scientific reports

OPEN

Pattern of SARS-CoV-2 variant B.1.1.519 emergence in Alaska

Tracie J. Haan[1], Lisa K. Smith[2], Stephanie DeRonde[2], Elva House[2], Jacob Zidek[2], Diana Puhak[2], Matthew Redlinger[3], Jayme Parker[2], Brian M. Barnes[1], Jason L. Burkhead[3], Cindy Knall[4], Eric Bortz[3,4 ✉], Jack Chen[1,2,5 ✉] & Devin M. Drown[1,5 ✉]

Alaska has the lowest population density in the United States (US) with a mix of urban centers and isolated rural communities. Alaska's distinct population dynamics compared to the contiguous US may have contributed to unique patterns of SARS-CoV-2 variants observed in early 2021. Here we examined 2323 SARS-CoV-2 genomes from Alaska and 278,635 from the contiguous US collected from December 2020 through June 2021 because of the notable emergence and spread of lineage B.1.1.519 in Alaska. We found that B.1.1.519 was consistently detected from late January through June of 2021 in Alaska with a peak prevalence in April of 77.9% unlike the rest of the US at 4.6%. The earlier emergence of B.1.1.519 coincided with a later peak of Alpha (B.1.1.7) compared to the contiguous US. We also observed differences in variant composition over time between the two most populated regions of Alaska and a modest increase in COVID-19 cases during the peak incidence of B.1.1.519. However, it is difficult to disentangle how social dynamics conflated changes in COVID-19 during this time. We suggest that the viral characteristics, such as amino acid substitutions in the spike protein, likely contributed to the unique spread of B.1.1.519 in Alaska.

In late 2020 and early 2021, variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), began emerging worldwide. Variants were found to contain mutations that affected transmissibility, disease severity, diagnostic detection, vaccine efficacy and immune evasiveness compared to the original SARS-CoV-2 virus (Wuhan-1, 2019) lineage[1]. Sustained high levels of transmission, variable vaccine and infection-induced immunity, along with geographic and social factors all contribute to the evolution, emergence, and spread of variants in locales across the world.

Alaska is the largest state in the United States by area but has the lowest population density with approximately 1.2 people per square mile on average[2]. Population density is highly location dependent across the state with two densely populated and highly interconnected economic regions and many sparsely populated rural areas only accessible by boat or plane. The unique community dynamics in Alaska, paired with its 800 km physical distance from the contiguous United States, has the potential to contribute to unique patterns of SARS-CoV-2 variant emergence and spread.

Here, we report on the striking difference in prevalence of SARS-CoV-2 variant B.1.1.519[3] between Alaska and the contiguous United States in early 2021. Based on all publicly available genomes outlined in outbreak.info, B.1.1.519 has been detected in at least 72 countries. At its peak, B.1.1.519 comprised around 3% of sequenced cases globally[4] with the highest prevalence observed in Mexico where the lineage was first recognized and comprises 7,959 of 54,886 (14.5%) sequenced genomes[5]. B.1.1.519 was detected at relatively low levels across the contiguous United States but was never characterized as a variant of concern (VOC), variant of interest (VOI), or variant being monitored (VBM) by the Center for Disease Control and Prevention (CDC). In contrast, the World Health Organization designated B.1.1.519 a variant under monitoring (VUM) on June 20, 2021, implying the potential future risk until November 2021 when Delta became the dominant variant globally[6]. This classification was in part due to the several key mutations B.1.1.519 encodes including amino acid substitutions T478K, P681H, and T732A in the spike (S) protein, that may affect affinity to the SARS-CoV-2 receptor, human angiotensin I converting enzyme-2 (hACE-2). Notably, a study examining the clinical impact of B.1.1.519 infection in Mexico City found patients infected with B.1.1.519 displayed a significant increase in the severity of COVID-19 respiratory tract symptoms relative to patients infected with non-B.1.1.519 variants[7].

[1]Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK, USA. [2]Alaska Division of Public Health, State of Alaska, Fairbanks, AK, USA. [3]Department of Biological Sciences, University of Alaska Anchorage, Anchorage, AK, USA. [4]WWAMI School of Medical Education, University of Alaska Anchorage, Anchorage, AK, USA. [5]Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK, USA. ✉email: ebortz@ alaska.edu; jchen15@alaska.edu; dmdrown@alaska.edu

In this study, we report the emergence, prevalence, and viral traits of SARS-CoV-2 lineages circulating in Alaska from 2233 sequenced specimens collected between the first week of December 2020 through the last week of June 2021, after which B.1.1.519 was no longer detected. Using SARS-CoV-2 sequence data from the United States available on NCBI GenBank by the CDC and its contracted labs, we explore how the trend in B.1.1.519 and other variants observed in Alaska compared to trends across the contiguous United States. These results highlight the importance of continued genomic surveillance for aiding efforts in controlling the pandemic by providing a context to compare the potential for the next VOC emergence.

