

1. **PURPOSE:** to describe a standardized procedure for submitting a biosample to NCBI.

2. **SCOPE:** this procedure applies to all clinical strains that are whole genome sequenced under PulseNet surveillance.

3. **DEFINITIONS:**
 - 3.1. **Bioproject:** a collection of biological data related to a single initiative, originating from a single organization or a consortium
 - 3.2. **Biosample:** a description of biologically or physically unique specimens (=sequenced bacterial strains)
 - 3.3. **WGS ID:** Whole Genome Sequencing Identification number
 - 3.4. **CDC:** Centers for Disease Control and Prevention
 - 3.5. **NCBI:** National Center for Bio**te**chnology Information
 - 3.6. **SRA:** Sequence Read Archive
 - 3.7. **PFGE:** Pulse-Field Gel Electrophoresis
 - 3.8. **SAMN number:** an accession number assigned to each biosample (strain) by NCBI after biosample file has been submitted.
 - 3.9. **QA/QC:** Quality Assurance/Quality Control
 - 3.10. **SOP:** Standard Operating Procedure
 - 3.11. **FTP:** File Transfer Protocol
 - 3.12. **FASTQ:** a text-based format for storing both a biological sequence and its corresponding quality scores

4. **RESPONSIBILITIES:**
 - 4.1. All genomic sequences generated under PulseNet surveillance are uploaded in real time to the SRA located at NCBI. Before the sequences can be uploaded to SRA, a biosample file containing the minimum metadata for the sequenced strains needs to be submitted to the NCBI.
 - 4.2. Laboratory personnel performing the sequencing are responsible for submitting the biosample to NCBI and entering the SAMN numbers into the national PFGE BioNumerics databases before uploading the raw sequence read data (fastq-files) to the Illumina BaseSpace or the CDC ftp site. They are also responsible for notifying the PulseNet database managers at PFGE@cdc.gov that the raw data is available for SRA upload at BaseSpace/ftp site.

5. **PROCEDURE:**
 - 5.1. Before the very first submission:
 - 5.1.1. E-mail Eija Trees at EHyytia-Trees@cdc.gov to request an invitation to the NCBI PulseNet submission group
 - 5.1.2. **If you already have an account at NCBI:** you will receive an e-mail invitation to join the submission group. Click on the link in the invitation to join the submission group. You are now ready to submit a biosample through the PulseNet submission group.

- 5.1.3. **If you don't have an account at NCBI:** you will receive an e-mail invitation to join the submission group. Click on the link in the invitation. It will take you to the NCBI Sign In Page.
 - 5.1.3.1. Click on "Register for an NCBI account"
 - 5.1.3.2. In the "Register for an NCBI Account" page, fill out the required information (indicated by an asterisk) and click "Create account".
 - 5.1.3.3. You will be redirected to the NCBI Submission Portal
 - 5.1.3.4. Fill out all personal information fields. You will not be able to save the page until you perform the following steps:
 - 5.1.3.4.1. You will receive an email instructing you to activate your new NCBI account.
 - 5.1.3.4.2. Click the link in the email. This will take you to a page informing you that your email address has been confirmed and that your account is now active.
 - 5.1.3.5. Once you have completed filling out the fields and have saved them, sign out of the My NCBI page and return to the PulseNet group invitation email.
 - 5.1.3.6. Click the link again and click "Accept" in the new window. You will now be redirected to the Submission Portal homepage and are ready to submit a biosample through the PulseNet submission group.
- 5.1.4. Download the biosample batch submission templates at <https://submit.ncbi.nlm.nih.gov/biosample/template/> and save on your computer as txt files. (**NOTE:** the submission templates are also available in the PulseNet QA/QC manual as related documents to this SOP)
 - 5.1.4.1. There are three separate batch submission templates:
 - 5.1.4.1.1. Submissions containing clinical (human) isolates only: Pathogen.cl.1.0
 - 5.1.4.1.2. Submissions containing environmental, food or other pathogens only: Pathogen.env.1.0
 - 5.1.4.1.3. Submissions containing both clinical and non-clinical isolates: Pathogen.combined.1.0
 - 5.1.4.2. Open the templates in Excel, delete the top ten rows containing instructions leaving the 11th row containing information field headings in row 1. Save each template as an Excel file for future use.
- 5.2. Fill out the required fields in the appropriate NCBI submission template
 - 5.2.1. The fields that need to be filled out are indicated by an asterisk
 - 5.2.1.1. For clinical isolates, the PulseNet required minimum metadata fields are:
 - 5.2.1.1.1. Sample_name: WGS ID for the sequenced strains (this same ID can be entered also into the "Strain" field)
 - 5.2.1.1.2. Bioproject:
 - 5.2.1.1.2.1. *E. coli* and *Shigella spp.*: 218110
 - 5.2.1.1.2.2. *Salmonella spp.*: 230403
 - 5.2.1.1.2.3. *Listeria monocytogenes*: 212117
 - 5.2.1.1.2.4. *Campylobacter spp.*: 239251
 - 5.2.1.1.2.5. *Vibrio spp.*: 266293

