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 create a nationally representative system for baseline genomic surveillance. Through NS3, CDC has built a collection of representative SARS-CoV-2 specimens and sequences that laboratories use to characterize the virus and better inform public health decisions. The NS3 program provides critical information and provides a comprehensive surveillance system to track virus evolution over time and identify emerging variants.

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.

National and Jurisdictional Surveillance Targets

The NS3 program depends on jurisdictions to ship culturable viral specimens to CDC and for high quality, rapid sequencing and submission of sequence data to public repositories for relevant and timely data on circulating SARS-CoV-2 strains. Surveillance targets for each jurisdiction are driven by the need for geographically diverse specimens collected and processed at regular intervals to provide a timely report of variants and estimate viral population growth rate.

- **Phenotypic Characterization Request:** CDC requests that jurisdictions collect recent positive SARS-CoV-2 specimens and ship them bi-weekly to CDC.
    - Prioritize shipping specimens that have not been sequenced yet. Physical specimens are used for phenotypic assessment of variants to help meet jurisdictional goals of physical specimen submission; laboratories can send residual specimens that have already been sequenced in-house.
    - Specimens submitted for this component are routinely sequenced by CDC, and the CDC will submit them to public repositories on behalf of the submitter, if data were not already added by the submitting laboratory.

    - Specimens submitted by CDC on the jurisdiction's behalf will also count towards the genetic sequencing targets for a given jurisdiction.

- **Genetic Surveillance Request**: In addition to specimen shipping, CDC requests that jurisdictions collect recent specimens, generate high-quality sequences and submit data to open, public repositories to support national and jurisdictional level genomic surveillance efforts. Collaboration between federal, state, local, territorial, academic, and commercial laboratories is critical in meeting these national and state goals and improves the availability and public health impact of sequencing data.

Biweekly targets for each jurisdiction to meet the viral characterization, national, and jurisdictional surveillance goals are listed in Table 1. The targets presented in Table 1 are thresholds, additional specimen submissions and sequencing are valued and can improve novel variant detection within the jurisdiction.

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 on a bi-weekly basis. In order to meet sequencing goals, these should be specimens that have not been sequenced, however if all available specimens have already been sequenced, the laboratory should still ship them and note it on the GFAT submission form so that CDC will not resubmit 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
To meet the national genetic surveillance goal of population-level genetic surveillance, we need a system with the capacity to track variants from early in their introduction, and to recognize those that gain a growth advantage. Approximately 3000 high quality, timely SARS-CoV-2 sequences must be generated and submitted to public repositories in order to detect 1 or more rare/novel variants at a prevalence of 0.1% (1/1000) with a 95% confidence interval. Reaching the national target of 3000 specimens from a representative set of jurisdictions on a bi-weekly basis will help to ensure the

specimens collected, sequenced, and analyzed represent the population of viruses currently circulating nationally and are not heavily weighted by individual studies or outbreak events.

The bi-weekly national surveillance target number of genomes for each jurisdiction listed in Table 1 (column 3) is based on both the population invariant aspects of novel variant detection and a jurisdictional population estimate factor. National surveillance target numbers for jurisdictions are capped at 90. The national bi-weekly target of 3000 is a year-round minimum sequencing goal. This number may be challenging to achieve during times of low prevalence, and for some jurisdictions, this target number may be difficult to reach even during times of high transmission. The target numbers are provided in order to describe the data needed to power the system as stated, however, no punitive action will be taken if goals are not met. Jurisdictions are encouraged to make efforts toward reaching these goals to support national surveillance.

Jurisdictional Target Sequencing Numbers for Novel Variant Detection

The reference number for sampling the virus population within any given jurisdiction to provide a novel variant detection threshold at 3% with a confidence of 95% is 90 genomes per jurisdiction. This number is invariant to human population, so this target is the same for all jurisdictions. The bi-weekly jurisdictional surveillance target numbers in Table 1, column 4 are provided as a guide, should jurisdictions want the ability to determine the presence or absence of a variant circulating at 3% in their jurisdiction (situational awareness). Additional sequencing beyond these targets is encouraged, particularly for jurisdictions with larger populations, and will enhance the ability to determine the presence or absence of a variant circulating below 3% in their jurisdiction. The two bi-weekly sequencing target numbers for national surveillance and jurisdictional surveillance (listed in columns 3 and 4 of Table 1) represent separate goals and should not be combined or added together to form a single total. These goals can be met by sequencing within the jurisdiction, or by shipping NS3 specimens to CDC for sequencing.

Partner Data

Jurisdictions with multiple laboratories (e.g., local laboratories) or partners (including academic centers) performing SARS-CoV-2 sequencing may use that data to count towards sequencing goals.

All facilities performing sequencing should adhere to all components of High-Quality Data for Genomic Surveillance as outlined including the quality control thresholds. Jurisdictional partners should submit 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

Specimens shipped to CDC for phenotypic characterization and sequencing will be counted using CDC's internal laboratory information management systems. These counts will only be shared with individual jurisdictions and will be used to ensure feedback is available throughout the year. Unacceptable specimens (e.g., specimens that are not at an acceptable temperature upon arrival) will not count towards jurisdictional counts to meet their submission goals. Jurisdictions supporting the NS3 program can request feedback from CDC to understand their specimen counts and ensure the specimens they are submitting are contributing towards their submission goals.

For jurisdictions sequencing and submitting data, CDC will use BioProject accessions as the primary tool for counting baseline surveillance submissions.

Summary of Numbers in Table 1

In summary, CDC needs specimens in column 2 to meet our needs for virus isolation and characterization. CDC needs the sequences in column 3 to meet our goals for nowcast modeling, and the sequences can be generated within the jurisdiction or by partners. Column 4 are numbers of sequences to be generated if jurisdictions want to be able to detect novel variants at 3% threshold within their own jurisdiction, but are not needed for CDC national genetic surveillance goals.

For example, from a public health lab in a medium sized jurisdiction, CDC might request that they ship 10 specimens to CDC bi-weekly for phenotypic characterizations and generate 64 sequences per bi-week for CDC's national nowcast goals (the 10 specimens shipped will be sequenced by CDC and can count towards the total goal of 64). In addition, if the lab would like to detect a novel variant circulating around 3% within their jurisdiction, the lab may choose to reach their jurisdictional surveillance goal and generate 90 or more sequences per bi-week, instead of the national surveillance goal of 64.

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

	Bi-Weekly Shipping #	Bi-Weekly Sequencing #	
State/Territory	CDC NS3 Viral Characterization	National Surveillance*	Jurisdictional Surveillance**
Alabama	10	64	90
Alaska	10	22	90
Arizona	15	86	90
Arkansas	10	45	90
California	35	90	90
Colorado	10	71	90
Connecticut	10	51	90
Delaware	10	25	90
District of Columbia	10	22	90
Florida	25	90	90
Georgia	20	90	90
Hawaii	10	29	90
Idaho	10	32	90
Illinois	20	90	90
Indiana	10	82	90

Iowa	10	46	90
Kansas	10	44	90
Kentucky	10	59	90
Louisiana	10	62	90
Maine	10	28	90
Maryland	10	75	90
Massachusetts	15	84	90
Michigan	15	90	90
Minnesota	15	71	90
Mississippi	10	45	90
Missouri	15	76	90
Montana	10	26	90
Nebraska	10	34	90
Nevada	10	45	90
New Hampshire	10	28	90
New Jersey	15	90	90
New Mexico	10	36	90
New York	20	90	90
North Carolina	15	90	90
North Dakota	10	23	90
Ohio	20	90	90
Oklahoma	10	54	90
Oregon	10	56	90
Pennsylvania	20	90	90
Rhode Island	10	26	90
South Carolina	10	65	90
South Dakota	10	24	90
Tennessee	15	82	90
Texas	35	90	90
Utah	10	46	90
Vermont	10	21	90
Virginia	15	90	90
Washington	15	89	90
West Virginia	10	33	90
Wisconsin	15	73	90
Wyoming	10	21	90
American Samoa	5	5	90

Guam	5	5	90
Marshall Islands	5	5	90
Micronesia	5	5	90
Northern Mariana Islands	5	5	90
Palau	5	5	90
Puerto Rico	10	49	90
Virgin Islands	5	5	90
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.
** Additional sequencing beyond the jurisdictional target will enhance the ability to determine the presence or absence of a variant circulating in their jurisdiction. This value is invariant to human population and therefore the same for each jurisdiction.

