

1. **PURPOSE:** To describe the procedure for certifying PulseNet personnel to enable full participation in PulseNet WGS activities, including on-line access to the PulseNet National Databases.
2. **SCOPE:** This procedure applies to all PulseNet personnel generating sequence data, checking data quality, and/or performing analysis and uploading to the PulseNet databases and NCBI using BioNumerics.
3. **DEFINITIONS/TERMS:**
 - 3.1 **Allele:** One of two or more alternative forms of a gene that arise by mutation and are found at the same place on a chromosome
 - 3.2 **Analysis-Certified:** An individual who is certified for checking WGS data quality, performing analysis and uploading to the PulseNet National Database and NCBI using BioNumerics
 - 3.3 **APHL:** Association of Public Health Laboratories
 - 3.4 **BAP:** Blood Agar Plate
 - 3.5 **BaseSpace:** Illumina cloud-based computing environment for next generation sequencing data analysis and management, including data sharing
 - 3.6 **BHIA:** Brain Heart Infusion Agar
 - 3.7 **BioNumerics:** Analysis software used by PulseNet, developed by Applied Maths (Sint-Martens-Latem, Belgium)
 - 3.8 **Bioproject:** a collection of biological data related to a single initiative, originating from a single organization or a consortium
 - 3.9 **Biosample:** Submitter-supplied descriptive information (metadata) about the biological materials from which the NCBI stored data are derived.
 - 3.10 **Bundle file:** A packed BioNumerics file with a .bdl extension that contains a selection of database entries along with one or more experiments and information fields, as specified by the user
 - 3.11 **Calculation Engine (CE):** Calculation Engine. A server application on a high performance computing cluster at CDC that is used by the BioNumerics client applications for doing calculation-intensive tasks, including de novo assembly, reference mapping and whole genome multi-locus sequence typing (wgMLST) allele calling.
 - 3.12 **CDC:** Centers for Disease Control and Prevention
 - 3.13 **cgMLST:** Core Genome Multi-Locus Sequence Typing
 - 3.14 **Certification file evaluator:** An individual who evaluates certification files
 - 3.15 **Certification files:** Fastq and/or excel document (completed WGS analysis template) files submitted by PulseNet participants for WGS certification evaluation
 - 3.16 **CLRW:** Clinical Laboratory Reagent Water
 - 3.17 **Core genome:** Genes shared by a vast majority (e.g., >99% of strains *Listeria monocytogenes* strains) of strains of the same species
 - 3.18 **Critical Quality metrics:** Average de novo coverage, average quality (Q score), assembly length, secondary species abundance (contamination), percent core present, allele and hqSNP difference compared to reference (wet lab certification only), average read length (wet lab certification only). Failing to meet any one of these metrics will result in rejection of the certification submission.

- 3.19 De Novo Assembly:** A sequence assembly generated from the short raw reads without the use of a reference genome
- 3.20 DNA:** Deoxyribonucleic Acid
- 3.21 FASTQ:** a text-based format for storing both a biological sequence and its corresponding quality scores
- 3.22 Fastq-Certified:** An individual who is certified in laboratory methods for WGS
- 3.23 FTP:** File Transfer Protocol; a standard network protocol used to transfer computer files from one host to another host via Internet
- 3.24 hqSNP:** High Quality Single Nucleotide Polymorphism
- 3.25 MB:** Mega Base
- 3.26 NCBI:** National Center for Biotechnology Information, part of the National Institutes of Health (NIH). NCBI houses several databases relevant to biotechnology, including GenBank for DNA sequence assemblies and Sequence Read Archive (SRA) for raw reads.
- 3.27 Organism-Specific Database:** A BioNumerics database, v 7.6 or higher, used for comparing isolate sequences for surveillance. Part of the standard PulseNet workflow.
- 3.28 PulseNet Area Laboratory:** Laboratory, designated by CDC, which has agreed to assume responsibility for additional PulseNet duties for laboratories within their support zone. The current Area Laboratories include CO, MI, MN, NY, TN, VA and WA.
- 3.29 PulseNet Central:** PulseNet team at CDC comprising of the Outbreak Detection and Surveillance Unit (PulseNet@cdc.gov) and the Next Generation Subtyping Methods Unit (PulseNetNGSlab@cdc.gov)
- 3.30 QA/QC:** Quality assurance/quality control
- 3.31 RefID Database:** A BioNumerics database, v 7.6 or higher, used for quality control of raw sequence data, assembly of sequences and species identification. Part of the standard PulseNet workflow.
- 3.32 SOP:** Standard Operating Procedure
- 3.33 WGS Certification Report:** A report that contains the evaluation and results of the participant's WGS certification files
- 3.34 WGS:** Whole Genome Sequencing

4. RESPONSIBILITIES:

- 4.1 Individuals performing PulseNet WGS-related work must submit certification file(s) and have them reviewed before being able to submit raw reads to the PulseNet bioprojects housed at NCBI or perform BioNumerics analyses using the CDC high performance computing resources and upload to the PulseNet National Databases.
- 4.1.1 Submitted certification files must document the submitter's highest level of competence in producing raw sequence data and/or performing analysis and checking data quality in BioNumerics.
- 4.1.2 Individuals can be certified for *Salmonella*, *Escherichia/Shigella*, *Listeria*, *Campylobacter*, and *Vibrio* in one of three ways:
- 4.1.2.1 **Fastq-Certified only.** An individual can generate fastq files and submit them to an analysis-certified person for subsequent quality check, analysis and uploading to the PulseNet bioprojects housed at NCBI and the PulseNet National Database.
- 4.1.2.2 **Analysis-Certified only.** An individual can analyze fastq files generated by fastq-certified individuals and upload data that pass the quality thresholds required by

PulseNet to the PulseNet bioprojects housed at NCBI and the PulseNet National Database.

4.1.2.3 **Both Fastq- and Analysis-Certified.** An individual can generate and analyze fastq files, and upload data that pass the quality thresholds required by PulseNet to the PulseNet bioprojects housed at NCBI and the PulseNet National Database.

4.1.3 For each PulseNet organism, at least one person from each PulseNet participating laboratory should be fastq-certified and one person should be analysis-certified. A person can be both fastq- and analysis-certified. Laboratories cannot have analysis-certified personnel without fastq-certified personnel; fastq certification must occur before or concurrent with analysis certification.

4.2 Individuals performing PulseNet-related WGS work at CDC laboratories must submit certification files(s) and have them reviewed before being able to submit fastq files, perform quality checks, analyze the data, or upload data to the PulseNet National Databases.

4.2.1 All PulseNet fastq files generated at CDC must be uploaded directly from a client database that is housed on the CDC network. Those at CDC without access to a client database on the network drive may not upload fastq files.

5. PROCEDURE:

5.1 PulseNet participants request the WGS laboratory certification strain set (if they do not already have it) and/or analysis certification material (a bundle file and submission template) from CDC (PulseNet@cdc.gov).

5.1.1 The WGS Laboratory Certification set includes *Escherichia*, *Salmonella*, *Listeria*, *Campylobacter* and *Vibrio*.

5.1.1.1 CDC sends the certification set and the WGS Certification SOP (PNQ08) with detailed instructions on how to reconstitute the cultures, perform the certification run, perform the quality check and submit results (fastq files) to CDC. See Appendices PNQ08-1 and PNQ08-2 for details.

NOTE: *Escherichia* certification also covers *Shigella*

5.1.1.2 Acceptable library preparation kits are the following: Illumina Nextera XT (PNL34), Illumina DNA Prep (PNL35), Qiagen QIAseq FX (PNL36), KAPA HyperPlus (PNL37) and NEBNext Ultra II FS (PNL42).

5.1.1.3 Acceptable sequencing platforms are the following: Illumina MiSeq (PNL38), Illumina MiniSeq (PNL39), Illumina iSeq (PNL40) and Illumina NextSeq.

NOTE: sequencing chemistries **below 300 cycles are not accepted.**

5.1.2 The WGS Analysis Certification currently includes *Escherichia*, *Salmonella*, *Listeria*, *Campylobacter* and *Vibrio*.

5.1.2.1 Upon receiving a request for WGS Analysis Certification, CDC will send the requestor the WGS Certification SOP (PNQ08) that includes supplemental instructions for submitting WGS analysis certification results (Appendix PNQ08-3), an organism specific bundle file (except for *Vibrio*) and an analysis submission template. An example of the analysis submission template is provided in Appendix PNQ08-4.

5.1.2.2 Participants are asked to review critical quality metrics provided by CDC to identify which entries should be submitted to the organism-specific database, correctly identify NCBI bioproject and which metadata should be submitted to NCBI, correctly format the metadata to be submitted to the national database and perform local cluster detection on the contents of an organism-specific bundle file (quality and wgMLST experiments). Using the provided submission template, all results and responses must be emailed to CDC (PulseNet@cdc.gov) for review.

NOTE1: *Escherichia* certification also covers *Shigella*

NOTE2: *Vibrio* allele database is not yet available so no cluster detection section (V) is included for *Vibrio*.

5.1.2.3 Analysis certification must be performed using BioNumerics version 7.6 or higher. **It is strongly recommended that individuals seeking to become analysis certified consult/work with the more experienced analysis certified personnel in their lab so that no routine database settings or experiment types are affected when the certification bundle files are opened (appendix PND08-3).**

5.2 Submitters will receive a WGS Certification Report summarizing the results of their certification file(s) evaluation (See Appendix PNQ08-5 for an example of a WGS Certification Report).

5.2.1 If the submitted certification file passes the certification evaluation, the submitter is considered certified as long as they remain in their current laboratory and that laboratory successfully completes the annual proficiency testing (SOP PNQ09). All fastq-certified individuals should perform routine sequencing at least every six months.

5.2.1.1 Upon becoming WGS Analysis Certified, the person who submitted the file will be issued a SecurID device (key fob) if they do not already have one, credentials to access the NCBI uploader and password to access the CDC high performance computing resources (calculation engine) for WGS data analysis and to upload data to the national database.

5.2.1.2 If a person relocates to a different PulseNet laboratory, they must be re-certified for fastq file generation. Analysis certification can be transferred over to the new laboratory as long as someone else in the new laboratory is fastq-certified and the arrangement is acceptable to the laboratory leadership.

5.2.1.3 If a person leaves the PulseNet laboratory and then returns to the same laboratory within six months, no re-certification is required for fastq file generation or analysis. Persons returning after six months of absence are required to get re-certified for both fastq file generation and analysis.

5.2.1.4 Supervisors of PulseNet laboratories must notify PulseNet Central about any personnel changes by emailing to PulseNet@cdc.gov.

