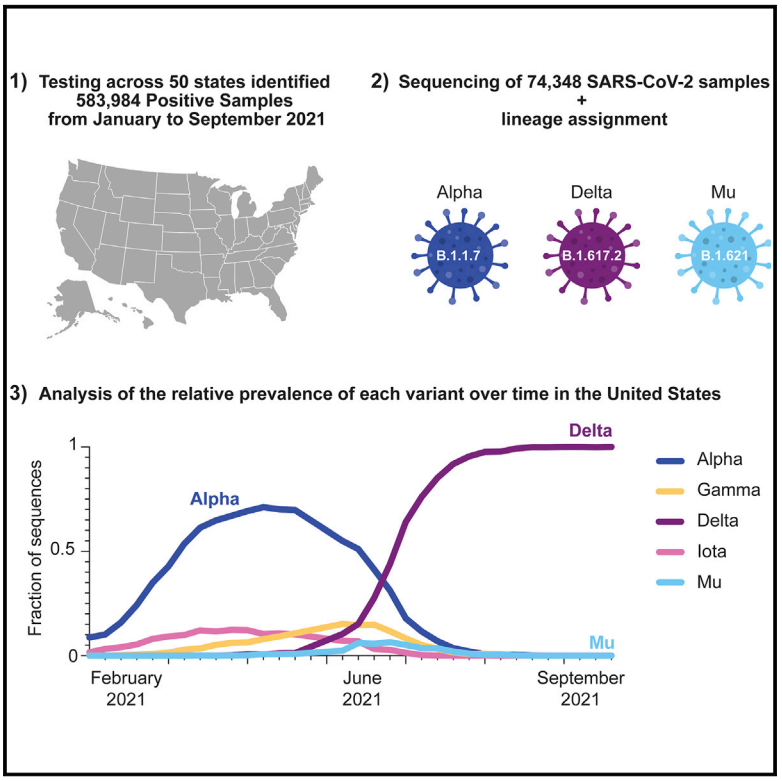
Cell Reports
Medicine

Report

SARS-CoV-2 variant Delta rapidly displaced variant Alpha in the United States and led to higher viral loads

Graphical abstract



Highlights

- Alpha was dominant in spring 2021 and went extinct in fall 2021 in the United States
- Delta also displaced Gamma, Iota, and Mu variants
- On average, viral load was ~1.7× higher in Delta infections versus Alpha infections

Authors

Alexandre Bolze, Shishi Luo, Simon White, ..., Magnus Isaksson, Nicole L. Washington, William Lee

Correspondence

alexandre.bolze@helix.com (A.B.), william.lee@helix.com (W.L.)

In brief

Bolze et al. show that the Delta variant of SARS-CoV-2 rapidly grew to become the dominant variant by July 2021 in the United States. Delta displaced variants of concern Alpha and Gamma and the variants of interest Iota and Mu.

Bolze et al., 2022, Cell Reports Medicine 3, 100564
March 15, 2022
https://doi.org/10.1016/j.xcrm.2022.100564

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Report

SARS-CoV-2 variant Delta rapidly displaced variant Alpha in the United States and led to higher viral loads

Alexandre Bolze,[1,2,3,4,*] Shishi Luo,[1,2] Simon White,[1] Elizabeth T. Cirulli,[1] Dana Wyman,[1] Andrew Dei Rossi,[1] Henrique Machado,[1] Tyler Cassens,[1] Sharoni Jacobs,[1] Kelly M. Schiabor Barrett,[1] Francisco Tanudjaja,[1] Kevin Tsan,[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] and William Lee[1,*]

[1]Helix, San Mateo, CA 94401, USA
[2]These authors contributed equally
[3]Twitter: @alexbolze
[4]Lead contact
*Correspondence: alexandre.bolze@helix.com (A.B.), william.lee@helix.com (W.L.)
https://doi.org/10.1016/j.xcrm.2022.100564

SUMMARY

We report on the sequencing of 74,348 SARS-CoV-2 positive samples collected across the United States and show that the Delta variant, first detected in the United States in March 2021, made up the majority of SARS-CoV-2 infections by July 1, 2021 and accounted for >99.9% of the infections by September 2021. Not only did Delta displace variant Alpha, which was the dominant variant at the time, it also displaced the Gamma, Iota, and Mu variants. Through an analysis of quantification cycle (Cq) values, we demonstrate that Delta infections tend to have a 1.7× higher viral load compared to Alpha infections (a decrease of 0.8 Cq) on average. Our results are consistent with the hypothesis that the increased transmissibility of the Delta variant could be due to the ability of the Delta variant to establish a higher viral load earlier in the infection as compared to the Alpha variant.

