

1 Founder effect contributes to the unique 2 pattern of SARS-CoV-2 variant 3 B.1.1.519 emergence in Alaska

4
5 Tracie J. Haan¹, Lisa K. Smith², Stephanie DeRonde², Elva House², Jacob Zidek², Diana
6 Puhak², Matthew Redlinger³, Jayme Parker², Brian M. Barnes¹, Jason L. Burkhead³, Cindy
7 Knall⁴, Eric Bortz^{3,4,*}, Jack Chen^{1,2,5,*}, and Devin M. Drown^{1,5*}

8
9 ¹ Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, Alaska, USA

10 ² Alaska Division of Public Health, State of Alaska, Fairbanks, Alaska, USA

11 ³ Department of Biological Sciences, University of Alaska Anchorage, Anchorage, Alaska, USA

12 ⁴ WWAMI School of Medical Education, University of Alaska Anchorage, Anchorage, Alaska,
13 USA

14 ⁵ Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, Alaska, USA

15

16 * Corresponding authors: Devin M. Drown, dmdrown@alaska.edu; Eric Bortz
17 ebortz@alaska.edu; Jack Chen, jchen15@alaska.edu

18

19 Running title: Emergence of B.1.1.519 in Alaska

20 Key Words: SARS-CoV-2; COVID-19; genomic epidemiology; variant

Emergence of B.1.1.519 in Alaska

21 Abstract

22 Alaska is the largest geographic state in the United States with the lowest population
23 density and a mix of urban centers and isolated rural communities. The differences in population
24 dynamics in Alaska from the contiguous United States may have contributed to a unique pattern
25 of emergence and spread of SARS-CoV-2 variants observed in early 2021. Here we examined
26 2,323 virus genomes from Alaska and 278,635 virus genomes from the contiguous United States
27 collected between the first week of December 2020 through the last week of June 2021. We
28 focused on this timeframe because of the notable emergence and spread of the SARS-CoV-2
29 lineage B.1.1.519 observed in Alaska. We found that this variant was consistently detected in
30 Alaska from the end of January through June of 2021 with a peak prevalence in April of 77.9%
31 unlike the rest of the United States with a peak prevalence of 4.6%. In Alaska, the earlier
32 emergence of B.1.1.519 coincided with a later peak of Alpha (B.1.1.7) when compared to the rest
33 of the United States. We also observed differences in the composition of lineages and variants
34 over time between the two most populated regions of Alaska. Although there was a modest
35 increase in COVID-19 cases during the peak incidence of B.1.1.519, it is difficult to disentangle
36 how social dynamics conflated changes in COVID-19 during this time. We suggest that the viral
37 characteristics, such as amino acid substitutions in the spike protein, and a founder effect likely
38 contributed to the unique spread of B.1.1.519 in Alaska.

Emergence of B.1.1.519 in Alaska

39 Introduction

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

48 Alaska is the largest state in the United States by area but has the lowest population
49 density with approximately 1.2 people per square mile on average (US Census, 2020). Population
50 density is highly location dependent across the state with two densely populated and highly
51 interconnected economic regions and many sparsely populated rural areas only accessible by
52 boat or plane. The unique community dynamics in Alaska, paired with its 800 km physical distance
53 from the contiguous United States (hereafter, Lower 48), has the potential to contribute to unique
54 patterns of SARS-CoV-2 variant emergence and spread.

55 Here, we report on the striking difference in prevalence of SARS-CoV-2 variant B.1.1.519
56 (Rambaut et al., 2020) between Alaska and the Lower 48 in early 2021. Based on all publicly
57 available genomes outlined in outbreak.info, B.1.1.519 has been detected in at least 72 countries.
58 At its peak, B.1.1.519 comprised around 3% of sequenced cases globally (Mullen et al., 2020)
59 with the highest prevalence observed in Mexico where the lineage was first recognized and
60 comprises 7,959 of 54,886 (14.5%) sequenced genomes (Rodriguez-Maldonado 2021).
61 B.1.1.519 was detected at relatively low levels across the Lower 48 but was never characterized
62 as a variant of concern (VOC), variant of interest (VOI), or variant being monitored (VBM) by the
63 Center for Disease Control and Prevention (CDC). In contrast, the World Health Organization

Emergence of B.1.1.519 in Alaska

64 designated B.1.1.519 a variant under monitoring (VUM) on June 20, 2021, implying the potential
65 future risk until November 2021 when Delta became the dominant variant globally (World Health
66 Organization [WHO], 2021). This classification was in part due to the several key mutations
67 B.1.1.519 encodes including amino acid substitutions T478K, P681H, and T732A in the spike (S)
68 protein, that may affect affinity to the SARS-CoV-2 receptor, human angiotensin I converting
69 enzyme-2 (hACE-2). Notably, a study examining the clinical impact of B.1.1.519 infection in
70 Mexico City found patients infected with B.1.1.519 displayed a significant increase in the severity
71 of COVID-19 respiratory tract symptoms relative to patients infected with non-B.1.1.519 variants
72 (Cedro-Tanda et al., 2021).

73 In this study, we report the emergence, prevalence, and viral traits of SARS-CoV-2
74 lineages circulating in Alaska from 2,233 sequenced specimens collected between the first week
75 of December 2020 through the last week of June 2021, after which B.1.1.519 was no longer
76 detected. Using SARS-CoV-2 sequence data from the United States available on NCBI GenBank
77 by the CDC and its contracted labs, we explore how the trend in B.1.1.519 and other variants
78 observed in Alaska compared to trends across the Lower 48. These results are used to make
79 predictions about the future of SARS-CoV-2 in Alaska and highlight the importance of continued
80 genomic surveillance for aiding efforts in controlling the pandemic.

81 Materials and Methods

82 *Retrieving SARS-CoV-2 Sequence Data for Alaska*

83 The data presented here were generated as part of the *Alaska SARS-CoV-2 Sequencing*
84 *Consortium* which is a partnership between the University of Alaska and the Alaska Division of
85 Public Health (AKDPH) with the aim to increase genomic surveillance of variants. On February
86 14th, 2022, we downloaded 2,323 sequences from samples collected between 2020-11-29 to
87 2021-06-26 available on GISAID for subsequent analysis (Shu & McCauley, 2017; Elbe & Merret,

Emergence of B.1.1.519 in Alaska

88 2017). Genome sequencing in Alaska is from a non-targeted sample of cases, which is the best
89 available approximation of random samples despite potential disparate coverage across Alaska's
90 economic regions. We used these data to estimate the prevalence of lineages per week on dates
91 beginning on Sunday.