5.2.2 If the submitted certification files do not pass the certification evaluation, the individual will need to review the troubleshooting comments received from the evaluator and resubmit once results have improved.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 08

Effective Date:

Page 5 of 20

5.2.2.1 If the submitter fails certification three times, the individual will not be allowed to submit again for six months. Before resubmitting, the individual will be expected to work with CDC and/or their area laboratory until satisfactory results are achieved. This includes but is not limited to troubleshooting and training in the PulseNet WGS protocols.

5.2.3 If a new PulseNet Participating Laboratory does not submit certification files and routine data within six months of joining PulseNet, CDC and/or APHL will contact them to check on their status.

6. FLOW CHART: N/A**7. RELATED DOCUMENTS**

Document Number	Title
PNL33	DNA Extraction and QC SOP
PNL34	Illumina Nextera XT Library Preparation SOP
PNL35	Illumina DNA Prep Library Preparation SOP
PNL36	Qiagen QIAseq FX Library Preparation SOP
PNL37	KAPA HyperPlus Library Preparation SOP
PNL38	Sequencing on the MiSeq SOP
PNL39	Sequencing on the MiniSeq SOP
PNL40	Sequencing on the iSeq SOP
PNL42	NEBNext Ultra II FS Library Preparation SOP
PNQ07	Illumina Sequence Data QC SOP
PND20	BioNumerics RefID Database Workflow SOP
PND05	BioNumerics Surveillance Database Workflow SOP
PND06	Cluster Detection SOP
SharePoint: Library of PulseNet Documents/BioNumerics Data Management/Fields and Metadata	Source Definition
SharePoint: Library of PulseNet Documents/BioNumerics Data Management/Fields and Metadata	BN Dictionary
SharePoint: Library of PulseNet Documents/Quick Tips/Database_BioNumerics/WGS	PulseNet WGS Cluster Coding Guidelines

8. REFERENCES: N/A

9. CONTACTS:

9.1 PulseNet Outbreak Detection and Surveillance Team

PulseNet@cdc.gov

(404) 639-4558

9.2 PulseNet NGS Laboratory

Pulsenetngslab@cdc.gov**10. AMENDMENTS:**

- 10.1 1/26/2016 : Updated appendix PNQ08-1 with instructions for reconstituting Campylyophils, previously live culture stabs.
- 10.2 10/25/2016 : Added instructions for WGS analysis certification and updated the PulseNet email address to PulseNet@cdc.gov.
- 10.3 02/15/2017 : Included reference to “Sequencing Quality Metrics in BN 7” document on SharePoint in Step 8 of Appendix 8-3.
- 10.4 03/23/2018: Updated SOP to include Certification of *Vibrio* and added appendix PNQ08-5 to show an example of the evaluation template. Added 5.4.1.1 indicating WGS certified personnel need to request access to PulseNet Biosample Submission Group on NCBI.
- 10.5 10/12/2018: Updated SOP to include *Escherichia*, *Salmonella* and *Campylobacter* WGS Analysis Certification. Updated WGS Analysis Certification Submission Template, WGS Certification Report and WGS Quality Metrics Table.
- 10.6 08/28/2019: Updated SOP to include *Vibrio* analysis certification. Removed contact information for Eija Trees and entered general contact info for PulseNet NGS Laboratory.
- 10.7 12/23/2019: Updated instructions and templates to remove the Fastq.gz file data quality check section from the analysis certification. Fastq.gz files will be replaced with a RefID quality check. Files will no longer be pulled from the PulseNetQA FTP site, they will be emailed to the requestor. Edited *Vibrio parahaemolyticus* certification strain number. Updated Appendix PNQ08-6.
- 10.8 04/11/2022: Updated the laboratory certification section to include acceptable library prep kits and sequencing platforms. Updated the analysis certification section to include formatting of metadata to be submitted to the national database and local cluster detection. Included language on what happens to the certification if the certified individual transfers to a different laboratory or leaves and then returns to the same laboratory. Added the Related Documents section. Updated the supplemental instructions for the analysis certification (PNQ08-3) to reflect the updated and new components. Median insert size was removed as a critical quality metric and replaced by average read length (updates to PNQ08-5 and PNQ08-6).

11. APPROVAL SIGNATURES:

Approved By: _____ Date: _____
QA/QC Personnel

Approved By: _____ Date: _____
PulseNet Outbreak Detection and Surveillance Unit Chief

Approved By: _____ Date: _____
PulseNet Next Generation Subtyping Methods Unit Chief

Approved By: _____ Date: _____
PulseNet Reference Outbreak Surveillance Team Lead

Appendix PNQ08-1**Laboratory Protocol to Reconstitute Lyophilized (Freeze-Dried) *E. coli* O157:H7, *Salmonella enterica* serotype Heidelberg, *Listeria monocytogenes*, *Campylobacter jejuni* and *Vibrio parahaemolyticus* Cultures**

Biological Safety Warning: *E. coli* O157:H7, *S. enterica* serotype Heidelberg, *L. monocytogenes*, *C. jejuni* and *V. parahaemolyticus* strains are considered Level 2 biological agents by the U.S. Department of Health and Human Services. Use appropriate precautions when handling the vial or culture. Carry out laboratory work in a biological safety cabinet when applicable to ensure aseptic conditions and personal safety.

