

# Distinct viral clades of SARS-CoV-2: Implications for modeling of viral spread

Adam Brufsky MD, PhD 

Department of Medicine, Division of Hematology-Oncology, UPMC Hillman Cancer Center, Magee Women's Hospital, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

## Correspondence

Adam Brufsky, MD, PhD, Department of Medicine, Division of Hematology-Oncology, UPMC Hillman Cancer Center, Magee Women's Hospital, University of Pittsburgh School of Medicine, Suite 4628, 300 Halket St, Pittsburgh, PA 15213.

Email: [brufskyam@upmc.edu](mailto:brufskyam@upmc.edu)

A current model of the coronavirus disease 2019 (COVID-19) epidemic<sup>1</sup> appears to have differing degrees of projected total deaths in various areas of the United States despite the close timing of social distancing measures. For example, California with a population of approximately 40 000 000, began social distancing measures on 19 March 2020, yet is projected to plateau at 1616 COVID-19 related deaths by 15 May 2020. New Jersey, by contrast, with a population of approximately 9 000 000 began social distancing measures on 18 March 2020 yet is projected to plateau at 3915 COVID-19-related deaths by 1 May 2020. Factors other than social distancing may in part explain this discrepancy.

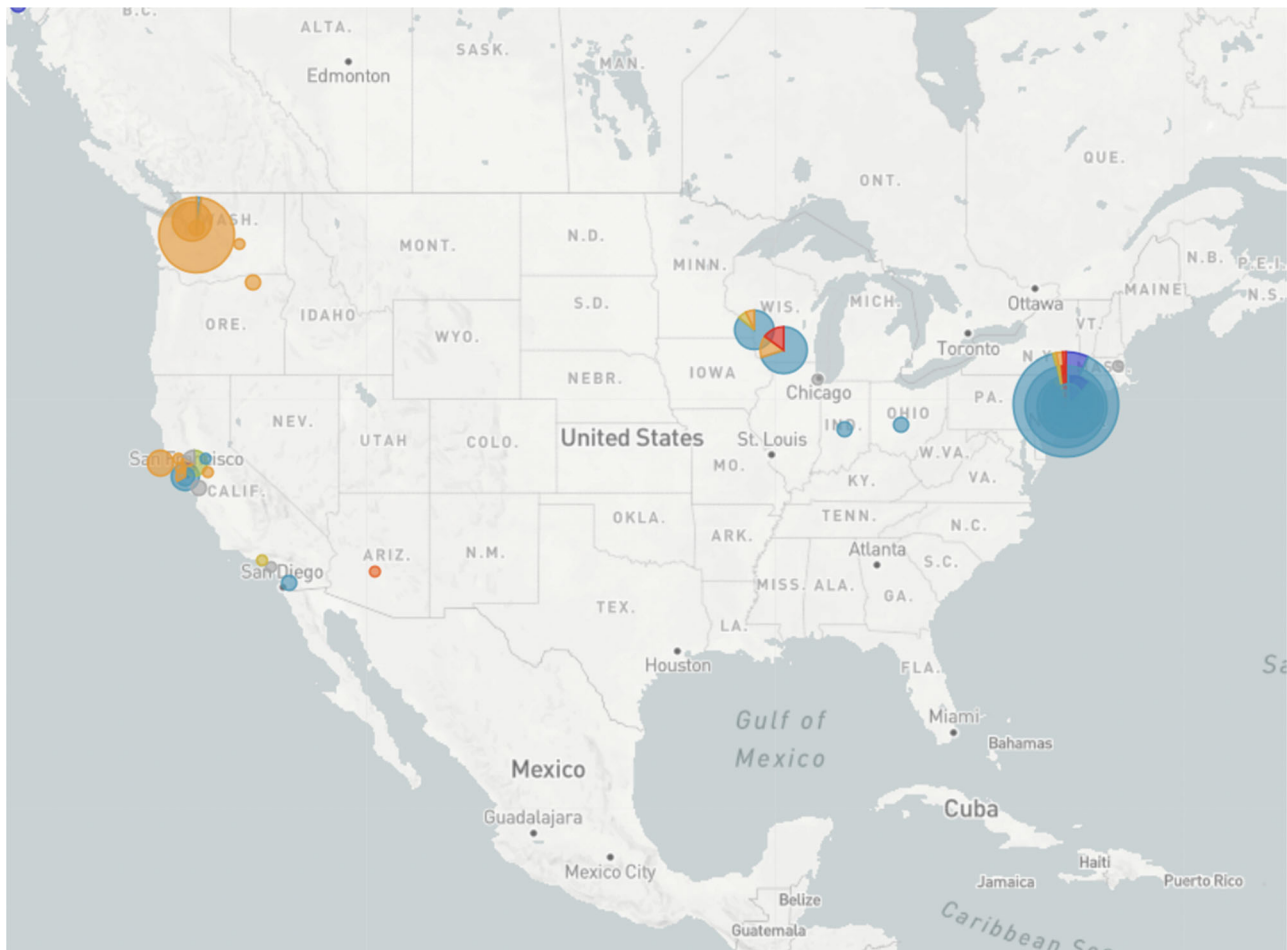
These and other COVID-19 epidemic models appear to assume a universal basic reproductive number ( $R_0$ ) for virus spread, as well as a uniform degree of pathogenicity. A recent report<sup>2</sup> suggests the  $R_0$  of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) may be as high as 5.7, increased from the original estimated  $R_0$  of 2.28.<sup>3</sup> These reports also appear to assume uniform pathogenicity and transmission of the virus through time and do not account for mutation of the virus either towards or away from a more virulent strain.