INTRODUCTION

The severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) Delta variant, which includes variant B.1.617.2 and all variants AY.x, has been classified as a variant of concern (VOC) by Public Health England (PHE), the World Health Organization (WHO), and the US Centers for Disease Control and Prevention (CDC).[1] The Delta variant was the predominant variant in India during the peak in cases in April and May 2021.[2,3] Delta is defined by a set of SNPs that include S:L452R and S:P681R.[2] When it spread to England, the Delta variant led to a new wave of cases and displaced the then-dominant Alpha variant, which includes B.1.1.7 and the Q.x variants.[4] The context in the United States was different from that in England or India. In addition to differences in public health policies and vaccination rates, the variants that were co-circulating when Delta was introduced also differed between England and the United States.[2,3] In England, Alpha represented more than 90% of the SARS-CoV-2 sequences when Delta was first identified in the country (in approximately March 2021), and there were very few sequences of Gamma (also named P.1), another VOC. In the United States, Alpha peaked at just above 70%, and there was a greater diversity of variants when Delta started to emerge (also in approximately March 2021), including an increasing amount of Gamma.[5] The objectives of this study were therefore to (1) analyze the impact of the Delta variant on the prevalence of Alpha in the United States, (2) assess the impact of Delta on the prevalence of Gamma in the United States, and (3) test the hypothesis that the higher transmissibility of Delta is due to a higher viral load in the nose.

RESULTS

Samples came from all states, age groups, and races

We sequenced and assigned a lineage to 74,738 samples collected by the Helix laboratory across the United States (Table S1) between February and September 2021 for genomic surveillance purposes. These samples came from anterior nares swabs of different individuals with one viral sequence per person infected (similar to a cross-sectional analysis). The individuals tested skewed slightly more female (51.2%) (Table 1). The mean age of individuals tested was 37.5 years old, and all age groups were represented, with 20–29 years old being the largest group (22.3% of samples) and 80–≥89 years old the smallest age group (1.1% of samples) (Table 1). The individuals tested came from multiple races (Table 1). The large majority of samples were collected at a national retail pharmacy, and an analysis of the billing codes indicates that suspected exposure was the main reason for testing (International Classification of Diseases, 10th revision (ICD-10) codes Z20.828 and Z20.822). Samples

Table 1. Demographics of the samples sequenced

	All infections	Alpha infections	Delta infections
N total	74,738	23,884	25,485
Gender (%)			
Female	38,299 (51.2)	12,190 (51.0)	13,199 (51.8)
Male	36,064 (48.3)	11,642 (48.7)	12,189 (47.8)
Transgender female	6	0	6
Transgender male	2	0	2
Nonbinary	1	0	0
Unknown/decline to state	366	52	89
Mean age, y	37.5	36.7	36.6
Age group distribution (%)			
0–9	2,509 (3.4)	698 (2.9)	1,348 (5.3)
10–19	9,215 (12.3)	3,056 (12.8)	3,398 (13.3)
20–29	16,692 (22.3)	5,638 (23.6)	5,390 (21.1)
30–39	14,755 (19.7)	4,762 (19.9)	5,159 (20.2)
40–49	11,362 (15.2)	3,757 (15.7)	3,739 (14.7)
50–59	10,439 (14.0)	3,313 (13.9)	3,167 (12.4)
60–69	6,500 (8.7)	1,930 (8.1)	2,023 (7.9)
70–79	2,456 (3.3)	558 (2.3)	945 (3.7)
80–≥89	810 (1.1)	172 (0.7)	316 (1.2)
Race (%)			
American Indian or Alaska Native	372 (0.6)	85 (0.4)	206 (0.9)
Asian	2,976 (4.6)	851 (4.0)	1,067 (4.5)
Black or African American	9,231 (14.2)	3,502 (16.5)	3,153 (13.2)
Native Hawaiian or Other Pacific Islander	344 (0.5)	79 (0.4)	163 (0.7)
White	42,931 (65.8)	13,577 (64.1)	14,919 (62.6)
Other race	9,325 (14.3)	3,069 (14.5)	4,329 (18.2)
Mixed	40 (0.1)	11 (0.1)	0 (0.0)
Unspecified/unknown	9,519	2,710	1,648

from San Diego County (California) were collected as part of community testing organized by San Diego County. Information regarding the vaccination status of individuals was not available for this study. Importantly, the collection method and collection sites have been consistent since February 2021, and the samples analyzed should not be biased for localized outbreaks (see method details). We therefore made the assumption that there was no significant sampling bias between the testing and sequencing done by our lab from February to April 2021, when Alpha was rapidly increasing in the United States,[6,7] and the following months of 2021. Supporting this, the distributions of self-reported age, gender, and race were very similar between the waves of Alpha infections and Delta infections: 51.8% of individuals infected with Delta identified as female compared to 51.0% for Alpha infections, and the mean self-reported age of individuals infected with Delta was 36.6 years old compared to 36.7 years old for Alpha infections (Table 1).

