Los
77 Angeles to be positive for SARS-CoV-2 by reverse transcription-polymerase chain reaction (RT-
78 PCR) was performed as previously described (5).

79 **SARS-CoV-2 sequence and variant analysis, and emerging variant monitoring**

80 Full-length SARS-CoV-2 sequences had been periodically downloaded from GISAID (10, 11)
81 and NCBI GenBank. They were combined with SARS-CoV-2 sequences from CHLA patients,
82 annotated, and curated using a suite of bioinformatics tools, CHLA-CARD, as previously
83 described (3). A custom Surging Mutation Monitor (SMM) standardized and integrated the viral
84 genome and demographic data, in order to identify the trend of surging mutations and lineages
85 across state and country levels. The current study was based on the 1.33 million global viral
86 genomes that were available on May 1st, 2021.

87 **Phylogenetic analysis**

88 Phylogenetic analysis was conducted using the NextStrain phylogenetic pipeline (version 3.0.1)
89 (<https://nextstrain.org/>). Mafft (v7.4) was used in multiple sequence alignment (12), IQ-Tree
90 (multicore version 2.1.1 COVID-edition) and TreeTime version 0.7.6 were used to infer and
91 time-resolve evolutionary trees, and reconstruct ancestral sequences and mutations (13, 14).
92 Phylogenetic analysis.

93 **Protein structure prediction**

94 Structural predictions of mutant Spike proteins were carried out against the wild type protein
95 PDB qhd43416 (<https://zhanglab.ccmb.med.umich.edu/COVID-19/>) using Missense3D service
96 hosted online by the Imperial College London (<http://www.sbg.bio.ic.ac.uk/~missense3d/>) (15).

97 CoV3D was used as the Spike Protein Mutation Viewer for the multiple mutations in B.1.1.7
98 (<https://cov3d.ibbr.umd.edu/MutViewer/QTY83983>).

99 **Results**

100 **Identification of a rapidly emerging B.1.1.7 sub-lineage**

101 We evaluated 1,333,679 SARS-CoV-2 viral genomes available on May 1st, 2021, including
102 2,900 from our own institution and the rest from GISAID and NCBI GenBank. We searched for
103 SARS-CoV-2 mutations with significantly higher prevalence rate in both the U.S. and globally.
104 Candidate mutations were further partitioned by pangolin lineage to identify emerging mutations
105 in the context of a specific lineage, such as B.1.1.7 or B.1.351. We focused initially on the M
106 mutations that we previously identified, including V70L, that were spiking near the end of 2020
107 (7). Overall the percentage of isolates that carried the M:V70L mutation had been relatively
108 stable in the U.S. and globally with a gradual month to month increase (**Table 1**). In the vast
109 majority of cases, the M:V70L mutation occurred on the B.1.1.7 lineage. While the percentage of
110 B.1.1.7 isolates with the V70L mutation remained relatively stable across the world, the
111 percentage fluctuated significantly in the U.S., attributable largely to the initial small number of
112 B.1.1.7 cases in the U.S.

113 We identified the acquisition of another S mutation, D178H, in this B.1.1.7 sub-lineage (**Figure**
114 **1, Table 2**), which was estimated to have occurred on January 23, 2021. By April, the prevalence
115 of SARS-CoV-2 isolates carrying the S:D178H mutation increased to 1.05% nationally and as
116 high as 14.77% in Washington. When we examined the prevalence of S:D178H in the context of
117 the B.1.1.7 lineage and the B.1.1.7-M:V70L sub-lineage, the numbers were even more striking
118 (**Figure 2**). While the percentage of all B.1.1.7 isolates that carried the S:178H mutation
119 increased from 0% in January 2021 to 1.8% in April 2021, there was a significant increase from
120 0% to 64.8% in B.1.1.7-M:V70L sub-lineage isolates that carried the S:D178H mutation were
121 observed in the same time period. (**Figure 2A, Table 2**). The B.1.1.7-M:V70L-S:D178H sub-
122 lineage was exclusive to the U.S. (**Table 2**), and findings when examined globally were not as
123 striking (0% to 0.56% and 0% to 37%, respectively) (**Figure 2A, Table 2**).

124 The S:D178H mutation was also seen in other lineages including B.1.234 and B.1. In February, it
125 was detected 10 times in 6 states and the first two were reported in California on February 4th,
126 both within this B.1.1.7 sub-lineage. Overall, 98.93% of the S:D178H carrying viral isolates
127 belonged to this new B.1.1.7 sub-lineage.

128 Within the U.S., most isolates of this lineage were from Washington, California, and Ohio
129 (**Figure 3**). California and Washington showed the greatest increases in terms of the percentage
130 of B.1.1.7 isolates carrying the S:D178H mutation. In Washington, the S:D178H mutation was
131 absent until February 2021, where it quickly increased to account for 9.6% (296/3077) of all
132 viral isolates in March and 14.8% (364/2464) in April (**Table 3**). It accounted for 36.8%
133 (364/990) of the April B.1.1.7 isolates in Washington (**Figure 2C, Table 3**). In California,
134 S:D178H was first seen in December 2020, but it was not seen within the B.1.1.7 lineage until
135 February 4th, 2021 (**Table 3**). Its prevalence increased to 1.6% (45/2904) in April compared to

136 all isolates studied in California, and 3.3% (45/1353) of the B.1.1.7 isolates (**Figure 2C, Table**
137 **3**).

138 **Origin and transmission of the B.1.1.7-M:V70L-S:D178H sub-lineage**

139 To understand the origin and evolution of this B.1.1.7-M:V70L-S:D178H sub-lineage, we used
140 a) all 1,125 S:D178H carrying viral sequences reported across the globe, b) 1,600 subsampled
141 M:V70L carrying viral sequences across the world, and c) the NC_045512 reference genome.
142 These sequences were further filtered with requirements of being at least 27,000 bp and having
143 complete date information. 2,609 genomes were kept for final phylogenetic and transmission
144 analysis using Nextstrain, rooted by NC_045512 (**Figure 1**). The D178H branch divergence date
145 was estimated to be January 23, 2021, with a date confidence interval of January 17, 2021 to
146 February 3, 2021. It likely originated in California (66% probability) or Washington (34%
147 probability).

148 Estimated transmission routes are demonstrated in **Figure 3**, where the size of the circle
149 represents the number of cases from this sub-lineage in each state, and the line colors correspond
150 to the exporting locations, California, Washington, and Ohio contributed the majority of
151 transmissions compared to other states. This sub-lineage was clearly exclusive to the U.S., and
152 not reported in any other countries at the time of the study.

153 **Signature mutations of the B.1.1.7-M:V70L-S:D178H sub-lineage**

154 Within this sub-lineage, the most common signature mutations were the same as B.1.1.7
155 signature mutations, with the additional M:V70L and S:D178H mutations. There were no other
156 novel common spike mutations within this sub-lineage.

157 **Protein structure and mutation effect prediction**

158 The 3D structure of the Spike protein, as visualized using the CoV3D mutation viewer. Using
159 these results we were able to show that the S:D178H mutation is close structurally to two
160 signature deletions of B.1.1.7, HV69_70del and Y144del (**Figure 4**). They are all surface
161 exposed and likely alter N terminal domain (NTD) tertiary configuration.

162 **Discussion**

163
164
165 In March 2021, the B.1.1.7-M:V70L-S:D178H lineage appeared abruptly at high prevalence
166 (35.6%) in all reported B.1.1.7 isolates in Washington. This was a remarkable finding as it was
167 absent in 115 Washington B.1.1.7 isolates reported in February. This abrupt change could
168 reflect undersampling or potentially reflect superspreader events. Given the spike in the number
169 of B.1.1.7 cases, this sub-lineage clearly appears to be more transmissible than even the original
170 B.1.1.7 lineage given the spike in the number of B.1.1.7 cases. This finding warrants prompt and
171 further attention by public health authorities, as this mutation profile is closely linked to the
172 resurgence of cases in Washington in particular. It also strongly supports the now widely
173 recognized need for more extensive SARS-CoV-2 viral sequencing of PCR positive COVID-19
174 cases for detection of new mutations of concern as part of widespread genomic surveillance (16,
175 17).

176

177 The S:D178H mutation, while demonstrably associated here with the more pathogenic B.1.1.7
178 lineage, is not necessarily by itself more pathogenic. Dozens of SARS-CoV-2 genomes that carry
179 the S:D178H mutation were reported before February, but none of these demonstrated the
180 increased frequency seen when the mutation occurs in the context of the B.1.1.7 lineage.
181 Phylogenetic analysis revealed a distinct and long branch leading to the new S:D178 branch after
182 M:V70L. Together, these observations suggest that the S:D178H mutation is recurrent, but only
183 increased exponentially in the context of the more pathogenic B.1.1.7 lineage, which serves as an
184 argument for its fitness. This is the same observed with other deleterious mutations like the
185 N501Y and E484K mutations, both of which are now superimposed on distinct and separate
186 more pathogenic lineages. The S:D178H mutation arose independently again in the U.S. on the
187 B.1.1.7-M:V70L background. The rapid increase in its prevalence, only after its acquisition by
188 the B.1.1.7-M:V70L sub-lineage suggests this combination of mutations is associated with
189 increased transmissibility. It is also of interest that this mutation occurs in the NTD, unlike most
190 of the mutations associated with current VOC that are centered on the spike protein RBD,
191 implying that NTD mutations beyond the original 69-70del and the 144del are of concern. And
192 finally, it should be noted that this NTD mutation co-exists with the previously reported M
193 protein mutation M:V70L, suggesting that M protein mutations also contribute to enhanced
194 biologic ‘fitness’ or pathogenicity of this sub-lineage.

195
196 The appearance of the S:D178H mutation in the context of the B.1.1.7 lineage is temporally
197 associated with the increased incidence of COVID-19 in Washington. New cases in Washington
198 were higher than the national level at the time of the study. According to New York Time
199 COVID-19 dashboard, the 7-day average of new cases on May 2 was 1,379 in Washington,
200 which was only a 50% reduction compared to 2,757 cases on December 15. In comparison, the
201 numbers on May 2 was 49,270 in U.S., a 77.3% reduction from the December 15 number of
202 21,7325. The appearance of this new B.1.1.7 sub-lineage temporally linked to increased cases in
203 Washington warrants further investigation.

204
205 The potential effect of the S:D178H mutation on immunity and vaccine ‘escape’ also warrant
206 further analysis. Mutations in the Spike N-terminal domain have been associated with lack of
207 neutralization by NTD directed antibodies, especially when the N5 loop is affected (18, 19). The
208 NTD initiates viral binding to the ACE2 receptor expressing host cell. Since the D178H falls in
209 the NTD domain close to the N5 loop, it may alter NTD structure and antibody recognition. It
210 may thus have a similar immune evasion effect as the HV69-70del and Y144del mutations, or it
211 may further enhance that of the two other mutations, based on the 3D model. These findings
212 highlight the continued importance of active genomic surveillance to monitor the spread of this
213 B.1.1.7-M:V70L-S:178H lineage.

214 **Acknowledgments**

215 We would like to acknowledge all members of the Department of Pathology and Laboratory
216 Medicine (PLM) at Children’s Hospital Los Angeles for dedication towards providing excellent
217 patient care throughout the pandemic, including especially the rapid launch of both the COVID-
218 19 diagnostic test in the Clinical Microbiology and Virology Laboratory and the SARS-CoV-2
219 whole genome sequencing assay at the Center for Personalized Medicine in March 2020, both
220 with the support of Thermo Fisher and Paragon Genomics, with whom these assays were
221 developed. We would like to acknowledge the frontline healthcare workers who remain devoted

222 in the fight against COVID-19. We would also like to acknowledge all institutions that have
223 contributed the SARS-CoV-2 sequences timely, as well as NCBI, GISAID, and Nextstrain for
224 providing valuable resources for SARS-CoV-2 genomics.