- 5.2.1.1.3. Organism:
 - 5.2.1.1.3.1. “*Escherichia coli*” for all *E. coli* serotypes; *Shigella sonnei*, *Shigella flexneri*, *Shigella dysenteriae* or *Shigella boydii* for all *Shigella* serotypes
 - 5.2.1.1.3.2. “*Salmonella enterica*” for all *Salmonella* serotypes
 - 5.2.1.1.3.3. *Listeria monocytogenes*
 - 5.2.1.1.3.4. *Campylobacter jejuni*, *Campylobacter coli*, or *Campylobacter spp.* (for non-jejuni & non-coli species or the exact species if known)
 - 5.2.1.1.3.5. *Vibrio cholera*, *Vibrio parahaemolyticys*, *Vibrio vulnificus*, or *Vibrio spp.*
- 5.2.1.1.4. Collected by: “CDC” or “Missing”
- 5.2.1.1.5. Geo_Loc_Name: “USA” or the appropriate country if non-US strain
- 5.2.1.1.6. Attribute_Package (combined submission only): “pathogen.cl” for clinical (human) isolates, “pathogen.env” for non-clinical isolates
- 5.2.1.1.7. For any other required field: “Missing”
- 5.2.2. Save as an Excel file and a txt file (tab delimited text file) using the standardized file naming format: PulseNet lab ID_Organism_year-month-date, for example CDC_Salmonella_2014-07-21
 - 5.2.2.1.**NOTE:** Only the txt file is used for uploading.
- 5.3. Log into the NCBI submission portal at <https://submit.ncbi.nlm.nih.gov/subs/>
- 5.4. Click on “BioSample” on the upper right hand corner
- 5.5. Click on the “New submission” button
- 5.6. In the “Submitter” tab, replace your first and last name with the word “PulseNet”, then under the “Select group for this submission” section select “PulseNet Next Generation Subtyping Methods Unit”, and finally uncheck “Update my contact information in profile” check box next to the “Continue” button
- 5.7. Click on the “Continue” button
- 5.8. In the “General info” tab, keep the default option “Release immediately following curation (recommended)” and select “Batch/Multiple BioSamples”
- 5.9. Click on the “Continue” button
- 5.10. In the “Sample type” tab, select “Pathogen affecting public health” and then select one of the three appearing options:
 - 5.10.1. “Clinical or host-associated pathogen” – select this option if all isolates in this biosample submission are of human origin
 - 5.10.2. “Environmental, food or other pathogen” - select this option if all isolates in this biosample submission are of non-human origin
 - 5.10.3. “Combined pathogen submission” - select this option if this biosample submission contains isolates of both human and non-human origin
 - 5.10.3.1. **NOTE:** When preparing the biosample file for a combined submission, make sure to specify either “Pathogen.cl” (for human isolates) or “Pathogen.env” (for non-human isolates) in the “attribute_package” field
- 5.11. Click on the “Continue” button

- 5.12. In the “Attributes” tab, browse to the location of the biosample file, select the file and click “Open”
- 5.13. Click on the “Continue” button
- 5.14. In the “Comments” tab, you can provide comments to NCBI staff if you so wish or leave the field blank
- 5.15. Click on the “Continue” button
- 5.16. In the “Overview” tab review the submission information for accuracy
 - 5.16.1. **NOTE:** the Submitter should be “PulseNet PulseNet”
- 5.17. Click on the “Submit” button. This will take you back to the “Submissions” tab in the Submission portal. Your submission should be listed on top with “Pulsenet” appearing as the “Group” and “Submitted” as the “Status”.
- 5.18. You will receive an email from NCBI listing the SAMN numbers for the sequenced strains in the body of the email and in the attached biosampleobjects.txt file.
- 5.19. Enter the SAMN numbers in the “NCBI accession” information field in the appropriate national PFGE database

6. FLOW CHART:

7. BIBLIOGRAPHY:

8. CONTACTS:

- 8.1. CDC PFGE Inbox: PFGE@cdc.gov
- 8.2. Eija Trees: EHyytia-Trees@cdc.gov

9. AMENDMENTS: