SARS-CoV-2 sequencing report

The Sequencing Lab is sequencing SARS-CoV-2 from patients.

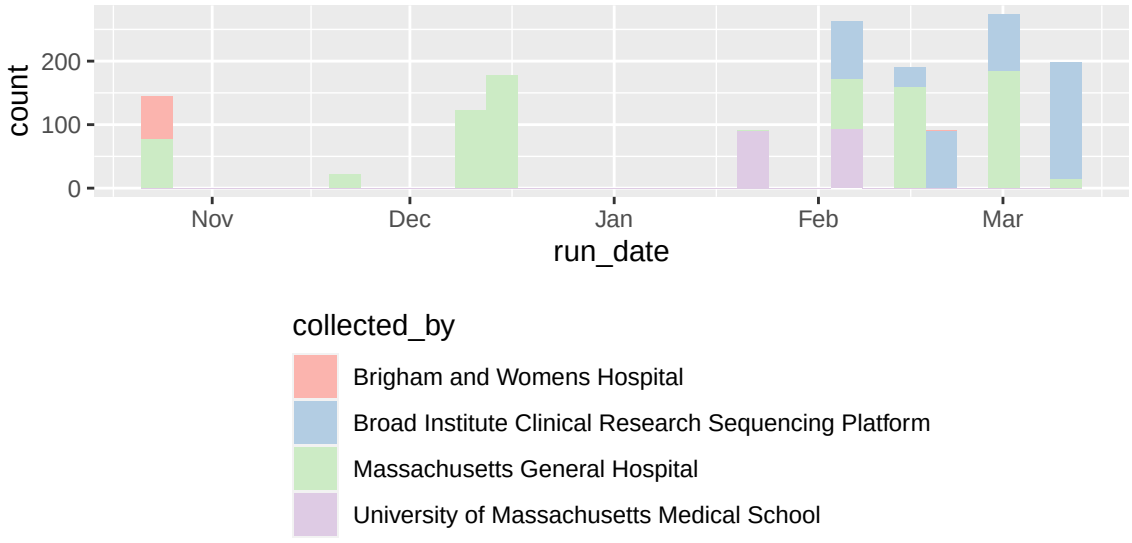
This report, generated on 2021-03-18, summarizes our sequencing activity for patient samples collected in **Massachusetts**. The samples included in this report were sequenced between 2020-10-25 and 2021-03-12. This report is current as of the epiweek ending **2021-03-13**.

Weekly summary

	week ending 2021-03-13	cumulative total
samples sequenced	198	1571
genomes assembled	185	1175
genomes submittable	179	1080
Variants of Interest (VoIs)	5	6
Variants of Concern (VoCs)	11	23

CDC definitions of Variants of Concern (VoCs) and Variants of Interest (VoIs) are available at: `https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html`

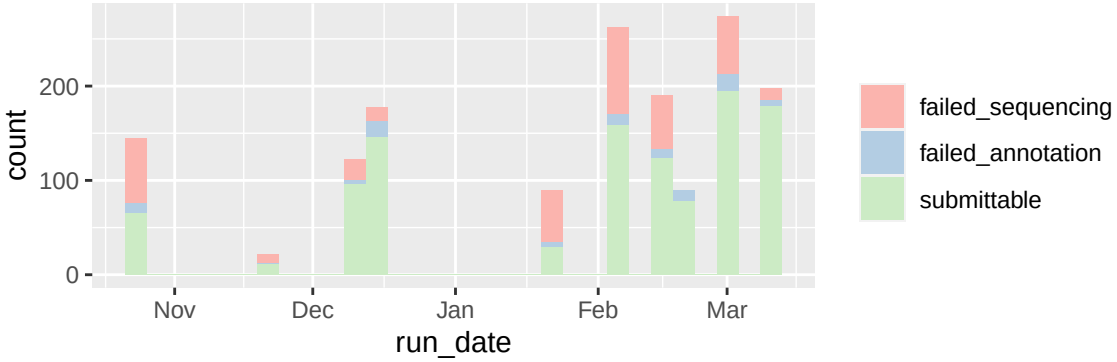
Sequencing activity over time



This describes sequencing attempts over time, broken down by the sample source laboratory.

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.

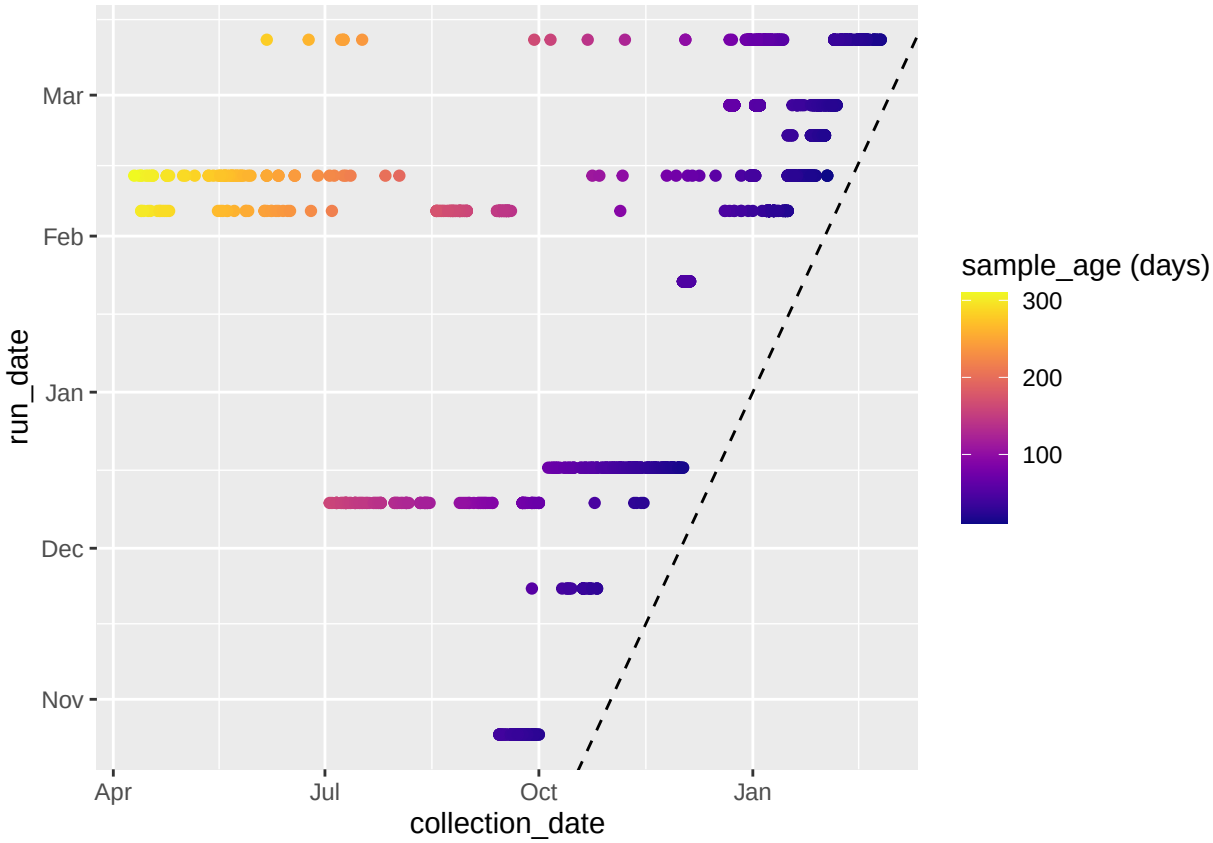
Submittable genomes pass all QC checks and are quickly released to public genome repositories. **Failed sequencing** are samples that failed to produce at least 24000 unambiguous base pairs of viral genome. Raw data from these samples are submitted to NCBI's SRA database, but the genomes are not used 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	144	76	65
2020-11-28	22	12	11
2020-12-12	123	100	96
2020-12-19	177	162	146
2021-01-23	90	35	29
2021-02-06	263	170	158
2021-02-13	190	133	123
2021-02-27	364	302	273
2021-03-13	198	185	179

