

SARS-CoV-2 sequencing report

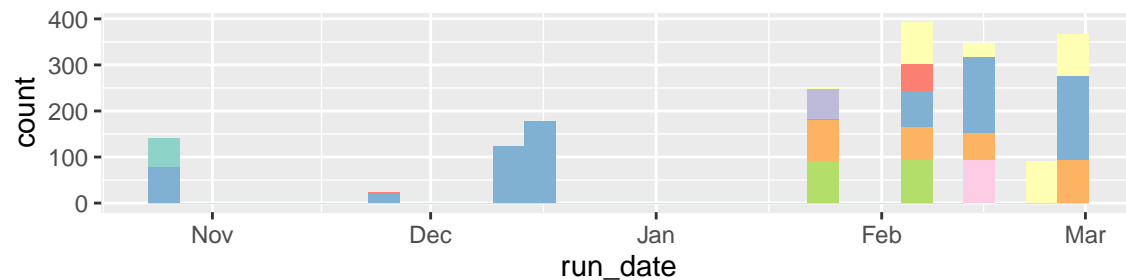
The Broad Institute Viral Genomics group, in partnership with the Genomics Platform and Data Sciences Platform, has been engaged in viral sequencing of COVID-19 patients since March 2020.

This report, generated on 2021-03-10, summarizes our sequencing activity for **all patient samples** across all collaborators and states. The samples included in this report were sequenced between 2020-10-25 and 2021-02-27. This report is current as of the epiweek ending **2021-02-27**.

Weekly summary

	week ending 2021-02-27	cumulative total
samples sequenced	474	2037
genomes assembled	378	1415
genomes submittable	340	1270
Variants of Concern	13	21

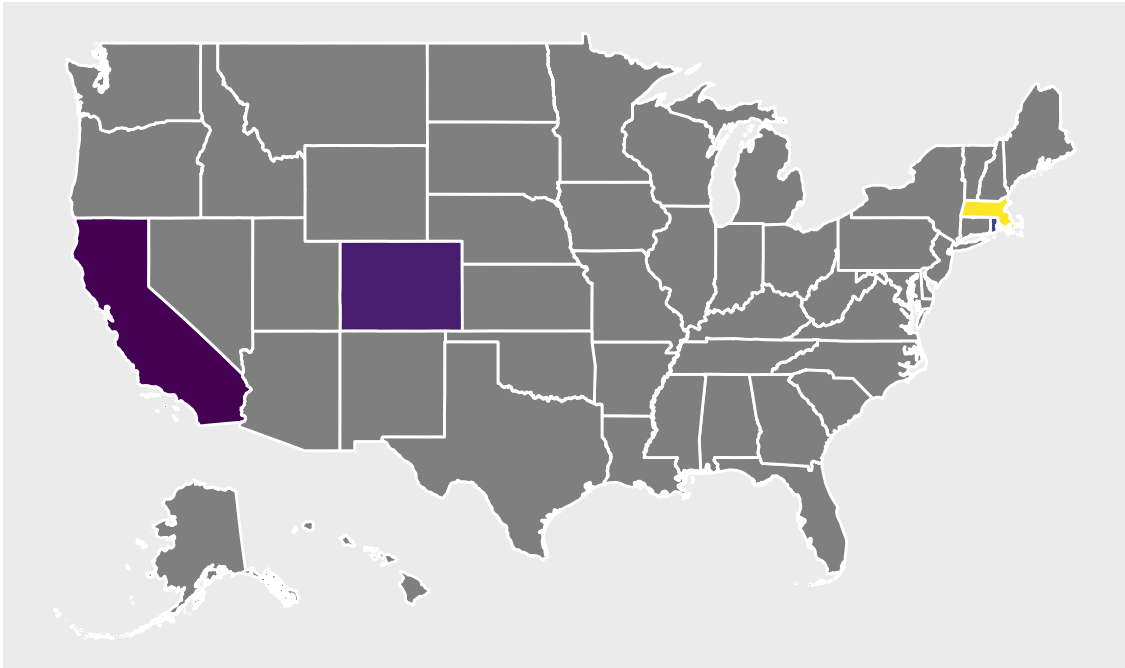
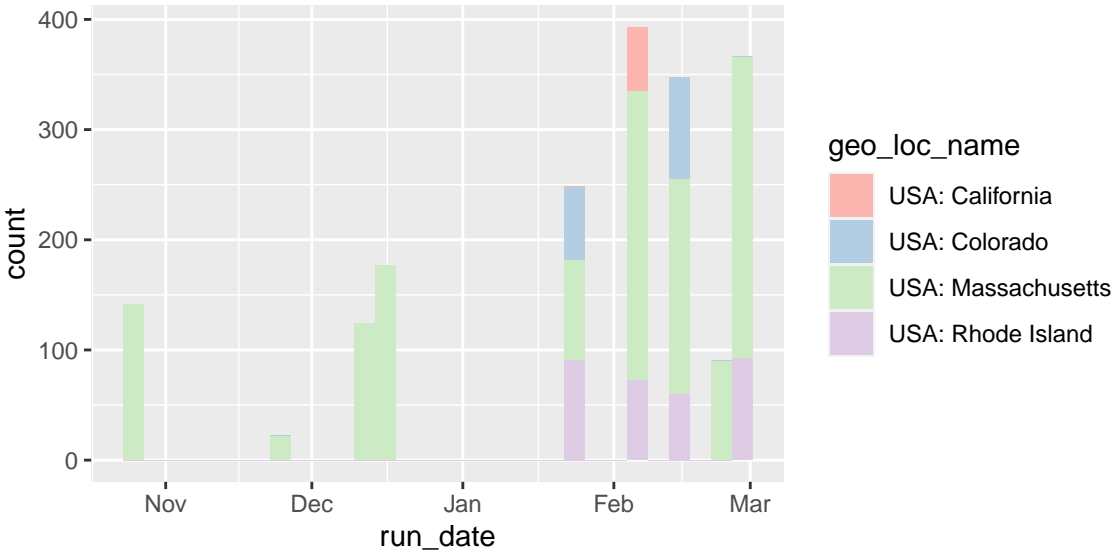
Sequencing activity over time by source



