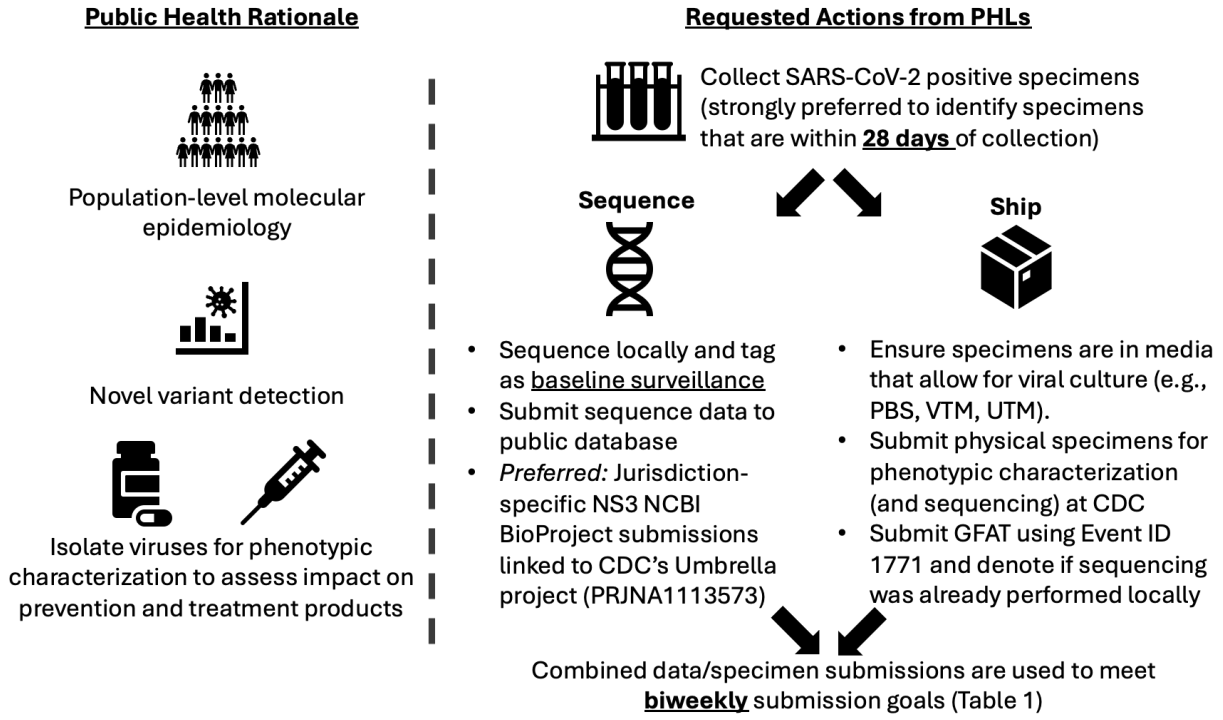
National SARS-CoV-2 Strain Surveillance (NS3) System Guide for Public Health Laboratories

The National SARS-CoV-2 Strain Surveillance (NS3) program was established in November 2020 to support nationally representative baseline genomic surveillance of SARS-CoV-2. Through NS3, CDC collects representative specimens and sequences to monitor viral evolution, identify emerging variants, and inform public health decisions.

General Overview:



There are multiple goals for routinely sequencing and characterizing clinical specimens that are positive for SARS-CoV-2. These can broadly be grouped into two primary objectives:

1. **Phenotypic Virus characterization:**
    - Select viruses isolated from specimens received from public health laboratories are evaluated in CDC laboratories to understand their potential impact on current vaccines, treatments, and diagnostics, and their overall risk to public health.
2. **Genetic Surveillance:**
    - *Population-level molecular epidemiology/virus monitoring:* By routinely acquiring sequences and associated metadata from a subset of COVID-19 cases, CDC can monitor the spread of viral lineages across time and within populations and more accurately describe which variants are growing and contributing to COVID-19 cases.

- *Novel Variant Detection*: Having a robust and recent set of virus genomic data will help rapidly identify new and emerging virus variants that might have vaccine and/or therapeutic resistance, different transmissibility, pathogenicity, or clinical outcomes.

This document describes opportunities for public health laboratories to participate in NS3 in support of national genetic surveillance and phenotypic characterization goals through specimen shipping and sequencing activities. The targets described are intended to serve as best-effort goals, and jurisdictions unable to meet them should not be discouraged from participating.

Section 1. National and Jurisdictional Surveillance Targets

NS3 depends on timely participation from jurisdictions in two ways: shipping culturable SARS-CoV-2–positive specimens to CDC and generating high-quality sequence data for submission to public repositories. Together, these activities provide timely, geographically diverse data on circulating SARS-CoV-2 strains. Targets for each jurisdiction are designed to support regular collection and analysis of geographically diverse specimens at the national level, including timely identification of emerging variants and estimates of viral population growth rate.