225 **Disclosure statement**

226 The authors declare no potential conflict of interest.

227 **References:**

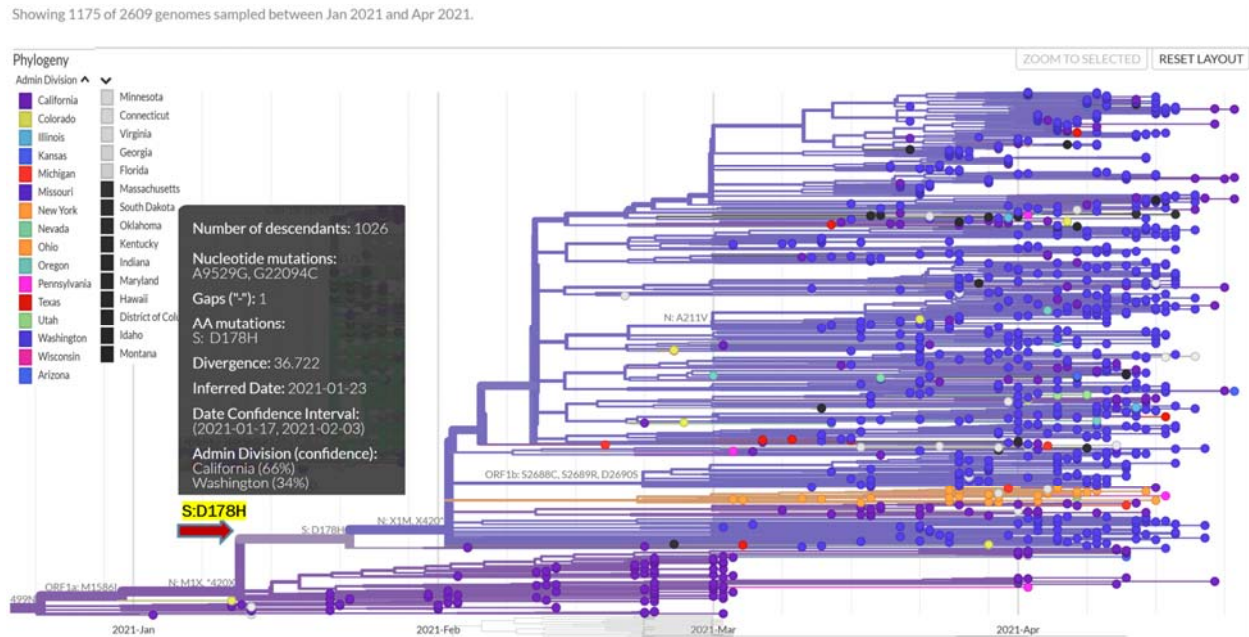
2281. Horby P, Huntley C, Davies N, Edmunds J, Ferguson N, Medley G, Calum Semple G. SAGE
229 meeting report. January 21, 2021.
2302. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD.
231 Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect
232 recognition by polyclonal human plasma antibodies. *Cell Host Microbe*. 2021 Mar 10;29(3):463-
233 476.e6. doi: 10.1016/j.chom.2021.02.003. Epub 2021 Feb 8. PMID: 33592168; PMCID:
234 PMC7869748.
2353. Shen L, Maglinte DT, Ostrow D, Pandey U, Bootwalla M, Ryutov A, Govindarajan A, Ruble D,
236 Han J, Triche TJ, Dien Bard J, Biegel JA, Judkins AR, Gai X. Children's Hospital Los Angeles
237 COVID-19 Analysis Research Database (CARD) - A Resource for Rapid SARS-CoV-2 Genome
238 Identification Using Interactive Online Phylogenetic Tools. *bioRxiv* 2020.05.11.089763; doi:
239 10.1101/2020.05.11.089763
2404. Shen L, Dien Bard J, Biegel JA, Judkins AR, Gai X. Comprehensive Genome Analysis of 6,000
241 U.S. SARS-CoV-2 Isolates Reveals Haplotype Signatures and Localized Transmission Patterns
242 by State and by Country. *Front Microbiol*. 2020 Sep 3;11:573430. doi:
243 10.3389/fmicb.2020.573430. PMID: 33013809; PMCID: PMC7509426.
2445. Pandey U, Yee R, Shen L, Judkins AR, Bootwalla M, Ryutov A, Maglinte DT, Ostrow D, Precit
245 M, Biegel JA, Bender JM, Gai X, Dien Bard J. High prevalence of SARS-CoV-2 genetic
246 variation and D614G mutation in pediatric patients with COVID-19. *Open Forum Infectious*
247 *Diseases*, ofaa551,
2486. Truong TT, Ryutov A, Pandey U, Yee R, Goldberg L, Bhojwani D, Aguayo-Hiraldo P, Pinsky
249 BA, Pekosz A, Shen L, Boyd SD, Wirz OF, Röltgen K, Bootwalla M, Maglinte DT, Ostrow D,
250 Ruble D, Han JH, Biegel JA, Li M, Huang C, Sahoo MK, Pannaraj PS, O'Gorman M, Judkins
251 AR, Gai X, Bard JD. Persistent SARS-CoV-2 infection and increasing viral variants in children
252 and young adults with impaired humoral immunity. *medRxiv* [Preprint]. 2021 Mar
253 2:2021.02.27.21252099. doi: 10.1101/2021.02.27.21252099. Update in: *EBioMedicine*. 2021
254 Apr 26;67:103355. PMID: 33688673; PMCID: PMC7941650.
2557. Shen L, Bard JD, Triche TJ, Judkins AR, Biegel JA, Gai X. Emerging variants of concern in
256 SARS-CoV-2 membrane protein: a highly conserved target with potential pathological and
257 therapeutic implications. *Emerg Microbes Infect*. 2021 Apr 26:1-24. doi:
258 10.1080/22221751.2021.1922097. Epub ahead of print. PMID: 33896413.
2598. Truong TT, Ryutov A, Pandey U, Yee R, Goldberg L, Bhojwani D, Aguayo-Hiraldo P, Pinsky
260 BA, Pekosz A, Shen L, Boyd SD, Wirz OF, Röltgen K, Bootwalla M, Maglinte DT, Ostrow D,
261 Ruble D, Han JH, Biegel JA, Li M, Huang C, Sahoo MK, Pannaraj PS, O'Gorman M, Judkins
262 AR, Gai X, Dien Bard J. Increased viral variants in children and young adults with impaired

- 263 humoral immunity and persistent SARS-CoV-2 infection: A consecutive case series.
264 EBioMedicine. 2021 Apr 26;67:103355. doi: 10.1016/j.ebiom.2021.103355. Epub ahead of print.
265 PMID: 33915337; PMCID: PMC8072072.
2669. Ryutov A, Gai X, Ostrow D, Maglinte DT, Flores J, Salas EJ, Glucoft M, Smit M, Dien Bard J.
267 Utility of viral whole genome sequencing for institutional infection surveillance during the
268 SARS-CoV-2 pandemic. *Infect Control Hosp Epidemiol.* 2021 Apr 19:1-12. doi:
269 10.1017/ice.2021.185. Epub ahead of print. PMID: 33866984.
27010. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to
271 global health. *Glob Chall.* 2017 Jan 10;1(1):33-46. doi: 10.1002/gch2.1018. PMID: 31565258;
272 PMCID: PMC6607375.
27311. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to
274 reality. *Euro Surveill.* 2017 Mar 30;22(13):30494. doi: 10.2807/1560-
275 7917.ES.2017.22.13.30494. PMID: 28382917; PMCID: PMC5388101.
27612. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence
277 alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002 Jul 15;30(14):3059-66. doi:
278 10.1093/nar/gkf436. PMID: 12136088; PMCID: PMC135756.
27913. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear
280 R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic
281 Era. *Mol Biol Evol.* 2020 May 1;37(5):1530-1534. doi: 10.1093/molbev/msaa015. Erratum in:
282 *Mol Biol Evol.* 2020 Aug 1;37(8):2461. PMID: 32011700; PMCID: PMC7182206.
28314. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis.
284 *Virus Evol.* 2018 Jan 8;4(1):vex042. doi: 10.1093/ve/vex042. PMID: 29340210; PMCID:
285 PMC5758920.
28615. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can Predicted
287 Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease
288 Associated? *J Mol Biol.* 2019 May 17;431(11):2197-2212. doi: 10.1016/j.jmb.2019.04.009.
289 Epub 2019 Apr 14. PMID: 30995449; PMCID: PMC6544567. [[Google Scholar](#)]
29016. Cyranoski D. Alarming COVID variants show vital role of genomic surveillance. *Nature.* 2021
291 Jan;589(7842):337-338. doi: 10.1038/d41586-021-00065-4. PMID: 33452508.
29217. Walensky RP, Walke HT, Fauci AS. SARS-CoV-2 Variants of Concern in the United States-
293 Challenges and Opportunities. *JAMA.* 2021 Feb 17. doi: 10.1001/jama.2021.2294. Epub ahead
294 of print. PMID: 33595644
29518. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD,
296 Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng
297 Z, Huang Y, Ho DD. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature.*
298 2021 May;593(7857):130-135. doi: 10.1038/s41586-021-03398-2. Epub 2021 Mar 8. PMID:
299 33684923.
30019. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, Lambson BE,
301 de Oliveira T, Vermeulen M, van der Berg K, Rossouw T, Boswell M, Ueckermann V, Meiring
302 S, von Gottberg A, Cohen C, Morris L, Bhiman JN, Moore PL. SARS-CoV-2 501Y.V2 escapes
303 neutralization by South African COVID-19 donor plasma. *Nat Med.* 2021 Apr;27(4):622-625.
304 doi: 10.1038/s41591-021-01285-x. Epub 2021 Mar 2. PMID: 33654292.