collected_by

- Brigham and Womens Hospital
- Broad Institute Clinical Research Sequencing Platform
- Colorado Mesa University
- Flow Health
- Massachusetts General Hospital
- Rhode Island Department of Health
- University of Massachusetts Medical School
- Warrior Diagnostics Laboratory

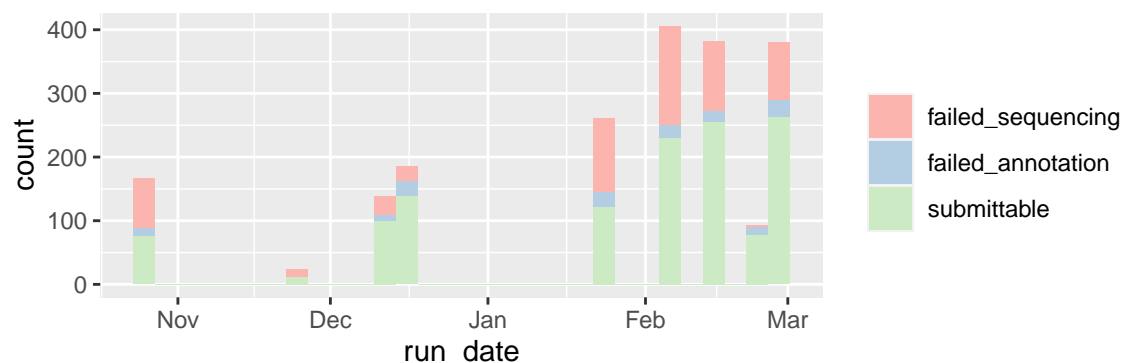
Sequencing activity by geography



This describes sequencing attempts over time, broken down by geography.

Sequencing performance over time

Samples sequenced weekly



This describes the total number of patient samples sequenced in this data set, plotted by the date of the sequencing run in our lab.