- **Phenotypic Characterization Request:** CDC requests that jurisdictions ship recent SARS-CoV-2 positive to CDC every two weeks.
    - When possible, prioritize shipping specimens that have not been sequenced yet. CDC needs physical specimens to perform phenotypic assessments of variants. Laboratories may send residual specimens that have already been sequenced in-house—this should be noted on the GFAT at the time of submission so that CDC does not duplicate entries in public sequence repositories.
    - CDC sequences specimens submitted through this component and submits sequence data to public repositories on behalf of the submitting jurisdiction, unless the GFAT notes that they are already submitted.
    - Specimens sequenced by CDC will count toward the submitting jurisdiction's genetic surveillance target.

- **Genetic Surveillance Request**: In addition to shipping specimens, CDC requests that jurisdictions with sequencing capacity collect recent specimens, generate high-quality sequences, and submit tagged sequence data to open public repositories. These activities support national and jurisdiction-level genomic surveillance. Collaboration among federal, state, local, territorial, academic, and commercial laboratories is critical to meeting these goals and improving the availability and public health impact of sequencing data.

Table 1 lists biweekly jurisdictional targets for specimen shipment to CDC and total sequencing for national surveillance. Additional specimen or sequence submissions are encouraged, as they can improve variant detection and situational awareness.

Public health laboratories that do not have the capability to sequence SARS-CoV-2 at their facility should send positive SARS-CoV-2 specimens to CDC for sequencing in order to meet sequencing goals.

Shipment of specimens to CDC for Phenotypic Characterization
Public health laboratories should ship the number of SARS-CoV-2 specimens requested in Table 1, column 2 to CDC every two weeks. To support sequencing goals, laboratories should prioritize specimens that have not been sequenced. If all available specimens have already been sequenced, the laboratory should still ship them and indicate prior sequencing on the GFAT submission form so CDC will not submit duplicate genomes to public repositories. See Appendix 3 for more information on the GFAT.

Jurisdictions should ship the requested number of specimens to CDC for sequencing and viral characterization (Table 1, column 2). If jurisdictions ship additional positive SARS-CoV-2 specimens to CDC for sequencing, all of these specimens will count towards meeting national genetic surveillance targets (Table 1, column 3). Additional specimens above the NS3 jurisdictional target in Table 1, column 2 may be shipped to CDC for sequencing using the test order CDC-10551 (Test Order | Submitting Specimens to CDC | Infectious Diseases Laboratories | CDC). Additional specimens will be sequenced based on availability of sequencing capacity at CDC. Shipping instructions can be found in Appendix 2.

National Target Sequencing Numbers for Genetic Surveillance
The national genetic surveillance target supports population-level genetic surveillance, including early detection of emerging variants and estimation of variant growth. Approximately 3000 high quality, timely SARS-CoV-2 sequences are needed every two weeks to detect 1 or more rare/novel variants at a prevalence of 0.1% (1/1000) with 95% confidence.

The sequencing targets in Table 1, column 3, are based on national variant detection thresholds and jurisdictional population estimates. Targets are capped at 90 per jurisdiction. The national biweekly target of 3000 is a year-round sequencing goal. Meeting these targets may be difficult during periods of low prevalence or in jurisdictions with limited specimen availability. These are best-effort targets, and jurisdictions are encouraged to contribute as many timely, high-quality sequences as possible.

Although Table 1 focuses on national surveillance targets, jurisdictions may choose to sequence beyond these targets to support additional local surveillance objectives. As a reference point, sequencing approximately 90 genomes every two weeks can support detection of a variant circulating within a jurisdiction at approximately 3% prevalence with 95% confidence. This 90-genome reference value is not a separate NS3 requirement and should not be added to the national surveillance target in Table 1.

Partner Data
Jurisdictions with multiple laboratories performing SARS-CoV-2 sequencing, such as multiple local public health laboratories or academic or commercial laboratories, will have that data integrated

automatically to count towards sequencing goals if they follow the quality control and submission guidelines.

All sequencing facilities should adhere to all components of High-Quality Data for Genomic Surveillance as outlined in this guide, including the quality control thresholds. Partner-generated data should be submitted under a SARS-CoV-2 US Genomic Surveillance BioProject as described in Appendix 1 to ensure the data will be included in the jurisdictional target sequencing count.

Determining Jurisdictional Sequencing Counts

CDC will count specimens shipped for phenotypic characterization and sequencing using internal laboratory information management systems. Counts will be shared only with the submitting jurisdiction and used to provide feedback throughout the year.

Specimens that are unacceptable upon arrival, such as specimens that arrive outside acceptable temperature conditions, will not count toward jurisdictional submission goals. Jurisdictions may request feedback from CDC to confirm specimen counts and understand whether submitted specimens are contributing toward their goals.

For jurisdictions that sequence and submit data directly to public repositories, CDC will use BioProject accessions as the primary method for counting baseline surveillance submissions. See **Section 2** for more information on setup and implementation of the BioProject method.

CDC and APHL will provide NS3 report cards on SARS-CoV-2 testing and submission. See Appendix 4 for frequently asked questions (FAQs).

