medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Rapid displacement of SARS-CoV-2 variant B.1.1.7 by B.1.617.2 and P.1 in the United States

Alexandre Bolze[1,*], Elizabeth T. Cirulli[1,*], Shishi Luo[1,*], Simon White[1,*], Dana Wyman[1], Andrew Dei Rossi[1], Tyler Cassens[1], Sharoni Jacobs[1], Jason Nguyen[1], Jimmy M. Ramirez III[1], Efren Sandoval[1], Xueqing Wang[1], David Wong[1], David Becker[1], Marc Laurent[1], James T. Lu[1], Magnus Isaksson[1], Nicole L. Washington[1], William Lee[1]

1: Helix, San Mateo, CA
*: Equal contributions

Correspondence: alexandre.bolze@helix.com, william.lee@helix.com

Abstract

The SARS-CoV-2 variant of concern B.1.617.2 displaced B.1.1.7 as the dominant variant in England and other countries. This study aimed to determine whether B.1.617.2 was also displacing B.1.1.7 in the United States. We analyzed PCR testing results and viral sequencing results of samples collected across the United States, and showed that B.1.1.7 was rapidly being displaced and is no longer responsible for the majority of new cases. The percentage of SARS-CoV-2 positive cases that are B.1.1.7 dropped from 67% in May 2021 to 33.4% in just 5 weeks. Our analysis showed rapid growth of variants B.1.617.2 and P.1 as the primary drivers for this displacement. Currently, the growth rate of B.1.617.2 was higher than P.1 in the US (0.66 vs. 0.34), which is consistent with reports from other countries.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Introduction

The SARS-CoV-2 B.1.617.2 variant, also named Delta, has recently been classified as a variant of concern (VOC) by Public Health England (PHE), the World Health Organization (WHO), and the U.S. Centers for Disease Control (CDC) [1]. The B.1.617.2 variant is the predominant variant in India and in the United Kingdom, and has been identified in 65 countries as of June 17, 2021 [2]. It has been shown to be more transmissible than the SARS-CoV-2 B.1.1.7 variant, also named Alpha, in England [3]. Moreover, a study by Public Health England showed that vaccine efficacy for AstraZeneca and Pfizer vaccines remained very good (>90%) against hospitalizations after two doses [4]. However, vaccine efficacy was lower against B.1.617.2 compared to B.1.1.7 after one dose.

In the United States, the first B.1.617.2 we sequenced was from a sample collected on March 13. The context in the United States is different compared to England in terms of vaccine strategies and the existing viral background. In England, B.1.1.7 represented more than 90% of the SARS-CoV-2 sequences when B.1.617.2 was first identified in the country, and there were very few sequences of P.1, also named Gamma, another variant of concern. In the United States, B.1.1.7 plateaued just above 70%, and there was a greater diversity of variants when B.1.617.2 started to emerge, including an increasing amount of P.1 [2].

The objectives of this study are therefore (i) to analyze the impact of the introductions of B.1.617.2 and P.1 variants of concern on the prevalence of B.1.1.7 in the United States, and (ii) to analyze the growth and transmissibility of B.1.617.2 and P.1 in the United States. To this end, we looked at the PCR testing results and sequencing results of samples collected by the Helix laboratory across the United States since March 2021. Importantly, the collection method and collection sites have not changed in the last few months, and the samples analyzed should not be biased for very localized outbreaks. We therefore make the assumption that there was no significant sampling bias between the testing and sequencing done by our lab in February and March 2021, when B.1.1.7 was rapidly increasing in the United States, and the months of May and June 2021.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Methods

Ethical statement

Helix data analyzed and presented here were obtained through IRB protocol WIRB#20203438, which grants a waiver of consent for a limited dataset for the purposes of public health under section 164.512(b) of the Privacy Rule (45 CFR § 164.512(b)).

Helix COVID-19 test data and sample selection

All viral samples in this investigation were collected by Helix through its COVID-19 diagnostic testing laboratory. The Helix COVID-19 Test (EUA 201636) was run on specimens collected across the US, and results were obtained as part of our standard test processing workflow using specimens from anterior nares swabs. The Helix COVID-19 Test is based on the Thermo Fisher TaqPath COVID-19 Combo Kit, which targets three SARS-CoV-2 viral regions (N gene, S gene, and ORF1ab). Test results from positive cases, together with a limited amount of metadata (including sample collection date, state, and RT-qPCR Cq values for all gene targets), were used to build the research database used here. Ongoing summary level data are viewable at https://www.helix.com/covid19db. Data used for analysis are based on samples that tested positive with N gene Cq value < 29.

SARS-CoV-2 sequencing and consensus sequence generation

Sequencing was performed by Illumina [6], and more recently by Helix, as part of the SARS-CoV-2 genomic surveillance program led by the Centers for Disease Control and Prevention (CDC). In the Helix workflow, RNA was extracted from 400 µl of patient anterior nares sample using the MagMAX Viral/Pathogen kit (ThermoScientific). All samples were subjected to total RNA library preparation using the Rapid RNA Library Kit Instructions (Swift Biosciences). SARS-COV-2 genome capture was accomplished using hybridization kit xGen COVID-19 Capture Panel (Integrated DNA Technologies). Samples were sequenced using the NovaSeq 6000 Sequencing system S1 flow cell, which included the NovaSeq 6000 Sequencing System S1 Reagent Kit v1.5 (300 cycles). Bioinformatic processing of this sequencing output was as follows. The flow cell output was demultiplexed with bcl2fastq (Illumina) into per-sample FASTQ sequences that were then run through the Helix klados-fastagenerator pipeline v1.6.0 to produce a sequence FASTA file. First, reads were aligned to a reference comprising the SARS-CoV-2 genome (NCBI accession NC_045512.2) and the human transcriptome (GENCODE v37) using BWA-MEM. Following duplicate-marking, SARS-CoV-2 variants were called using the Haplotyper algorithm (Sentieon, Inc). The per-base coverage from the alignment file (BAM) and per-variant allele depths from the variant call format (VCF) file were then used to build a consensus sequence according to the following criteria: if there are at least 5 unique reads covering a base, and at least 80% of the reads support a particular allele, that allele is reported. Otherwise, that base is considered uncertain, and an N is reported.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Viral lineage designation

Viral sequences were assigned a Pango lineage [5] using pangolearn (https://github.com/cov-lineages/pangoLEARN). We analyzed 54,294 sequences from samples collected in 2021 for this analysis.

Vaccination rates

Vaccination rate by county was downloaded from the CDC (https://covid.cdc.gov/covid-data-tracker/#county-view). The percent of individuals completely vaccinated on the given date was used. We used the percentage of individuals completely vaccinated as of May 1 2021, a date that would be relevant to the types of virus growth patterns seen in June. States with more than 25% of records missing county information were excluded (CO, GA, TX, VA, WV).

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Results

B.1.1.7 is rapidly decreasing in the US

One of the defining mutations of the B.1.1.7 variant of concern is the deletion of amino acids 69 and 70 in the spike protein. This deletion interferes with the PCR test target on the S gene in many COVID-19 tests [7], including the Helix COVID-19 Test, and causes S-gene target failure (SGTF). In January 2021, SGTF positives were found to be caused by B.1.1.7 variants, as well as a few other variants such as B.1.375. Moreover, the S-gene target may fail if viral load is low and Cq is high, usually above 30. To assess whether SGTF could be used to study the increase or decrease of the B.1.1.7 variant of concern in the United States, we looked at all 4,869 sequences from SGTF samples in May and June 2021. Of those, 99.3% (4,834 of 4,869) were B.1.1.7 (**Figure 1A**). The next lineage leading to SGTF in that time period in the United States was B.1.525 (11 of 4,869). Of note, there were two SGTF samples sequenced that were P.1. The other P.1 variants sequenced, as well as all other variants of concerns that are not B.1.1.7, did not lead to SGTF. And 96.1% (4,834 of 5,028) of the sequences of B.1.1.7 lineage in May and June 2021 were SGTF. SGTF is therefore a reliable test to look at the epidemiological dynamics of B.1.1.7 in the United States.