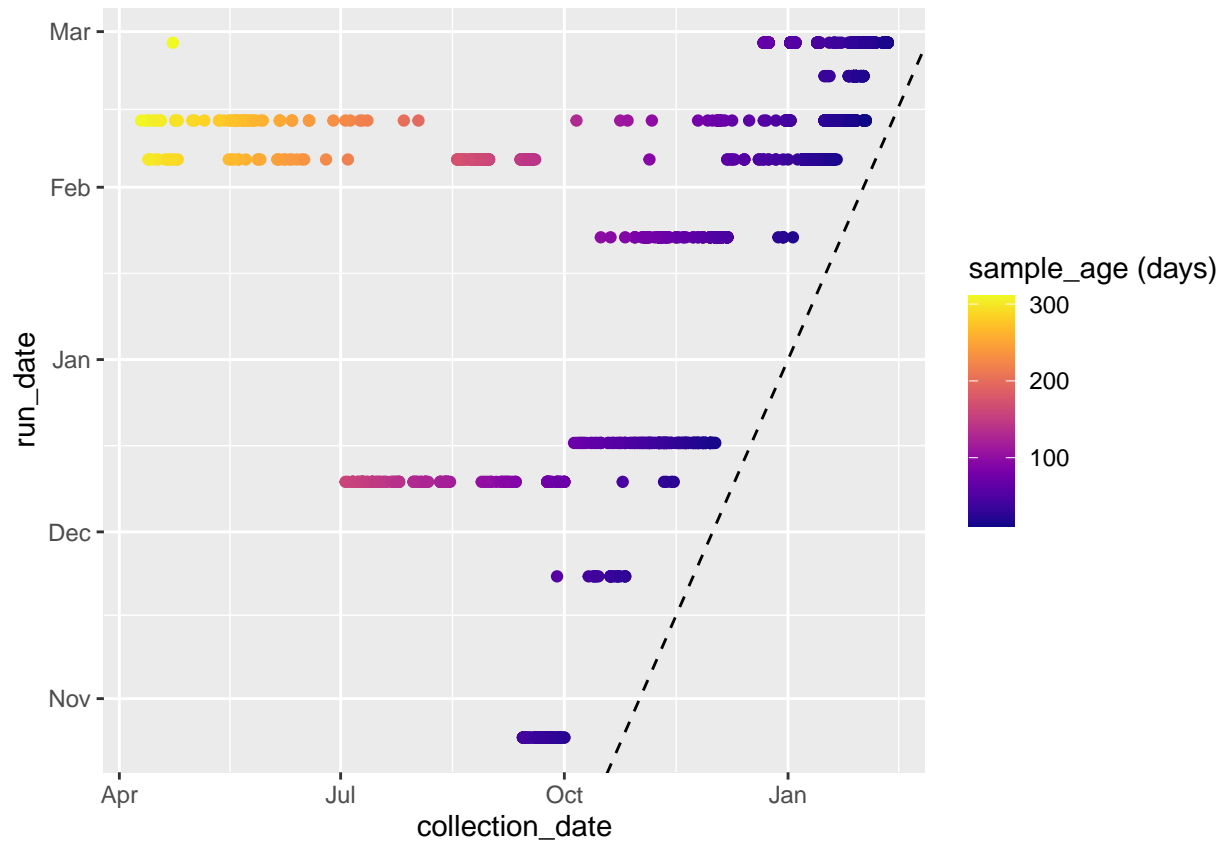
“Submittable” genomes pass all QC checks and are quickly released to public genome repositories. “Failed sequencing” are samples that failed to produce at least 24000 unambiguous base pairs of viral genome. Raw data from these samples are submitted to NCBI’s SRA database, but the genomes are not used for any analyses. “Failed annotation” are samples that produced a sufficiently complete genome, but did not pass NCBI’s VADR quality checks.