Summary of Numbers in Table 1

Table 1 includes two biweekly jurisdictional targets:

- **Column 1**: State or Territory
- **Column 2:** The number of specimens CDC requests each jurisdiction ship to CDC for virus isolation, sequencing, and phenotypic characterization.
- **Column 3:** The total number of sequences requested from each jurisdiction to support national genomic surveillance and nowcast modeling. These sequences may be generated by the jurisdiction, by sequencing partners, or by CDC from specimens shipped through NS3.

Specimens shipped to CDC and sequenced by CDC can count toward the submitting jurisdiction's total sequencing target in column 3. For example, if CDC requests that a jurisdiction ship 10 specimens every two weeks and contribute 64 total sequences for national surveillance, the 10 specimens sequenced by CDC may count toward the 64-sequence target.

Table 1. SARS-CoV-2 Biweekly Sequencing and Specimen Submission Goals by Jurisdiction

State/Territory	Ship to CDC for Viral Characterization and Sequencing	Total Sequencing for National Surveillance*
Alabama	10	64
Alaska	10	22
Arizona	15	86
Arkansas	10	45
California	35	90
Colorado	10	71
Connecticut	10	51
Delaware	10	25
District of Columbia	10	22
Florida	25	90
Georgia	20	90
Hawaii	10	29
Idaho	10	32
Illinois	20	90
Indiana	10	82
Iowa	10	46
Kansas	10	44
Kentucky	10	59
Louisiana	10	62
Maine	10	28
Maryland	10	75
Massachusetts	15	84
Michigan	15	90
Minnesota	15	71
Mississippi	10	45
Missouri	15	76
Montana	10	26
Nebraska	10	34
Nevada	10	45
New Hampshire	10	28
New Jersey	15	90
New Mexico	10	36
New York	20	90
North Carolina	15	90

North Dakota	10	23
Ohio	20	90
Oklahoma	10	54
Oregon	10	56
Pennsylvania	20	90
Rhode Island	10	26
South Carolina	10	65
South Dakota	10	24
Tennessee	15	82
Texas	35	90
Utah	10	46
Vermont	10	21
Virginia	15	90
Washington	15	89
West Virginia	10	33
Wisconsin	15	73
Wyoming	10	21
American Samoa	5	5
Guam	5	5
Marshall Islands	5	5
Micronesia	5	5
Northern Mariana Islands	5	5
Palau	5	5
Puerto Rico	10	49
Virgin Islands	5	5
Total:	725	3065

*The national targets described in Table 1 are for the purpose of identifying novel variants and estimating viral distribution at a national level.

Year-Round Surveillance

SARS-CoV-2 requires sustained, year-round surveillance. Notably, the targets in Table 1 do not increase during periods of high transmission. During high transmission periods, the variant population is typically dominated by one or two strains, and additional sequence beyond the targets may not substantially improve baseline surveillance. Sustained surveillance during low-transmission periods remains important because during that time, the SARS-CoV-2 virus population is more diverse, and it may be a critical time to observe an emerging variant.

Specimen Shortages

The biweekly targets listed in Table 1 may not be achievable year-round, especially during times of low transmission. When positive specimen are limited, jurisdictions should continue to sequence and ship as many specimens as possible on a regular, preferably two-week, schedule. Timeliness is a priority. Sequencing fewer recent specimens quickly is more useful than sequencing larger numbers of older specimens after significant delays. If in-house sequencing will be delayed, jurisdictions may send specimens to CDC for sequencing.

Specimen Diversity
For baseline surveillance, specimens should be selected without targeting specific patient characteristics, exposure histories, outbreak events, travel history, suspected variants, or other special circumstances. When possible, laboratories should submit a routine, broadly representative set of recent SARS-CoV-2–positive specimens from multiple geographic areas within the jurisdiction.

Specimens selected for a specific reason, such as an outbreak investigation or suspected variant case, can bias estimates of variant prevalence and growth if included as baseline surveillance. These specimens may still be useful for other public health purposes, but they should not be categorized or tagged as baseline surveillance.

For additional information on categorizing sequence selection for baseline surveillance, see Appendix 1. Refer to Appendix 2 for specimen selection criteria.

Turn-Around Time
Jurisdictions should prioritize recent specimens and rapid sequencing or shipment. Timely, high-quality data are more valuable for surveillance than larger volumes of delayed or lower-quality data.

Specimen Retention
If possible, retain residual specimens for 10 weeks to allow for further investigation, if needed.

SARS-CoV-2 Sequencing Methods and Equipment
Jurisdictions are not required to use standardized sequencing equipment or protocols, provided the quality standards in this guide are met. Methodological diversity is acceptable and may be beneficial, given existing investments in equipment, workflows, and expertise across laboratories.

Section 2. Components of High-Quality Data for Genomic Surveillance

High-quality genomic surveillance data depend on five components: recent specimens, high-quality sequences, timely submission to public repositories, appropriate sequence tagging, and consistent metadata.

Collection of Recent Specimens
Specimen collection and processing should be performed on recent specimens for the success of this program. The timeliness of data is critical and can impact suitability for downstream analyses.

- Specimens shipped to CDC for this program must be collected in media that allow for viral culture (See Appendix 2 for more information).
- Specimens collected for on-site sequencing to meet genetic surveillance goals do not have media restrictions.
- Specimens should be laboratory confirmed SARS-CoV-2 positive, deidentified specimens (with Ct values ≤ 28, if a Ct value is not available the specimens should have a positive result) and include standardized metadata on a representative selection of COVID-19 cases.
- Specimens should be collected within 28 days prior to shipment and properly stored.
- Prioritize specimens with the most recent collection date.

See Appendix 2 for additional information on acceptable specimen types and additional considerations for specimen selection.

Generation of high-quality sequences

Jurisdictions performing on-site sequencing should aim to meet the following quality controls for the generation of high-quality sequences for public surveillance. Sequence data that do not meet these quality thresholds may be excluded from nowcast analyses.

- Collection Date
  - Sequences should be uploaded within 28 days of sample collection
- Depth of coverage
  - 20X per nucleotide to call a base at a position (<20X use "N" to mark uncertain/indeterminate bases)
- Base representation
  - Ambiguous nucleotide calls at 25-75% should be marked using appropriate IUPAC ambiguity code
- Spike gene coverage
  - Complete spike gene coverage at 20X coverage per nucleotide
  - Sequences not passing the outlined spike specific criteria should be carefully reviewed to differentiate low quality sequencing events, indications of amplicon dropout, or novel mutation events. Aggregate analysis including previous runs may be required to detect systematic issues, such as variant-specific primer mismatches/failures.
  - Systematic issues can be reported to CDC via the sarsseq@cdc.gov inbox and CDC will provide advice.
- Primer Trimming
  - Primer sequences *must* be trimmed prior to consensus generation
- Length
  - ≥ 27700 unambiguous bases, represented as single contig
  - These partial genomes should be reviewed to differentiate low quality sequencing events vs. indications of primer dropout. Aggregate analysis including previous runs may be required to detect systematic issues, such as variant-specific primer mismatches/failures.

Submission and publication of sequences to public repositories

Laboratories should submit sequence data to public repositories as soon as possible after sequencing so the data are available before CDC analysis deadlines. Submission by jurisdictions requires:

- Required metadata - include in all sequence data submissions:
    - specimen type, collection date, and geolocation information including state.
- Publication deadline - The publication deadline for all public repositories is 10 days prior to the date of the next CDC SARS-CoV-2 Variant Proportions public update.
    - For instance, if the next update is Friday 7/6/2026 only data available on public repositories by 6/26/2026 will be used.
    - Please account for the time lag between submission and publication for different public repositories.

Sequence Tagging and BioProject Accessioning

SARS-CoV-2 sequence data submitted for national surveillance must be properly cataloged so CDC, submitting laboratories, and other users can identify sequences generated through baseline surveillance. Sequences generated through targeted activities, such as outbreak investigations, may bias baseline surveillance estimates and should **not** be tagged as baseline surveillance.

The preferred method for cataloging NS3 sequence data is through NCBI BioProjects. Each jurisdiction should create a BioProject for National SARS-CoV-2 Strain Surveillance that includes only specimens collected through baseline surveillance. Jurisdictional BioProjects should be associated with CDC's Umbrella BioProject, **PRJNA1113573**. To request association with the umbrella BioProject, email bioprojecthelp@ncbi.nlm.nih.gov. See Appendix 1 for detailed instructions.

CDC also supports consensus-level tagging in GenBank for sequences submitted as part of baseline surveillance. Consensus sequence records should be tagged as "baseline surveillance" according to the guidance in Appendix 1 so they can be included in CDC analyses. Standard, consistent tagging improves CDC's ability to identify, search, analyze, and share sequence data across jurisdictions. It also supports reproducible analyses by helping ensure that the same baseline surveillance dataset can be identified over time.

While only consensus sequences and required metadata are needed for submission, submitting raw reads to NCBI's SRA is encouraged. Before uploading raw reads, email bioprojecthelp@ncbi.nlm.nih.gov to enable automated human-read scrubbing for the BioProject.

Database naming conventions

Sequence names should follow established database naming conventions and should reflect the specimen's geographic collection location.

If a specimen is shipped to CDC and sequenced by CDC, CDC will be included in the isolate name to indicate where sequencing was performed. The jurisdiction where the specimen was collected will remain the geographic origin in the sequence name.

If a specimen is sequenced and submitted locally by the jurisdiction or another sequencing partner, do not include CDC in the isolate name. Instead, use an abbreviation that identifies the laboratory or entity that performed sequencing.

Use the following formats when submitting sequences:

NCBI/ICTV format:

SARS-CoV-2/host/country/isolate/year

Example: SARS-CoV-2/human/USA/XX-CDC-9898989/2021

GISAID format:

hCoV-19/country/isolate/date

Example: hCoV-19/USA/XX-CDC-9898989/2021

In these examples:

- **XX** represents the two-letter state abbreviation or other collection location.
- **CDC** should be replaced with the abbreviation for the sequencing laboratory or submitting entity.
- **9898989** should be replaced with a meaningful strain or isolate identifier.
- **2021** should be replaced with the appropriate collection year or date, depending on the database format.