305

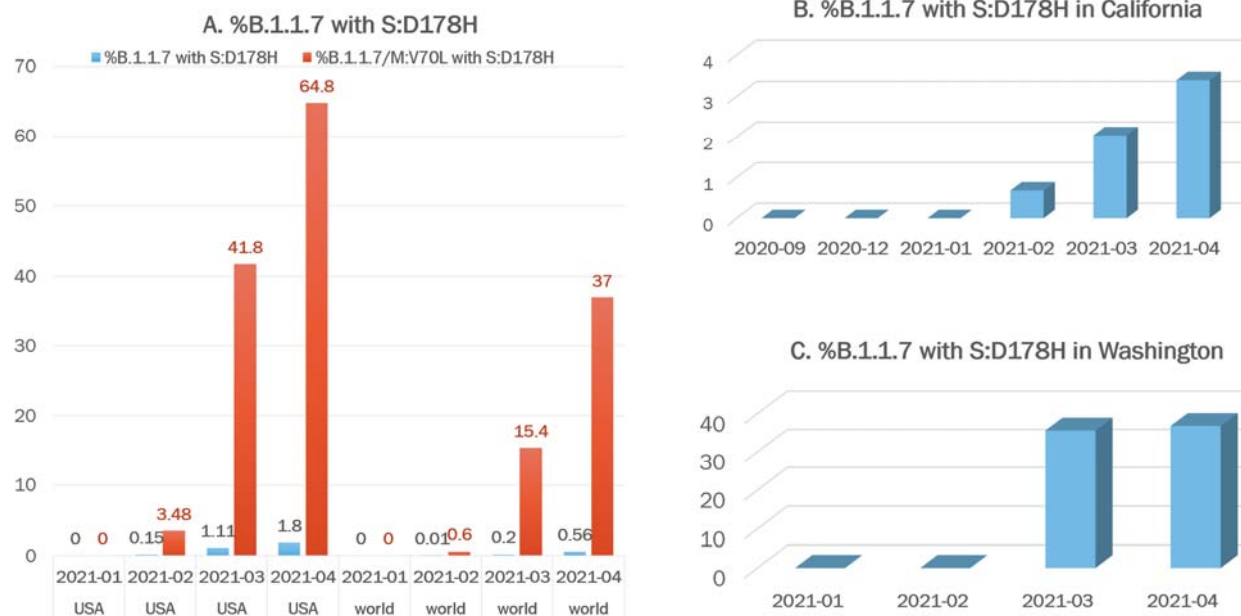
306

307 **Figure 1. Branching off a sub-lineage of B.1.1.7 that is distinguished by the M:V40L**
308 **mutation because of the emergence of the S:D178H mutation on January 23rd, 2021 as**
309 **estimated.** Each dot represents an isolate and is colored according to the state where it was
310 reported. The inset provides more information about the branching S:D178H mutation, namely
311 the genomic position, the inferred date, the data confidence interval, and the likely states of
312 origin.
313



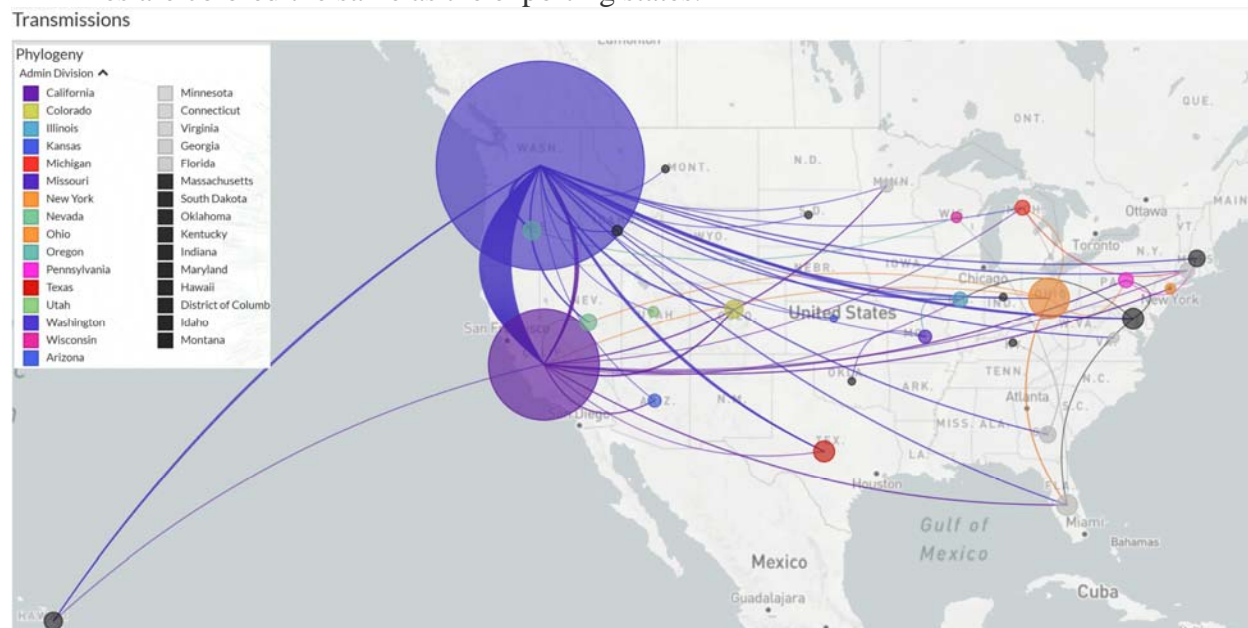
314
315
316
317

318 **Figure 2. Prevalence of the B.1.1.7 sub-lineage carrying mutation S:D178H Globally, U.S.,**
 319 **Washington, and California.** A. Percentages of B.1.1.7 (blue) or B.1.1.7-M:V70L (orange)
 320 isolates that carried the S:D178H mutation from January to April 2021 in U.S. and globally,
 321 respectively. B. Percentages of B.1.1.7 isolates that carried the S:D178H mutation from
 322 September 2020 to April 2021 in California. C. Percentages of B.1.1.7 isolates that carried the
 323 S:D178H mutation from January 2021 to April 2021 in Washington.



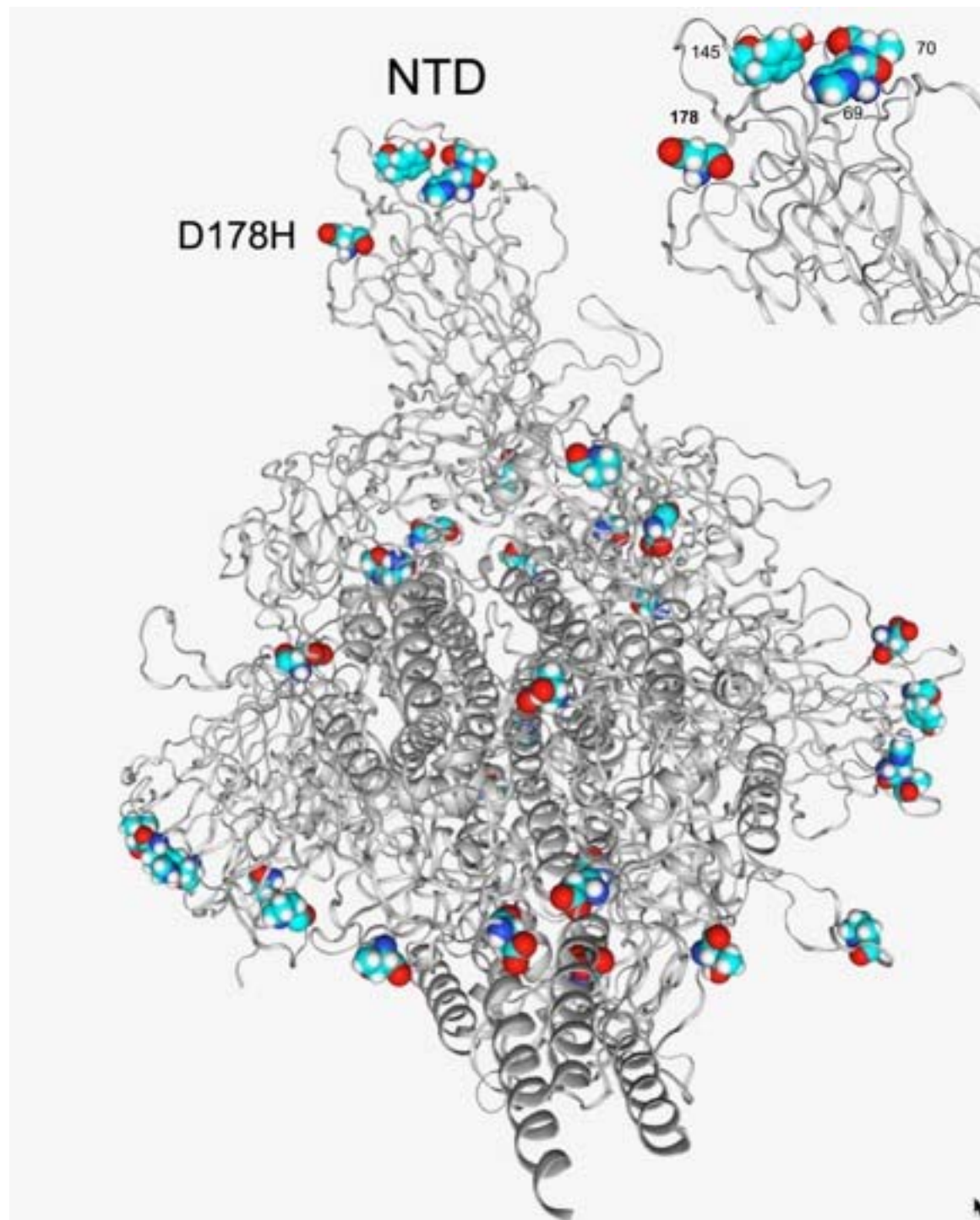
324
 325
 326

327 **Figure 3. Transmission of the 1026 B.1.1.7-MV70L-S:D178H isolates across the U.S..**
328 The size of the circles are drawn proportional to the number of cases in the specific state.
329 Lines are colored the same as the exporting states.



330
331
332

333 **Figure 4. Structural location of the D178H and other spike protein mutations in the B.1.1.7**
334 **lineage.** D178H is in close proximity to other amino acids in the NTD that are affected in the
335 B.1.1.7 lineage, HV69-70del and Y144/Y145del (top right).



336
337
338
339

340

341 **Table 1. Number of isolates from the B.1.1.7 lineage and with the M:V70L mutation in the**
 342 **U.S. and globally**

Month	Country	All isolates	B.1.1.7 isolates	All isolates w/ M:V70L	B.1.1.7 isolates w/ M:V70L	% all isolates w/ M:V70L	% B.1.1.7 isolates w/ M:V70L	% M:V70L in B.1.1.7
2021-04	U.S.	44878	25837	720	716	1.60	2.77	99.44
2021-03	U.S.	91468	36034	965	953	1.06	2.64	98.76
2021-02	U.S.	56669	6859	309	287	0.55	4.18	92.88
2021-01	U.S.	52532	1365	193	150	0.37	10.99	77.72
2020-12	U.S.	23593	117	46	43	0.50	36.75	93.48
2020-11	U.S.	17878	2	4	1	0.02	50.00	25
2021-04	World	109886	82392	1257	1253	1.14	1.52	99.68
2021-03	World	291238	195853	2607	2592	0.90	1.32	99.42
2021-02	World	219853	107841	1845	1816	0.84	1.68	98.43
2021-01	World	195282	69514	1005	932	0.51	1.34	92.74
2020-12	World	107503	18356	257	236	0.24	1.29	91.83
2020-11	World	81587	2222	39	29	0.05	1.31	74.36

343

344

345 **Table 2. Number of isolates form the B.1.1.7 lineage and with the S:D178H mutation in the**
 346 **U.S. and globally**

347
 348
 349

Month	Country	All isolates	B.1.1.7 isolates	All isolates w/ S: D178H	B.1.1.7 isolates w/ S:D178H	% all isolates w/ S: D178H	%B.1.1.7 isolates w/ S:D178H	% S:D178H from B.1.1.7 isolates
2021-04	U.S.	44878	25837	469	464	1.05	1.80	98.93
2021-03	U.S.	91468	36034	402	399	0.44	1.11	99.25
2021-02	U.S.	56669	6859	18	10	0.03	0.15	55.56
2021-01	U.S.	52532	1365	5	0	0.01	0.00	0.00
2021-04	World	109886	82392	469	464	0.43	0.56	98.93
2021-03	World	291238	195853	403	400	0.14	0.20	99.26
2021-02	World	219853	107841	20	11	0.01	0.01	55.00
2021-01	World	195282	69514	7	0	7	0.00	0.00

350
 351

352 **Table 3. Number of isolates from the B.1.1.7 lineage and with the S:D178H mutation in the**
353 **states of California and Washington**

State	Month	All isolates	B.1.1.7 isolates	Isolates S:D178H	w/ w/ S:D178H	% of all isolates w/ S:D178H	% B.1.1.7 isolates w/ S:D178H
California	2021-04	2904	1353	45	1.55	3.33%	
California	2021-03	9951	2254	46	0.46	2.04%	
California	2021-02	8889	589	10	0.11	0.68%	
California	2021-01	10441	214	4	0.04	0%	
Washington	2021-04	2464	990	364	14.77	36.8%	
Washington	2021-03	3077	832	296	9.62	35.6%	
Washington	2021-02	2152	115	0	0.0	0%	
Washington	2021-01	1510	33	0	0.0	0%	

354

355