We therefore analyzed 245,625 positive samples for SARS-CoV-2 with a Cq for the N gene <29. All of these samples were tested at the Helix laboratory between January 1 and June 23 2021. These samples were collected across the United States, but they do not proportionally represent the different areas of the United States by population, with 26% of our positives coming from Florida (**Table S1**). The other states that are most represented in this study are: California, Pennsylvania, and Georgia. Both SGTF and sequencing data indicate that the B.1.1.7 variant, after becoming the dominant SARS-CoV-2 lineage in the United States [6,8] has seen its prevalence plateau at around 70% at the end of April 2021 (**Figure 1B**). By looking at May and June test results in the US, we see a clear and rapid decrease of the fraction of SGTF among positive results, decreasing from 67% in week 20 (May 14 - 20) to 33.4% in week 25 (June 18 - 23) of 2021 (**Figure 1B**). To make sure that this result was not driven by a change in the states or regions with high number of cases, or other artefacts, we looked at the trend in Florida alone and observed the same rapid decrease of the fraction of SGTF representing 66% of positives in week 20 and only 33.2% in week 25 (**Figure 1C**). Overall, these results show that the variant of concern B.1.1.7 is rapidly being displaced in the United States.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

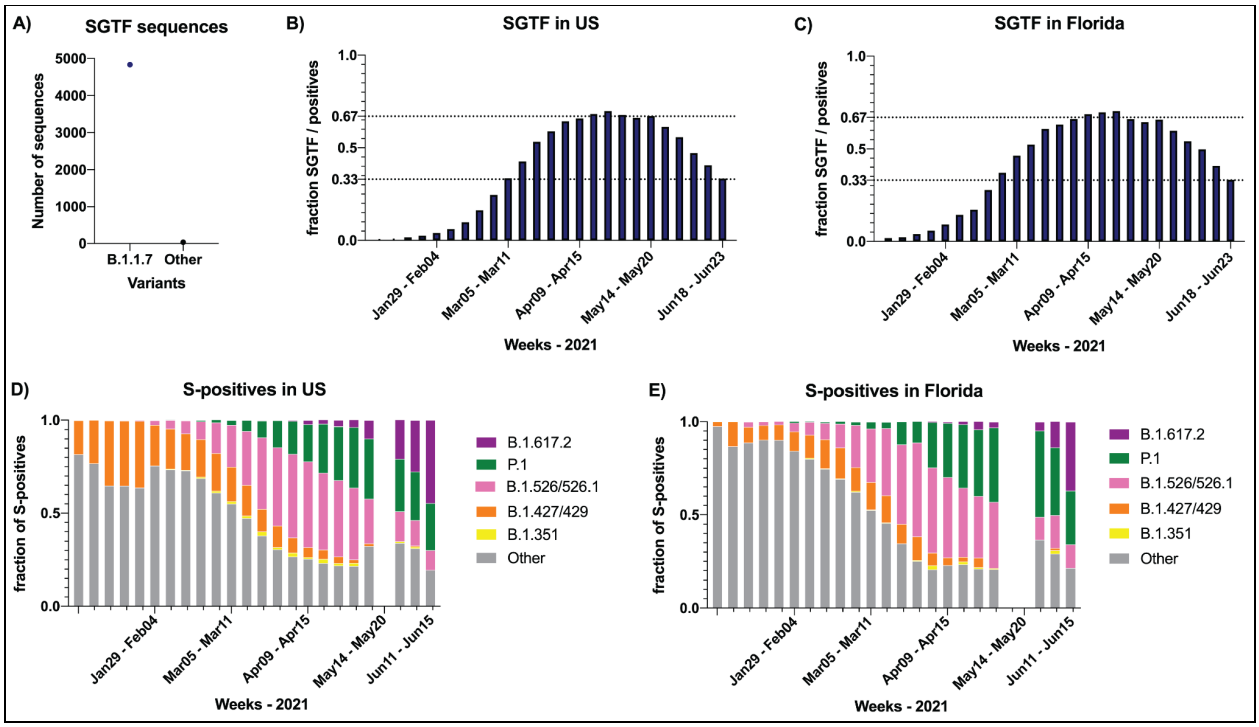


Figure 1: B.1.1.7 is being replaced by B.1.617.2 and P.1 in the United States. A) Counts of S-Gene Target Failure (SGTF) sequenced in May and June 2021 that were B.1.1.7 or Other variants. B) Fraction of SGTF to total positives per week in the US. C) Fraction of SGTF to total positives per week in Florida. D-E) Fraction of different variants of concerns and variants of interests that were S-positives in the US (panel D) and in Florida (panel E). Purple: B.1.617.2, Delta. Green: P.1, Gamma. Pink: B.1.526/B.1.526.1, Iota. Orange: B.1.427/B.1.429, Epsilon. Yellow: B.1.351, Beta.

B.1.617.2 and P.1 are responsible for B.1.1.7 decrease

We analyzed the Pango lineage associated with each sequence to investigate which variants might be displacing B.1.1.7 in the United States. Since SGTF is a near perfect proxy for B.1.1.7, we first looked at what variants comprised the growing S-positive fraction (the non-SGTF). We sequenced 2,782 samples that were S-positives and collected in May and June 2021, including 297 from June 11 to June 15, when the B.1.1.7 fraction was decreasing rapidly. From June 11 to June 15 (week 24), B.1.617.2 represented 44.6% and P.1 represented 25.2% of S-positive samples in the United States (**Figure 1D**). This was the first week where B.1.617.2 was the most prevalent lineage that was S-positive, and this fraction is increasing. By looking at the SGTF results, we observed that for week 25, S-positives represented 66.6% of the positives (855 of 1283). Using the proportion of S-positives from the week prior, we estimate that B.1.617.2 represented at least 29.7% and P.1 represented at least 16.7% of the cases in the US for the week of June 18 to June 23. In our more targeted look at Florida, the overall proportion of S-positives explained by B.1.617.2 and P.1 is similar to nationwide (66%). Similarly, B.1.617.2 is now the most prevalent S-positive lineage in Florida as it accounted for 37% of S-positive cases of week 24 and P.1 accounted for 28.9% of those (**Figure 1E**). Overall, these results

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

showed that the main variants replacing B.1.1.7 in the United States are the two variants of concern B.1.617.2 and P.1.

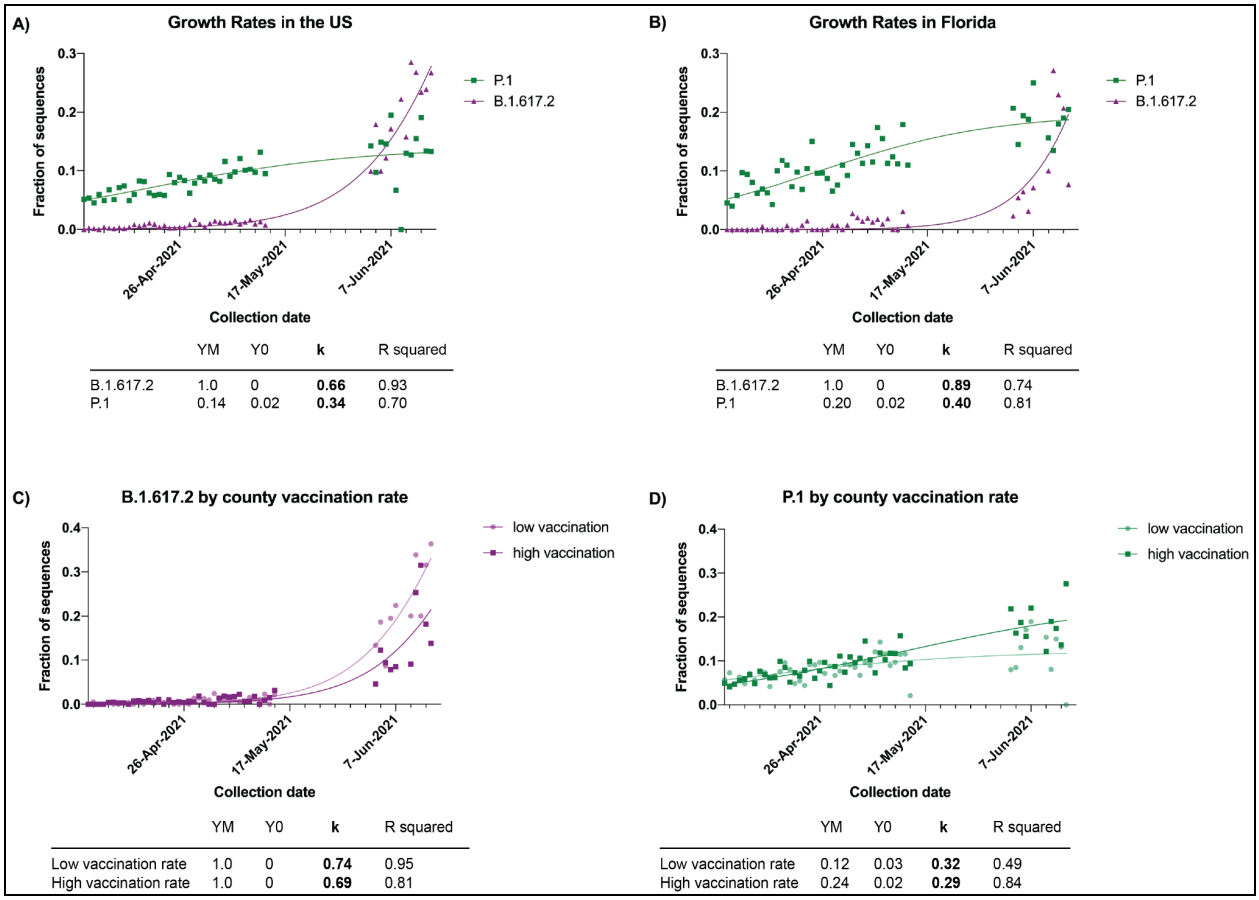
Growth rates of B.1.617.2 and P.1 in the United States

Our observation that both B.1.617.2 and P.1 are contributing to the displacement of B.1.1.7 in the United States stands in contrast to what was observed in England, where B.1.617.2 was the main variant replacing B.1.1.7. To better understand the dynamics between these two new variants of concern and B.1.1.7, we looked at growth rates by fitting a logistic growth curve on the fraction of all positives that are B.1.617.2 or P.1. In the United States, this analysis showed that the growth rate of B.1.617.2 was faster than P.1 ($k = 0.66$ vs. 0.34), and that the predicted maximum fraction of B.1.617.2 was higher than P.1 (**Figure 2A**). The numbers obtained In Florida were consistent with national numbers, highlighting that in states where both B.1.617.2 and P.1 were present, B.1.617.2 appeared to grow faster than P.1 in addition to outcompeting B.1.1.7 (**Figure 2B**).

The growth rates of B.1.617.2 and P.1 differ by county vaccination rate

The samples sequenced at Helix since March 15 2021 have spanned 687 US counties with vaccination data available from the CDC. Of the 27,717 samples sequenced from these counties during the study period, we divided them roughly evenly into those from counties with lower vaccination rates (<29% completely vaccinated on May 1: 13,263 samples across 379 counties) and those with higher vaccination rates (14,454 samples across 308 counties). The growth curve for B.1.617.2, which is more transmissible but against which vaccines are highly effective, may show moderately faster growth in counties with lower vaccination rates (**Figure 2C**). In contrast, P.1, which is less transmissible but against which vaccines have somewhat less efficacy, appears to have a higher prevalence in counties with higher vaccination rates (**Figure 2D**).

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .



	YM	Y0	k	R squared
B.1.617.2	1.0	0	**0.66**	0.93
P.1	0.14	0.02	**0.34**	0.70

	YM	Y0	k	R squared
B.1.617.2	1.0	0	**0.89**	0.74
P.1	0.20	0.02	**0.40**	0.81

	YM	Y0	k	R squared
Low vaccination rate	1.0	0	**0.74**	0.95
High vaccination rate	1.0	0	**0.69**	0.81

	YM	Y0	k	R squared
Low vaccination rate	0.12	0.03	**0.32**	0.49
High vaccination rate	0.24	0.02	**0.29**	0.84

Figure 2: Growth rates of B.1.617.2 and P.1 in the United States. A) and B) Fractions of total sequences (SGTF or not) by day in the US (panel A) and in Florida (panel B) that were B.1.617.2 (purple triangles) or P.1 (green squares). A logistic growth curve was then fitted and is represented by the continuous purple line for B.1.617.2 and green line for P.1 The table below the graphs shows the key values of the curve Y=YM*Y0/((YM-Y0)*exp(-k*x) +Y0). YM is the maximum population; Y0 is the starting population; k is the rate constant. R squared is a measure of the goodness of fit. C) and D) Growth curves of B.1.617.2 (panel C) and P.1 (panel D) by county vaccination rate. Light purple and light green represent counties with a low vaccination rate (below 29% fully vaccinated on May 1). Dark purple and dark green represent counties with a high vaccination rate (above 29% completely vaccinated on May 1). Each symbol indicates the fraction of B.1.617.2 or P.1 to the number of samples sequenced per day. Characteristics of the curves are below each panel.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Discussion

Here, we use viral sequence data from 54,294 Helix COVID-19 tests and 245,625 SGTF values from Helix COVID-19 tests collected since January 2021 to show the trajectories of different variants of concern in the United States. The total percentage of positive COVID-19 tests attributed to B.1.1.7 in the United States fell from a peak of 70% in April down to 33% in the most recent week (3rd week of June 2021). We show that most of the displacement of B.1.1.7 can be attributed to B.1.617.2 and P.1. Both of these variants of concern are growing in the United States and explain the rapid proportional decrease of the B.1.1.7 variant. Preliminary growth rates show that both B.1.617.2 and P.1 are growing faster than B.1.1.7, and that B.1.617.2 is growing faster than P.1 in the United States (k= 0.66 vs. k=0.34). Our results are consistent with those from Public Health England, which found that compared to B.1.1.7, B.1.617.2 had a growth rate of 0.93 and P.1 had a growth rate of 0.34 [3].

The expectation is that B.1.617.2 will soon be the dominant variant in the United States. However, questions remain whether it will entirely take over as it is doing in England, or whether it will plateau at a lower level like B.1.1.7 did in the US. One reason to argue that B.1.617.2 may not reach levels as high in the US compared to England is the more diverse sets of policies between US states with regard to vaccinations and other public health measures. With this in mind, we showed that B.1.617.2 may be growing more rapidly in counties with lower vaccination rates (**Figure 2C**).

One important limitation to this study is the relatively small number of positives analyzed in the last 2 months. This is partly due to the much lower number of cases in the United States and the decrease in test positivity rate. Another limitation is that the data is not homogeneous across the United States. There were many additional limitations to our analysis of variant growth based on county vaccination rate. We had to remove some states from the analysis because of missing vaccination rate data. We did not take into account the number of introductions of each variant, and we did not adjust for prevalence and vaccination rate of neighboring counties. All of these factors could have a role in the growth of B.1.617.2 and P.1. We will continue to test and sequence positive samples in order to characterize these variants. We also continuously update our public dashboard tracking SGTF and sequences by state and collection date at https://public.tableau.com/profile/helix6052#!/.

medRxiv preprint doi: https://doi.org/10.1101/2021.06.20.21259195; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Acknowledgements

We thank the employees of Helix, employees of Illumina, members of the CDC SPHERES consortium and California CovidNET, and members of the Andersen Lab at Scripps Research for discussion and help with logistics. We thank the healthcare workers, frontline workers, and patients who made the collection of this SARS-CoV-2 dataset possible. This work has been funded in part by CDC BAA contract 75D30121P10258 (Illumina, Helix).

Declarations of Interest

A.B., E.T.C., S.L., S.W., D.W., A.D.R., T.C., S.J., J.N., J.M.R., E.S., X.W., D.W., D.B., M.L., J.T.L., M.I., N.L.W. and W.L. are employees of Helix.

References

1. CDC. SARS-CoV-2 Variant Classifications and Definitions. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fcases-updates%2Fvariant-surveillance%2Fvariant-info.html (2021).
2. outbreak.info. https://outbreak.info/situation-reports?pango=B.1.617.2&loc=IND&loc=GBR&loc=USA&selected=IND.
3. Public Health England. Investigation of SARS-CoV-2 variants of concern: technical briefings. https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201 (2020).
4. Public library. https://khub.net/web/phe-national/public-library/-/document_library/v2WsRK3ZlEig/view/479607266.
5. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).
6. Washington, N. L. *et al.* Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* **184**, 2587–2594.e7 (2021).
7. Bal, A. *et al.* Two-step strategy for the identification of SARS-CoV-2 variant of concern 202012/01 and other variants with spike deletion H69-V70, France, August to December 2020. *Euro Surveill.* **26**, (2021).
8. Alpert, T. *et al.* Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. *Cell* **184**, 2595–2604.e13 (2021).