When submitting the same sequence to multiple databases, the shared stem should be identical across submissions. For example, **USA/XX-CDC-9898989/2021** should be used consistently in both NCBI and GISAID records, with only the database-specific prefix differing.

Questions and Technical Guidance

For technical guidance, questions about forms or shipments, or sequencing-related questions, please contact SARSSEQ@cdc.gov.

For assistance in submitting data to public repositories, labs may find it helpful to use SeqSender (https://github.com/CDCgov/seqsender)

Appendix 1: Technical Assistance and Instructions for Public Health Laboratories on Categorizing Sequence Data as "Baseline Surveillance" for Inclusion in CDC's National SARS-CoV-2 Genomic Surveillance

Objective: Ensure Representative Surveillance Data
The primary goal of this program is to monitor what SARS-CoV-2 variants are circulating broadly within a jurisdiction. To achieve this, CDC relies on data that represents the general population, not specific events or high-risk groups.

Sequences that meet the criteria for **baseline surveillance** analyses include those:

- Sampled randomly for genomic surveillance
- Not identified in a targeted sampling effort (targeted efforts defined below)
- Sampled from across targeted sequencing efforts to be representative of the circulating viral population of SARS-CoV-2 as necessary to meet submission thresholds.

Sequences from **targeted efforts** include, but are not limited to, those:

- Sampled based on cluster/outbreak investigations
- Longitudinally or repeatedly sampled from the same individual
- Sampled based on pre-screening for a particular variant (e.g., S-gene target failure)
- Sampled for the purpose of vaccine escape studies
- Sampled based on travel history
- Sampled based on disease severity

Identifying a representative subset of sequences from targeted sequencing efforts:
Inclusion of all sequences from targeted sequencing efforts in baseline surveillance could bias estimates of circulating SARS-CoV-2 lineages by overrepresenting lineages. However, sampling sequences from targeted sequencing efforts that are representative of the community can and should be included in baseline surveillance.

To achieve a representative sample:

- Sample a similar proportion of sequences from a targeted sequencing effort as what is sampled for general surveillance efforts.
- For targeted efforts involving longitudinal or repeated sampling of the same individual, tag only one sequence per individual as baseline surveillance.

For CDC to correctly identify and ingest SARS-CoV-2 sequences generated by your laboratory in the baseline surveillance analysis, the sequences need to be tagged as such in online databases:

- For NCBI submissions, this is done by including a keyword: "purposeofsampling:baselinesurveillance".
- For GISAID submissions, this is accomplished by selecting "Baseline surveillance" in the sampling strategy field.

- See further instructions below for how to tag new and former submissions as baseline surveillance.
- Use the standard file formats available from each data source to improve the timeliness of data ingestion and analyses, which CDC performs daily.
- Where possible, use the database tag instead of directly emailing sequences/accession to CDC.

Instructions for NCBI Baseline Surveillance

Below we outline two methods for tagging sequence data in NCBI. The BioProject/BioSample method of tagging is preferred; however, if your lab cannot submit BioSamples, the GenBank method is acceptable. If you or your sequencing partner(s) are already marking baseline surveillance BioSamples using the purpose of sequencing field with the "Baseline surveillance (random sampling)" option outlined in the PHA4GE metadata specification, you do not have to make any changes to your sequence data tagging. PHA4GE compliant instructions for marking baseline surveillance samples through BioSample appear below.

Submitting SARS-CoV-2 metadata to BioSample (Preferred)

BioSample is a central location in which to store normalized, descriptive information about biological source materials used to generate experimental data. Metadata included in the archival BioSample database are reciprocally linked with BioProjects as well as with derived experimental data in NCBI's primary archives, including the Sequence Read Archive (SRA) and GenBank.

1. Start your BioSample submission.
    a. Submission of BioSamples can be done in batches using a tab-delimited text file that describes each of the samples and attributes.
    b. Template files can be downloaded from the attributes tab within the submission portal wizard (link within portal wizard).
    c. Please use the following template for clinical SARS-CoV-2 sequence data: SARS-CoV-2: clinical or host-associated.

2. Once you choose the correct attribute package, you will have the option of using a built-in table editor or uploading a spreadsheet that includes the attributes for each of your BioSamples.
    a. Required attributes are marked with an asterisk within the built-in table editor and spreadsheet.
    b. The value for the following optional "purpose of sequencing" attribute should be filled in to specify "baseline surveillance (random sampling)".

3. Once you have finished registering your BioSamples, they will be assigned BioSample accession numbers that you can include within your Source.src file in the BioSample Column. These have the following format: SAMNXXXXXXXX.

Submitting SARS-CoV-2 metadata to GenBank

Your submission is important to NCBI and the global research community! Get GenBank accessions when you submit assembled SARS-CoV-2 reads with FASTA files and source metadata. NCBI annotates all SARS-CoV-2 submissions on your behalf to ease submission.

Important: To tag your submission as part of CDC's baseline surveillance effort and make it more readily searchable once released, please follow these steps:

Updating existing submissions

If you need to add or edit the keyword in submissions, follow these instructions:

- If all records are to have the same keyword, provide a list of the relevant accessions and the keyword (purposeofsampling:baselinesurveillance) to be added to: gb-admin@ncbi.nlm.nih.gov
- If adding a mix of keywords to your records, send a two-column table where the first column is the GenBank accession number and the second column is the keyword text.

