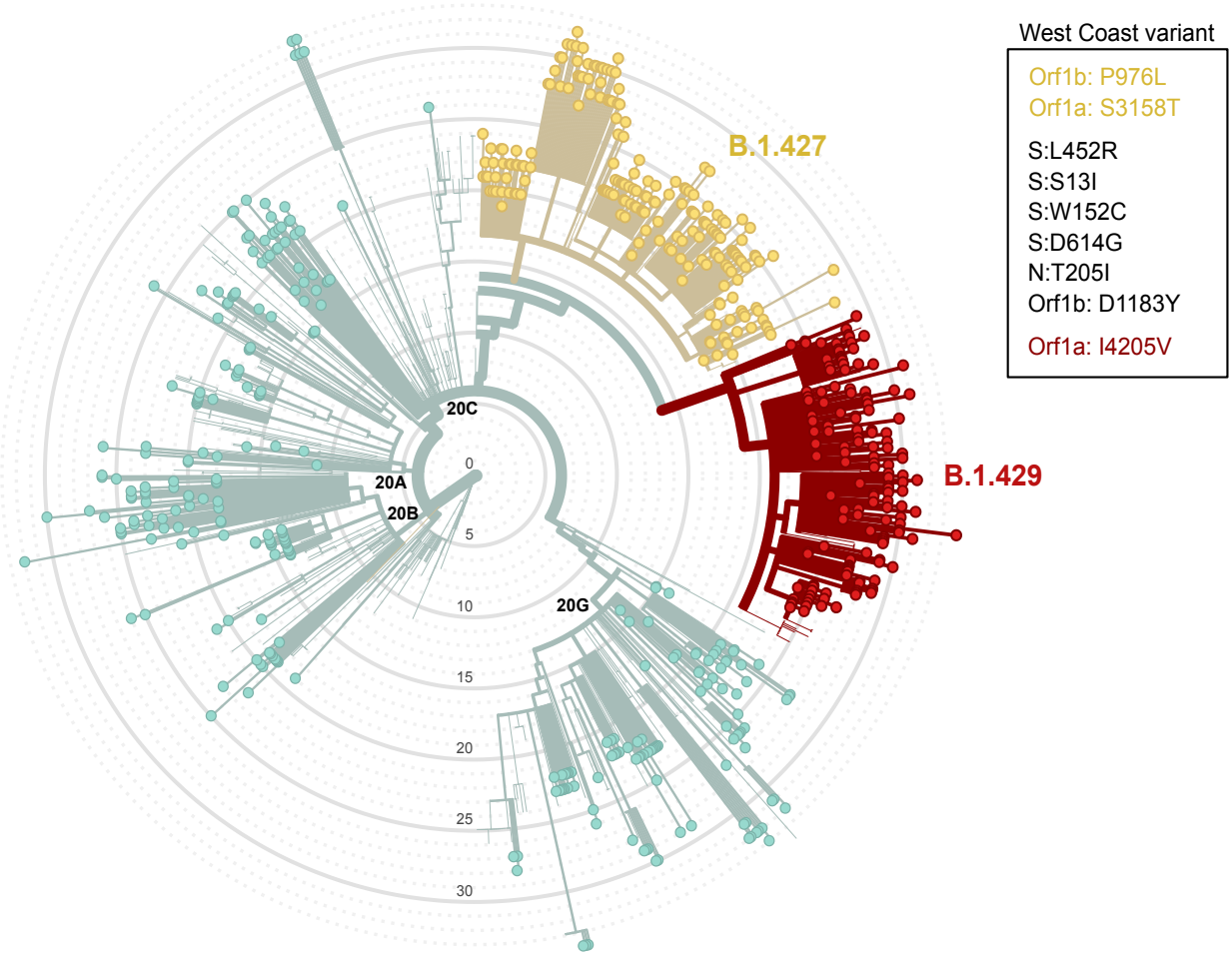
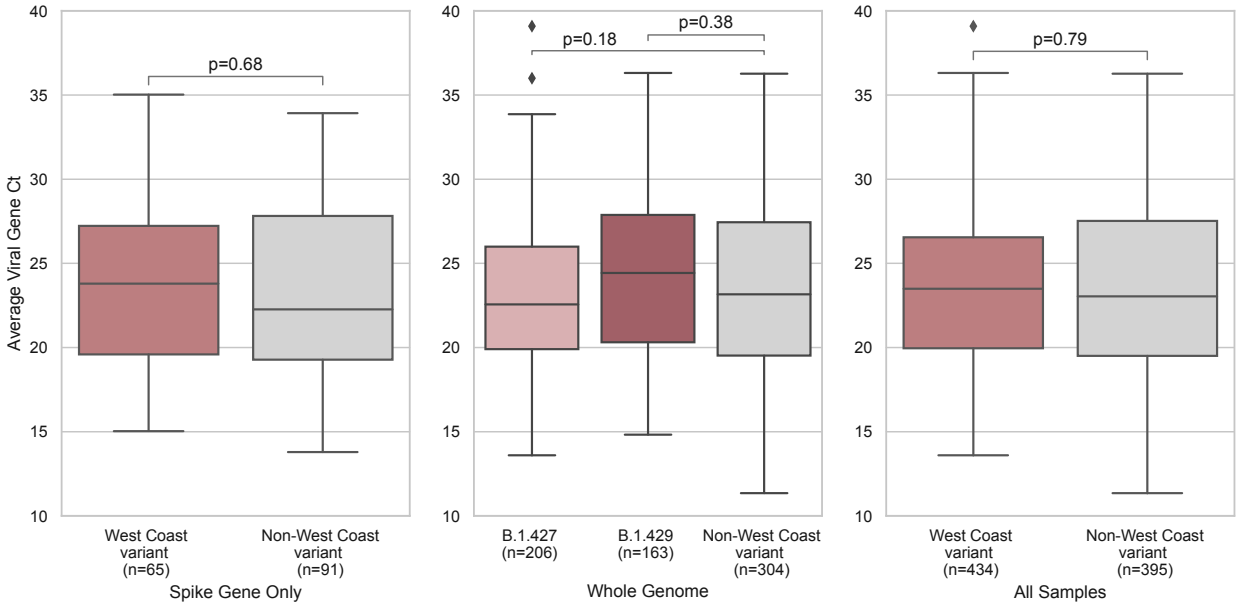
medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Supplemental Figure 1



medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Supplemental Figure 2



medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Table 1: Characteristics of all persons (N=8,822; 10.0% with positive result) tested at the community-based testing site between January 10-29.

	Negative (N=7937)	Positive (N=885)	Total (N=8822)
Gender			
Female	3595 (90.2%)	391 (9.8%)	3986 (100%)
Male	4190 (89.7%)	481 (10.3%)	4671 (100%)
Other	152 (92.1%)	13 (7.9%)	165 (100%)
Age Group			
Age ≤ 12	488 (83.1%)	99 (16.9%)	587 (100%)
Age 13-35	3032 (88.8%)	382 (11.2%)	3414 (100%)
Age 36-64	3844 (91.5%)	355 (8.5%)	4199 (100%)
Age ≥ 65	573 (92.1%)	49 (7.9%)	622 (100%)
Race/Ethnicity			
Hispanic/Latinx	5700 (88.5%)	744 (11.5%)	6444 (100%)
Asian	610 (93.7%)	41 (6.3%)	651 (100%)
White/Caucasian	818 (96.3%)	31 (3.7%)	849 (100%)
Black or African American	228 (96.6%)	8 (3.4%)	236 (100%)
Other	581 (90.5%)	61 (9.5%)	642 (100%)
Occupation*			
Food & Beverage	987 (90.5%)	104 (9.5%)	1091 (100%)
Tradesperson	423 (87%)	63 (13%)	486 (100%)
Day laborer	227 (89.4%)	27 (10.6%)	254 (100%)
Healthcare	277 (93.9%)	18 (6.1%)	295 (100%)
Student	780 (84.4%)	144 (15.6%)	924 (100%)
Other	2557 (92.4%)	211 (7.6%)	2768 (100%)
No employment	1126 (88.8%)	142 (11.2%)	1268 (100%)
Day of Test Symptoms			
Asymptomatic	6089 (94.6%)	347 (5.4%)	6436 (100%)
Symptomatic	1848 (77.5%)	538 (22.5%)	2386 (100%)

* Occupation information missing from 1,736 persons.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Table 2: Sample sequencing summary and viral cycle threshold characteristics.

		Spike Gene Only	**Whole Genome**		**Total Count (Jan 10-Jan 30)**
Total Samples Run		165	674		839
Total High Quality (>92% coverage)		96	620		716
Average Reads/Sample		96,237	2,437,080		-
Average % Gene/Genome Covered		83.35	97.09		94.39
Mean Viral Ct		23.78	23.63		23.65
Mean Viral Ct West Coast variant	B.1.427	24.00	23.60	22.96	23.62
	B.1.429			24.40	
Mean Viral Ct Non-West Coast variant		23.63	23.76		23.67
Median Viral Ct		23.65	23.26		23.26
Median Viral Ct West Coast variant	B.1.427	24.00	23.42	22.56	23.49
	B.1.429			24.42	
Median Viral Ct Non-West Coast variant		22.83	23.16		23.04
Viral Ct Interquartile Range (IQR)		8.29	7.16		7.42
Viral Ct IQR West Coast variant	B.1.427	7.39	6.40	6.09	6.59
	B.1.429			7.57	
Viral Ct IQR Non-West Coast variant		8.88	7.93		8.01

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Table 3: Amino acid substitutions observed in the spike gene and count of sequences per mutation in each study.

Spike domain	Position in S gene	Mutation	Count Nov. 23-Dec. 1 2020 (n=191)	Count Jan. 10-Jan.29 2021 (n=716)	Total count (n=907)
	5	L5F	4	4	**8**
	8	L8W	0	3	**3**
	9	P9L	0	3	**3**
N-terminal domain	13	S13I	30	381	**411**
	18	L18F	0	2	**2**
	19	T19I	0	3	**3**
	21	R21T	0	2	**2**
	22	T22I	0	9	**9**
	26	P26S	9	46	**55**
	26	P26L	2	0	**2**
	27	A27V	0	1	**1**
	49	H49Y	0	4	**4**
	67	A67V	2	15	**17**
	69	H69Q	1	6	**7**
	70	V70F	3	0	**3**
	76	T76I	0	1	**1**
	78	R78K	0	1	**1**
	95	T95I	7	4	**11**
	111	D111N	0	2	**2**
	138	D138Y	1	2	**3**
	141	L141F	0	1	**1**
	142	G142-	0	1	**1**
	143	V143F	0	1	**1**
	143	V143-	0	1	**1**
	152	W152C	30	354	**384**
	153	M153I	1	2	**3**
	155	S155N	0	4	**4**
	173	Q173K	0	1	**1**
	178	D178G	4	2	**6**
	185	N185T	0	1	**1**
	215	D215Y	7	1	**8**
	222	A222V	0	1	**1**
	251	P251S	1	1	**2**
	252	G252V	0	2	**2**
	254	S254F	1	0	**1**
	257	G257S	0	3	**3**
	258	W258L	0	14	**14**
	261	G261V	0	1	**1**
Receptor binding domain	357	R357K	0	7	**7**
	367	V367I	7	3	**10**
	372	A372V	0	1	**1**
	449	Y449-	0	1	**1**
	452	L452R	30	381	**411**
	475	A475S	0	1	**1**
	478	T478K	0	26	**26**
	484	E484K	0	1	**1**
	485	G485V	0	1	**1**
	494	S494P	0	6	**6**
	520	A520S	1	1	**2**
	574	D574Y	3	0	**3**

Spike domain	Position in S gene	Mutation	Count Nov. 23-Dec. 1 2020 (n=191)	Count Jan. 10-Jan.29 2021 (n=716)	Total count (n=907)
	583	E583D	5	11	**16**
	613	Q613H	7	4	**11**
	614	D614G	191	716	**907**
	622	V622I	0	1	**1**
	657	N657D	1	0	**1**
Adjacent to S1/S2 junction	675	Q675H	2	5	**7**
	677	Q677P	0	10	**10**
	677	Q677H	11	10	**21**
	679	N679K	0	1	**1**
	681	P681H	0	29	**29**
	681	P681R	1	0	**1**
	688	A688V	5	8	**13**
	701	A701S	0	1	**1**
	716	T716I	0	1	**1**
	719	T719I	0	3	**3**
	732	T732A	7	52	**59**
	769	G769V	3	4	**7**
	771	A771D	0	2	**2**
	779	Q779H	0	2	**2**
	780	E780Q	0	1	**1**
	792	P792L	0	1	**1**
	812	P812S	0	1	**1**
	813	S813G	0	2	**2**
Fusion peptide	818	I818V	0	1	**1**
	845	A845S	0	3	**3**
	846	A846V	0	1	**1**
	879	A879S	2	0	**2**
Heptad repeat 1	936	D936H	3	1	**4**
	936	D936Y	1	0	**1**
	938	L938F	2	25	**27**
	954	Q954K	0	2	**2**
	1026	A1026S	1	0	**1**
	1063	L1063F	0	1	**1**
	1072	E1072K	0	10	**10**
	1078	A1078S	0	3	**3**
	1078	A1078V	0	1	**1**
	1103	F1103L	1	0	**1**
	1121	F1121V	1	0	**1**
Heptad repeat 2	1163	D1163V	0	2	**2**
	1168	D1168G	1	0	**1**
	1171	G1171S	0	2	**2**
	1176	V1176F	0	1	**1**
	1191	K1191N	9	4	**13**
	1195	E1195Q	0	1	**1**
	1201	Q1201K	0	2	**2**
Trans-membrane domain	1228	V1228L	0	1	**1**
	1235	C1235F	5	4	**9**
	1244	L1244F	1	0	**1**
	1252	S1252P	1	0	**1**
	1252	S1252F	0	1	**1**

Co-variant group key

- B.1.243
- Q677P
- Q677P
- P681H
- B.1.427
- B.1.429
- P.2

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Table 4: Individual characteristics of all persons tested (both positive and negative, and including index case) living in one of the 318 households meeting inclusion criteria for household secondary attack rate analyses, stratified by strain classification of the household.

	Non-West coast (N=571)	West coast			Total (N=1184)
		B.1.427 (N=284)	B.1.429 (N=246)	All West coast (N=613)	
Sex					
Female	275 (48.2%)	123 (43.3%)	120 (48.8%)	286 (46.7%)	561 (47.4%)
Male	288 (50.4%)	156 (54.9%)	124 (50.4%)	320 (52.2%)	608 (51.4%)
Other	8 (1.4%)	5 (1.8%)	2 (0.8%)	7 (1.1%)	15 (1.3%)
Age Group					
Age <= 12	78 (13.7%)	24 (8.5%)	42 (17.1%)	79 (12.9%)	157 (13.3%)
Age 13-35	253 (44.3%)	117 (41.2%)	105 (42.7%)	257 (41.9%)	510 (43.1%)
Age 36-64	216 (37.8%)	126 (44.4%)	90 (36.6%)	245 (40.0%)	461 (38.9%)
Age >= 65	24 (4.2%)	17 (6.0%)	9 (3.7%)	32 (5.2%)	56 (4.7%)
Race/Ethnicity					
Hispanic/Latinx	511 (89.5%)	233 (82.0%)	217 (88.2%)	518 (84.5%)	1029 (86.9%)
Asian	17 (3.0%)	10 (3.5%)	12 (4.9%)	28 (4.6%)	45 (3.8%)
White/Caucasian	16 (2.8%)	13 (4.6%)	3 (1.2%)	18 (2.9%)	34 (2.9%)
Black or African American	4 (0.7%)	7 (2.5%)	1 (0.4%)	12 (2.0%)	16 (1.4%)
Other	23 (4.0%)	21 (7.4%)	13 (5.3%)	37 (6.0%)	60 (5.1%)
Occupation*					
Food & Beverage	68 (11.9%)	48 (16.9%)	23 (9.3%)	80 (13.1%)	148 (12.5%)
Tradesperson	35 (6.1%)	16 (5.6%)	4 (1.6%)	24 (3.9%)	59 (5.0%)
Day laborer	19 (3.3%)	8 (2.8%)	9 (3.7%)	18 (2.9%)	37 (3.1%)
Healthcare	13 (2.3%)	7 (2.5%)	2 (0.8%)	11 (1.8%)	24 (2.0%)
Student	110 (19.3%)	34 (12.0%)	62 (25.2%)	110 (17.9%)	220 (18.6%)
Other	133 (23.3%)	70 (24.6%)	62 (25.2%)	148 (24.1%)	281 (23.7%)
No employment	86 (15.1%)	55 (19.4%)	43 (17.5%)	115 (18.8%)	201 (17.0%)

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Day of Test Symptoms					
Asymptomatic	320 (56.0%)	166 (58.5%)	131 (53.3%)	348 (56.8%)	668 (56.4%)
Symptomatic	251 (44.0%)	118 (41.5%)	115 (46.7%)	265 (43.2%)	516 (43.6%)

* Occupation information missing for 214 persons.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Table 5: Adjusted attack rates from sensitivity analysis using Targeted Maximum Likelihood and Super Learning.

	Adjusted	
	aRR (95% CI)	**p-value**
Class		
Non-West Coast	-	-
West Coast	1.17 (0.93-1.49)	0.188
Lineage		
B.1.427	1.17 (0.87-1.56)	0.294
B.1.429	1.34 (1.01-1.78)	0.045

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

SUPPLEMENTARY METHODS

SARS-CoV-2 genomic sequence recovery

Swab samples from individuals testing positive by BinaxNOW were placed in DNA/RNA shield and processed as previously described [1]. Extracted total nucleic acid was diluted based on average SARS-CoV-2 N and E gene cycle threshold (Ct) values; samples with a Ct range 12-15 were diluted 1:100, 15-18 1:10 and >18 no dilution. For high throughput scaling, library preparation reaction volumes and dilutions were miniaturized utilizing acoustic liquid handling. Library preparation followed either the modified versions of the Primal-Seq Nextera XT version 2.0 protocol [2,3], or the modified version of SARS-CoV-2 Tailed Amplicon Illumina Sequencing V.2 [4], both using the ARTIC Network V3 primers [5]. A subset of initial samples were library prepared using the Tailed Amplicon Sequencing V.2 with only primer pairs 71-84 of the ARTIC V3 primers to tile all of the S gene. Final libraries were sequenced by paired-end 2 x 150bp sequencing on an Illumina NovaSeq platform, or for the S-gene-only set, 2 x 300bp on an Illumina MiSeq.

SARS-CoV-2 consensus genome generation

Raw .fastq files were imported into IDseq and consensus genomes were generated automatically using the embedded SARS-CoV-2 pipeline [6]. Specifically, minimap2 was used to align raw reads to the reference genome MN908947.2 [7], then the consensus sequence was generated using samtools [8], mpileup and ivar [9]. The IDseq consensus genome pipeline is implemented in WDL [10]. Viral genomes with at least 92% (27,500nt) recovery were uploaded to GISAID [11], a worldwide repository for SARS-CoV-2 genomes and Genbank [12].

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Phylogenetic analysis was conducted and results were visualized in Nextclade (https://clades.nextstrain.org) [13].