Materials/Equipment Needed:

- Sterile sturdy forceps
- 1 ml pipetman
- 1 ml sterile pipet tips
- 1 µl and 10 µl sterile inoculating loops
- Incubator (37°C; microaerobic conditions for *Campylobacter*)

Reagents Needed:

- Trypticase Soy + 5% Sheep Blood Agar plates (BAP) or equivalent media
- Brain Heart Infusion agar plates (BHIA) for *Listeria*
- Brain Heart Infusion agar + 5% Rabbit Blood or equivalent for *Campylobacter*
- Sterile Clinical Laboratory Reagent Water (CLRW) or 0.85% sterile saline (*Vibrio parahaemolyticus*)
- 70% isopropyl alcohol

Procedure for Reviving Cultures:***Salmonella, E. coli, Listeria and Vibrio* lyophilized cultures:****Day 1**

1. Document the isolate number(s) and the date(s) lyophilized for your records. Wipe the aluminum cover and outside of the vial with isopropyl alcohol. Using sturdy forceps, aseptically remove the aluminum cover and rubber stopper from the vial containing the lyophilized culture. Wipe the outside of the rubber stopper and neck of the vial with isopropyl alcohol before removing the stopper.
2. Re-suspend the lyophilized cells with 1.0 ml of CLRW (0.85% sterile saline for *Vibrio parahaemolyticus*). Allow to stand for a few minutes and/or mix gently to produce a uniform suspension. With a 10 µl inoculating loop, streak a small amount of this suspension onto a blood agar plate (or BHIA plate for *Listeria*) and incubate at 37°C overnight in aerobic conditions.

Days 2 and 3

1. Check the plate to confirm the culture is viable and appears to be pure. If the package does not arrive in satisfactory condition or if cultures are non-viable please contact PulseNet QA/QC Team (PulseNet@cdc.gov). If the culture appears pure, pick an isolated colony from the first plate, and inoculate a second fresh plate. Incubate at 37°C overnight.
2. Use the growth from this plate to make DNA templates of the certification strains. Transfer culture to a fresh medium and incubate at 37°C overnight; this will ensure that the same culture can be retested, if necessary.
3. **Freeze (-70°C) or stock these strains according to your laboratory's policy within one week of receiving them to ensure that your laboratory will have stock cultures of this PulseNet certification set for future use, including the PulseNet certification of additional personnel.**

Campylobacter* lyophilized culture:*Day 1**

1. Document the isolate number(s) and the date(s) lyophilized for your records. Wipe the aluminum cover and outside of the vial with isopropyl alcohol. Using sturdy forceps, aseptically remove the aluminum cover and rubber stopper from the vial containing the lyophilized culture. Wipe the outside of the rubber stopper and neck of the vial with isopropyl alcohol before removing the stopper.
2. Re-suspend the lyophilized cells with 250 µl of CLRW. Allow to stand for a few minutes and/or mix gently to produce a uniform suspension. Pipet 100 µl of *Campylobacter* cell suspension onto a BHI 5% Rabbit Blood agar plate and with a 10 µl loop streak for growth. Incubate microaerobically at 37°C or 42°C for 48 hours.

Days 2 and 3

1. Check the plate to confirm the culture is viable and appears to be pure. If the package does not arrive in satisfactory condition or if cultures are non-viable please contact PulseNet QA/QC Team (PulseNet@cdc.gov). If the culture appears pure, pick an isolated colony from the first plate, and inoculate a second fresh plate for heavy growth; incubate microaerobically for 24 hours at 37°C or 42°C.
2. Use the growth from this plate to make DNA templates of the certification strain. Transfer culture to a fresh medium and incubate at 37°C to 42°C microaerobically overnight; this will ensure that the same culture can be retested, if necessary.
3. **Freeze (-70°C) or stock these strains according to your laboratory's policy within one week of receiving them to ensure that your laboratory will have stock cultures of this PulseNet certification set for future use, including PulseNet certification of additional personnel.**

Appendix PNQ08-2

Supplemental Instructions for PulseNet WGS Laboratory Certification

The WGS Lab Certification Set contains five (5) lyophilized culture vials.

Please let us know if this package does not arrive in satisfactory condition, or if the cultures are not viable. **Make a stock culture (freeze at -70°C) of each of the strains according to your laboratory's policy within 1 week from receiving them. Long-term storage of these cultures will ensure the availability of the PulseNet certification set for future use, including WGS certification of additional personnel.**

The strain numbers of the cultures are as follows:

- 03-98 Shiga toxin-producing *Escherichia coli* O157:H7
 - C6472 *Salmonella enterica* serovar Heidelberg
 - H8394 *Listeria monocytogenes*
 - D5480 *Campylobacter jejuni*
 - K3430* *Vibrio parahaemolyticus*
- *Alternate ID F5828, please use K3430 to label your submission

Please follow these supplemental instructions for testing the certification isolates by WGS. Refer to the specific SOPs for detailed instructions.

1. Perform the DNA extraction (PNL33), library preparation, normalization and pooling (PNL34 for Nextera XT, PNL35 for Illumina DNA Prep, PNL36 for QIAseq, PNL37 for KAPA HyperPlus, PNL42 for NEBNext Ultra) and loading of the cartridge (PNL38, PNL39, PNL40) following the instructions of the standardized protocol.
 - 1.1. Files must be named according to the PulseNet standardized naming scheme outlined in SOPs PNL34, PNL35, PNL36, PNL37 and PNL42). For certification, this includes **strain number, initials, LabID, instrument ID and run date, i.e. 03-98jk-GA-M0947-220624.**
 - 1.2. It is acceptable to include routine isolates on the same cartridge with the certification strains in order not to waste reagents. **NOTE:** do not upload the routine sequences to the National Database and NCBI until you know you have passed the certification.
 - 1.3. Perform quality check on the fastq files following the procedures outlined in PNQ07 and the quality thresholds outlined in Appendix PNQ08-7. **NOTE:** Do not share the fastq files until you know that the files pass the quality check. It is recommended to initially stream the files to a separate QC project on BaseSpace that is not shared with CDC. Passing files can then be copied to a project shared with CDC.
2. Transfer the fastq files to CDC for evaluation either using the Illumina BaseSpace or the PulseNet1 FTP-site. In BaseSpace, the sample data can be shared with PulseNet@cdc.gov using an existing organism-specific project already shared with CDC. Alternatively, you can