Timeliness of surveillance

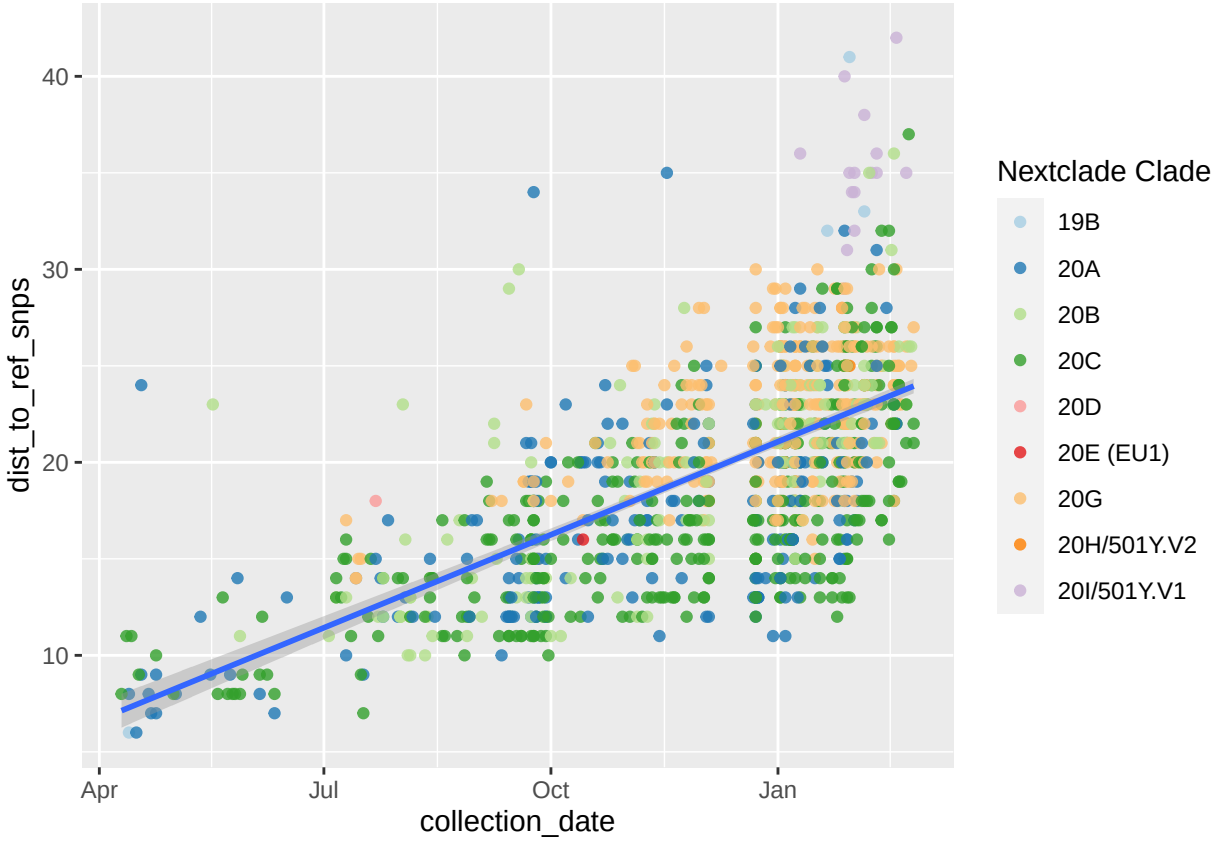
Sequencing date vs collection date



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-surveilance 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

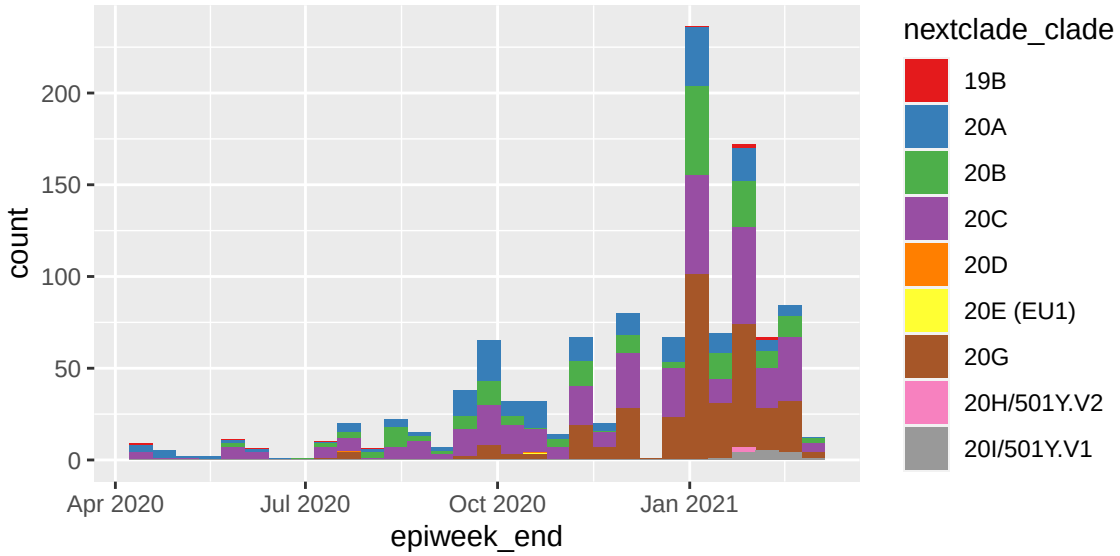
Collection epiweek ending	B.1.1.7	B.1.351	B.1.427	B.1.429
2021-01-09	0	0	1	0
2021-01-16	1	0	0	0
2021-01-30	4	3	0	1
2021-02-06	5	0	0	0
2021-02-13	3	0	0	2
2021-02-20	1	0	0	1
2021-02-27	1	0	0	0

Variants of Interest (VoIs) by CDC epiweek of sample collection

Collection epiweek ending	B.1.526	P.2
2021-01-09	0	1
2021-02-13	2	0
2021-02-20	1	0
2021-02-27	2	0

CDC definitions of Variants of Concern (VoCs) and Variants of Interest (VoIs) are available at: `https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html`

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. At this time, no other VoC or VoI lineages have dedicated Nextclade definitions.