92 *Analysis of SARS-CoV-2 Sequence Data for Alaska*

93 Lineages were determined by running sequences through PANGO-v 1.2.124, Pangolin v
94 3.1.20, and pangoleARN v 2/2/22, and Scorpio v 0.3.16 (O'Toole et al., 2021). We estimated the
95 prevalence of genomes grouped by the emerging lineages including B.1.617.2 and AY
96 sublineages (Delta), B.1.1.519, B.1.427/429 (Epsilon), B.1.525 (Eta), B.1.617.1 (Kappa),
97 C.37(Lambda), B.1.526 (Iota), B.1.621 (Mu), B.1.351 (Beta), P.1 (Gamma), and B.1.1.7 (Alpha).
98 Genomes that did not fall into these lineages were grouped together into the category 'Not
99 Emerging Lineage.' The number of Spike protein amino acid substitutions was determined using
100 Nextclade (v 1.13.2) with 'bad' quality genomes, as determined by Nextclade's algorithm for
101 quality, removed prior to analyses (Aksamentov et al., 2021). Nextclade's quality control
102 incorporates six individual rules based on missing data, mutations, and amino acid changes as
103 described in the online documentation.

104 We collected metadata on the number of cases for Alaska from the AKDPH COVID-19
105 CSV Files Database Cases Data B dataset on 2022-2-14 (AKDPH, 2022). We collected 2020
106 population estimates for economic regions of Alaska from the Alaska Department of Labor and
107 Workforce Development population estimates page (Alaska Department of Labor and Workforce
108 Development, 2021). Information on the timeline of COVID-19 policies (in terms of restrictions,
109 opening, and vaccines) in Alaska was collected from the Johns Hopkins Coronavirus Resource
110 Center (John Hopkins, 2022)

Emergence of B.1.1.519 in Alaska

111 *SARS-CoV-2 Sequence Data for the Lower 48*

112 On February 14th, 2022, we downloaded metadata available on GenBank for 1,331,799
113 sequences including Pangolin assignment, Isolate, and GeoLocation for all samples from the
114 United State of America collected between 2020-11-29 to 2021-06-26 sequenced by the CDC.
115 We then filtered out cases from Alaska, Hawaii, and US territories to limit our comparisons to the
116 lower 48 contiguous states. We collected daily case data with seven day rolling averages from
117 the CDC COVID Data Tracker site for the entire United States (CDC, 2022).

118 *Visualizations and Statistical Analyses*

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

123 Results and Discussion

124 *Higher prevalence of B.1.1.519 in Alaska versus the Lower 48*

125 We found a striking difference in the percentage of sequenced cases, a metric that can be
126 used as an estimate of prevalence, assigned to the lineage B.1.1.519 between Alaska (Figure
127 1B) and the rest of the United States (Figure 1D) in early 2021. While the earliest detection of
128 B.1.1.519 in the United States occurred 27 December 2020 in sequence data from New York, it
129 was not until five weeks later, 4 February 2021, that B.1.1.519 was detected in sequence data
130 from Alaska. By this time in the United States, other VOCs including Alpha (B.1.1.7) and Epsilon
131 (B.1.427/429) had already emerged and were being consistently detected in sequence data. In
132 contrast, Alaska had detected a single case of Alpha and no cases of Epsilon in December of
133 2020.

Emergence of B.1.1.519 in Alaska

134 Unlike the Lower 48 where Alpha comprised approximately 5% of sequenced cases by
135 the end of January, in Alaska, Alpha was not consistently detected in sequenced cases until
136 March of 2021 (Figure 1B). By the first week of March in Alaska, B.1.1.519 had already reached
137 an estimated 60.8% of sequenced cases compared to 4.6% in the rest of the US, which was the
138 peak prevalence for the Lower 48 (Figure 1B). During B.1.1.519's peak in the Lower 48, Alpha
139 made up 30.5% of sequenced cases. In Alaska, B.1.1.519 reached a peak prevalence of 77.9%
140 by the week of 4 April 2021, 5 weeks after the peak in the Lower 48.

141 In terms of COVID-19 cases reported in Alaska, there was a sharp decline the first several
142 weeks of December 2020 followed by an increase in cases through the first week of January 2021
143 and then another decline (Figure 1A). During this time, cases in the United States increased until
144 the second week of January 2021 followed by a decline through the third week of February 2021
145 (Figure 1C). In both Alaska and the rest of the United States new COVID-19 cases remained at
146 a relatively stable rate while the cases attributed to VOC, VOI, and VBMs were rising (Figure
147 1A,C). Given the complexity of factors that affect case rates, it is difficult to attribute the
148 stabilization in cases to biological mechanisms driven by variants or behavioral and social
149 dynamics such as increased vaccination rates, social restrictions, or adherence to respiratory
150 hygiene measures during this time.

151 When B.1.1.519 first emerged in Alaska, several restrictions were in place as a result of
152 COVID-19 Public Health Disaster Emergency Declaration enacted 11 December 2020 and then
153 renewed 14 January 2021 (Figure 1A). In December, Alaska also received its initial allocations of
154 the Comirnaty vaccine (Pfizer-BioNTech BNT162b2) for frontline workers. By 11 January 2021,
155 Alaskans 65 years of age and older were eligible for both Comirnaty and Spikevax (Moderna
156 mRNA-1273). Vaccine availability was expanded to the entire public by 10 March 2021 and
157 included primarily Spikevax and Comirnaty, at which time B.1.1.519 comprised 60.8% of the
158 sequenced cases. Alaska was ahead of the Lower 48 in terms of vaccine availability and percent

Emergence of B.1.1.519 in Alaska

159 of the population vaccinated in early 2021, which confounds the role host immunity may have had
160 in variant emergence and infection dynamics.