The Delta variant grew to represent more than 99.9% of SARS-CoV-2 infections in the United States in September 2021

We first observed the Delta variant in March 2021. Delta variant infections accounted for >50% of SARS-CoV-2 infections in the United States by July 1, 2021 and represented >99.9% (8,600 of 8,604) of the SARS-CoV-2 we sequenced across the United States in September 2021 (Figure 1). The speed at which the Delta fraction rose was similar in Florida (Figure 1A), where we sequenced samples from all counties, in California (Figure 1B), where the majority of samples came from San Diego County, and in the other states sampled (Figure 1C), despite differences in health policies, weather, and circulating variants. In addition to displacing the Alpha variant, which was the dominant variant across the United States in spring 2021, Delta also directly displaced the Gamma VOC and the Mu (B.1.621) variant of interest, which were variants that grew in their proportion in May and June 2021 in Florida (Figure 1A). Overall, these results clearly showed that the SARS-CoV-2 Delta variant was more transmissible than Alpha and the other SARS-CoV-2 concurrently circulating variants in the United States. These results are similar to how Delta emerged and spread in the New England region in the United States,[8] India, England, and other countries.

Quantification cycle (Cq) values were lower for Delta infections compared to Alpha infections when the two variants co-circulated in the population

We next tried to understand what could be driving this advantage of Delta compared to Alpha. One hypothesis is that Delta virions replicate more rapidly and to higher viral loads in the nose, so that individuals exposed to the Delta variant become infectious earlier compared to individuals exposed to other variants. In a study of quarantined subjects, researchers observed that subjects infected with the Delta variant had detectable virus (by qRT-PCR) 2 days earlier and at a 1,000-fold higher titer (~10 Ct [cycle threshold]) than subjects infected with the initial SARS-CoV-2 virus from early 2020.[9] Before testing our central hypothesis, we first assessed whether other factors influenced viral loads. Cq values were lower for male samples compared to female samples for both Alpha infections (mean Cq_{female} = 20.48 versus mean Cq_{male} = 20.26, $p < 0.0001$, Mann-Whitney *U*) and Delta infections (mean Cq_{female} = 19.92 versus mean Cq_{male} = 19.75, $p = 0.0002$, Mann-Whitney *U*) (Figure S1A). This result meant that, on average, viral load was higher in males compared to females.

We then tested our hypothesis and compared Cq values of Alpha and Delta infections from samples collected at the same time, when the two variants were co-circulating in the population, and therefore on the same qRT-PCR plates. We restricted our analysis to dates for which there were at least 10 Alpha sequences and 10 Delta sequences. In total, 40 days could be analyzed for a direct comparison of the Cq values for 1,283 Alpha infections (Cq_{Alpha}) and 4,428 Delta infections (Cq_{Delta}). We observed a higher mean Cq_{Alpha} compared to Cq_{Delta} for 35 of the 40 days (Figures 2A and S1B), with an overall difference of ~0.8 (mean Cq_{Alpha}) − (mean Cq_{Delta}), corresponding to approximately 1.7× more virus in the nose for Delta infections compared to Alpha infections. This was during the months of June and July 2021 when the two variants were co-circulating

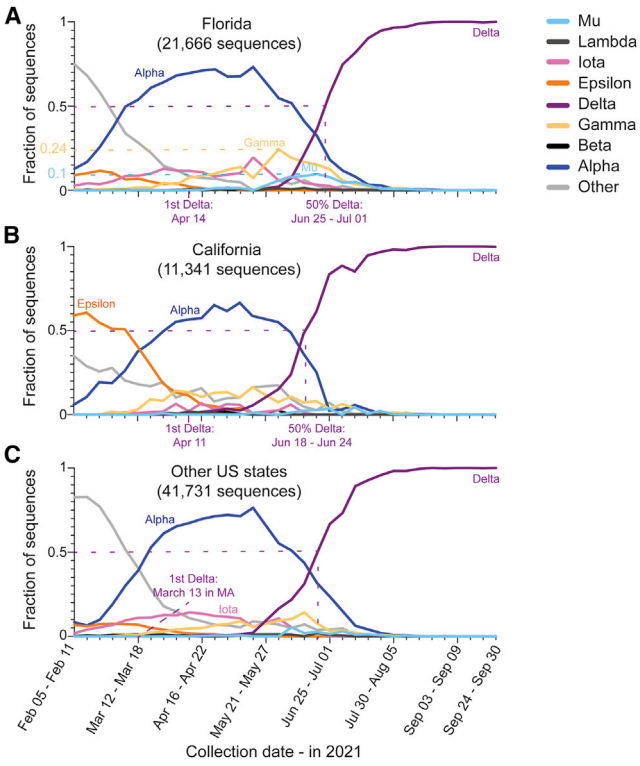


Figure 1. Delta displaced Alpha and all other variants in the United States in the summer of 2021

For each variant of concern or of interest, the fraction of the total sequences each variant represented at a given time is plotted. There is 1 point per week starting from February 5 and ending September 30, 2021.

(A) Sequences from samples collected in Florida.

(B) Sequences from samples collected in California.

(C) Sequences from samples collected in the rest of the United States.

Each variant is represented by 1 line with a unique color. Alpha: dark blue; Beta: black; Gamma: light orange; Delta: purple; Epsilon: dark orange; Iota: pink; Lambda: dark gray; Mu: light blue; and all of the other variants by light gray.

See also Table S1.

but the number of Alpha infections was declining, while the number of Delta infections was increasing.