Materials and methods

Retrieving SARS-CoV-2 sequence data for Alaska. The data presented here were generated as part of the *Alaska SARS-CoV-2 Sequencing Consortium* which is a partnership between the University of Alaska and the Alaska Division of Public Health (AKDPH) with the aim to increase genomic surveillance of variants. On February 14th, 2022, we downloaded 2,323 sequences from samples collected between 2020-11-29 to 2021-06-26 available on GISAID for subsequent analysis[8,9]. Genome sequencing in Alaska is from a non-targeted sample of cases, which is the best available approximation of random samples despite potential disparate coverage across Alaska's economic regions. We used these data to estimate the prevalence of lineages per week on dates beginning on Sunday.

Analysis of SARS-CoV-2 sequence data for Alaska. Lineages were determined by running sequences through PANGO-v 1.2.124, Pangolin v 3.1.20, and pangoLEARN v 2/2/22, and Scorpio v 0.3.16[10]. We estimated the prevalence of genomes grouped by the emerging lineages including B.1.617.2 and AY sublineages (Delta), B.1.1.519, B.1.427/429 (Epsilon), B.1.525 (Eta), B.1.617.1 (Kappa), C.37(Lambda), B.1.526 (Iota), B.1.621 (Mu), B.1.351 (Beta), P.1 (Gamma), and B.1.1.7 (Alpha). Genomes that did not fall into these lineages were grouped together into the category 'Not Emerging Lineage.' The number of Spike protein amino acid substitutions was determined using Nextclade (v 1.13.2) with 'bad' quality genomes, as determined by Nextclade's algorithm for quality, removed prior to analyses[11]. Nextclade's quality control incorporates six individual rules based on missing data, mutations, and amino acid changes as described in the online documentation.

We collected metadata on the number of cases for Alaska from the AKDPH COVID-19 CSV Files Database Cases Data B dataset on 2022-2-14[12]. We collected 2020 population estimates for economic regions of Alaska from the Alaska Department of Labor and Workforce Development population estimates page[13]. Information on the timeline of COVID-19 policies (in terms of restrictions, opening, and vaccines) in Alaska was collected from the Johns Hopkins Coronavirus Resource Center[14].

SARS-CoV-2 sequence data for the contiguous United States. On February 14th, 2022, we downloaded metadata available on GenBank for 1,331,799 sequences including Pangolin assignment, Isolate, and GeoLocation for all samples from the United State of America collected between 2020-11-29 to 2021-06-26 sequenced by the CDC. We then filtered out cases from Alaska, Hawaii, and US territories to limit our comparisons to the contiguous United States. We collected daily case data with seven day rolling averages from the CDC COVID Data Tracker site for the entire United States[15].

Visualizations and statistical analyses. We generated visualizations in RStudio (v 1.4.1106) using packages ggplot2 (v 3.3.5), ggpubr (v 0.4.0), tidyverse (v 1.3.1), and lubridate (v 1.7.10). To test for significant differences in spike glycoprotein amino acid substitutions between emerging lineages or sub-lineages, we used the Wilcoxon test.

Results and discussion

Higher prevalence of B.1.1.519 in Alaska versus the contiguous United States. We found a striking difference in the percentage of sequenced cases, a metric that can be used as an estimate of prevalence, assigned to the lineage B.1.1.519 between Alaska (Fig. 1B) and the rest of the United States (Fig. 1D) in early 2021. While the earliest detection of B.1.1.519 in the United States occurred 27 December 2020 in sequence data from New York, it was not until five weeks later, 4 February 2021, that B.1.1.519 was detected in sequence data from Alaska. By this time in the United States, other VOCs including Alpha (B.1.1.7) and Epsilon (B.1.427/429) had already emerged and were being consistently detected in sequence data. In contrast, Alaska had detected a single case of Alpha and no cases of Epsilon in December of 2020.

Unlike the contiguous United States where Alpha comprised approximately 5% of sequenced cases by the end of January, in Alaska, Alpha was not consistently detected in sequenced cases until March of 2021 (Fig. 1B). By the first week of March in Alaska, B.1.1.519 had already reached an estimated 60.8% of sequenced cases compared to 4.6% in the rest of the US, which was the peak prevalence for the contiguous United States (Fig. 1B). During B.1.1.519's peak in the contiguous United States, Alpha made up 30.5% of sequenced cases. By contrast, in Alaska, B.1.1.519 reached a peak prevalence of 77.9% by the week of 4 April 2021, 5 weeks after the peak in the contiguous United States.

