


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

Lishuang Shen , Jennifer Dien Bard, Timothy J. Triche, Alexander R. Judkins, Jaclyn A. Biegel and Xiaowu Gai

Children's Hospital Los Angeles, Department of Pathology and Laboratory Medicine, Keck School of Medicine of University of Southern California, Los Angeles, CA, USA

ABSTRACT

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



ARTICLE HISTORY Received 14 May 2021; Revised 10 June 2021; Accepted 10 June 2021

KEYWORDS SARS-CoV-2; D178H; V70L; B.1.1.7 lineage; COVID-19

Introduction

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

Using the Children's Hospital Los Angeles (CHLA) COVID-19 Analysis Research Database (CARD) [3], and viral sequences submitted to GISAID and NCBI GenBank, we have routinely performed genomic epidemiology and genomic surveillance studies of local, national and international databases [4–9]. This allowed us to identify a new rapidly expanding SARS-CoV-2 lineage (B.1.575) with a signature mutation I82T in the M gene [7]. In the same study, we identified multiple other M mutations including V70L that are currently being encountered with significantly increased frequency. We have identified the M:V70L mutation in multiple SARS-CoV-2 lineages but primarily in the B.1.1.7 lineage. The B.1.1.7-M:V70L sub-lineage has been circulating at consistently moderate prevalence since November 2020. Through continuous genomic surveillance, we identified the subsequent acquisition of yet another Spike mutation, D178H, within this lineage. This new B.1.1.7 sub-lineage, carrying both M:V70L and S:D178H mutations, appeared in February 2021 but by April 2021 quickly increased to account for 36.8% and 1.8% of all reported B.1.1.7 genomes in

CONTACT Xiaowu Gai  xgai@chla.usc.edu  Center for Personalized Medicine, Division of Genomic Medicine, Department of Pathology and Laboratory Medicine, 4650 Sunset Blvd., Mailstop #173, Los Angeles 90027, CA, USA

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Washington and the US, respectively. This B.1.1.7-M:V70L-S:D178H sub-lineage is currently exclusive to the US as of 8 May 2021, and had been detected in 31 states, with the majority of cases found in Washington, California and Ohio. Here, we report its detection, characterization, transmission, and evolution.

Materials and methods

Ethics approval

The study design conducted at Children's Hospital Los Angeles was approved by the Institutional Review Board under IRB CHLA-16-00429.

SARS-CoV-2 whole genome sequencing

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

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

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

Phylogenetic analysis

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

Protein structure prediction

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

[ic.ac.uk/~missense3d/](https://www.imperial.ac.uk/~missense3d/)) [15]. CoV3D was used as the Spike Protein Mutation Viewer for the multiple mutations in B.1.1.7 (<https://cov3d.ibbr.umd.edu/MutViewer/QTY83983>).

Results

Identification of a rapidly emerging B.1.1.7 sub-lineage

We evaluated 1,333,679 SARS-CoV-2 viral genomes available on 1 May 2021, including 2900 from our own institution and the rest from GISAID and NCBI GenBank. We searched for SARS-CoV-2 mutations with a significantly higher prevalence rate in both the US and globally. Candidate mutations were further partitioned by pangolin lineage to identify emerging mutations in the context of a specific lineage, such as B.1.1.7 or B.1.351. We focused initially on the M mutations that we previously identified, including V70L, that was spiking near the end of 2020 [7]. Overall the percentage of isolates that carried the M:V70L mutation had been relatively stable in the US and globally with a gradual month to month increase (Table 1). In the vast majority of cases, the M:V70L mutation occurred on the B.1.1.7 lineage. While the percentage of B.1.1.7 isolates with the V70L mutation remained relatively stable across the world, the percentage fluctuated significantly in the US, attributable largely to the initial small number of B.1.1.7 cases in the U.S.