Cq values were lower for Delta infections compared to Alpha infections when accounting for the phase of the epidemic

The period of the epidemic could influence Cq values on average across a population.[10] When an epidemic is growing, it is more likely to test a sample with a recent infection, and when an epidemic is declining, it is more likely to test someone with an old infection. In addition, it is known that on average viral load decreases more slowly after hitting peak viral load compared to the rate at which it increases during the incubation period after infection. Thus, it is predicted that Cq values would be lower with higher transmission rates, and that Cq values would be higher with lower transmission rates.[10] To take this potential bias into account, we decided to compare the Cq_{Alpha} when the Alpha wave was growing to Cq_{Delta} when the Delta wave was growing (the total number of cases). We also made the comparison when each respective wave was declining. At first, we restricted our analysis to Florida, the state where we have the most sequences, to be as precise as possible with the dates and the dynamics of the epidemic. Florida also had both an Alpha and a Delta wave, whereas California had a very small Alpha wave, which could interfere with this analysis. We then repeated the analysis on the rest of the United States, excluding Florida.

In Florida, the Alpha cases were increasing between February 1 and April 13, 2021 (average estimated R = 1.12), while the Delta cases were increasing between June 18 and August 13 (average estimated R = 1.39) (Figure 2B; method details). Using these dates, we could compare the Cq values of 3,616 Alpha infections with 3,682 Delta infections. On average, Cq_{Delta} = 19.91 was significantly lower compared to Cq_{Alpha} = 20.22 by −0.307 ($p = 0.0002$, Mann-Whitney *U* test) (Figure 2C; Table S2), whereas the mean Cq for the qRT-PCR positive controls (Cq_{Ctrl+}) was actually 0.17 higher during the Delta increase period (Figure S1C). Similarly, comparisons of 1,337 Cq_{Alpha} and 1,802 Cq_{Delta} during the periods of decreasing cases (estimated R was 0.83 for the Alpha declining phase and 0.82 for the Delta declining phase) showed that Cq_{Delta} = 19.98 was significantly lower compared to Cq_{Alpha} = 20.50 by −0.525 ($p = 0.0002$, Mann-Whitney *U* test) (Figure 2C; Table S2), whereas Cq_{Ctrl+} was 0.79 higher during the Delta declining wave compared to the Alpha declining wave (Figure S1C).

We repeated the analysis on the rest of the United States, comparing 5,560 Alpha infections with 9,014 Delta infections during the growing phases of the epidemics. On average, Cq_{Delta} = 19.63 was significantly lower compared to Cq_{Alpha} = 20.45 by −0.82 ($p < 0.0001$, Mann-Whitney *U* test) (Figure 2D; Table S2), whereas Cq_{Ctrl+} was 0.23 higher during the Delta growing wave compared to the Alpha growing wave (Figure S1D). Similarly, comparisons of 3,028 Cq_{Alpha} and 6,765 Cq_{Delta} during the declining phases of the epidemics showed that Cq_{Delta} = 19.96 was significantly lower compared to Cq_{Alpha} = 20.63 by −0.67 ($p < 0.0001$, Mann-Whitney *U* test) (Figure 2D; Table S2), whereas Cq_{Ctrl+} was 0.79 higher during the Delta declining wave compared to the Alpha declining wave (Figure S1D). Overall, these results indicate the mean Cq_{Delta} was lower compared to the mean Cq_{Alpha}. Based on these analyses, we estimate that the Cq difference was approximately 0.8, which corresponds to Delta infections with 1.7× more viruses in the nose as compared to Alpha infections.

DISCUSSION

Here, we used the results from 74,738 viral sequences with basic demographic information from Helix COVID-19 tests collected between January and September 2021 to show the trajectories of different SARS-CoV-2 variants in the United States. Our analysis showed that the SARS-CoV-2 Delta variant displaced the Alpha variant in the United States between May and August 2021. Our results are consistent with reports from England[4] and other countries. While it is reasonable to expect variants to behave similarly in different countries, it is worth noting that the Alpha variant grew to account for more than 90% of infections in England early in 2021, but peaked at 70% in the United States.