In terms of COVID-19 cases reported in Alaska, there was a sharp decline the first several weeks of December 2020 followed by an increase in cases through the first week of January 2021 and then another decline (Fig. 1A). During this time, cases in the United States increased until the second week of January 2021 followed by a decline through the third week of February 2021 (Fig. 1C). In both Alaska and the rest of the United States, new COVID-19 cases remained at a relatively stable rate while the cases attributed to VOC, VOI, and VBMs were rising (Fig. 1A,C). Given the complexity of factors that affect case rates, it is difficult to attribute the stabilization
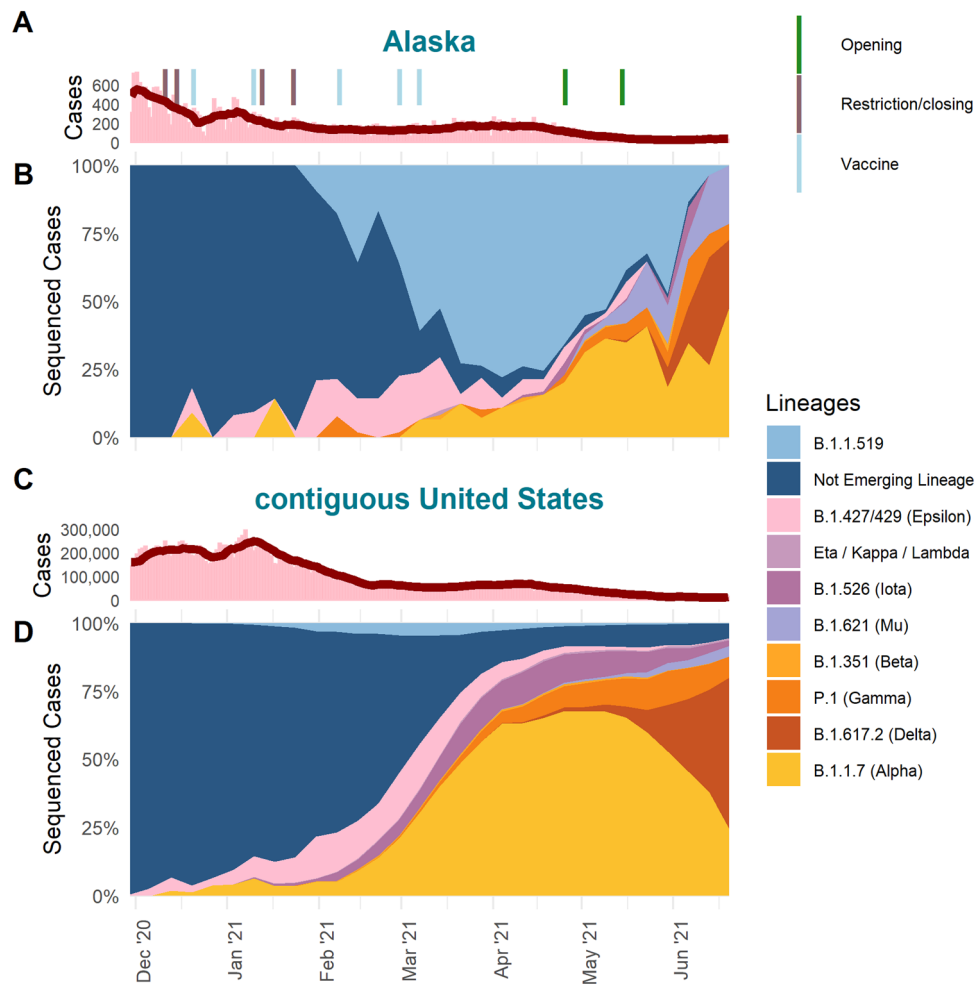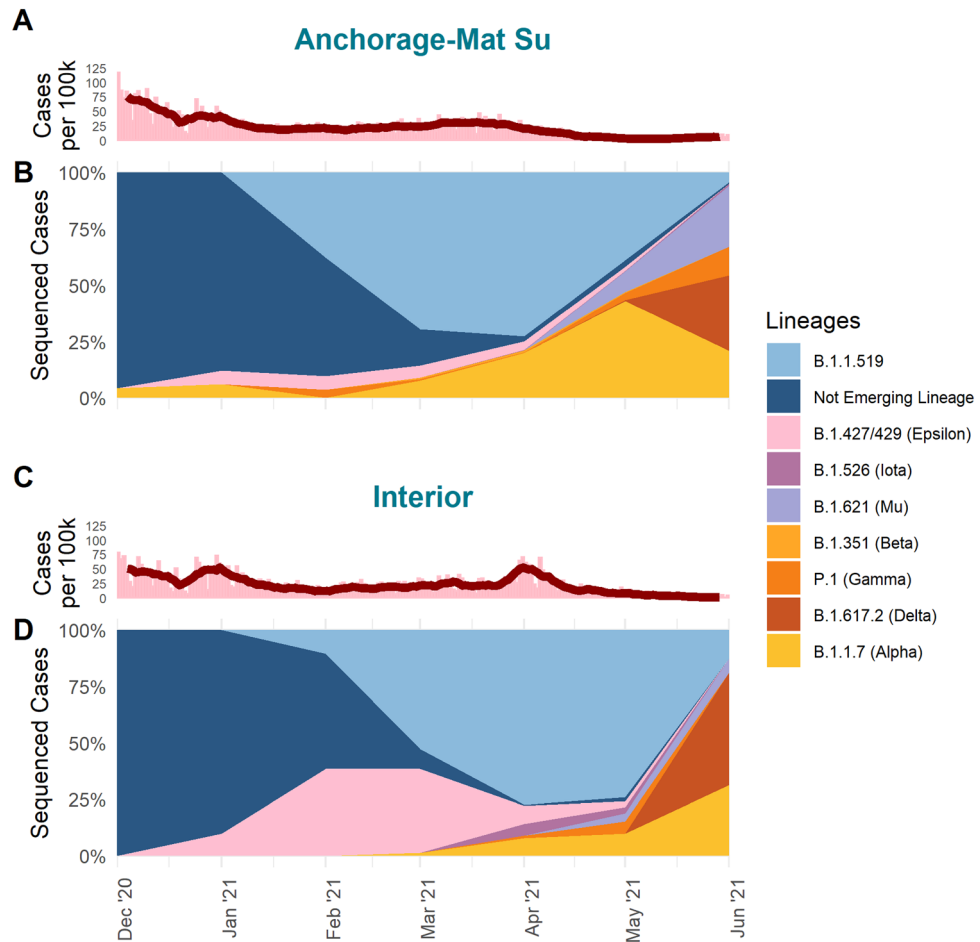
Figure 1. COVID-19 cases and variants in Alaska versus the contiguous United States. (**A,C**) Daily case count (bars) and seven day rolling average (red line) and (**B,D**) percent of sequences (estimated prevalence) by PANGO lineages detected by week from 2020-11-29 to 2021-06-26 for Alaska (**A,B**) and the contiguous United States (**C,D**).

in cases to biological mechanisms driven by variants or behavioral and social dynamics such as increased vaccination rates, social restrictions, or adherence to respiratory hygiene measures during this time.

When B.1.1.519 first emerged in Alaska, several restrictions were in place as a result of COVID-19 Public Health Disaster Emergency Declaration enacted 11 December 2020 and then renewed 14 January 2021 (Fig. 1A). The State of Alaska recommended wearing a face covering in most situations, working remotely when possible, and avoiding gatherings with non-household members. In December, Alaska also received its initial allocations of the Comirnaty vaccine (Pfizer-BioNTech BNT162b2) for frontline workers. By 11 January 2021, Alaskans 65 years of age and older were eligible for both Comirnaty and Spikevax (Moderna mRNA-1273). Vaccine availability was expanded to the entire public by 10 March 2021 and included primarily Spikevax and Comirnaty, at which time B.1.1.519 comprised 60.8% of the sequenced cases. Alaska was ahead of the contiguous United States in terms of vaccine availability and percent of the population vaccinated in early 2021, which confounds the role host immunity may have had in variant emergence and infection dynamics. Citing the the rapid vaccination rates, Alaska ended the COVID-19 Disaster Declaration, signally and end to mitigation measures on April 30, 2021.

Difference in the emergence of B.1.1.519 within Alaska's two most populated regions. Given Alaska's vast geographic expanse and limited travel options, populations of people tend to interact more frequently within economic regions of the state. In Alaska there are six economic regions defined by the Department of Labor and Workforce Development: the Anchorage-Mat Su, Interior, Gulf Coast, Southeast, Southwest, and Northern regions in order from highest to lowest population. Based on available but limited data, we observed distinct trends in SARS-CoV-2 across economic regions of Alaska (Fig. S1). However, in our analysis, we excluded all but the two most populous economic regions, the Anchorage-Mat Su and Interior, because of the low proportion of sequenced cases in the other economic regions of the state. We can draw more robust conclu-

Figure 2. Variants in Alaska's two most populated regions. (**A,C**) Daily case count (bars) and seven day rolling average (red line) and (**B,D**) percent of sequences (estimated prevalence) by PANGO lineages detected by week from 2020-11-29 to 2021-06-26 for the Anchorage-Mat Su (**A,B**) and Interior (**C,D**) regions of Alaska from December 2020 through June 2021.

sions about estimates of prevalence because these two regions sequenced greater than 5% of the newly reported cases February through June of 2021 (Fig. S2).