Example:
MZxxxxx1 purposeofsampling:baselinesurveillance
MZxxxxx2 purposeofsampling:environmentalsurvey

Submitting new SARS-CoV-2 data to Genbank using the web-based Submission Portal

1. Begin your submission with all required materials. Click the "New submission" button under the Submission Portal header to get started.
2. Select SARS-CoV-2 on the "Submission Type" tab.
3. Once you reach the "Sequences" tab, you will upload your FASTA file and include the following CDC-requested keyword **in this exact format in the location for keyword**: purposeofsampling:baselinesurveillance.
4. Your FASTA file should contain the following in the FASTA definition line, separated from the Sequence ID by a space
   [keyword=purposeofsampling:baselinesurveillance]

   Note: this tag should appear in ***each*** FASTA definition line.

   Examples:
   >Seq1 [keyword=purposeofsampling:baselinesurveillance]
   CTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG
   >Seq2 [keyword=purposeofsampling:baselinesurveillance]
   CTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG

5. Before you complete the submission, go to the "Review & Submit" tab which will show your submission's details including a preview of the GenBank record on the right side of the screen where you can see the keyword.
6. Please see the section for "Updating existing submissions" above to update your records after submission.

Submitting SARS-CoV-2 data to GenBank using FTP

1. Contact gb-admin@ncbi.nlm.nih.gov to receive an account and brief instructions from the NCBI team. They will help you get started.
2. Once you can submit via FTP, ensure that the files you place on FTP with your FASTA sequences include the following CDC-requested keyword **in this exact format in the location for keyword**: purposeofsampling:baselinesurveillance.
3. Your FASTA file should contain the following in the FASTA definition line, separated from the Sequence ID by a space

   [keyword=purposeofsampling:baselinesurveillance]

   Note: this tag should appear in ***each*** FASTA definition line.

   Examples:
   >Seq1 [keyword=purposeofsampling:baselinesurveillance]
   CTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG
   >Seq2 [keyword=purposeofsampling:baselinesurveillance]
   CTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG
4. You will receive an error report in your FTP folder if there is a problem with your submission. You do not need to resubmit previous SARS-CoV-2 data to add this keyword but can contact gb-admin@ncbi.nlm.nih.gov with any updates required for your data.

Your keyword will appear in the GenBank record and be indexed for searching.

```
LOCUS       EU865993               29903 bp    RNA     linear   VRL 03-MAY-2021
DEFINITION  Severe acute respiratory syndrome coronavirus 2 isolate
            SARS-CoV-2/human/USA/CDC-xyz/2020, complete genome.
ACCESSION   EU865993
VERSION     EU865993
KEYWORDS    purposeofsampling:usbaselinesurveillance.
```

5. Please see the section for "Updating existing submissions" above to update your records after submission.

Instructions for GISAID EpiCov™ Baseline Surveillance

If jurisdictions choose to also upload to GISAID, "Baseline surveillance" should be selected in the Sampling Strategy (covv_sampling_strategy) field.

In the event you experience any difficulties with your upload or have additional questions, please contact GISAID for assistance at hCoV-19@gisaid.org

Appendix 2: Shipping Specimens to CDC

1. **Acceptable Specimen Types:**
    a. Acceptable specimen types for sequencing and potential virus characterization are the same as for the CDC SARS-CoV-2 diagnostic assays that were authorized by FDA under an EUA:
        i. Upper and lower respiratory specimens, including nasopharyngeal, oropharyngeal, nasal mid-turbinate, and anterior nares (nasal swab) specimens.
        ii. Nasopharyngeal wash/aspirate or nasal wash/aspirate specimen collected by a healthcare professional is acceptable, as is a naturally expectorated sputum.
    b. Acceptable specimens will be limited to those collected in media that allow for viral culture (e.g., PBS, VTM, UTM).
        i. Specimens collected in Hologic Aptima buffer and Molecular Transport Media are excluded from submission.
    c. For more information, see the interim specimen collection guidelines: Interim Guidelines for Collecting and Handling of Clinical Specimens for COVID-19 Testing | COVID-19 | CDC
2. **Considerations for selecting NS3 specimens:**
    a. The quality of the specimen directly affects sequencing and virus culture success.
        i. It is preferred for specimens to have an RT-PCR Ct value of ≤28. However, if sequencing goals cannot be met for this criteria, shipment of any known SARS-CoV-2 positive specimens is acceptable.
        ii. Testing with the CDC Flu-SC2 assay is recommended, however other testing platforms for SARS-CoV-2 are acceptable.
        iii. If Ct values are not available, specimens that are positive for SARS-CoV-2 may be included in the random selection of samples to be sent.
    b. The time from specimen collection to sequence characterization has a large impact on CDC's ability to quickly detect and track proportions of emerging variants.
        i. Send specimens that have been collected within the last 28 days whenever possible.
        ii. If the number of specimens collected within the last 28 days is insufficient to meet your jurisdiction's requested number for NS3 specimen submission (Table 1), please send the most recent specimens possible
    c. Specimens should be selected for baseline surveillance without targeting specific characteristics or events.
        i. Specimens should not be selected based on specific patient characteristics, exposure histories, outbreak events, travel history, suspected variants, or other special circumstances.
        ii. When feasible, submit specimens from multiple geographic areas within the jurisdiction.
        iii. Do not delay shipment to achieve an ideal specimen distribution. Timely submission of recent, eligible specimens is the priority.
        iv. Specimens associated with targeted activities, such as outbreak investigations or suspected variant cases, may be useful for other public health purposes but should not be categorized or tagged as baseline surveillance.