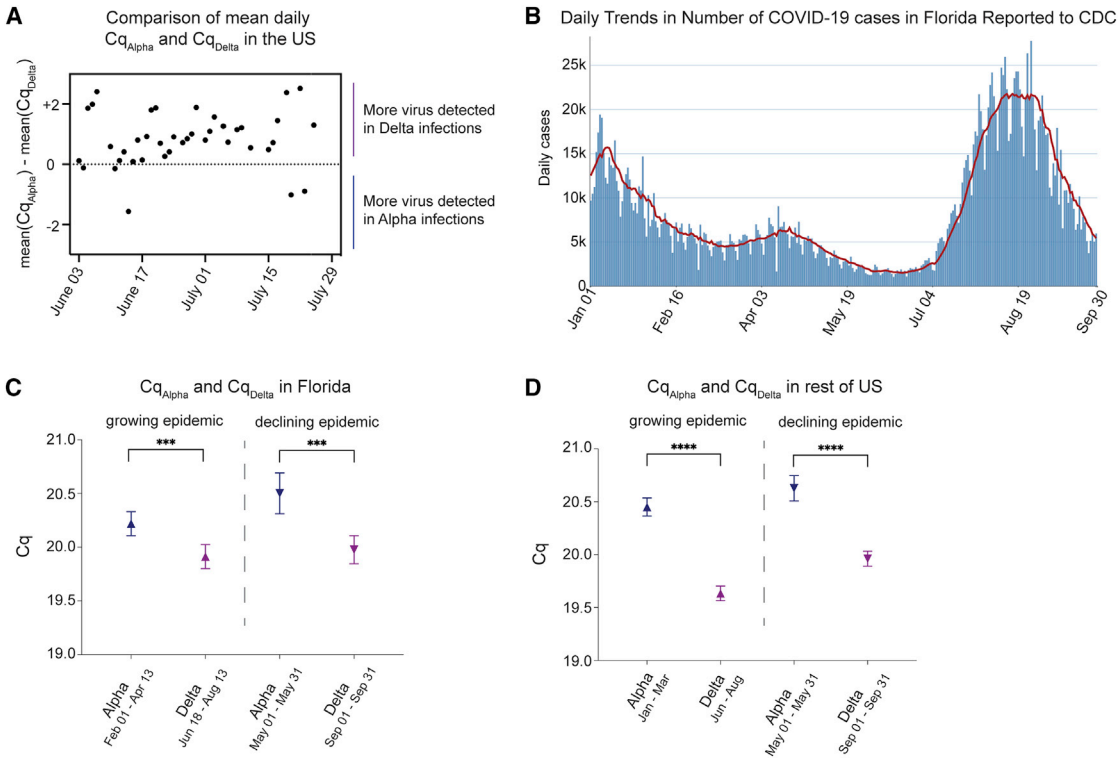


Figure 2. Higher viral loads in Delta infections compared to Alpha infections (all dates are in 2021)

(A) Mean Cq_{Alpha} minus mean Cq_{Delta} levels over time in 2021. Each dot corresponds to a value of this difference for 1 day. Sequences from samples collected from all over the United States were used.

(B) Epidemic phases in Florida. Chart downloaded from https://covid.cdc.gov/covid-data-tracker/#trends_dailycases after restricting the sample to cases in Florida. The blue bars show daily cases. The red line is the 7-day moving average of cases.

(C) Mean Cq_{Alpha} and Cq_{Delta} by phase of the epidemic in Florida. The left 2 columns are while the epidemic was growing. The right 2 columns are while the epidemic was declining. Dark blue: Alpha; purple: Delta. The error bars represent the 95% confidence interval (CI).

(D) Mean Cq_{Alpha} and Cq_{Delta} by phase of the epidemic in the United States, except Florida. The left 2 columns are while the epidemic was growing. The right 2 columns are while the epidemic was declining. Dark blue: Alpha; purple: Delta. The error bars represent the 95% CI. Mann-Whitney *U* tests were performed. ***$p < 0.001$, ****$p < 0.0001$.

See also Figure S1 and Table S2.

Here, we showed that the Delta variant also displaced the Gamma, Iota, and Mu variants in the United States. An analysis of sequences circulating in South America showed that in Brazil, the Delta variant also displaced the Gamma variant, and in Colombia, the Delta variant also displaced the Mu variant.[3]

One biological explanation for the dominance of Delta could be that Delta replicates faster than Alpha, and Delta infections led to higher viral loads earlier after infection. Our analysis of the Cq is consistent with this hypothesis. The difference we observed of 1.7× higher viral load in Delta infections as compared to Alpha was in the same direction and of similar magnitude to what was reported in other US studies[8] or in France[11] in a similar setting. Our results were also similar to those observed in a study that analyzed the viral dynamics of Alpha and Delta infections in 72 individuals affiliated with the (US) National Basketball Association for which longitudinal qRT-PCR results were available.[12] However, our observed difference was small compared to the ~1,000× difference observed in a study of quarantined individuals.[9] This large discrepancy could be explained by vastly different study designs. First, Li et al. compared Delta infections to infections with the earliest form of the virus (19A/B) present in China in early 2020.[9] It is very likely that the S:D614G mutation and other recently acquired mutations have since led to a faster replication of the virus.[13] Therefore, the difference in the viral load of samples in the Li et al. study may in reality be less than 1,000× if the comparison was between Delta infections and Alpha infections. Second, the samples studied here came from a community setting, with the large majority from a national retail pharmacy. For the majority of these samples, it is unlikely that they were collected at the first moment when they would be detectable by qRT-PCR.

The next question is whether SARS-CoV-2 can continue to evolve to be intrinsically more transmissible, or whether it will evolve to better escape the existing population immunity. Tracking the growth and behavior of different Delta sublineages (different AY.x variants) and/or altogether new lineages, as well as tracking the differences of Cq values over time, will continue to be important methods to understand the evolution of SARS-CoV-2 in the future.