Between the Anchorage-Mat Su and Interior economic regions there were distinct dynamics in the prevalence of variants over time (Fig. 2). For example, Alpha was consistently detected earlier in the Anchorage-Mat Su (Fig. 2B) than the Interior (Fig. 2D), 360 miles apart. However, in the Interior, Epsilon comprised a high percentage of sequenced cases from January through March 2021 (Fig. 2D), and the emergence and persistence of B.1.1.519 differed between these two regions of the state.

The first case of B.1.1.519 in Alaska was detected from a sample collected 4 February 2021 in the Southwest region of the state. In both the Anchorage-Mat Su and Interior regions, B.1.1.519 was detected in specimens with collection dates one day apart from one another, 7 and 8 February 2021, respectively. By the end of February, B.1.1.519 comprised 37.8% of sequenced cases in the Anchorage-Mat Su region and 10.5% of cases in the Interior (Fig. 2). By February VBMs detected in the Anchorage-Mat Su included Alpha, Epsilon, and Gamma. On the other hand, in the Interior, Epsilon was the only VBM detected and comprised 38.5% of sequenced cases in February (Fig. 2D). In the Interior, it appears that Epsilon, which was first detected 6 January 2021, was able to establish itself before B.1.1.519. However, one month after B.1.1.519 was first detected in the Interior, B.1.1.519 overtook Epsilon. By April, the majority of sequenced cases were assigned to B.1.1.519 in the Interior (77.2%) and the Anchorage-Mat Su (72.6%). In both regions, there is a noticible increase in cases around this peak in B.1.1.519 prevlaence (Fig. 2A,C). After peaking in April, prevalence of B.1.1.519 in May only marginally decreased for the Interior to 74.1% but sharply decreased for the Anchorage-Mat Su region to 38.8% (Fig. 2).

The sharp decline of B.1.1.519 in the Anchorage-Mat Su region may have occurred earlier than in the Interior due to the emergence of other variants that had established themselves in the Anchorage-Mat Su region well before the Interior. For example, in the Anchorage-Mat Su region, Alpha was detected on 20 December 2020—97 days before it was first detected in the Interior. This gave Alpha more time to establish itself within the population and consequently reached 42.9% of sequenced cases by May in the Anchorage-Mat Su region versus 9.8% in the Interior. Delta also emerged in the Anchorage-Mat Su region weeks before it was first detected in
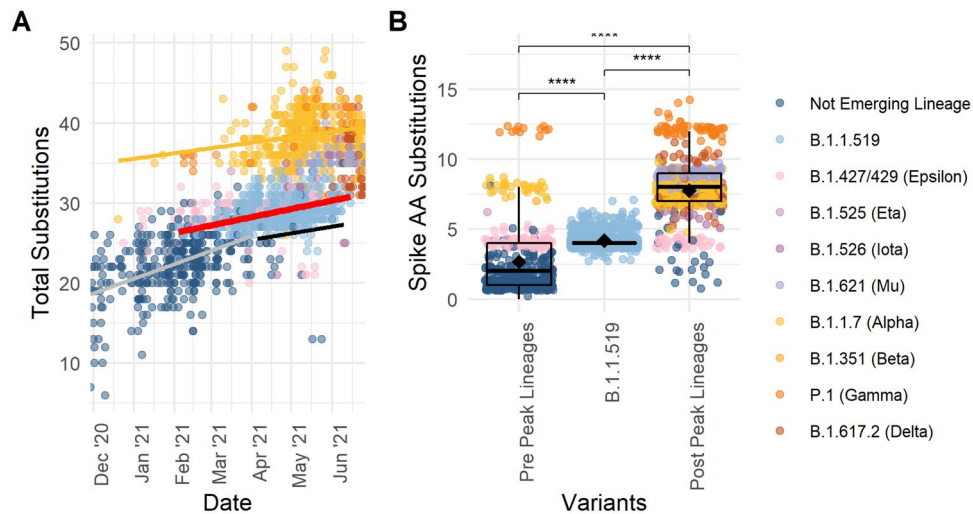
Figure 3. Amino acid substitutions in SARS-CoV-2 genomes from Alaska. Scatterplots depicting the (**A**) total AA substitutions over time. Data includes each genome from an Alaskan confirmed case, colored by variant. The lines show a generalized linear model regression of not emerging lineage cases before 2021-04-04 (gray), after 2021-04-04 (black), Alpha (yellow), and B.1.1.519 (red). (**B**) Boxplot of number of S protein AA substitutions per genome from sequences pre-peak prevalence of B.1.1.519 on 2021–04-04, B.1.1.519, and post peak prevalence genomes. Diamonds depict the mean number of AA substitutions. Points are individual genomes colored by lineage. Wilcoxon test between group significance $p < 0.0001$ ****, $p < 0.05$ *, ns > 0.1.

the Interior, highlighting why B.1.1.519 may have persisted in Interior populations with the last case of B.1.1.519 detected 2 weeks after last detection in the Anchorage-Mat Su.