Tabular view by CDC epiweek

epiweek ending	samples sequenced	genomes assembled	genomes submittable
2020-10-31	167	88	75
2020-11-28	24	12	11
2020-12-12	138	109	99
2020-12-19	185	162	138
2021-01-23	261	145	122
2021-02-06	406	249	230
2021-02-13	382	272	255
2021-02-27	474	378	340

Timeliness of surveillance

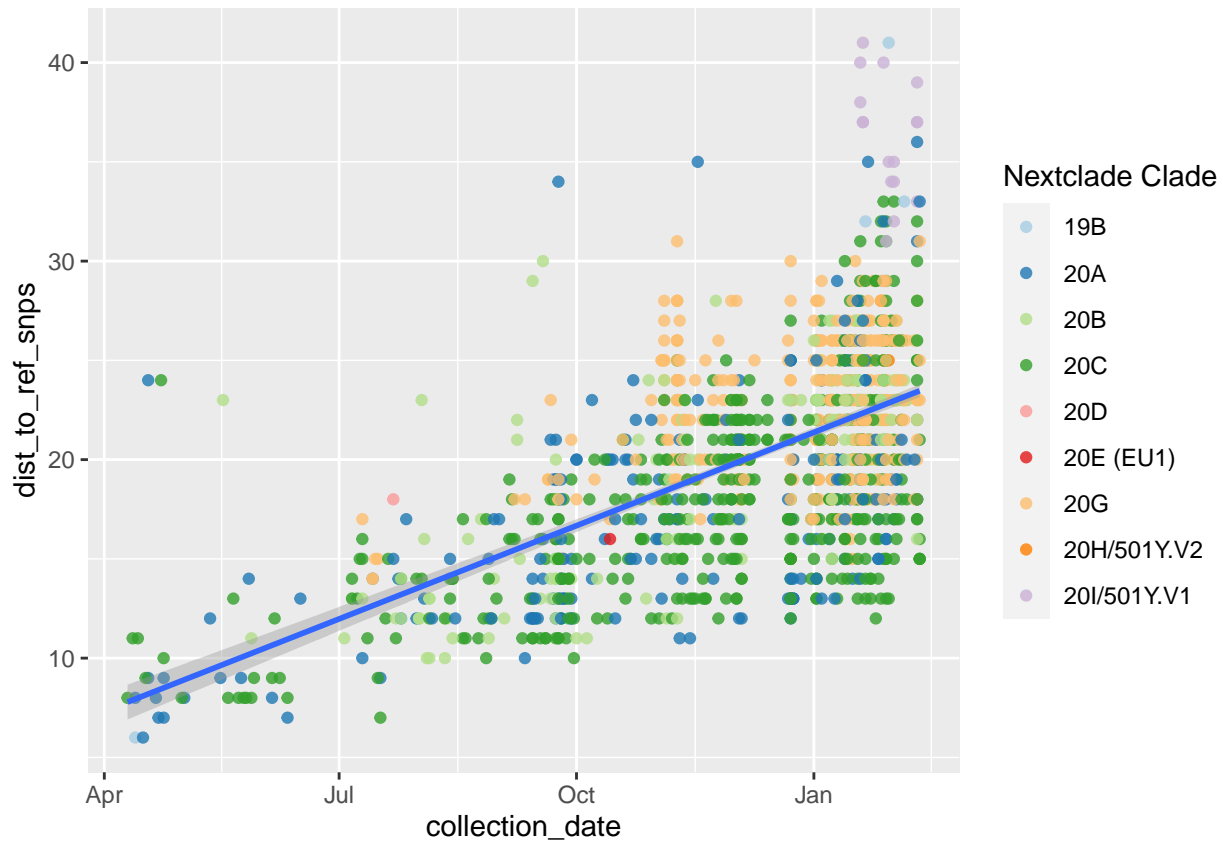
Sequencing date vs collection date



This plot describes the “timeliness” of the sequencing run for the purpose of real-time surveillance of circulating lineages and variants of interest. Note that this plot likely includes many samples that were sequenced for non-surveillance purposes.