161 *Difference in the emergence of B.1.1.519 within Alaska's two most populated regions*

162 Given Alaska's vast geographic expanse and limited travel options, populations of people
163 tend to interact more frequently within economic regions of the state. In Alaska there are six
164 economic regions defined by the Department of Labor and Workforce Development: the
165 Anchorage-Mat Su, Interior, Gulf Coast, Southeast, Southwest, and Northern regions in order
166 from highest to lowest population. Based on available but limited data, we observed distinct trends
167 in SARS-CoV-2 across economic regions of Alaska (Figure S1). However, in our analysis, we
168 excluded all but the two most populous economic regions, the Anchorage-Mat Su and Interior,
169 because of the low proportion of sequenced cases in the other economic regions of the state. We
170 can draw more robust conclusions about estimates of prevalence because these two regions
171 sequenced greater than 5% of the newly reported cases February through June of 2021 (Figure
172 S2).

173 Between the Anchorage-Mat Su and Interior economic regions there were distinct
174 dynamics in the prevalence of variants over time (Figure 2). For example, Alpha established itself
175 earlier in the Anchorage-Mat Su than the Interior, 360 miles apart. In the Interior, Epsilon
176 comprised a high percentage of sequenced cases from January through March 2021, and the
177 emergence and persistence of B.1.1.519 differed between these two regions of the state.

178 The first case of B.1.1.519 in Alaska was detected from a sample collected 4 February
179 2021 in the Southwest region of the state. In both the Anchorage-Mat Su and Interior regions,
180 B.1.1.519 was detected in specimens with collection dates one day apart from one another, 7 and
181 8 February 2021, respectively. By the end of February, B.1.1.519 comprised 37.8% of sequenced
182 cases in the Anchorage-Mat Su region and 10.5% of cases in the Interior (Figure 2). By February
183 VBMs detected in the Anchorage-Mat Su included Alpha, Epsilon, and Gamma. On the other

Emergence of B.1.1.519 in Alaska

184 hand, in the Interior, Epsilon was the only VBM detected and comprised 38.5% of sequenced
185 cases in February (Figure 2). In the Interior, it appears that Epsilon, which was first detected 6
186 January 2021, was able to establish itself before B.1.1.519. However, one month after B.1.1.519
187 was first detected in the Interior, B.1.1.519 overtook Epsilon. By April, the majority of sequenced
188 cases were assigned to B.1.1.519 in the Interior (77.2%) and the Anchorage-Mat Su (72.6%).
189 After peaking in April, prevalence of B.1.1.519 in May only marginally decreased for the Interior
190 to 74.1% but sharply decreased for the Anchorage-Mat Su region to 38.8% (Figure 2).

191 The sharp decline of B.1.1.519 in the Anchorage-Mat Su region may have occurred earlier
192 than in the Interior due to the emergence of other variants that had established themselves in the
193 Anchorage-Mat Su region well before the Interior. For example, in the Anchorage-Mat Su region,
194 Alpha was detected on 20 December 2020 — 97 days before it was first detected in the Interior.
195 This gave Alpha more time to establish itself within the population and consequently reached
196 42.9% of sequenced cases by May in the Anchorage-Mat Su region versus 9.8% in the Interior.
197 Delta also emerged in the Anchorage-Mat Su region weeks before it was first detected in the
198 Interior, highlighting why B.1.1.519 may have persisted in Interior populations with the last case
199 of B.1.1.519 detected two weeks after last detection in the Anchorage-Mat Su.

200 *High prevalence of B.1.1.519 in Alaska driven by founder effect and mutational advantage*

201 Mutations in the SARS-CoV-2 genomes from Alaskan cases occurred in an almost
202 clockwise fashion (Figure 3B). Although most of these mutations are expected to have neutral
203 effects, some changes can confer selective advantages that enhance fitness by increasing
204 immune evasion, transmissibility, and/or infectivity (Korber et al., 2020). Mutations that cause
205 amino acid changes to the spike (S) glycoprotein may confer selective advantages given the role
206 of the S protein in COVID-19 pathogenesis and tropism. The mature S protein is cleaved into a
207 fusion domain (S2), and a S1 domain containing an N-terminal head and the receptor binding
208 domain (RBD) that binds to host cell human angiotensin converting enzyme-2 (hACE2) receptor.

Emergence of B.1.1.519 in Alaska

209 The S1 domain and RBD particularly are key targets for binding of neutralizing antibodies induced
210 by infection or vaccines, principally IgG (Walls et al., 2020). Given the role of the S protein in
211 receptor binding and subsequent membrane fusion and viral entry into host cells, mutations
212 affecting S1, especially the RBD, can impact hACE2 affinity, viral entry, infectivity, and immune
213 evasion (Berger & Schaffitzel, 2020; Hadi-Alijanvand & Rouhani, 2020).

214 B.1.1.519 has 11 shared amino acid (AA) mutations relative to the original SARS-CoV-2
215 (Wuhan-1) genome, with 4 in S: T478K and D614G in the RBD, P681H in the S1/S2 cleavage
216 site, and T732A in the S2 domain (Rodriguez-Maldonado 2021). T478K also arose in B.1.617.2
217 (Delta) (Jhun, 2021) and may contribute to antibody evasion (Di Giacomo et al., 2021). P681H
218 also occurs in B.1.1.7 (Alpha) and has a weak ability to increase proteolytic cleavage of S1/S2 *in*
219 *vitro* (Lubinskiet et al., 2021). The functional significance of B.1.1.519 harboring these mutations
220 in S is unclear and in a study from the Netherlands there was no evidence of increased viral load
221 conferred by B.1.1.519 (Gorgels et al., 2022).

222 Interestingly, based on the regression of total AA changes over time, which exhibited an
223 increase at a steady rate, B.1.1.519 genomes had fewer total AA changes across the genome
224 than predicted (Figure 3A) but aligned with the number of spike protein AA changes (Figure 3B).
225 B.1.1.519 had significantly more AA substitutions in the spike protein than lineages detected in
226 Alaska prior to the week of B.1.1.519's peak prevalence, 4 April 2021, suggesting a potential
227 competitive advantage over previously circulating lineages (B.1.1.519 = 4.18 ± 0.01 , Pre-peak
228 lineages = 2.65 ± 0.10 S AA substitutions, Wilcoxon $W = 42302$, p-value < $2e-16$; Figure 3C).
229 Genomes sequenced after B.1.1.519's peak prevalence had significantly more AA changes in the
230 spike protein than B.1.1.519, dominated by more divergent Alpha and Delta VOC that emerged
231 after B.1.1.519's peak (Post-peak lineages = 7.75 ± 0.07 S AA substitutions, Wilcoxon $W = 1e+05$,
232 p-value < $2e-16$; Figure 3C). These VOC likely held a selective advantage that allowed them to
233 outcompete B.1.1.519 in Alaska, a phenomenon of variant replacement observed across Alpha
234 and Delta waves in the USA (Lambrou et al., 2022).

Emergence of B.1.1.519 in Alaska

235 B.1.1.519 established itself within the circulating population of SARS-CoV-2 viruses in
236 Alaska weeks before other variants, such as Alpha, were consistently detected (Figure 1B). The
237 difference in prevalence of B.1.1.519 during Alpha's emergence in Alaska versus the Lower 48,
238 paired with the finding of more S protein amino acid changes in this variant than previous lineages
239 in Alaska, suggests that in Alaska B.1.1.519 likely emerged and became dominant due to a
240 founder effect. In this context, the founder effect results in reduction of genetic diversity from a
241 founding lineage of SARS-CoV-2 outcompeting the previously circulating lineages because of a
242 selective advantage that the founding population already possessed. The same trend was likely
243 not possible in the other states because Alpha, which proved to outcompete B.1.1.519, had
244 already established itself within the population, outcompeting B.1.1.519 when it emerged.
245 However, other locations displayed similar patterns to Alaska in terms of B.1.1.519's emergence
246 and spread, most notably Mexico. B.1.1.519 was first detected in Mexico in November 2020,
247 about three months prior to Alaska's first detection, and was found to rapidly outcompete existing
248 variants to become the dominant variant in the country, comprising 51.5% of sequenced genomes
249 by January 2021 (Rodríguez-Maldonado et al., 2020). By February, 90.9% of Mexico City's
250 sequenced cases were B.1.1.519. In a study examining the clinical impacts of B.1.1.519 in Mexico
251 City, Cedro-Tanda (2020) found that B.1.1.519-infected patients had a significant increase in
252 adjusted odds ratio for developing severe symptoms of COVID-19 including dyspnea, chest pain
253 and cyanosis, and hospitalizations compared to non-B.1.1.519 infected patients. Unfortunately,
254 we lacked the availability of the necessary data for Alaska to draw conclusions about the clinical
255 impacts of B.1.1.519 in Alaska as shown in the Mexico-City study. Another study, from the
256 Netherlands, describes a cluster of cases associated with B.1.1.519 demonstrating a high within
257 facility attack rate (Gorgels et al., 2022). Interestingly, that same study did not find differences in
258 B.1.1.519 viral loads compared with other variants, and sample size limited conclusions about
259 disease severity.

Emergence of B.1.1.519 in Alaska

260 *Conclusions*

261 Using genomic data, we identified the unique dynamics of B.1.1.519 spread in Alaska
262 compared to the contiguous United States. The distinct pattern of emergence and spread in
263 Alaska emphasizes both how founder events impact regional differences in circulating SARS-
264 CoV-2 lineages and the significance of having robust sequencing efforts in place to detect this
265 variation. There are several resources that compile SARS-CoV-2 sequence data to help examine
266 patterns of emergence and spread. One such resource, outbreak.info, uses genomic data from
267 the sequence repository GISAID to estimate prevalence, emergence, and spread of SARS-CoV-
268 2 lineages globally (Mullen et al., 2020). When considering data from this source, the highest
269 worldwide prevalence of B.1.1.519 was in early March 2021, a month before Alaska's peak
270 prevalence. At a country level, B.1.1.519 was found to comprise the highest percentage of
271 sequenced genomes in Mexico (15% of sequenced cases compared to 1% of cases sequenced
272 in the United States) over the course of the pandemic. On 14 February 2022, when this analysis
273 was conducted, B.1.1.519 comprised 10.4% of all of Alaska's SARS-CoV-2 sequences to date.
274 While this is high compared to other geographical locations, spatial and temporal heterogeneity
275 in sequencing efforts is known to bias genomic surveillance and complicate comparisons between
276 locations.

277 In Alaska, disparate sequencing coverage affected our ability to generate robust analyses
278 in rural regions of the state where sampling and subsequent sequencing efforts were lower than
279 more populated regions. Despite this disparate coverage, genomic surveillance continues to be
280 an essential tool for informing public health decision-making throughout the state. Genomic
281 surveillance paired with epidemiological findings will be key to detecting new threats as they
282 emerge and is a critical tool in shaping policy to address the clinical impacts of local variation in
283 circulating virus lineages. At a patient level, variant surveillance can inform which monoclonal
284 antibodies will be most effective for early treatment to directly reduce individual mortality and

Emergence of B.1.1.519 in Alaska

285 morbidity. At a community level, variant proportion data paired with epidemiological findings can
286 inform the broader population of their risks of reinfection, evaluate vaccine efficacy, and assess
287 the need for appropriate treatment stockpiles and quantities. This information will be of great
288 importance for public health policy decisions as SARS-CoV-2 becomes endemic in Alaska and
289 around the globe.

290

291

292 **Data Availability Statement:** All data used in this study are available online at GISAID
293 (<https://www.gisaid.org/>) and NCBI GenBank (<https://www.ncbi.nlm.nih.gov/sars-cov-2/>).

294 Funding

295 Work presented here was supported by Epidemiology and Laboratory Capacity for Prevention
296 and Control of Emerging Infectious Diseases (ELC) from the Centers for Disease Control and
297 Prevention and a supplement award to Alaska INBRE, an Institutional Development Award (IDeA)
298 from the National Institute of General Medical Sciences of the National Institutes of Health under
299 grant number 2P20GM103395.

300

301 Acknowledgments

302 We gratefully acknowledge the researchers responsible for obtaining the specimens and the
303 laboratories where genetic sequence data were generated and shared via the GISAID Initiative
304 and NCBI GenBank, on which this research is based.

305

Emergence of B.1.1.519 in Alaska

306 References.

- 307 Aksamentov, I., Roemer, C., Hodcroft, E. B., & Neher, R. A., (2021). Nextclade: clade
308 assignment, mutation calling and quality control for viral genomes. *Journal of Open*
309 *Source Software*, 6(67), 3773.
- 310 Alaska Department of Labor and Workforce Development, *2020 Population Estimates by*
311 *Borough, Census Area, and Economic Region*. Nov 5, 2021.
312 <https://live.laborstats.alaska.gov/pop/index.cfm>
- 313 Alaska Department of Health and Social Services (Alaska DHSS) *COVID-19 CSV Files*
314 *Database*. Nov 5, 2021.
315 <https://experience.arcgis.com/experience/b2f5a105a41c4ca88024efbc2c2e868f/>
- 316 Berger, I., & Schaffitzel, C. (2020). The SARS-CoV-2 spike protein: balancing stability and
317 infectivity. *Cell research*, 30(12), 1059-1060.
- 318 Centers for Disease Control and Prevention (CDC) *United States COVID-19 Cases, Deaths,*
319 *and Laboratory Testing (NAATs) by State, Territory, and Jurisdiction*. Nov 5, 2021.
320 https://covid.cdc.gov/covid-data-tracker/#cases_community
- 321 Cedro-Tanda, A.; Gómez-Romero, L.; Alcaraz, N.; de Anda-Jauregui, G.; Peñaloza, F.; Moreno,
322 B.; Escobar-Arrazola, M.A.; Ramirez-Vega, O.A.; Munguia-Garza, P.; Garcia-Cardenas,
323 F.; Cisneros-Villanueva, M.; Moreno-Camacho, J.L.; Rodriguez-Gallegos, J.; Luna-Ruiz
324 Esparza, M.A.; Fernández Rojas, M.A.; Mendoza-Vargas, A.; Reyes-Grajeda, J.P.;
325 Campos-Romero, A.; Angulo, O.; Ruiz, R.; Sheinbaum-Pardo, C.; Sifuentes-Osornio, J.;
326 Kershenovich, D.; Hidalgo-Miranda, A.; Herrera, L.A. (2021). The Evolutionary
327 Landscape of SARS-CoV-2 Variant B.1.1.519 and Its Clinical Impact in Mexico City.
328 *Viruses*, 13, 2182.
- 329 Di Giacomo, S., Mercatelli, D., Rakhimov, A., & Giorgi, F. M. (2021). Preliminary report on
330 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K.
331 *Journal of medical virology*, 93(9), 5638-5643.
- 332 Elbe, S., & Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative
333 contribution to global health. *Global challenges (Hoboken, NJ)*, 1(1), 33–46.
- 334 Gorgels, K. M., Dingemans, J., van der Veer, B. M., Hackert, V., Hensels, A. Y., den Heijer, C.
335 D., den Heijer, C.D., van Alphen, L.B., Savelkoul, P.H., & Hoebe, C. J. (2022). Linked
336 nosocomial COVID-19 outbreak in three facilities for people with intellectual and
337 developmental disabilities due to SARS-CoV-2 variant B. 1.1. 519 with spike mutation
338 T478K in the Netherlands. *BMC infectious diseases*, 22(1), 1-9.
- 339 Hadi-Alijanvand, H., & Rouhani, M. (2020). Studying the effects of ACE2 mutations on the
340 stability, dynamics, and dissociation process of SARS-CoV-2 S1/hACE2 complexes.
341 *Journal of proteome research*, 19(11), 4609-4623.
- 342 John Hopkins Coronavirus Resource Center. *Impact of Opening and Closing Decisions by*
343 *State*. March 9, 2022. [https://coronavirus.jhu.edu/data/state-timeline/new-confirmed-](https://coronavirus.jhu.edu/data/state-timeline/new-confirmed-cases/alaska/)
344 [cases/alaska/](https://coronavirus.jhu.edu/data/state-timeline/new-confirmed-cases/alaska/)
- 345 Jhun, H., Park, H. Y., Hisham, Y., Song, C. S., & Kim, S. (2021). SARS-CoV-2 Delta (B. 1.617.
346 2) variant: a unique T478K mutation in receptor binding motif (RBM) of spike gene.
347 *Immune Network*, 21(5).

Emergence of B.1.1.519 in Alaska

- 348 Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., ... & Montefiori,
349 D. C. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases
350 infectivity of the COVID-19 virus. *Cell*, *182*(4), 812-827.
- 351 Lambrou, A. S., Shirk, P., Steele, M. K., Paul, P., Paden, C. R., Cadwell, B., Reese, H.E., Aoki,
352 Y., Hassell, N., Zheng, X.Y., & Emerging Variants NS3 Working Group. (2022). Genomic
353 Surveillance for SARS-CoV-2 Variants: Predominance of the Delta (B. 1.617. 2) and
354 Omicron (B. 1.1. 529) Variants—United States, June 2021–January 2022. *Morbidity and
355 Mortality Weekly Report*, *71*(6), 206.
- 356 Lubinski, B., Fernandes, M. H., Frazier, L., Tang, T., Daniel, S., Diel, D. G., Jaimes, J.A., &
357 Whittaker, G. R. (2022). Functional evaluation of the P681H mutation on the proteolytic
358 activation of the SARS-CoV-2 variant B. 1.1. 7 (Alpha) spike. *Iscience*, *25*(1), 103589.
- 359 Mullen, J.L., Tsueng, G., Latif, A.A., Alkuzweny, M., Cano, M., Haag, E., Zhou, J., Zeller, M.,
360 Hufbauer, E., Matteson, N., Andersen, K.G., Wu, C., Su, A.I., Gangavarapu, K., Hughes,
361 L.D. (2020) Center for Viral Systems Biology *outbreak.info*. Available online:
362 <https://outbreak.info/>
- 363 O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., & Rambaut, A.
364 (2021). Assignment of epidemiological lineages in an emerging pandemic using the
365 pangolin tool. *Virus Evolution*, *7*(2), veab064.
- 366 Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., & Ruis, C. A dynamic
367 nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat
368 Microbiol.* 2020 Jul 15; 1–5.
- 369 Rodríguez-Maldonado, A. P., Vázquez-Pérez, J. A., Cedro-Tanda, A., Taboada, B., Boukadida,
370 C., Wong-Arámbula, C., Nuñez-García, T. E., Cruz-Ortiz, N., Barrera-Badillo, G.,
371 Hernández-Rivas, L., López-Martínez, I., Mendoza-Vargas, A., Reyes-Grajeda, J. P.,
372 Alcaraz, N., Peñalosa-Figueroa, F., Gonzalez-Barrera, D., Rangel-DeLeon, D., Herrera-
373 Montalvo, L. A., Mejía-Nepomuceno, F., Hernández-Terán, A., ... Ramírez-González, J.
374 E. (2021). Emergence and spread of the potential variant of interest (VOI) B.1.1.519 of
375 SARS-CoV-2 predominantly present in Mexico. *Archives of virology*, *166*(11), 3173–
376 3177.
- 377 Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data—from
378 vision to reality. *Eurosurveillance*, *22*(13), 30494.
- 379 United States Census Bureau. *QuickFacts Alaska*. July 1, 2021.
380 <https://www.census.gov/quickfacts/fact/table/AK#>
- 381 Walensky, R. P., Walke, H. T., & Fauci, A. S. (2021). SARS-CoV-2 variants of concern in the
382 United States—Challenges and opportunities. *Jama*, *325*(11), 1037-1038.
- 383 World Health Organization (WHO) Tracking SARS-CoV-2 Variants. Nov 5, 2021.
384 <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants>
- 385 Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Velesler, D. (2020).
386 Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, *181*(2),
387 281-292.

Emergence of B.1.1.519 in Alaska

Figure Legends

Figure 1. (A and C) Daily case count (bars) and seven day rolling average (red line) and (B and D) percent of sequences (estimated prevalence) by PANGO lineages detected by week from 2020-11-29 to 2021-06-26 for Alaska (A and B) and the Lower 48 (C and D).

Figure 2. Percent of sequenced cases (estimation of prevalence) by PANGO lineage by month for the (A) Anchorage-Mat Su and (B) Interior regions of Alaska from December 2020 through June 2021.

Figure 3. Scatterplots depicting the A) S protein amino acid (AA) substitutions and the (B) total AA substitutions over time. Data includes each genome from an Alaskan confirmed case, colored by variant. The gray line shows a generalized linear model regression of data.(C) 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 .

Emergence of B.1.1.519 in Alaska

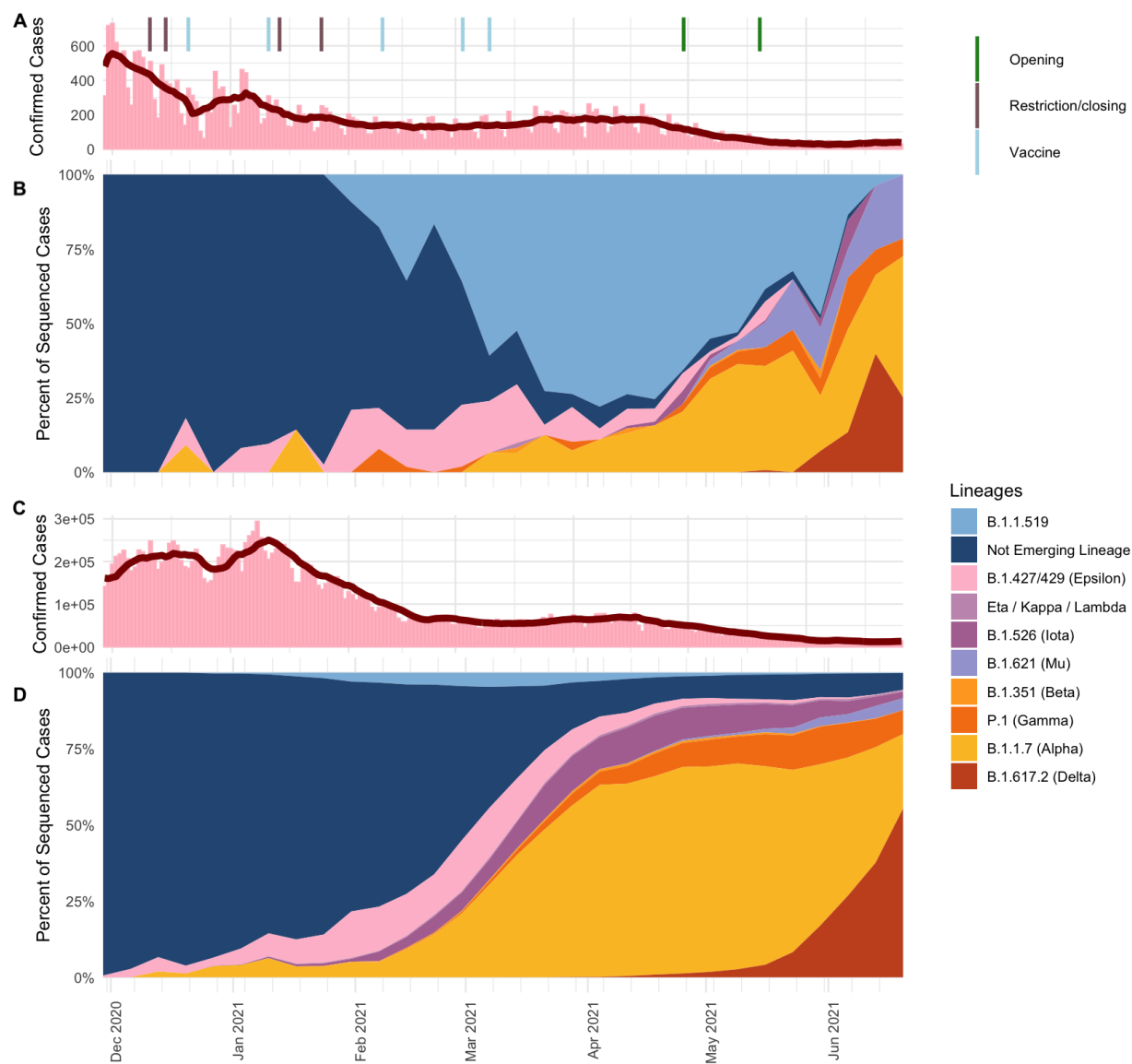


Figure 1. (A and C) Daily case count (bars) and seven day rolling average (red line) and (B and D) percent of sequences (estimated prevalence) by PANGO lineages detected by week from 2020-11-29 to 2021-06-26 for Alaska (A and B) and the Lower 48 (C and D).

Emergence of B.1.1.519 in Alaska

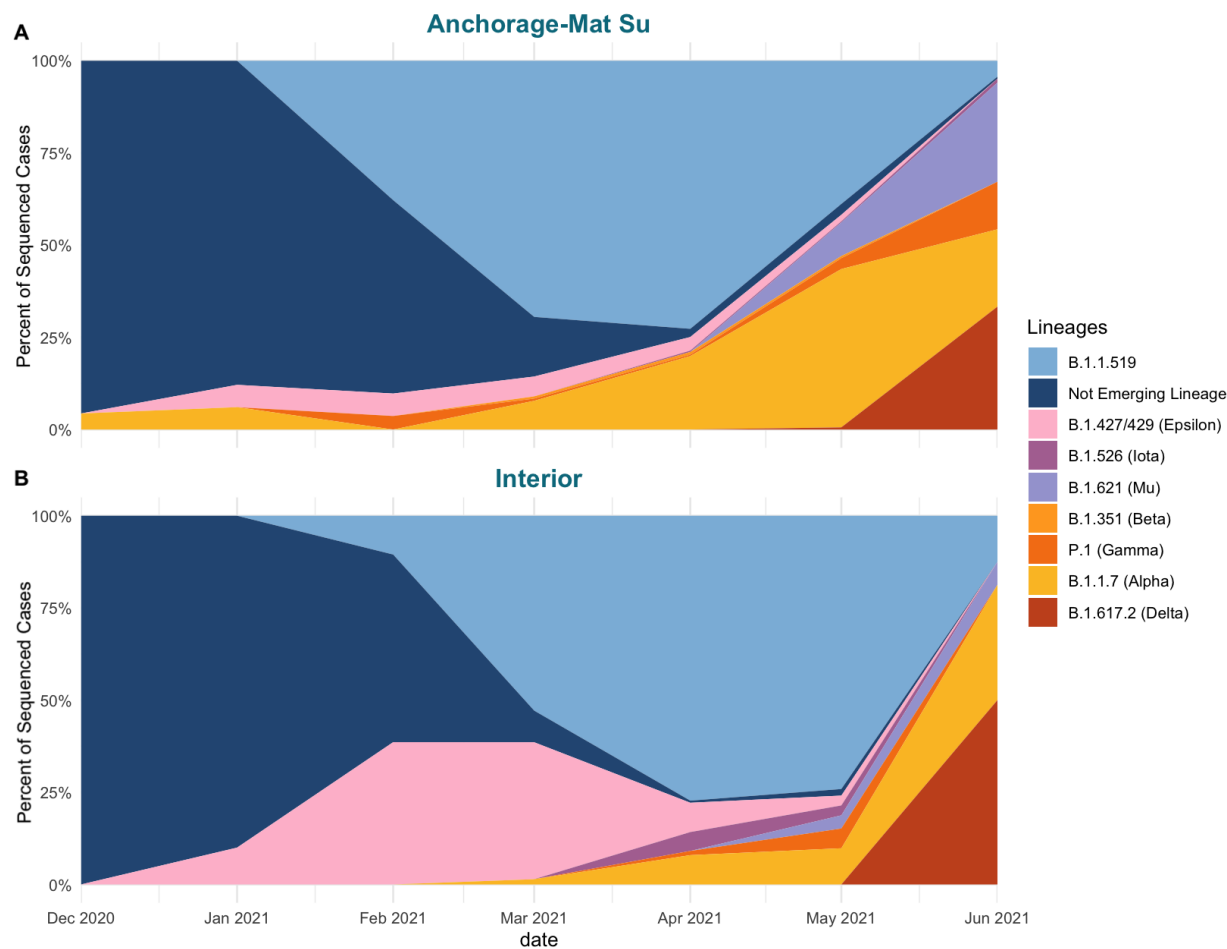


Figure 2. Percent of sequenced cases (estimation of prevalence) by PANGO lineage by month for the (A) Anchorage-Mat Su and (B) Interior regions of Alaska from December 2020 through June 2021.

Emergence of B.1.1.519 in Alaska

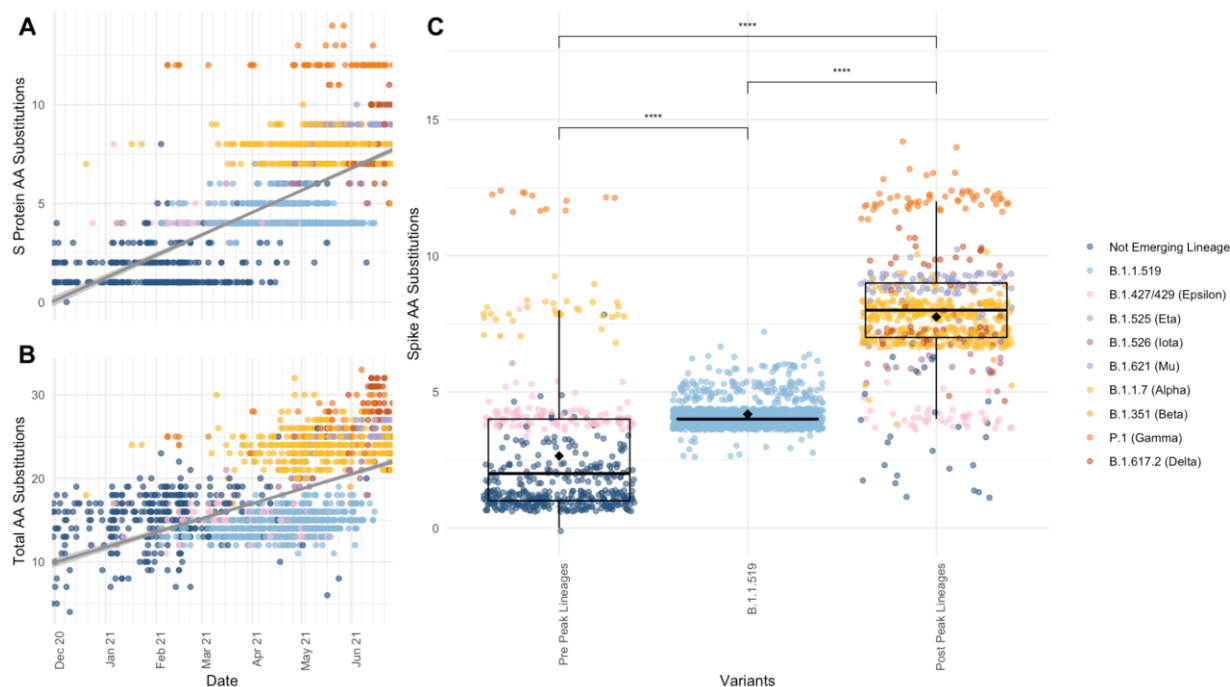


Figure 3. Scatterplots depicting the (A) S protein amino acid (AA) substitutions and the (B) total AA substitutions over time. Data includes each genome from an Alaskan confirmed case, colored by variant. The gray line shows a generalized linear model regression of data. (C) Boxplot of number of S protein 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 .

Emergence of B.1.1.519 in Alaska

Supplemental Figures.

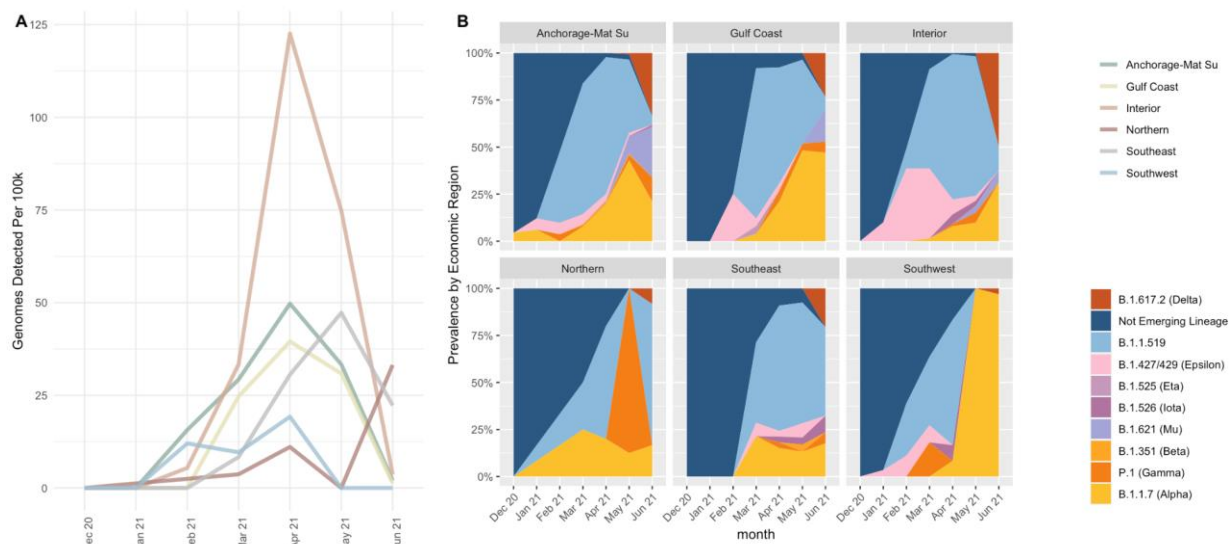


Figure S1. (A) Number of detected B.1.1.519 genomes per 100k people by Alaska economic region over time. Lines are colored by Alaska economic region. (B) Estimated prevalence of emerging lineages by Alaska economic region from December 2020 through June 2021.

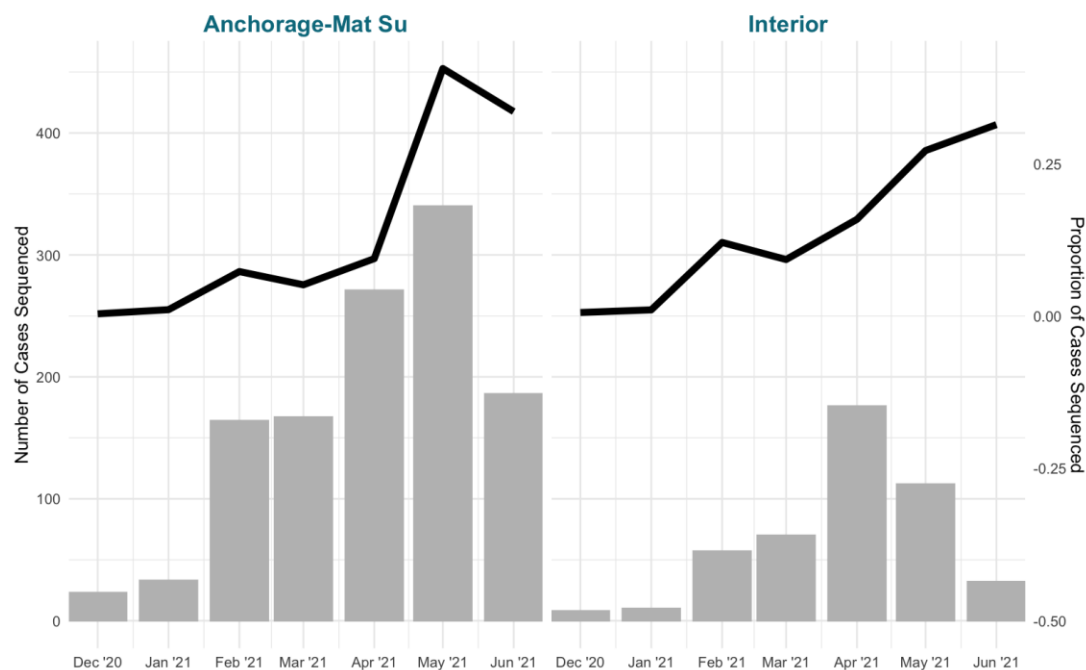


Figure S2. Number of cases sequenced (bars) and the proportion of total cases sequenced in the Anchorage-Mat Su and Interior economic regions of Alaska.