High prevalence of B.1.1.519 in Alaska driven mutational advantage.
Mutations in the SARS-CoV-2 genomes from Alaskan cases occurred in an almost clockwise fashion (Fig. 3A). Although most of these mutations are expected to have neutral effects, some changes can confer selective advantages that enhance fitness by increasing immune evasion, transmissibility, and/or infectivity[16]. Mutations that cause amino acid changes to the spike (S) glycoprotein may confer selective advantages given the role of the S protein in COVID-19 pathogenesis and tropism. The mature S protein is cleaved into a fusion domain (S2), and a S1 domain containing an N-terminal head and the receptor binding domain (RBD) that binds to host cell human angiotensin converting enzyme-2 (hACE2) receptor. The S1 domain and RBD particularly are key targets for binding of neutralizing antibodies induced by infection or vaccines, principally IgG[17]. Given the role of the S protein in receptor binding and subsequent membrane fusion and viral entry into host cells, mutations affecting S1, especially the RBD, can impact hACE2 affinity, viral entry, infectivity, and immune evasion[18,19].

B.1.1.519 has 11 shared amino acid (AA) mutations relative to the original SARS-CoV-2 (Wuhan-1) genome, with 4 in S: T478K and D614G in the RBD, P681H in the S1/S2 cleavage site, and T732A in the S2 domain[5]. T478K also arose in B.1.617.2 (Delta)[20] and may contribute to antibody evasion[21]. P681H also occurs in B.1.1.7 (Alpha) and has a weak ability to increase proteolytic cleavage of S1/S2 in vitro[22]. The functional significance of B.1.1.519 harboring these mutations in S is unclear. In a study from the Netherlands, there was no evidence of increased viral load conferred by B.1.1.519[23].

Interestingly, based on the regression of total substitutions over time, which exhibited an increase at a steady rate, B.1.1.519 genomes had more total substitutions across the genome (Fig. 3A, red line) than the background of non-emerging lineages (Fig. 3A, gray and black lines). B.1.1.519 had significantly more AA substitutions in the spike protein than lineages detected in Alaska prior to the week of B.1.1.519's peak prevalence, 4 April 2021, suggesting a potential competitive advantage over previously circulating lineages (B.1.1.519 = 4.18 ± 0.01, Pre-peak lineages = 2.65 ± 0.10 S AA substitutions, Wilcoxon W = 42,302, p-value < 2e−16; Fig. 3B). The period of time after B.1.1.519's peak prevalence was dominated by the emergence of many VOC. Genomes sequenced after B.1.1.519's peak prevalence had significantly more AA changes in the spike protein than B.1.1.519, dominated by more divergent Alpha (Fig. 3A, yellow line) and Delta VOC (Post-peak lineages = 7.75 ± 0.07 S AA substitutions, Wilcoxon W = 1e+05, p-value < 2e−16; Fig. 3B). These VOC likely held a selective advantage that allowed them to outcompete B.1.1.519 in Alaska, a phenomenon of variant replacement observed across Alpha and Delta waves in the USA[24].

B.1.1.519 established itself within the circulating population of SARS-CoV-2 viruses in Alaska weeks before other variants, such as Alpha, were consistently detected (Fig. 1B). The difference in prevalence of B.1.1.519 during Alpha's emergence in Alaska versus the contiguous United States, paired with the finding of more S protein amino acid changes in this variant than previous lineages in Alaska, suggests that in Alaska B.1.1.519 likely emerged and became dominant due the unique diversity of variants. In this context, the initial transmission events of B.1.1.519 led to this variant outcompeting the previously circulating lineages because of a selective

advantage that that variant already possessed. The same trend was likely not possible in the other states because Alpha, which proved to outcompete B.1.1.519, had already established itself within the population, outcompeting B.1.1.519 when it emerged.

Other locations displayed similar patterns to Alaska in terms of B.1.1.519's emergence and spread, most notably Mexico. B.1.1.519 was first detected in Mexico in November 2020, about three months prior to Alaska's first detection, and was found to rapidly outcompete existing variants to become the dominant variant in the country, comprising 51.5% of sequenced genomes by January 2021[5]. By February, 90.9% of Mexico City's sequenced cases were B.1.1.519. In Mexico, B.1.1.519 was not displaced by ongoing transmission of Alpha[25]. The pattern of circulation and displacement is similar to Alaska. In the comprehensive survey of variants in Mexico, Taboada et al. attributed the success of B.1.1.519 to the unique combination of circulating lineages at the time of arrival[26].