Year-Round Surveillance
SARS-CoV-2 requires sustained, year-round surveillance. Notably, the targets in Table 1 will not increase during periods of high transmission. During times of high transmission, there is typically a dominant strain in circulation, so additional sequence testing is not necessary or required. Most SARS-CoV-2 variants emerge during times of low prevalence, so it is necessary to maintain routine surveillance testing even when transmission is low.

Specimen Shortages
The bi-weekly targets listed in Table 1 may not be achievable year-round, especially during times of low transmission. When there are not enough positive specimens to meet jurisdictional goals, to maintain both phenotypic characterization and sensitive genetic surveillance it is important to continue to sequence and ship as many specimens as possible on a regular (preferably 2-week) basis. Timeliness is a priority. If in-house sequencing will be delayed during times of low prevalence, jurisdictions may send specimens to CDC for sequencing.

To help meet jurisdictional shipping goals, laboratories can send residual specimens that have already been sequenced in-house to CDC for complete genotype to phenotype characterization.

- See Appendix 3 for information to include on the GFAT.
- Residual positive specimens will be sequenced at CDC but data will not be uploaded to public repositories if the specimen was already sequenced at the submitting laboratory.

Specimen Diversity
To ensure accurate national monitoring and understanding of variant proportions, it is crucial to collect a diverse set of specimens representing multiple geographic locations. Avoid multiple samples from single outbreak events, as this can bias the prevalence of the variant. This approach

helps maintain the integrity of SARS-CoV-2 surveillance by providing unbiased data on variant prevalence and growth rates, which are vital for assessing variant emergence and planning mitigation strategies. By selecting specimens from every jurisdiction and minimizing the impact of highly connected outbreaks on the overall fractional growth rate estimation, a representative set of sequences can be generated for baseline surveillance. For additional information on categorizing sequence selection for baseline surveillance, see Appendix 1. Refer to Appendix 2 for additional information on specimen selection criteria.

Turn-Around Time
It is essential that jurisdictions place priority on turn-around times and focus on sequencing and shipping the most recent samples. It is more important to sequence fewer specimens faster than to sequence many specimens slowly, or with poor data quality.

Specimen Retention
If possible, retain residual specimens for 10 weeks (about 2 and a half months) to allow for further investigation, if needed.

SARS-CoV-2 Sequencing Methods and Equipment
Laboratories have invested significant time and money purchasing equipment and developing methods and expertise to perform SARS-CoV-2 sequencing in their laboratories. There are benefits to having some diversity of methods and equipment so standardizing the equipment and protocols is not currently required, as long as the quality standards outlined in this document are met.

Components of High-Quality Data for Genomic Surveillance

Below are descriptions of five components required for generating high-quality genomic surveillance data.

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 14 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. If sequence data does not meet these quality thresholds, they are not able to be included in nowcast analysis.

- Collection Date
  - Sequences should be uploaded within 21 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 coverage
  - Complete spike gene coverage at 20X coverage per base
  - Sequences not passing the outlined spike specific criteria should be carefully reviewed to differentiate low quality sequencing events versus indications of amplicon dropout. 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 versus 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

Submission of data to public repositories post-sequencing needs to be performed rapidly to ensure the data are available to CDC prior to analysis deadlines. Submission by jurisdictions includes the following requirements:

- 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 5 days prior to the date of the next COVID Data Tracker update.
  - For instance, if the next update is Friday 7/5/2024 only data available on public repositories by 7/1/2024 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 that is uploaded into public databases and intended to be used for national surveillance efforts must be properly cataloged. SARS-CoV-2 sequences generated as part of "targeted" efforts (e.g. outbreak investigation) can bias baseline surveillance estimates so should not be tagged as baseline surveillance. Properly tagging sequence data allows the submitting facility, CDC, and other users to quickly identify unbiased samples sequenced as part of the national surveillance effort. Sequence data should be uploaded to NCBI, other databases are optional.

The preferred method of tagging relies on NCBI BioProjects. Each jurisdiction should create a BioProject for National SARS-CoV-2 Strain Surveillance with their jurisdiction's name. This BioProject should *only* contain specimens that can be considered baseline surveillance.

- These should be associated with CDC's Umbrella BioProject: **PRJNA1113573**.
- Please email bioprojecthelp@ncbi.nlm.nih.gov to associate the local BioProject.
- Review Appendix 1 for additional detailed instructions.

Consensus level data may also be tagged as "baseline surveillance" for inclusion in CDC analyses using the guidance in Appendix 1.