d. Specimens that are being sequenced by your laboratory or your partners may be submitted to CDC if this will help to meet your jurisdiction's NS3 submission target (Table 1).
    i. If submitting specimens being sequenced and submitted locally, please enter "yes" in the GFAT form's column "Alpha Numeric 01" (See Appendix 3) and provide the GISAID or GenBank accession numbers if they are available at the time of shipping in the GFAT form's column "Additional ID 1" (See Appendix 3).
    ii. For sequence data generated locally, add a tag to the submission as described in the guidelines. For example, a submission to GenBank would have the keyword (purposeofsampling:baselinesurveillance) (See Appendix 1).
    iii. If you or your sequencing partner(s) are already marking baseline surveillance samples using the purpose of sequencing field with the "Baseline surveillance (random sampling)" option outlined in the PHA4GE metadata specification, you do not have to change to the surveillance tagging system described above (See Appendix 1).

3. Storage and Shipping Conditions:

a. Please submit original clinical specimens with at least 500 µL volume and no more than 1 mL unless confirmed beforehand.
b. Please use 1.0–2.0 mL O-ring screw cap microcentrifuge tubes labeled with the de-identified specimen ID.
c. Specimens can be stored at 2–8°C for no more than 72 hours from the time of collection, prior to shipping preparation.
    i. The 72- hour timeframe is a strict requirement for sequencing to be completed successfully.
    ii. Specimens that require storage longer than 72 hours should instead be frozen at ≤ -70°C.
d. Prior to shipping, specimens should be frozen at ≤ - 70°C and shipped overnight on dry ice.
e. Please ship routinely selected SARS-CoV-2 positive specimens every other Monday for overnight delivery to CDC on Tuesday.
    i. If Monday is an observed holiday, please ship on the next available business day (Tuesday).
    ii. Ship overnight using your usual courier, such as FedEx or UPS.
    iii. Please do not send shipments on Fridays or weekend days.

4. Global File Accessioning Template (GFAT) and Shipping Instructions:

a. Please fill in the electronic Global File Accessioning Template (GFAT) form.
b. Each specimen must be labeled with a unique identifier also included on the GFAT using the SPHL Submitter Specimen ID or the Original Submitter Specimen ID field (if no SPHL ID) (See Appendix 3 for more information).
c. Please fill out all GFAT fields marked as required, and if possible for the requested fields for which you have data.
d. The fields highlighted in Appendix 3 are required or requested for the processing of specimens and downstream uses of the sequence data for public health surveillance.

e. **If submitting specimens being sequenced locally**, please enter "yes" in the GFAT column "Alpha Numeric 01" and provide the GISAID or GenBank accession numbers, if available at the time of shipping, in the GFAT form's column "Additional ID 1" (See Appendix 3).
f. Do not include Personally Identifiable Information including "Patient Names, Birthdates".
g. In the GFAT form, please select or enter "NS3 - National SARS-CoV-2 Strain Surveillance" in the Event Name field and "1771" in the Event ID field.
h. For additional support in filling out the GFAT or questions please contact sarsseq@cdc.gov
i. Specimens should be packaged and shipped as Category B infectious substances, and all requirements for proper packaging and shipping should be observed (see Interim Guidelines for Clinical Specimens for COVID-19 | CDC).
j. Email the GFAT form along with tracking information to sarsseq@cdc.gov
k. Please include a printed manifest of your specimens with your shipment.
l. If possible, please ship specimens on dry ice biweekly Monday through Thursday for overnight delivery using your usual courier, such as FedEx or UPS, to CDC to the following address:

ATTN: STATT Lab: Unit 232 MIST
Centers for Disease Control and Prevention
1600 Clifton Road, NE
Atlanta, Georgia, 30329
Telephone: 404-639-3931
Email: sarsseq@cdc.gov

Appendix 3: SARS-CoV-2 GFAT Column Description of Required and Requested Fields

- Additional requested and required columns have been added to account for data that may have been collected using the supplementary form. The NS3 Supplementary form is no longer required. Please fill out the requested columns instead.
- Periodic adjustments to the GFAT structure will occur, the GFAT column letters are provided to help identify the appropriate column in the GFAT file but may be incorrect in newer versions. The column names can be used to search for the updated column.
- Column description of required and requested fields detailed below are provided based on the GFAT version 7.3 template, effective date of March 30, 2026.

Requested/ Required	GFAT Column Letter	Column Name	Acceptable Values	Required For:
Required	D	Origin	Human, Animal, Environmental	Shipping and Accessioning Non-CLIA
Required	E	Test Order Name	SARS-CoV-2 Surveillance Sequencing- Non-CLIA	Determining Sample Acceptability
Required	F	Suspected Agent	SARS-CoV-2	Determining Sample Acceptability
Required	G	Date Sent to CDC	Approximate date samples shipped or expected to be shipped to CDC	Sample Accessioning and Processing
Required	H	At CDC, bring to the attention of	STATT Lab: Unit 232 MIST	Sample Accessioning and Processing
Requested	P	Patient Age	Numeric, **Blank**	Data Analysis
Requested	Q	Age Units	Years, Months, Days	Data Analysis
Required	AC	Specimen Collected Date	Date of Specimen Collection	Determining Sample Acceptability
Required	AE	Material Submitted	Values must be available in the dropdown menu option	Determining Sample Acceptability
Required	AF	Specimen Source (Type)	Values must be available in the dropdown menu option	Determining Sample Acceptability
Requested	AH	Specimen Source Site	Values must be available in the dropdown menu option	Determining Sample Acceptability
Required	AL	Transport Medium/ Specimen Preservative	Options for NS3: • Viral Transport Media • Universal Transport Media • Sterile Saline	Determining Sample Acceptability

			Exceptions for Similar Names (contact sarsseq@cdc.gov for additional exceptions): • M4RT = Universal Transport Media	
One of these two is required to be complete	AN,BD	SPHL Submitter [Information Set]	SPHL Submitters need to include the SPHL Submitter ID that matches with their institution, and include the SPHL Submitter Specimen ID	Shipping and Accessioning Non-CLIA
One of these two is required to be complete	BF, BI	Original Submitter [Information Set]	Original Submitters need to include their provided Original Submitter ID and use the Original Submitter Specimen ID to include the Specimen ID If you do not have an Original Submitter ID, or do not know your organizations Original Submitter ID, please contact SARSSEQ@cdc.gov	Shipping and Accessioning Non-CLIA
One of these two is required to be complete	DX, DY	Vaccine Status information	Column based descriptions, **Blank** AND a date in MM/DD/YYYY format	Internal Data Analysis
Requested	EF	Previous Laboratory Results	Include Diagnostic PCR results using the following pattern with semi-colons as the delimiter If multiple targets, or multiple results sets are available – provide only a single assay and the lowest target available. • **Assay;Target;Ct value** If only a Ct value is known, just include it as a single number. • **24.5** If an assay only generates a Positive or Negative value, only include that • **Positive** If sequencing has been performed previously, instead include the lineage information in this column • **Pangolin Lineage**	Sample Accessioning and Processing

Requested	ES	Additional ID 1	If the specimen has been, or will be sequenced and submitted – Include the Sequencing Identifier or other ID of the strain name • **Genbank Accession** • **Genbank Strainname** • **GISAID Accession** • **GISAID Strainname** • **Submitter ID used in sequencing**	External Data Matching
Required	FD	Alpha Numeric 01	This specimen has been, or will be sequenced and submitted to public repositories by the jurisdiction – Answer • **Yes or No**	Submission
Required	FW	CDC EVENT ID	NS3 SC2 Submissions • **1771**	Sample Accessioning and Processing
Required	FX	EVENT NAME	NS3 SC2 Submissions • **NS3 - National SARS-CoV-2 Strain Surveillance**	Sample Accessioning and Processing

Appendix 4: NS3 Report Cards FAQs

Q: Why don't my numbers match what was on the report cards?

A: Many factors affect the final numbers including fringe collection/submission dates that fall close to the edges of the quarter might cause a sample to be counted for a previous/future quarter. In addition, shipping or QC issues, missing/incorrect metadata, or incomplete specimen metadata may cause a specimen to be excluded from analysis.

Q: What should I do if our numbers are significantly different than expected?

A: Start by confirming your GFAT was submitted and accepted. Next, confirm that your dates align with the given reporting period. Finally, check to see if your specimens are present in the online database you expect them to be on (GISAID/NCBI, etc.). If entire batches are missing then there was likely an issue with the submission, but if only certain samples are missing then these were likely QC failures.

Q: Who should I contact with questions and concerns regarding my submissions?

A: If large, unexplainable discrepancies are found, then contact APHL (infectious.diseases@aphl.org) for additional troubleshooting. Include context regarding expected/reported numbers, sequencing responsibility (CDC sequenced or Jurisdiction sequenced), and specimen IDs for tracking.

Q: Why is Turnaround Time (TAT) important and how is it calculated on report cards?

A: For effective surveillance of SARS-CoV-2 variants, sequences should be evaluated as quickly after collection as possible. Significant delays between specimen collection and sequencing can prevent the sequence from contributing towards sequence surveillance efforts. For the purpose of the report cards, TAT is calculated as the time between the specimen collection date and the "Received date" or "Accessioned date" (whichever is earlier). Note that the TAT isn't about the sequencing date but when CDC gains access to the sequencing results.