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 08	Effective Date:	Page 11 of 20

set up a separate project for certification named “LabID_certification”. Separate instructions will be provided for accessing the PulseNet1 FTP-site. Email PulseNet@cdc.gov to request the access credentials.

3. Send an email to the PulseNet inbox (PulseNet@cdc.gov) with “WGS Certification” in the subject line to notify PulseNet Central of the certification files’ availability. In the email, please include, fastq filenames, BaseSpace project name (if shared through BaseSpace), and information on which library prep kit and sequencing platform was used.
4. After the fastq files are submitted, the PulseNet certification file evaluator will analyze the files and inform your laboratory of your results (“Satisfactory” or “Needs Improvement”) within four weeks of receiving the files. If the certification files are satisfactory, the person who submitted the file(s) will be eligible to send fastq files to PulseNet Central or to an in-house analysis-certified individual for analysis, and upload to the National Database and NCBI. If the submitted certification files are not satisfactory, the individual will need to review the troubleshooting comments received from the evaluator and resubmit once results have improved. If the submitter fails certification three times, the individual will not be allowed to submit again for six months. Before resubmitting, the individual will be expected to work with CDC or his/her PulseNet Area Laboratory until satisfactory results are achieved.

Thank you,

PulseNet Central

CDC

PulseNet@cdc.gov

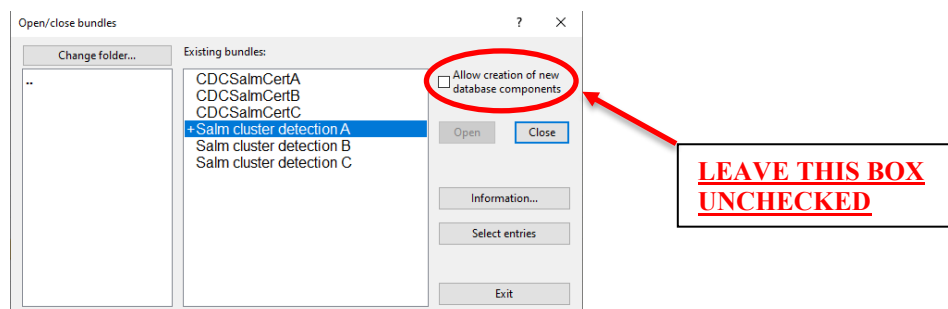
(404) 639-4558

Appendix PNQ08-3
Supplemental Instructions for PulseNet WGS Analysis Certification

After at least one person in your laboratory has been certified for generating fastq files (either prior to analysis certification or concurrently) according to the directions in Appendix PNQ08-2, individuals may submit files for analysis certification using BioNumerics 7.6 or higher.

Please follow these supplemental instructions for submitting WGS analysis certification results.

1. Upon sending a request to PulseNet@cdc.gov for the WGS analysis certification, you will receive the SOP PNQ08 (WGS Certification), an organism-specific bundle file (except for *Vibrio*) and a submission template (example template Appendix PNQ08-4). There are multiple certification sets, please use the set that is emailed to you.
2. Review the critical quality metrics listed in section I of the submission template. Based on the quality, determine if you would import the data to the organism-specific BioNumerics database. Please refer to PNQ08-6 for minimum quality thresholds and/or acceptable ranges. If you determine that the quality of the sequence data is acceptable, note this in the submission template for the appropriate isolate. If you determine that the sequence quality is not acceptable, then you must cite your reasoning (i.e., which quality metrics failed) in the submission template.
3. In section II provide the name of the organism-specific PulseNet BioProject to which you would upload the data.
4. Section III of the submission template contains three questions pertaining to Metadata Submission to NCBI. The first two questions contain metadata for one clinical strain (Table 1) and one non-clinical strain (Table 2). For both strains, determine what metadata is required for submission to NCBI by reviewing the pre-filled NCBI biosample submission templates. Select the NCBI submission template (A, B, C, or D) that correctly depicts how information should be set-up in BioNumerics. For the third question, determine what information should not be shared on NCBI. Refer to the PulseNet SOP PND05 (Organism-Specific Surveillance Database Workflow) Appendix PND05-5 to complete all three questions in this section.
5. Metadata for three strains are provided in section IV, Table 3. Enter the metadata in Table 4 in the format required for PulseNet national database upload. Please refer to the "Source Definitions" PDF and the "BN7 Data Dictionary" excel file available on the PulseNet SharePoint site within the Library of PulseNet Documents/BioNumerics Data Management/Fields and Metadata.
6. To complete section V, **consult/work with the individuals already certified in the BioNumerics analysis in your laboratory so that no experiment types or information fields in your laboratory's organism-specific databases get adversely affected when the analysis certification bundle files are opened.** Save the organism-specific bundle file in the "Bundles" folder under your laboratory's organism-specific database. Go to "Files" → "Open bundle". Click "Change folder", navigate to the location of the file, highlight it when it is viewable in your "Existing bundles" window and choose "Open" → "Select entries". **NOTE: DO NOT** check the box for "Allow creation for new database components".



7. Create a comparison by clicking on the green plus sign in the comparison window (Alt+C).
8. Click on the eye next to the quality experiment in the "Experiments" panel. Under "Experiment Data", click on the first 123 icon (Show values) and change "Character name" to "Character Description". Now you can see the quality metrics for the isolates selected. Proceed to cluster detection with all sequences that pass the critical quality metrics.
9. Create a dendrogram following the instructions outlined in the SOP PND06 (Cluster Detection). Make the assumption that all isolates have been collected within the past 60 or 120 (*Listeria* only) days. Interpret the findings in accordance with the PulseNet general cluster detection rules for cgMLST data. A document giving guidance for PulseNet cluster detection is available on SharePoint: "20190924_PulseNet WGS cluster coding guidelines" (Library of PulseNet Documents/Quick Tips/Database_BioNumerics/WGS). In section V, please record the results for each of the isolates under the "Cluster Name" column as follows: if the isolate belongs to a cluster, name the detected cluster A, B, C, etc; if the isolate does not belong to a detected cluster, enter "NA" (not applicable). Copy and paste the dendrogram in the designated space on the submission template.
10. Once you have completed your submission template, please save the file as a PDF with your initials at the end (Ex. **Submission Template Salmonella_A_JK**) and submit the template for evaluation to the PulseNet Inbox (PulseNet@cdc.gov). In the email, please include "WGS Analysis Certification" in the subject line.

After the submission template is emailed to CDC, the PulseNet certification file evaluator will analyze the submission and inform you of your results ("Satisfactory" or "Needs Improvement") within four weeks of receiving the files. If the results are satisfactory, the person who submitted the file will be issued a SecurID device (key fob, if you do not already have one), credentials to access the NCBI uploader and password to access the CDC high performance computing resources (calculation engine) for WGS data analysis and to upload data to the national database. SecurID devices cannot be shared and must be returned to CDC if the certified individual leaves his or her position in the laboratory. If the submitted certification files are not satisfactory, the individual will need to review the troubleshooting comments received from the evaluator and resubmit once results have improved. If the submitter fails certification three times, the individual will not be allowed to submit again for six months. Before resubmitting, the individual will be expected to work with CDC or his/her PulseNet Area Laboratory until satisfactory results are achieved.

Appendix PNQ08-4

Example: WGS Analysis Submission Template

Note: This is an example submission template. Please use the appropriate (organism specific) template emailed to you upon certification set request.

Section I. <i>Salmonella</i> BioNumerics Critical Quality Check:			
WGS_ID	BioNumerics Critical Quality Metrics	Would you import this entry into the <i>Salmonella</i> database? Yes/No	If you would NOT import this data to the <i>Salmonella</i> database, you must explain why by listing any failed quality metrics.
PNUSASCert1	<u>RefID Database:</u> - Genus/species: <i>Salmonella enterica</i> - Avg. quality: 36.6 - Secondary species abundance: 0.0 - Avg. denovo coverage: 32.6 - Length: 4820981 <u>Salmonella Database:</u> Percent Core: 95%		
PNUSASCert2	<u>RefID Database:</u> Genus/species: <i>Salmonella enterica</i> Avg. quality: 35.5 Secondary species abundance: 0.0 Avg. denovo coverage: 51.3 Length: 4692124 <u>Salmonella Database:</u> Percent Core: 89%		
PNUSASCert5	<u>RefID Database:</u> Genus/species: <i>Salmonella enterica</i> Avg. quality: 35.8 Secondary species abundance: 0.0 Avg. denovo coverage: 13.5 Length: 4713665 <u>Salmonella Database:</u> Percent Core: 65%		
PNUSASCert7	<u>RefID Database:</u> Genus/species: <i>Salmonella enterica</i> Avg. quality: 27.8 Secondary species abundance: 0.3 Avg. denovo coverage: 103.7 Length: 4933641 <u>Salmonella Database:</u> Percent Core: 82%		
Section II. <i>Salmonella</i> fastq.gz files BioProject selection: What is the PulseNet BioProject for <i>Salmonella</i> ? _____			

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 08

Effective Date:

Page 15 of 20

Section III. Salmonella NCBI Metadata Check:

Please review the metadata provided in the two tables below and select which NCBI submission settings template (A, B, C, or D) correctly depicts how information should be set-up in BioNumerics. Refer to the PulseNet SOP for the Organism-Specific Surveillance Database Workflow, Appendix PND05-5 to complete this section.

Question 1:

Table 1: Salmonella Metadata

Strain (WGS_ID)	Organism	Collected By	Geographic origin (Source Country)	Host (Source Type)	Isolation Source (Source Site)	Collection Date (IsolatDate)
PNUSASert1	Salmonella	GA	USA	Human	blood	1/10/2022

Please review the choices below and enter the template that is correctly filled out here (A, B, C or D) _____

A. NCBI submission settings ? X

BioSample (continued)
Information fields to be used as BioSample metadata.

Collected by: CDC

Collection / Isolation date: IsolatDate (field content)

Collection / Isolation date format: YYYY

Geographical origin: USA

Geographical coordinates: <Missing>

Isolation source: <Missing>

Host: <Missing>

Host disease: <Missing>

< Back Next > Cancel

B. NCBI submission settings ? X

BioSample (continued)
Information fields to be used as BioSample metadata.

Collected by: CDC

Collection / Isolation date: IsolatDate (field content)

Collection / Isolation date format: YYYY-mm-dd

Geographical origin: USA

Geographical coordinates: <Missing>

Isolation source: SourceType (field content)

Host: <Missing>

Host disease: <Missing>

< Back Next > Cancel

C. NCBI submission settings ? X

BioSample (continued)
Information fields to be used as BioSample metadata.

Collected by: LabID (field content)

Collection / Isolation date: IsolatDate (field content)

Collection / Isolation date format: YYYY-mm

Geographical origin: USA:GA

Geographical coordinates: <Missing>

Isolation source: <Missing>

Host: <Missing>

Host disease: <Missing>

< Back Next > Cancel

D. NCBI submission settings ? X

BioSample (continued)
Information fields to be used as BioSample metadata.

Collected by: LabID (field content)

Collection / Isolation date: IsolatDate (field content)

Collection / Isolation date format: YYYY

Geographical origin: USA:GA

Geographical coordinates: <Missing>

Isolation source: SourceSite (field content)

Host: <Missing>

Host disease: <Missing>

< Back Next > Cancel

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 08

Effective Date:

Page 16 of 20

Question 2:

Table 2: Salmonella Metadata

Strain (WGS_ID)	Organism	Collected By	Geographic origin (Source Country)	Host (Source Type)	Isolation Source (Source Site)	Collection Date (IsolatDate)
PNUSASert2	Salmonella	MA	USA	Food	Eggs	1/5/2022

Please review the choices below and enter the template that is correctly filled out here (A, B, C or D) _____

A. NCBI submission settings ? X

BioSample (continued)
Information fields to be used as BioSample metadata.

Collected by: LabID (field content) v

Collection / Isolation date: IsolatDate (field content) v

Collection / Isolation date format: YYYY v

Geographical origin: USA:MA v

Geographical coordinates: <Missing> v

Isolation source: SourceSite (field content) v

Host: <Missing> v

Host disease: <Missing> v

< Back Next > Cancel

B. NCBI submission settings ? X

BioSample (continued)
Information fields to be used as BioSample metadata.

Collected by: CDC v

Collection / Isolation date: IsolatDate (field content) v

Collection / Isolation date format: YYYY v

Geographical origin: USA v

Geographical coordinates: <Missing> v

Isolation source: <Missing> v

Host: <Missing> v

Host disease: <Missing> v

< Back Next > Cancel

C. NCBI submission settings ? X

BioSample (continued)
Information fields to be used as BioSample metadata.

Collected by: LabID (field content) v

Collection / Isolation date: IsolatDate (field content) v

Collection / Isolation date format: YYYY-mm v

Geographical origin: USA:MA v

Geographical coordinates: <Missing> v

Isolation source: SourceSite (field content) v

Host: <Missing> v

Host disease: <Missing> v

< Back Next > Cancel

D. NCBI submission settings ? X

BioSample (continued)
Information fields to be used as BioSample metadata.

Collected by: CDC v

Collection / Isolation date: IsolatDate (field content) v

Collection / Isolation date format: YYYY-mm v

Geographical origin: USA:MA v

Geographical coordinates: <Missing> v

Isolation source: TypeDetails (field content) v

Host: <Missing> v

Host disease: <Missing> v

< Back Next > Cancel

Question 3:

What type of information should not be shared on NCBI?

Answer: ____

- A. Allele Codes
- B. Outbreak Codes
- C. Neither Allele Codes or Outbreak Codes

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 08

Effective Date:

Page 17 of 20

Section IV: Metadata Cleanup for BioNumerics

The following database entries must be "cleaned up" so that the metadata is standardized and correctly entered into the BioNumerics database. Please refer to the Source definitions PDF and the BN7 Data Dictionary excel file available on the PulseNet SharePoint site within the Library of PulseNet Documents/BioNumerics Data Management/Fields and Metadata.

Table 3. Salmonella metadata

Salm11	Food	Lettuce	bagged/shredded		1/5/2022
Salm22		Stool	Female	36 months old	1/20/2022
Salm33	Human	Stool	sex: M	67 years old	1/12/2022

Table 4. Enter the metadata listed above in the format it should be uploaded to BioNumerics:

Key	SourceType	SourceSite	TypeDetails	Patientageyears	Patientagemonths	Patientagedays	Patientsex	Isolatdate

Section V: Cluster Detection

Please open the provided bundle file in your organism specific database under the supervision of an experienced WGS Analysis Certified individual. See Appendix PNQ08-3 Step 6 for details on ensuring no experiment files/ information fields are adversely affected while opening bundle files. Proceed to check the quality metrics and perform cluster detection by using methods available to you locally (as outlined in the SOP PND06) and make the assumption that all isolates have been collected within the past 60 days. Interpret the findings in accordance with the PulseNet general cluster detection rules for cgMLST data. A document giving guidance for PulseNet cluster detection has been posted on SharePoint: "20190924_PulseNet WGS cluster coding guidelines" (Library of PulseNet Documents/Quick Tips/Database_BioNumerics/WGS). Please record the results for each of the isolates under the "Cluster name" column as follows: if the isolate belongs to a cluster, name the detected cluster A, B, C, etc; if the isolate does not belong to a detected cluster, enter "NA" (not applicable).