The use of standard and consistent tagging information on the submitted sequences improves CDC's ability to search, analyze, and share the data generated across all jurisdictions. Additionally, it enables reproducible analyses using the same, identifiable dataset.

While only the consensus sequence and metadata are required for submission, submission of raw reads to NCBI's SRA is encouraged. Please email bioprojecthelp@ncbi.nlm.nih.gov to enable automated human read scrubbing for your BioProject prior to upload of raw reads.

Database naming conventions

Sequences are named according to their geographical collection location, as per established conventions (e.g., SARS-CoV-2/human/USA/XX (state acronym)-[Jurisdiction ID]-xxxxxxxx (unique identifier)/YYYY). If specimens are shipped to CDC, then CDC will be included in the name to reference that it was sequenced at CDC and not by the state public health laboratory or other entity. If the specimen was sequenced locally, CDC should not be included in the name.

If submitting to GISAID as well as NCBI, note that there are different naming conventions:

ICTV (NCBI):

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

e.g. SARS-CoV-2/human/USA/XX-CDC-9898989/2021

GISAID:

hCoV-19/country/isolate/date

e.g. hCoV-19/USA/XX-CDC-9898989/2021

In these examples "XX" represents the location, or two letter state abbreviation, of collection location, and "CDC" should be replaced with an abbreviation indicating where it was sequenced. "9898989" should be replaced with a meaningful strain/isolate ID.

Proficiency Panels

Proficiency panels are a useful tool to ensure quality standards are being met by jurisdictions performing sequencing. Proficiency panels for SARS-CoV-2 sequencing assays are a known need and may be provided in the future to help evaluate the laboratory, metadata, and computational processes to ensure end-to-end quality is being maintained for the sequence data generated. In silico proficiency panels may be used to test bioinformatic workflows, including the use of publicly available raw read sets to compare generated consensus sequences.

Questions and Technical Guidance

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

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

Baseline surveillance is achieved by sequencing specimens that represent geographic, demographic (e.g., age), and clinical (e.g., disease severity or outcome) diversity across a jurisdiction through a random selection of SARS-CoV-2-positive, diagnostic specimens.

The goals of including public health laboratory sequencing data in CDC's National SARS-CoV-2 Genomic Surveillance are to:

1. Allow for more robust state-level estimates of circulating SARS-CoV-2 lineages, including variants of interest and concern.
2. Monitor viral evolution within jurisdictions at a more granular level.
3. Provide comprehensive data for public health decision makers at the jurisdictional and national levels
4. More accurately represent sequencing efforts and contributions made by jurisdictions to the broader scientific community

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 across targeted sequencing efforts to be representative of the community

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 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 FASTA file. 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 in 2 hours (average) 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]
   CTAGCTAGCTAGCTAGCTAGCTAGCTAG
   >Seq2 [keyword=purposeofsampling:baselinesurveillance]
   CTAGCTAGCTAGCTAGCTAGCTAGCTAG
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

As a courtesy, jurisdictions also uploading to GISAID should select "Baseline surveillance" 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. Ideally, specimens should have an RT-PCR Ct value of ≤28.
        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 14 days whenever possible.
        ii. If the number of specimens collected within the last 14 days is insufficient to meet your jurisdiction's requested number for NS3 specimen submission (Table 1), please send the most recent specimens possible (i.e., collected up to 28 days prior to shipment).
    c. Select a diverse set of specimens that represent multiple geographic locations not associated with a single outbreak event.
    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 goal (Table 1).
        i. If submitting specimens being sequenced 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 randomly 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 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 form's column "Alpha Numeric 01" 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).
   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 sarsseqshipping@cdc.gov.
   k. Please include a printed manifest of your specimens with your shipment.

l. If possible, please ship specimens on dry ice bi-weekly Monday for overnight delivery using your usual courier, such as FedEx or UPS, to CDC to the following address:

ATTN: STATT Lab: Unit 66 COORS
Centers for Disease Control and Prevention
1600 Clifton Road, NE
Atlanta, Georgia, 30329
Telephone: 404-639-3931
Email: sarsseqshipping@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 6 template, effective date of September 16, 2024.

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	"Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)"	Determining Sample Acceptability
Required	G	Date Sent to CDC	Date Samples shipped to CDC	Sample Accessioning and Processing
Required	H	At CDC, bring to the attention of	STATT Lab: Unit 66 COORS	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 (Confirm with GS's): • 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