We identified the acquisition of another S mutation, D178H, in this B.1.1.7 sub-lineage (Figure 1 and Table 2), which was estimated to have occurred on 23 January 2021. By April, the prevalence of SARS-CoV-2 isolates carrying the S:D178H mutation increased to 1.05% nationally and as high as 14.77% in Washington. When we examined the prevalence of S:D178H in the context of the B.1.1.7 lineage and the B.1.1.7-M:V70L sub-lineage, the numbers were even more striking (Figure 2). While the percentage of all B.1.1.7 isolates that carried the S:D178H mutation increased from 0% (0/1365) in January 2021 to 0.15% (18/6859) in February 2021, to 1.11% (402/36034) in March 2021, and to 1.8% (469/25837) in April 2021, there was a significant increase from 0% (0/150) to 3.48% (10/287) to 41.8% (399/953) and 64.8% (464/716) in B.1.1.7-M:V70L sub-lineage isolates that carried the S:D178H mutation observed in the same 4-month time period. (Figure 2(A) and Table 2). The B.1.1.7-M:V70L-S:D178H sub-lineage was exclusive to the U.S. (Table 2), and findings, when examined globally, were not as striking, going in the same 4 months from 0% (0/69514) to 0.01% (20/107841) to 0.2% (403/195853) and to 0.56% (469/82392) in B.1.1.7 lineage and 0% (0/932) to 0.6% (11/1816) to 15.4% (400/2592) and to 37% (464/1253) in B.1.1.7-

Table 1. Number of isolates from the B.1.1.7 lineage and with the M:V70L mutation in the US 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

M:V70L sub-lineage, respectively (Figure 2(A) and Table 2).

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

Within the US, most isolates of this lineage were from Washington, California, and Ohio (Figure 3). California and Washington showed the greatest

increases in terms of the percentage of B.1.1.7 isolates carrying the S:D178H mutation. In Washington, the S:D178H mutation was absent until February 2021, where it quickly increased to account for 9.6% (296/3077) of all viral isolates in March and 14.8% (364/2464) in April (Table 3). It accounted for 36.8% (364/990) of the April B.1.1.7 isolates in Washington (Figure 2(C) and Table 3). In California, S:D178H was first seen in December 2020, but it was not seen within the B.1.1.7 lineage until 4 February 2021 (Table 3). Its prevalence increased to 1.6% (45/2904) in April compared to all isolates studied in California,