In a study examining the clinical impacts of B.1.1.519 in Mexico City, Cedro-Tanda (2021) found that B.1.1.519-infected patients had a significant increase in adjusted odds ratio for developing severe symptoms of COVID-19 including dyspnea, chest pain and cyanosis, and hospitalizations compared to non-B.1.1.519 infected patients. Unfortunately, the necessary data to draw conclusions about the clinical impacts of B.1.1.519 in Alaska, as shown in the Mexico-City study, were not available. Another study, from the Netherlands, describes a cluster of cases associated with B.1.1.519, demonstrating a high within-facility attack rate[23]. Interestingly, that same study did not find differences in B.1.1.519 viral loads compared with other variants, and sample size limited conclusions about disease severity.

Conclusion

Using genomic data, we identified the unique dynamics of B.1.1.519 spread in Alaska compared to the contiguous United States. The distinct pattern of emergence and spread in Alaska emphasizes both how novel introductions impact regional differences in circulating SARS-CoV-2 lineages and the significance of having robust sequencing efforts in place to detect this variation. There are several resources that compile SARS-CoV-2 sequence data to help examine patterns of emergence and spread. One such resource, outbreak.info, uses genomic data from the sequence repository GISAID to estimate prevalence, emergence, and spread of SARS-CoV-2 lineages globally[4]. When considering data from this source, the highest worldwide prevalence of B.1.1.519 was in early March 2021, a month before Alaska's peak prevalence. At a country level, B.1.1.519 was found to comprise the highest percentage of sequenced genomes in Mexico (15% of sequenced cases compared to 1% of cases sequenced in the United States) over the course of the pandemic. On 14 February 2022, when this analysis was conducted, B.1.1.519 comprised 10.4% of all of Alaska's SARS-CoV-2 sequences to date. While this is high compared to other geographical locations, spatial and temporal heterogeneity in sequencing efforts is known to bias genomic surveillance and complicate comparisons between locations.

In Alaska, disparate sequencing coverage affected our ability to generate robust analyses in rural regions of the state where sampling and subsequent sequencing efforts were lower than more populated regions. Despite this disparate coverage, genomic surveillance continues to be an essential tool for informing public health decision-making throughout the state. Genomic surveillance paired with epidemiological findings will be key to detecting new threats as they emerge and is a critical tool in shaping policy to address the clinical impacts of local variation in circulating virus lineages. At a patient level, variant surveillance can inform which monoclonal antibodies will be most effective for early treatment to directly reduce individual mortality and morbidity. At a community level, variant proportion data paired with epidemiological findings can inform the broader population of their risks of reinfection, evaluate vaccine efficacy, and assess the need for appropriate treatment stockpiles and quantities. Isolated populations are vulnerable to variants with significant clinical outcomes that rise to local dominance. While the WHO did monitor the abundance of B.1.1.519, this variant did not receive the VOC classification as it lacked a globally spread. It is cases like these that stress the importance of genomic surveillance in places like Alaska. Genomic surveillance information will continue to be of great importance for public health policy decisions as SARS-CoV-2 becomes endemic in Alaska and around the globe.

Data availability

All data used in this study are available online through publicly open databases. SARS-CoV-2 genomes for Alaska were retrieved from the global initiative on sharing all influenza data (GISAID) (https://www.gisaid.org/). GISAID Accession Numbers are provided in the Supplementary file. United States SARS-CoV-2 metadata was retrieved from NCBI GenBank (https://www.ncbi.nlm.nih.gov/sars-cov-2/). Metadata on the number of cases for Alaska is available on the Alaska COVID-19 Summary Dashboard (https://covid19.alaska.gov/).

References
1. Walensky, R. P., Walke, H. T. & Fauci, A. S. SARS-CoV-2 variants of concern in the United States—Challenges and opportunities. *JAMA* **325**(11), 1037–1038 (2021).
2. United States Census Bureau. *QuickFacts Alaska*. https://www.census.gov/quickfacts/fact/table/AK# (2021).
3. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**(11), 1403–1407 (2021).
4. Mullen, J.L. *et al. Center for Viral Systems Biology. outbreak.info*. https://outbreak.info/ (2020).
5. Rodríguez-Maldonado, A. P. *et al.* Emergence and spread of the potential variant of interest (VOI) B.1.1.519 of SARS-CoV-2 predominantly present in Mexico. *Arch. Virol.* **166**(11), 3173–3177 (2021).
6. World Health Organization. *Tracking SARS-CoV-2 Variants*. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants (2021).