LIMITATIONS OF THE STUDY

The main limitations of the present study are that we did not have access to (1) individual vaccination information, (2) the amount of time between infection and collection of the sample for testing, or (3) the human genome of the individuals being infected. Both vaccination status and time after infection may be confounders in the viral load analysis. Studies have shown that viral loads decrease faster in vaccinated individuals and that viral load varies with time after infection, first increasing and then decreasing.[12,14] Our confidence in our results is based on two assumptions. The first assumption is that in the samples we tested, Delta infections would have had a larger fraction of vaccinated individuals compared to Alpha infections because, unlike Alpha, the Delta wave occurred when vaccines were available to the public. Taking into account a greater fraction of vaccinated individuals among Delta infections would lead to an even greater difference in viral load between Delta infection compared to Alpha infection (higher viral load for Delta infections). The second assumption we made was that a significant percentage of the individuals in this study got tested because they presented symptoms or because of a suspected exposure. It has been shown that comparing Cq values between variants obtained from symptomatic testing (non-random surveillance) is more likely to reflect the true differences in underlying viral kinetics.[15] Host genetics including any potential defect in the type I interferon immune response could also be a confounder in our viral load analysis as different people will clear the virus at different speeds.[16] Here, our assumption is that these human genetic factors would be distributed similarly in individuals infected by the Alpha variant and individuals infected by the Delta variant. Our study should be more robust to biases due to individual host genetic factors compared to studies with a smaller number of individuals that do not take into account the human genome.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
  - Ethical statement
- METHOD DETAILS
  - Helix COVID-19 test data and sample selection
  - SARS-CoV-2 sequencing and consensus sequence generation
  - Viral lineage designation
  - Cq analysis
  - Estimated R analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2022.100564.

ACKNOWLEDGMENTS

We thank the employees of Helix, the employees of Illumina, the members of the CDC SPHERES consortium and California CovidNET, and the members of the Andersen lab at Scripps Research for discussions 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). We also want to give special thanks to Public Health England for their inspiring work and technical reports, as well as the team behind outbreak.info, a resource that we used throughout this study.

AUTHOR CONTRIBUTIONS

Conceptualization, A.B. and S.L.; software, S.L., D. Wyman, A.D.R., and M.I.; investigation, A.B., S.L., and E.T.C.; resources, H.M., T.C., S.J., K.M.S.B., K.T., J.N., J.M.R., E.S., X.W., D. Wong, D.B., and M.L.; data curation, S.W., F.T., and M.I.; writing – original draft, A.B.; writing – review & editing, A.B., S.L., E.T.C., K.M.S.B., J.T.L., N.L.W., and W.L.; supervision, W.L.; funding acquisition, J.T.L. and W.L.

DECLARATION OF INTERESTS

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

Received: October 21, 2021
Revised: January 10, 2022
Accepted: February 17, 2022
Published: March 15, 2022

REFERENCES

1. Konings, F., Perkins, M.D., Kuhn, J.H., Pallen, M.J., Alm, E.J., Archer, B.N., Barakat, A., Bedford, T., Bhiman, J.N., Caly, L., et al. (2021). SARS-CoV-2 variants of interest and concern naming scheme conducive for global discourse. Nat. Microbiol. *6*, 821–823.
2. Mlcochova, P., Kemp, S.A., Dhar, M.S., Papa, G., Meng, B., Ferreira, I.A.T.M., Datir, R., Collier, D.A., Albecka, A., Singh, S., et al. (2021). SARS-CoV-2 B.1.617.2 delta variant replication and immune evasion. Nature *599*, 114–119.
3. Mullen, J.L., Tsueng, G., Latif, A.A., Alkuzweny, M., Cano, M., Haag, E., Zhou, J., Zeller, M., Matteson, N., Andersen, K.G., et al. (2021). Outbreak.info.https://outbreak.info/.
4. Allen, H., Vusirikala, A., Flannagan, J., Twohig, K.A., Zaidi, A., Chudasama, D., Lamagni, T., Groves, N., Turner, C., Rawlinson, C., et al. (2022). Household transmission of COVID-19 cases associated with SARS-CoV-2 delta variant (B.1.617.2): national case-control study. Lancet Reg. Health Eur. *12*, 100252.
5. Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D. da S., Mishra, S., Crispim, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science *372*, 815–821.
6. Washington, N.L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E.T., Schiabor Barrett, K.M., Larsen, B.B., Anderson, C., White, S., Cassens, T., et al. (2021). Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. Cell *184*, 2587–2594.e7.
7. Alpert, T., Brito, A.F., Lasek-Nesselquist, E., Rothman, J., Valesano, A.L., MacKay, M.J., Petrone, M.E., Breban, M.I., Watkins, A.E., Vogels, C.B.F.,

et al. (2021). Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. Cell *184*, 2595–2604.e13.