Evolutionary Clock

Genetic distance root-to-tip vs sample collection date



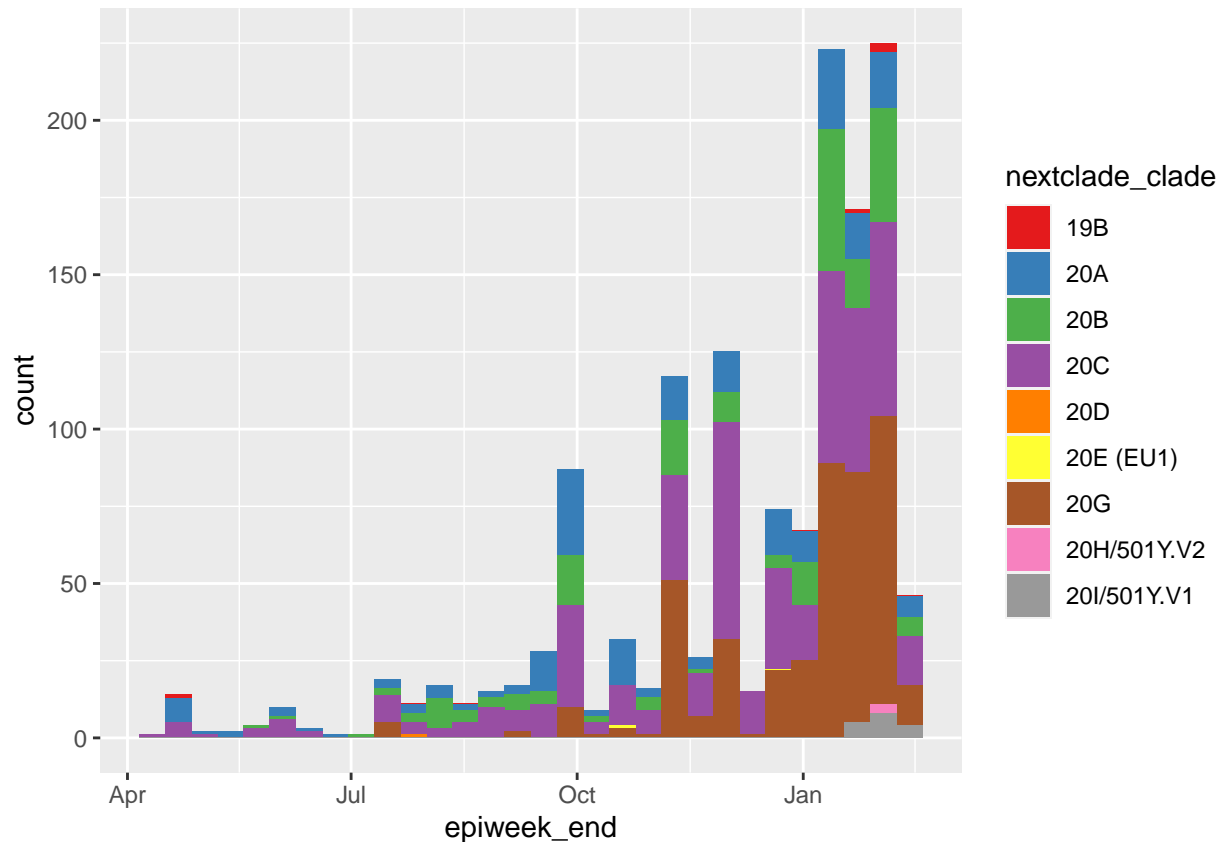
A “root-to-tip plot” plots the genetic distance of each sample from Wuhan Hu-1 against the date it was collected. It is generally somewhat linear. Outliers on this plot may be indicative of laboratory or metadata errors, or of evolutionarily unusual lineages (such as B.1.1.7).

Phylogenetic Clades and Variants

Reportable Variants of Concern (VoCs) by CDC epiweek of sample collection

Collection epiweek ending	B.1.1.7	B.1.351
2021-01-23	5	0
2021-01-30	4	3
2021-02-06	4	0
2021-02-13	4	0

Nextclade classifications vs sample collection date



This shows the breakdown of major phylogenetic clades over time, using the Nextclade naming system. Variants of Concern (VoCs) are highlighted as specially named Nextclade clades. Nextclade clade 20I/501Y.V1 corresponds to PANGO lineage B.1.1.7, 20H/501Y.V2 corresponds to B.1.351, and 20J/501Y.V3 corresponds to P.1.