As is typical of coronavirus, SARS-CoV-2 has rapidly acquired mutations allowing for tracking its ancestry and spread.<sup>4,5</sup> For example, an S type of the virus accounted for only 3.7% of viral isolates in Wuhan compared to 96.3% of the L type yet isolates outside of Wuhan were 61.3% L type and 38.4% S type.<sup>4</sup> Examining clades of SARS-CoV-2 in 2310 geographically distributed viral DNA sequences using Nexstrain<sup>5</sup> (Figure 1) reveal that the B1 clade predominates on the West Coast of the United States, and the A2a clade, which apparently spread to New York through Europe and Italy if this limited viral sequence phylogeny analysis is correct, predominates on the East Coast.

It has not escaped notice that these clades may vary in virulence. For example, analysis in Nexstrain in 2310 viral isolates of distribution of a nonsynonymous mutation in the viral spike protein at codon 614 (Figure 2) reveals that aspartic acid (D) at this residue is predominant on the West Coast and glycine (G) is predominant on the East Coast. SARS-CoV-2 genome phylogeny analysis reveals that this D614G mutation appeared to arise from an ancestral D residue (Figure 3). This mutation resides in a highly glycosylated region of the viral spike protein. A theory of viral

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Medical Virology* published by Wiley Periodicals LLC

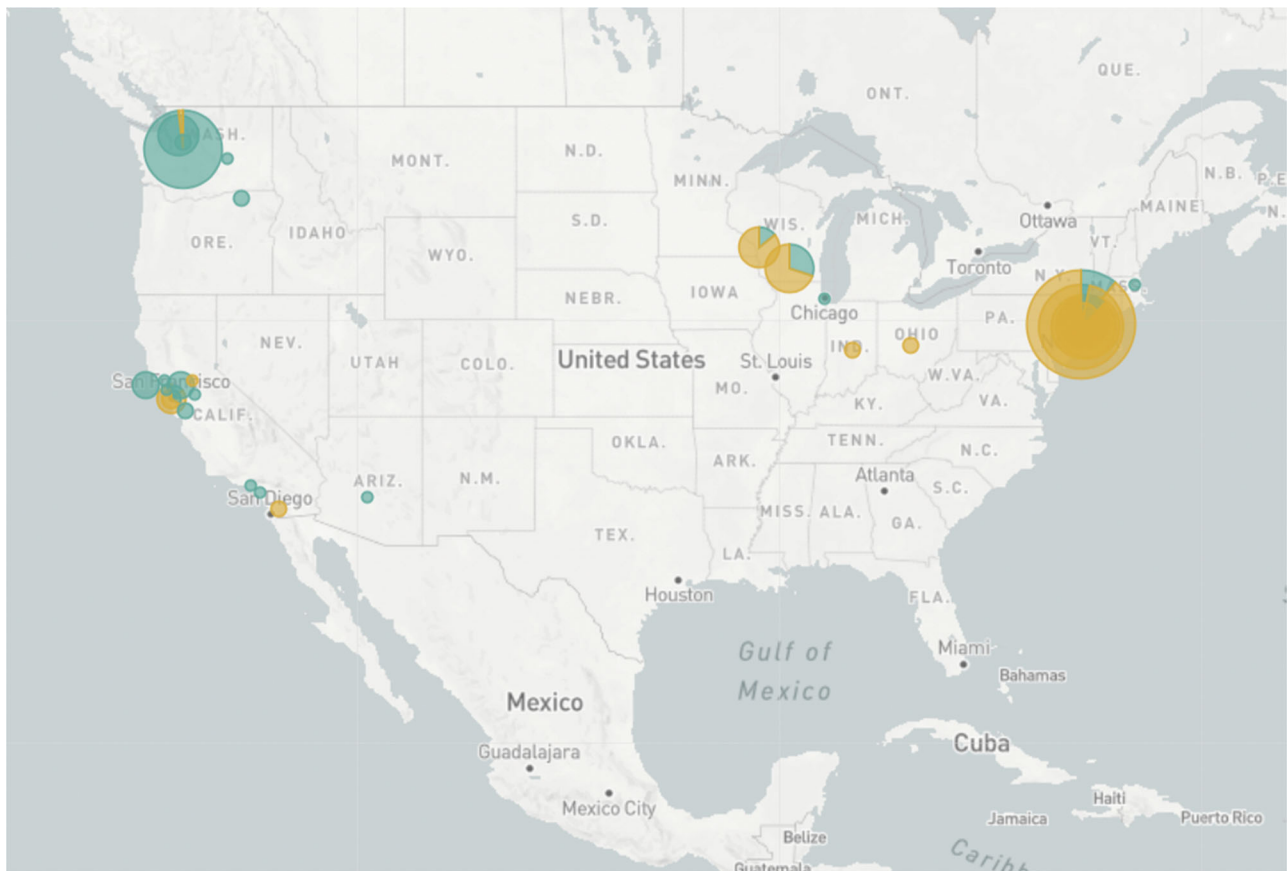


**FIGURE 1** Dominant viral clades in United States clade B1 (orange) and clade A2a (blue) [https://nextstrain.org/ncov/north-america?c=clade\\_membership&r=location](https://nextstrain.org/ncov/north-america?c=clade_membership&r=location). Accessed 15 April 2020

pathogenesis<sup>6</sup>, as well as cryogenic cryogenic electron microscopy<sup>7</sup> and structure and function experiments<sup>8</sup>, suggest the possibility that mutations in this region could alter this heavily glycosylated viral spike and alter membrane fusion in tissues, resulting in more pathogenicity and more human to human spread. This residue is highly conserved in coronaviruses.<sup>8</sup> There are many examples of mutation in this spike protein region resulting in

changes in virulence,<sup>9,10</sup> and, in particular, a single serine to glycine change at residue 310 in a murine coronavirus spike protein resulted in increased virulence<sup>11</sup> through decreased stability of the viral fusion machinery.

This analysis has clear limitations. Virus phylogeny studies, for example, can have sampling bias as well as statistical bias in the inference model used.<sup>12</sup> However, it should be relatively



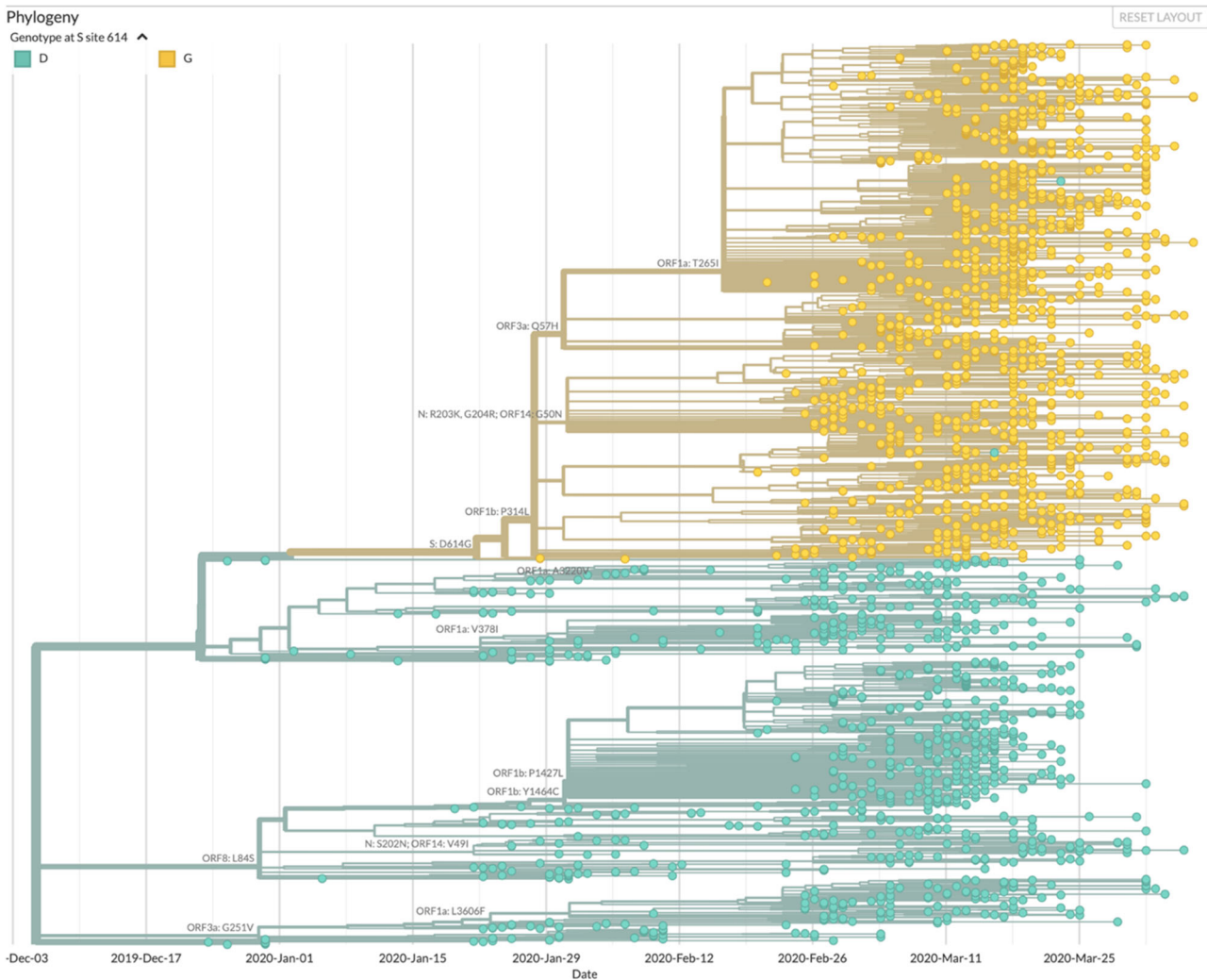
**FIGURE 2** Distribution of codon 614 in the SARS-CoV-2 coronavirus aspartic acid (aqua), glycine (orange) [https://nextstrain.org/ncov/north-america?c=gt-S\\_614&r=location](https://nextstrain.org/ncov/north-america?c=gt-S_614&r=location). Accessed 15 April 2020. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

straightforward to test the virulence of various viral strains of SARS-CoV-2 in vitro, in particular those that have variability at residue 614 of the viral spike protein.

This suggests that competition between viral strains of varying virulence may be at play during the COVID-19 pandemic. It is unclear whether current serologic or viral polymerase chain reaction-based assays are able to detect this variability.<sup>8</sup> If this hypothesis is correct,

it may be important to develop assays based on local viral clades, to determine the distribution of the virus and its spectrum of effects. It may also be important for modelers to consider these possible variations in SARS-CoV-2 viral pathogenicity as models are developed for a gradual relaxation of social restrictions. In particular, differences and timing of the activation of innate or adaptive immunity to both viral strains as the pandemic progresses could be considered.

Showing 2310 of 2310 genomes sampled between Dec 2019 and Apr 2020.



**FIGURE 3** Inferred phylogeny of codon 614 spike protein mutation in SARS-CoV-2 [https://nextstrain.org/ncov/north-america?branchLabel=aa&c=gt-S\\_614](https://nextstrain.org/ncov/north-america?branchLabel=aa&c=gt-S_614). Accessed 15 April 2020. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

### CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

### ORCID

Adam Brufsky  <http://orcid.org/0000-0001-8080-7960>

### REFERENCES

- Murray CJL. IHME COVID-19 health service utilization forecasting team. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *medRxiv*. 2020. <https://doi.org/10.1101/2020.03.27.20043752>
- Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg Infect Dis J*. 2020;26(7). <https://doi.org/10.3201/eid2607.200282>
- Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: a data-driven analysis. *Int J Infect Dis*. 2020;93:201-204. <https://doi.org/10.1016/j.ijid.2020.02.033>
- Tang X, Wu C, Li X, et al. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev*. 2020. <https://doi.org/10.1093/nsr/nwaa036>
- Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121-4123. <https://doi.org/10.1093/bioinformatics/bty407>
- Brufsky A. Hydroxychloroquine, hyperglycemia, and the COVID 19 epidemic. *J Med Virol*. 2020. <https://doi.org/10.1002/jmv.25887>
- Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020; 367(6483):1260-1263. <https://doi.org/10.1126/science.abb2507>
- Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. 2020;181:281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>

9. Krueger DK, Kelly SM, Lewicki DN, Ruffolo R, Gallagher TM. Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J Virol.* 2001;75(6):2792-2802.
10. Geoghegan JL, Holmes EC. The phylogenomics of evolving virus virulence. *Nat Rev Genet.* 2018;19(12):756-769.
11. Ontiveros E, Kim TS, Gallagher TM, Perlman S. Enhanced virulence mediated by the murine coronavirus, mouse hepatitis virus strain JHM, is associated with a glycine at residue 310 of the spike glycoprotein. *J Virol.* 2003;77(19):10260-10269.
12. Liò P, Goldman N. Phylogenomics and bioinformatics of SARS-CoV. *Trends Microbiol.* 2004;12(3):106-111.