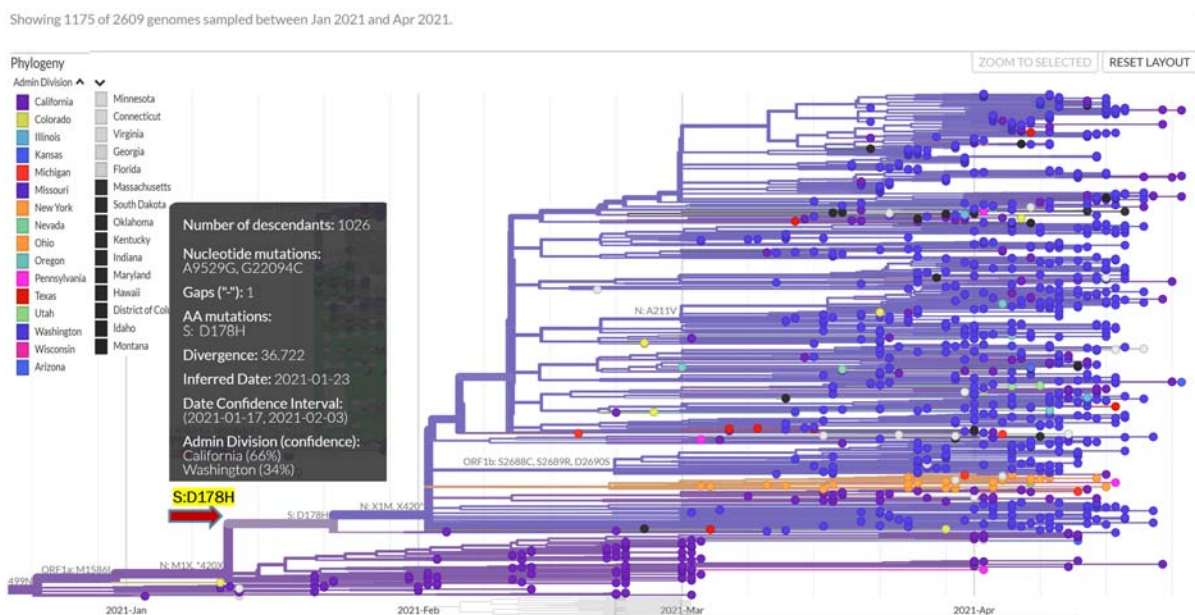


Figure 1. Branching off a sub-lineage of B.1.1.7 that is distinguished by the M:V40L mutation because of the emergence of the S:D178H mutation on 23 January 2021 as estimated. Each dot represents an isolate and is colored according to the state where it was reported. The inset provides more information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

Table 2. Number of isolates form the B.1.1.7 lineage and with the S:D178H mutation in the US and globally.

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

and 3.3% (45/1353) of the B.1.1.7 isolates (Figure 2(C) and Table 3).

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

To understand the origin and evolution of this B.1.1.7-M:V70L-S:D178H sub-lineage, we used (a) all 1125 S:D178H carrying viral sequences reported across the globe, (b) 1600 subsampled M:V70L carrying viral sequences across the world, and (c) the NC_045512 reference genome. These sequences were further filtered with requirements of being at least 27,000 bp and having complete date information. 2609 genomes were kept for final phylogenetic and transmission analysis using Nextstrain, rooted by NC_045512 (Figure 1). The D178H branch divergence date was estimated to be 23 January 2021, with a date confidence interval of 17 January 2021 to February 3, 2021. It likely originated in California (66% probability) or Washington (34% probability).

Estimated transmission routes are demonstrated in Figure 3, where the size of the circle represents the

number of cases from this sub-lineage in each state, and the line colours correspond to the exporting locations, California, Washington, and Ohio contributed the majority of transmissions compared to other states. This sub-lineage was clearly exclusive to the US, and not reported in any other countries at the time of the study.

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

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

Protein structure and mutation effect prediction

The 3D structure of the Spike protein, as visualized using the CoV3D mutation viewer. Using these results we were able to show that the S:D178H mutation is close structurally to two signature deletions of