Bayesian Phylogenetic Analysis

To compare the viral diversity of the two variants of interest, B.1.427 and B.1.429, we identified two other SARS-CoV-2 lineages, B.1.232 and B.1.243 that were prevalent in the state of California from July 2020 onwards. Both of these lineages contained more sequences on GISAID from California than any other location, based on the PANGO lineage assignment [14] on GISAID. For each of the 4 lineages spreading in California, we randomly subsampled all available genomes from GISAID and this study to 500 or fewer genomes. For subsampled genomes, we aligned them against the reference genome (Genbank accession: MN996528.1) using MAFFT v7.471 [15] with default settings. Each multi-sequence alignment was used to build a separate maximum likelihood tree in IQ-TREE v.1.6.12 [16] with default options. The trees were rooted at the reference genome. The maximum likelihood tree was used to visually identify outlier sequences which could have been misclassified as that PANGO lineage. The resulting number of genomes included in the downstream analysis were: B.1.232: 368; B.1.243: 500; B.1.427: 495; B.1.429: 443. The multi-sequence alignment of the coding region for each lineage was analyzed in BEAST v.1.10.4 [17] with unlinked molecular clocks between the S gene and other genes, uncorrelated relaxed molecular clock (lognormal distribution), GTR substitution model with 4 rate categories (selected by the BIC value in ModelTest [18]), and the Bayesian Skygrid population model [19]. Default prior values and operator values were used. The MCMC chains were 100M in length, sampled every 50,000, and the first 50% of samples

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

discarded as burnin. We ran 2 replicate MCMC chains for each analysis, and used all samples to summarize the results.

We fit an exponential model to the median SkyGrid estimate between 2020-07-01 and 2021-01-01 to calculate the growth rate per day *r*, and calculated the reproductive number *R* using the formula [20] $R = (1+r/b)^a$, where a=1.39 is the shape parameter and b=0.14 is the scale parameter of a gamma generation time distribution with a mean of 5 days a standard deviation of 1.9 days [21]. The 95% confidence interval (95% CI) around *R* was calculated based on the 95% CI around the *r* estimate. Samples used for the Bayesian Analysis are detailed in the supplementary file.

REFERENCES

1. Crawford ED, Acosta I, Ahyong V, et al. Rapid deployment of SARS-CoV-2 testing: The CLIAHUB. PLOS Pathog **2020**; 16:e1008966.
2. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc **2017**; 12:1261–1276.
3. R&d DP. COVID-19 ARTIC v3 Illumina library construction and sequencing protocol. **2020**; Available at: https://www.protocols.io/view/covid-19-artic-v3-illumina-library-construction-an-bibtkann. Accessed 23 February 2021.
4. Gohl DM, Garbe J, Grady P, et al. A rapid, cost-effective tailed amplicon method for sequencing SARS-CoV-2. BMC Genomics **2020**; 21:863.
5. artic-network/artic-ncov2019. Available at: https://github.com/artic-network/artic-ncov2019. Accessed 25 February 2021.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

6. Kalantar KL, Carvalho T, de Bourcy CFA, et al. IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. GigaScience **2020**; 9.

7. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinforma Oxf Engl **2018**; 34:3094–3100.

8. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl **2009**; 25:2078–2079.

9. Grubaugh ND, Gangavarapu K, Quick J, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol **2019**; 20:8.

10. chanzuckerberg/idseq-workflows. Chan Zuckerberg Initiative, 2021. Available at: https://github.com/chanzuckerberg/idseq-workflows. Accessed 25 February 2021.

11. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall Hoboken NJ **2017**; 1:33–46.

12. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res **2016**; 44:D67-72.

13. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinforma Oxf Engl **2018**; 34:4121–4123.

14. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol **2020**; 5:1403–1407.

15. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **2013**; 30:772–780.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.01.21252705; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

16. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**; 32:268–274.

17. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol **2018**; 4:vey016.

18. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinforma Oxf Engl **1998**; 14:817–818.

19. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol **2013**; 30:713–724.

20. Fraser C, Donnelly CA, Cauchemez S, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. Science **2009**; 324:1557–1561.

21. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science **2020**; 368.