7. Cedro-Tanda, A. *et al.* The evolutionary landscape of SARS-CoV-2 variant B.1.1.519 and its clinical impact in Mexico City. *Viruses* **13**, 2182 (2021).
8. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data—From vision to reality. *Eurosurveillance* **22**(13), 30494 (2017).
9. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Chall.* **1**(1), 33–46 (2017).
10. O'Toole, Á. *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolut.* **7**(2), veab064. https://doi.org/10.1093/ve/veab064 (2021).
11. Aksamentov, I., Roemer, C., Hodcroft, E. B. & Neher, R. A. Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *J. Open Sour. Softw.* **6**(67), 3773. https://doi.org/10.21105/joss.03773 (2021).
12. Alaska Department of Health and Social Services. *COVID-19 CSV Files Database.* https://experience.arcgis.com/experience/b2f5a105a41c4ca88024efbc2c2e868f/ (2021).
13. Alaska Department of Labor and Workforce Development. *2020 Population Estimates by Borough, Census Area, and Economic Region.* https://live.laborstats.alaska.gov/pop/index.cfm (2021).
14. John Hopkins Coronavirus Resource Center. *Impact of Opening and Closing Decisions by State.* https://coronavirus.jhu.edu/data/state-timeline/new-confirmed-cases/alaska/ (2022).
15. Centers for Disease Control and Prevention (CDC). *United States COVID-19 Cases, Deaths, and Laboratory Testing (NAATs) by State, Territory, and Jurisdiction.* https://covid.cdc.gov/covid-data-tracker/#cases_community (2021).
16. Korber, B. *et al.* Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**(4), 812–827 (2020).
17. Walls, A. C. *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**(2), 281–292 (2020).
18. Berger, I. & Schaffitzel, C. The SARS-CoV-2 spike protein: Balancing stability and infectivity. *Cell Res.* **30**(12), 1059–1060 (2020).
19. Hadi-Alijanvand, H. & Rouhani, M. Studying the effects of ACE2 mutations on the stability, dynamics, and dissociation process of SARS-CoV-2 S1/hACE2 complexes. *J. Proteome Res.* **19**(11), 4609–4623 (2020).
20. Jhun, H., Park, H. Y., Hisham, Y., Song, C. S. & Kim, S. SARS-CoV-2 Delta (B.1.617.2) variant: A unique T478K mutation in receptor binding motif (RBM) of spike gene. *Immune Netw.* **21**(5), e32. https://doi.org/10.4110/in.2021.21.e32 (2021).
21. Di Giacomo, S., Mercatelli, D., Rakhimov, A. & Giorgi, F. M. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike mutation T478K. *J. Med. Virol.* **93**(9), 5638–5643 (2021).
22. Lubinski, B. *et al.* Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (alpha) spike. *Iscience* **25**(1), 103589. https://doi.org/10.1016/j.isci.2021.103589 (2022).
23. Gorgels, K. M. *et al.* Linked nosocomial COVID-19 outbreak in three facilities for people with intellectual and developmental disabilities due to SARS-CoV-2 variant B.1.1.519 with spike mutation T478K in the Netherlands. *BMC Infect. Dis.* **22**(1), 1–9 (2022).
24. Lambrou, A. S. *et al.* Genomic surveillance for SARS-CoV-2 variants: Predominance of the Delta (B.1617.2) and Omicron (B.1.1.529) Variants—United States, June 2021–January 2022. *Morbidity Mortality Wkly. Rep.* **71**(6), 206–211 (2022).
25. Zárate, S. *et al.* The Alpha variant (B.1.1.7) of SARS-CoV-2 failed to become dominant in Mexico. *Microbiol. Spectr.* **10**(2), 4021 (2022).
26. Taboada, B. *et al.* Genetic analysis of SARS-CoV-2 variants in Mexico during the first year of the COVID-19 pandemic. *Viruses* **13**(11), 2161 (2021).
27. Haan, T. J. *et al.* Founder effect contributes to the unique pattern of SARS-CoV-2 variant B.1.1.519 emergence in Alaska. *medRxiv* 2022.03.17.22272446. https://doi.org/10.1101/2022.03.17.22272446 (2022).

Acknowledgements

Author contributions

T.J.H., B.M.B., J.L.B., C.K., E.B., J.C., D.M.D. were responsible for the conception of the project. L.K.S, S.D., E.H., J.Z., D.P., M.R., J.P., J.C., D.M.D. were responsible for acquiring and making publicly available the sequence data from Alaska samples. T.J.H. and D.M.D. conducted the analysis and interpreted the data. T.J.H. prepared the figures and wrote the initial draft of the manuscript. T.J.H. and D.M.D. substantially revised the manuscript. L.K.S., J.P., B.M.B., J.L.B., E.B., and J.C. provided comments and feedback on manuscript drafts. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-25373-1.

Correspondence and requests for materials should be addressed to E.B., J.C. or D.M.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.