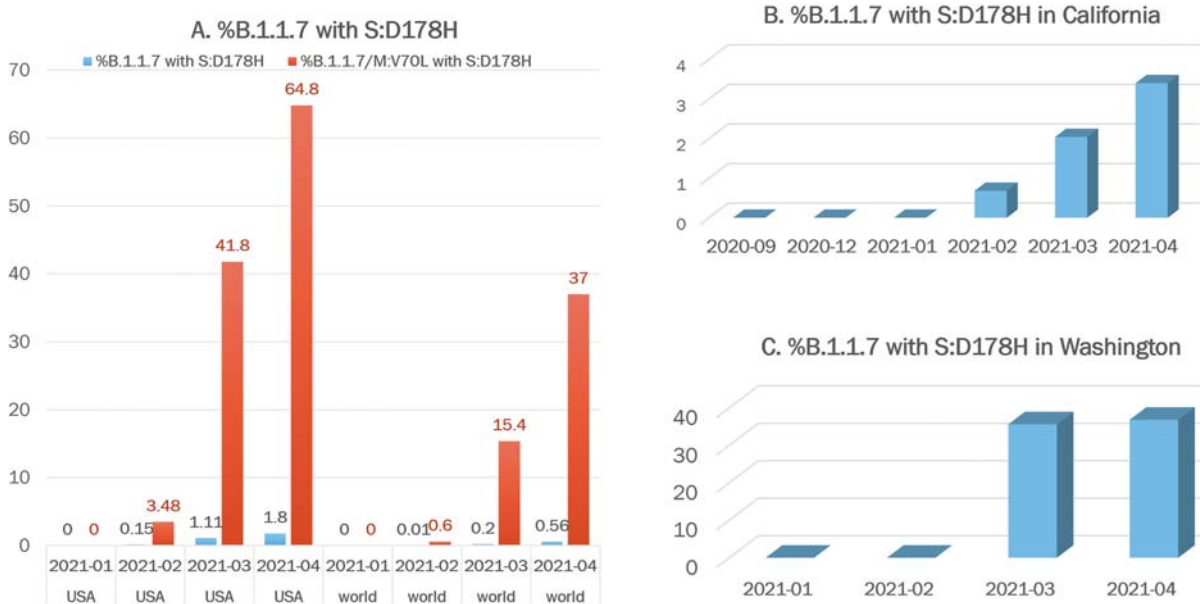


Figure 2. Prevalence of the B.1.1.7 sub-lineage carrying mutation S:D178H Globally, US, Washington, and California. (A) Percentages of B.1.1.7 (blue) or B.1.1.7-M:V70L (orange) isolates that carried the S:D178H mutation from January to April 2021 in US and globally, respectively. (B) Percentages of B.1.1.7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

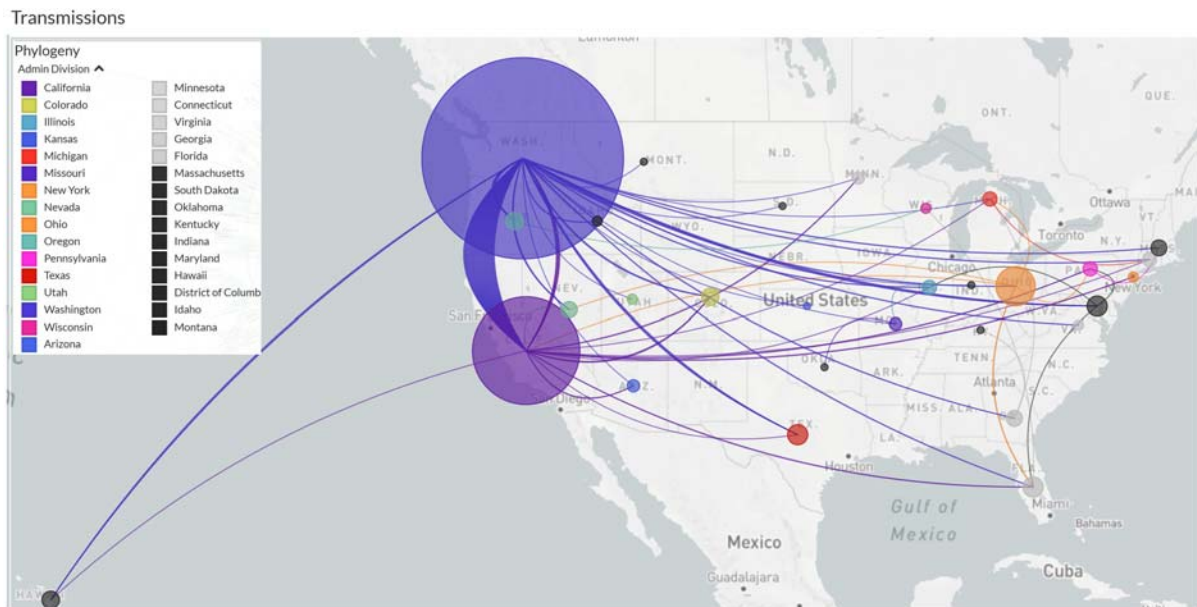


Figure 3. Transmission of the 1026 B.1.1.7-MV70L-S:D178H isolates across the US. The size of the circles are drawn proportional to the number of cases in the specific state. Lines are coloured the same as the exporting states.

B.1.1.7, HV69_70del and Y144del (Figure 4). They are all surface exposed and likely alter N terminal domain (NTD) tertiary configuration.

Discussion

It should be noted that our study relied primarily upon sequence data deposited at GISAID, which represents a limitation and could also introduce potential bias, as state public health laboratories have varied sequencing capacity and nonuniform data sharing and reporting practices. It is less likely to be biased significantly by the practice of a single laboratory though. As an example, we included 2900 SARS-CoV-2 sequences that we obtained at Children's Hospital Los Angeles since March 2020. We found only 29 sequences belonging to the B.1.1.7 lineage, among which only one carried M:V70L and S:D178H mutations. Inclusion of this data set did not impact the summary statistics of the state of California significantly.

In March 2021, the B.1.1.7-M:V70L-S:D178H lineage appeared abruptly at high prevalence (35.6%) in all reported B.1.1.7 isolates in Washington. This was a remarkable finding as it was absent in 115 Washington B.1.1.7 isolates reported in February. This abrupt change could reflect undersampling or

potentially reflect superspreader events. Given the spike in the number of B.1.1.7 cases, this sub-lineage clearly appears to be more transmissible than even the original B.1.1.7 lineage. This finding warrants prompt and further attention by public health authorities, as this mutation profile is closely linked to the resurgence of cases in Washington in particular. It also strongly supports the now widely recognized need for more extensive SARS-CoV-2 viral sequencing of PCR positive COVID-19 cases for detection of new mutations of concern as part of widespread genomic surveillance [16,17].

The S:D178H mutation, while demonstrably associated here with the more pathogenic B.1.1.7 lineage, is not necessarily by itself more pathogenic. Dozens of SARS-CoV-2 genomes that carry the S:D178H mutation were reported before February, but none of these demonstrated the increased frequency seen when the mutation occurs in the context of the B.1.1.7 lineage. Phylogenetic analysis revealed a distinct and long branch leading to the new S:D178 branch after M:V70L. Together, these observations suggest that the S:D178H mutation is recurrent, but only increased exponentially in the context of the more pathogenic B.1.1.7 lineage, which serves as an argument for its fitness. This is the same observed

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

State	Month	All isolates	B.1.1.7 isolates	Isolates 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%

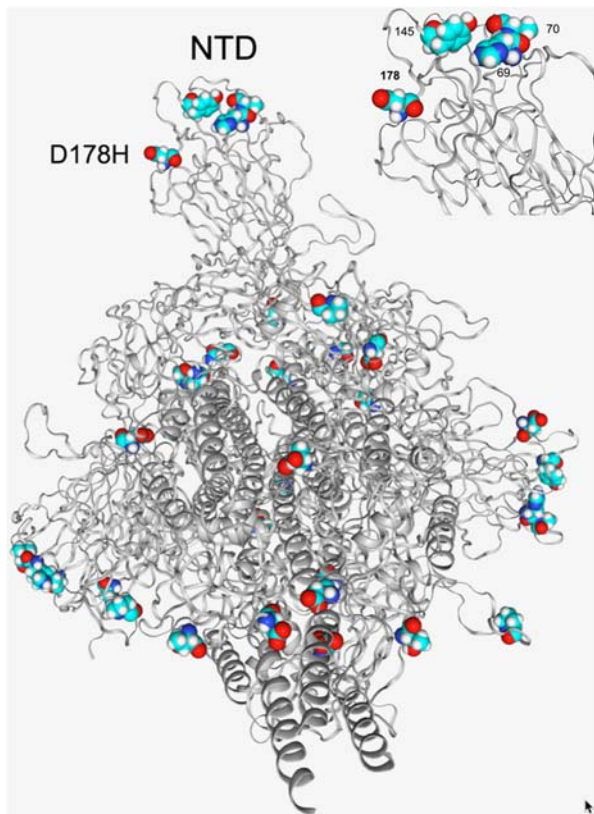


Figure 4. Structural location of the D178H and other spike protein mutations in the B.1.1.7 lineage. D178H is in close proximity to other amino acids in the NTD that are affected in the B.1.1.7 lineage, HV69-70del and Y144/Y145del (top right).

with other deleterious mutations like the N501Y and E484K mutations, both of which are now superimposed on distinct and separate more pathogenic lineages (<https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>). The S:D178H mutation arose independently again in the US on the B.1.1.7-M:V70L background. The rapid increase in its prevalence, only after its acquisition by the B.1.1.7-M:V70L sub-lineage suggests this combination of mutations is associated with increased transmissibility. It is also of interest that this mutation occurs in the NTD, unlike most of the mutations associated with current VOC that are centred on the spike protein RBD, implying that NTD mutations beyond the original 69-70del and the 144del are of concern. And finally, it should be noted that this NTD mutation co-exists with the previously reported M protein mutation M:V70L, suggesting that M protein mutations also contribute to enhanced biologic “fitness” or pathogenicity of this sub-lineage.

The appearance of the S:D178H mutation in the context of the B.1.1.7 lineage is temporally associated with the increased incidence of COVID-19 in Washington. New cases in Washington were higher than the national level at the time of the study. According to New York Time COVID-19 dashboard, the 7-day average of new cases on May 2 was 1379 in

Washington, which was only a 50% reduction compared to 2757 cases on December 15. In comparison, the numbers on May 2 was 49,270 in US, a 77.3% reduction from the December 15 number of 21,7325. The appearance of this new B.1.1.7 sub-lineage temporally linked to increased cases in Washington warrants further investigation.

The potential effect of the S:D178H mutation on immunity and vaccine “escape” also warrant further analysis. Mutations in the Spike N-terminal domain have been associated with a lack of neutralization by NTD directed antibodies, especially when the N5 loop is affected [18,19]. The NTD initiates viral binding to the ACE2 receptor-expressing host cell. Since the D178H falls in the NTD domain close to the N5 loop, it may alter NTD structure and antibody recognition. It may thus have a similar immune evasion effect as the HV69-70del and Y144del mutations [20], or it may further enhance that of the two other mutations, based on the 3D model. These findings highlight the continued importance of active genomic surveillance to monitor the spread of this B.1.1.7-M:V70L-S:178H lineage.

Acknowledgments

The authors would like to acknowledge all members of the Department of Pathology and Laboratory Medicine (PLM) at Children’s Hospital Los Angeles for dedication towards providing excellent patient care throughout the pandemic, including especially the rapid launch of both the COVID-19 diagnostic test in the Clinical Microbiology and Virology Laboratory and the SARS-CoV-2 whole genome sequencing assay at the Center for Personalized Medicine in March 2020, both with the support of Thermo Fisher and Paragon Genomics, with whom these assays were developed. The authors would also like to acknowledge the frontline health-care workers who remain devoted to the fight against COVID-19. We would also like to acknowledge all institutions that have contributed the SARS-CoV-2 sequences timely, as well as NCBI, GISAID, and Nextstrain for providing valuable resources for SARS-CoV-2 genomics.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Lishuang Shen  <http://orcid.org/0000-0002-0436-0199>

References

- [1] Horby P, Huntley C, Davies N, Edmunds J, Ferguson N, Medley G, Calum Semples G. SAGE meeting report. January 21, 2021.
- [2] Greaney AJ, Loes AN, Crawford KHD, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell*

- Host Microbe. 2021;29(3):463–476.e6. doi:10.1016/j.chom.2021.02.003.
- [3] Shen L, Maglinte DT, Ostrow D, et al. Children's Hospital Los Angeles COVID-19 Analysis Research Database (CARD) – a resource for rapid SARS-CoV-2 genome Identification using interactive online phylogenetic tools. *bioRxiv*. 2020. doi:10.1101/2020.05.11.089763
- [4] Shen L, Dien Bard J, Biegel JA, et al. Comprehensive genome analysis of 6,000 USA SARS-CoV-2 isolates reveals haplotype signatures and localized transmission patterns by state and by country. *Front Microbiol*. 2020;11:573430. doi:10.3389/fmicb.2020.573430. PMID: 33013809; PMCID: PMC7509426.
- [5] Pandey U, Yee R, Shen L, et al. High prevalence of SARS-CoV-2 genetic variation and D614G mutation in pediatric patients with COVID-19. *Open Forum Infect Dis*. ofaa551.
- [6] Truong TT, Ryutov A, Pandey U, et al. Persistent SARS-CoV-2 infection and increasing viral variants in children and young adults with impaired humoral immunity. *medRxiv [Preprint]*. 2021. Update in: *EBioMedicine*. 2021 Apr 26;67:103355. PMID: 33688673; PMCID: PMC7941650.
- [7] Shen L, Bard JD, Triche TJ, et al. Emerging variants of concern in SARS-CoV-2 membrane protein: a highly conserved target with potential pathological and therapeutic implications. *Emerg Microbes Infect*. 2021: 1–24. doi:10.1080/22221751.2021.1922097. Epub ahead of print. PMID: 33896413.
- [8] Truong TT, Ryutov A, Pandey U, et al. Increased viral variants in children and young adults with impaired humoral immunity and persistent SARS-CoV-2 infection: A consecutive case series. *EBioMedicine*. 2021;67:103355, doi:10.1016/j.ebiom.2021.103355 Epub ahead of print. PMID: 33915337; PMCID: PMC8072072.
- [9] Ryutov A, Gai X, Ostrow D, et al. Utility of viral whole-genome sequencing for institutional infection surveillance during the coronavirus disease 2019 (COVID-19) pandemic. *Infect Control Hosp Epidemiol*. 2021: 1–2. doi:10.1017/ice.2021.185. Epub ahead of print. PMID: 33866984.
- [10] Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall*. 2017;1(1):33–46. doi:10.1002/gch2.1018 PMID: 31565258; PMCID: PMC6607375.
- [11] Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 2017 Mar 30;22(13):30494. PMID: 28382917; PMCID: PMC5388101.
- [12] Katoh K, Misawa K, Kuma K, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059–3066. doi:10.1093/nar/gkf436 PMID: 12136088; PMCID: PMC135756.
- [13] Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–1534. doi:10.1093/molbev/msaa015. Erratum in: *Mol Biol Evol*. 2020 Aug 1;37(8):2461. PMID: 32011700; PMCID: PMC7182206.
- [14] Sagulenko P, Puller V, Neher RA. Treetime: maximum-likelihood phylodynamic analysis. *Virus Evol*. 2018;4(1):vex042, doi:10.1093/ve/vex042. PMID: 29340210; PMCID: PMC5758920.
- [15] Ittisoponpisan S, Islam SA, Khanna T, et al. Can predicted protein 3D structures provide reliable insights into whether Missense variants Are disease associated? *J Mol Biol*. 2019;431(11):2197–2212. doi:10.1016/j.jmb.2019.04.009 Epub 2019 Apr 14. PMID: 30995449; PMCID: PMC6544567. [Google Scholar].
- [16] Cyranoski D. Alarming COVID variants show vital role of genomic surveillance. *Nature*. 2021;589(7842):337–338. doi:10.1038/d41586-021-00065-4. PMID: 33452508.
- [17] Walensky RP, Walke HT, Fauci AS. SARS-CoV-2 variants of concern in the United States-challenges and opportunities. *JAMA*. 2021; 325(11):1037. Epub ahead of print. PMID: 33595644
- [18] Wang P, Nair MS, Liu L, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*. 2021;593(7857):130–135. doi:10.1038/s41586-021-03398-2. Epub 2021 Mar 8. PMID: 33684923.
- [19] Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med*. 2021 Apr;27(4):622–625. doi:10.1038/s41591-021-01285-x. Epub 2021 Mar 2. PMID: 33654292.
- [20] McCarthy KR, Rennick LJ, Nambulli S, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science*. 2021 Mar 12;371(6534):1139–1142. doi:10.1126/science.abf6950. Epub 2021 Feb 3. PMID: 33536258; PMCID: PMC7971772.