8. Earnest, R., Uddin, R., Matluk, N., Renzette, N., Siddle, K.J., Loreth, C., Adams, G., Tomkins-Tinch, C.H., Petrone, M.E., Rothman, J.E., et al. (2021). Comparative transmissibility of SARS-CoV-2 variants delta and Alpha in new England, USA. Preprint at medRxiv. https://doi.org/10.1101/2021.10.06.21264641.
9. Li, B., Deng, A., Li, K., Hu, Y., Li, Z., Xiong, Q., Liu, Z., Guo, Q., Zou, L., Zhang, H., et al. (2021). Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. Preprint at bioRxiv. https://doi.org/10.1038/s41467-022-28089-y.
10. Hay, J.A., Kennedy-Shaffer, L., Kanjilal, S., Lennon, N.J., Gabriel, S.B., Lipsitch, M., and Mina, M.J. (2021). Estimating epidemiologic dynamics from cross-sectional viral load distributions. Science *373*, eabh0635.
11. Teyssou, E., Delagrèverie, H., Visseaux, B., Lambert-Niclot, S., Brichler, S., Ferre, V., Marot, S., Jary, A., Todesco, E., Schnuriger, A., et al. (2021). The Delta SARS-CoV-2 variant has a higher viral load than the Beta and the historical variants in nasopharyngeal samples from newly diagnosed COVID-19 patients. J. Infect. *83*, e1–e3.
12. Kissler, S.M., Fauver, J.R., Mack, C., Ta, C.G., Breban, M.I., Watkins, A.E., Samant, R.M., Anderson, D.J., Metti, J., Khullar, G., et al. (2021). Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated persons. N. Engl. J. Med. *385*, 2489–2491.
13. Jones, T.C., Biele, G., Mühlemann, B., Veith, T., Schneider, J., Beheim-Schwarzbach, J., Bleicker, T., Tesch, J., Schmidt, M.L., Sander, L.E., et al. (2021). Estimating infectiousness throughout SARS-CoV-2 infection course. Science *373*, eabi5273.
14. Chia, P.Y., Xiang Ong, S.W., Chiew, C.J., Ang, L.W., Chavatte, J.-M., Mak, T.-M., Cui, L., Kalimuddin, S., Chia, W.N., Tan, C.W., et al. (2021). Virological and serological kinetics of SARS-CoV-2 Delta variant vaccine-breakthrough infections: a multi-center cohort study. Clin. Microbiol. Infect. S1198-743X 00638. https://doi.org/10.1016/j.cmi.2021.11.010.
15. Hay, J.A., Kennedy-Shaffer, L., and Mina, M.J. (2021). Viral loads observed under competing strain dynamics. Preprint at bioRxiv. https://doi.org/10.1101/2021.07.27.21261224.
16. Zhang, Q., Bastard, P., Bolze, A., Jouanguy, E., Zhang, S.-Y., COVID Human Genetic Effort; Cobat, A., Notarangelo, L.D., Su, H.C., Abel, L., et al. (2020). Life-threatening COVID-19: defective interferons unleash excessive inflammation. Med (N Y) *1*, 14–20.
17. Tartof, S.Y., Slezak, J.M., Fischer, H., Hong, V., Ackerson, B.K., Ranasinghe, O.N., Frankland, T.B., Ogun, O.A., Zamparo, J.M., Gray, S., et al. (2021). Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. Lancet *398*, 1407–1416.
18. Bruxvoort, K., Sy, L.S., Qian, L., Ackerson, B.K., Luo, Y., Lee, G.S., Tian, Y., Florea, A., Takhar, H.S., Tubert, J.E., et al. (2021). Real-world effectiveness of the mRNA-1273 vaccine against COVID-19: interim results from a prospective observational cohort study. Lancet Reg. Health Am. *6*, 100134.
19. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
20. Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., et al. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. *20*, 8.
21. Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., and Pybus, O.G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. *5*, 1403–1407.
22. Cori, A., Ferguson, N.M., Fraser, C., and Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. Am. J. Epidemiol. *178*, 1505–1512.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Helix ® COVID-19 test	Helix	EUA201636
TaqPath COVID-19 Combo Kit	Thermo Fisher	Cat#A47814
Illumina CovidSeq Test	Illumina	1000000128490 v01
MagMAX Viral/Pathogen II Nucleic Acid Isolation Kit	Thermo Fisher	Cat#A48383
xGen COVID-19 Capture Panel	Integrated DNA Technologies	Cat# 10006764
NovaSeq 6000 Sequencing system S1 flow cell	Illumina	
NovaSeq 6000 Sequencing System S1 Reagent Kit v1.5 (300 cycles)	Illumina	Cat#20028317
Deposited data		
Raw SARS-CoV-2 genomes	GISAID	Virus name includes 'CDC-STM' and/or 'originating lab field' is 'Helix'
Software and algorithms		
Bcl2fastq	Illumina	N/A
Klados-fastagenerator	Helix	N/A
BWA-MEM	https://github.com/lh3/bwa	N/A
Haplotyper algorithm	Sentieon, Inc	N/A
Pangolin v3.1.11	https://github.com/cov-lineages/pangoLEARN	N/A
EpiEstim	http://cran.r-project.org/web/packages/EpiEstim/index.html	N/A
PRISM v8	Graphpad	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for data and resources should be directed to and will be fulfilled by the Lead contact, Alexandre Bolze (alexandre.bolze@helix.com)

Materials availability

This study did not generate new unique reagents.

Data and code availability

The raw genomes used in this analysis were uploaded to GISAID. There are two ways to find them on GISAID. One way is to download all of the samples that have a collection date within the time period studied in this paper and filter for 'Helix' in the 'Originating lab' field. The other way is to search for all samples with 'CDC-STM' in the 'Virus name' field as only Helix uses this nomenclature for the name of the virus (and filter by collection date as above). An example is: {virus name: hCoV-19/USA/CA-CDC-STM-000720832/2021, GISAID accession: EPI_ISL_6489962}.

The raw data used to generate Figure 1 are available in Table S3. The raw data (all of the Cq values for Alpha and Delta infections, as well as positive controls) used for our viral load analyses are available in the supplemental information of this paper: Table S3. Additional supplemental items are available from Mendeley Data: https://doi.org/10.17632/c8kf2tmjwy.1

- This study does not generate custom code.
- Any additional information required to reanalyze the data reported in this paper is available from the Lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

This paper is based on the study of SARS-CoV-2 found and collected from individuals in the United States. The detailed demographics about these individuals can be found in Table 1 of this paper.

Ethical statement

The 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)). All samples were de-identified before receipt by the study investigators.

METHOD DETAILS

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.

For this study, we analyzed 583,984 positive samples collected from January 1 2021 to October 2021. We collected samples from all 50 states of the United States, though not in proportion to each state's population. Table S1 shows the number of positive tests, as well as the number of viruses sequenced per state. Importantly, samples sequenced for purposes other than genomic surveillance such as the samples sequenced for two vaccine-effectiveness studies[17,18] were not included in the analysis of this paper.

SARS-CoV-2 sequencing and consensus sequence generation

Sequencing was performed by Illumina[6] prior to June 2021, and subsequently 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). Samples then proceeded via either a hybrid capture or amplicon workflow. Under the hybrid capture method, SARS-CoV-2 genome capture was accomplished using hybridization kit xGen COVID-19 Capture Panel (Integrated DNA Technologies). Under the amplicon method, SARS-CoV-2 amplicons were generated using the Swift Normalase Amplicon SARS-CoV-2 Panel (Swift Biosciences). 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 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. If a sample was processed through the hybrid capture method, reads were marked for duplicates. If a sample was instead processed through the amplicon method, reads were trimmed for adapters before alignment using Trimmomatic v0.39[19] and for primer sequences after alignment using iVar v1.3.1.[20] Both methods – hybrid capture and amplicon – then use the Haplotyper algorithm (Sentieon, Inc) to call SARS-CoV-2 variants. 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: for the hybrid capture method, coverage from at least 5 unique reads is required with at least 80% of the reads supporting the allele; for the amplicon method, coverage from at least 5 reads is required with at least 60% of the reads supporting the allele. Otherwise, that base is considered uncertain, and an N is reported.

Viral lineage designation

Viral sequences were assigned a Pango lineage[21] using pangoLEARN (https://github.com/cov-lineages/pangoLEARN). For this analysis, pangoLEARN version 2021-08-24 with Pangolin software version 3.1.11 was used. We sequenced and were able to attribute a lineage to 74,738 sequences from samples collected between February and October 2021 for genomic surveillance purposes.

Samples sequenced in January were not randomly selected. They were enriched for samples with S-gene target failures (SGTF) with the original goal to assess how good of a proxy SGTF was for the Alpha variant. This is why these sequences were not included in the Figure 1 analysis.

Cq analysis

We used the Cq of the N gene target to compare the Cq of Alpha and infections and Delta infections. For comparisons based on the time of the epidemic (either growing phase or declining phase), we restricted the analysis to sequences from positive samples with a Cq N-gene <27. This is because our threshold for sequencing samples evolved over time. In the beginning of 2021, we used a Cq N

gene threshold of 27 to select samples to sequence in order to maximize the percentage of successful sequences. More recently, we improved our protocols and have been able to sequence samples with higher Cq values.

We selected the dates for the start and end of the growing and declining phases of the Alpha and Delta waves based on the CDC cases charts (such as the one in Figure 2B). Importantly, the start of the Alpha growth phase overlapped with an overall decline in cases across the United States. We were confident that Alpha numbers were growing from mid-January until mid-April based on our previous work analyzing SGTF positive tests, which were a very good proxy for Alpha infections.[6]

Estimated R analysis

We calculated the estimated R for Alpha and Delta at different periods of the epidemic in Florida using the EpiEstim software.[22] The incidence counts were obtained by multiplying the daily 7-day average case counts in Florida reported by the CDC (https://covid.cdc.gov/covid-data-tracker/#trends_dailycases) with the fraction of sequences that were Alpha or Delta at a given day based on Helix viral surveillance (data plotted Figure 1A).

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparison of Cq values between different weeks were done using Mann-Whitney U tests with PRISM version 8.

ADDITIONAL RESOURCES

US distribution dashboard of COVID-19 lineages for samples originating from Helix Viral Surveillance: https://public.tableau.com/app/profile/helix6052/viz/HelixSARS-CoV-2LineagesofInterest/Lineages.