SampleID	Cluster Name	Comments
A. PNUSASCert16		
B. PNUSASCert17		
C. PNUSASCert18		
D. PNUSASCert19		
E. PNUSASCert20		

Insert Dendrogram Below:

Appendix PNQ08-5

Example: WGS Certification Report Template

Note: Fastq (wet lab) certification evaluation reports will only have fastq sections on the report form. Analysis certification evaluation reports will only have analysis sections on the report form. The report form below has both fastq and analysis sections condensed into one report form to show examples of each section.

Lab ID – WGS

Enteric Diseases Laboratory Branch
 CDC/NCEZID/DFWED
 WGS Fastq File Certification Report
 Certification Strains for QA/QC



Certification Set:
Laboratory Submitting Files:
Submitted by:
Fastq File Names:
Sequencer Type:
Date Fastq Files Received by CDC:

2022 WGS Fastq Certification Set: *Escherichia*, *Salmonella*, *Listeria*, *Campylobacter*, and *Vibrio*

<i>Escherichia coli</i> O157:H7 strain 03-98		
QC statistic	QC value (acceptable value)	Comments
Average quality	(min 30)	
Average coverage	(min 40x)	
Genus/species/serotype	<i>Escherichia coli</i> O157:H7	
Average Read Length	(300 cycle chemistry: ≥135 bp; 500 cycle chemistry: ≥225 bp)	See <i>Escherichia</i> Insert Size Histogram Below
Secondary species abundance	(≤ 1.0)	
No. of hqSNPs compared to reference	(< 1/ MB)	
No. of allele differences compared to reference	(≤ 3)	
% core present	(≥ 85%)	
Assembled genome size	(5.5MB, 5% error margin)	

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 08

Effective Date:

Page 19 of 20

<i>Escherichia analysis</i>		
WGS Analysis Element	Pass/Fail	Comments
BioNumerics Critical Quality Check		
BioProject Selection		
NCBI Metadata Submission		
Metadata Cleanup for BioNumerics		
Cluster Detection		

(Wet Lab Certification only – Insert size Histograms)

Escherichia Insert Size Histogram:

Salmonella Insert Size Histogram:

Listeria Insert Size Histogram:

Campylobacter Insert Size Histogram:

Vibrio Insert Size Histogram:

QA/QC Certification Result:

Escherichia Fastq files: Satisfactory Needs Improvement

Escherichia analysis file: Satisfactory Needs Improvement

Salmonella Fastq files: Satisfactory Needs Improvement

Salmonella analysis file: Satisfactory Needs Improvement

Listeria Fastq files: Satisfactory Needs Improvement

Listeria analysis file: Satisfactory Needs Improvement

Campy Fastq files: Satisfactory Needs Improvement

Campy analysis file: Satisfactory Needs Improvement

Vibrio Fastq files: Satisfactory Needs Improvement

Vibrio analysis file: Satisfactory Needs Improvement

Certification is valid as long as laboratory successfully completes annual proficiency tests.

Comments or Suggestion(s) for Improvement:

Certification Evaluation Performed By: _____

Date of Evaluation:

Approved By: _____

Date of Approval:

Appendix PNQ08-6**Critical Quality Metric Thresholds/Acceptable Ranges for PulseNet Certification**

Organism	Average read length ¹ (bp)	hqSNP difference compared to reference ²	Average denovo coverage	Average quality	Assembly length (MB) ³	Secondary species abundance	% Core present ⁴	Allele difference compared to reference
<i>Listeria monocytogenes</i>	135/225	< 1/MB	≥ 20x	≥ 30	2.8-3.2	≤ 1.0	≥ 95	≤ 3
<i>E. coli</i> (most serotypes)	135/225	< 1/MB	≥ 40x	≥ 30	4.9-6.0	≤ 1.0	≥ 85	≤ 3
<i>Shigella</i> spp./Rare <i>E. coli</i>	135/225	< 1/MB	≥ 40x	≥ 30	4.2-4.9	≤ 1.0	≥ 85	≤ 3
<i>Salmonella</i> spp.	135/225	< 1/MB	≥ 30x	≥ 30	4.4-5.7	≤ 1.0	≥ 85	≤ 3
<i>Campylobacter</i> spp.	135/225	< 1/MB	≥ 20x	≥ 30	1.4-2.2	≤ 1.0	≥ 85 ⁵	≤ 3
<i>Vibrio cholerae</i>	135/225	< 1/MB	≥ 40x	≥ 30	3.8-4.3	≤ 1.0	NA ⁶	NA ⁶
<i>Vibrio parahaemolyticus</i>	135/225	< 1/MB	≥ 40x	≥ 30	4.9-5.5	≤ 1.0	NA ⁶	NA ⁶
<i>Vibrio vulnificus</i>	135/225	< 1/MB	≥ 40x	≥ 30	4.7-5.3	≤ 1.0	NA ⁶	NA ⁶

- Rows shaded with the same color represent organisms transferred to the same organism-specific database.
- ¹Average read length and ²hqSNP differences are critical quality metrics for WGS Lab Certifications only, not for proficiency testing or routine submissions.
- ³The sequence assembly lengths for each species listed in the table apply for routine work. For fastq certification, the assembly lengths are strain-specific based on the reference sequence generated by CDC with ± 5% margin of error.
- ⁴The goal for percent core genome present is 95% or above. For certification, a percent core of 85% or above will pass for all other species *except Listeria* if all other required quality thresholds are met.
- ⁵85% threshold applies to *C. jejuni* only. Other *Campylobacter* species may still pass quality with % core present <85%; please contact PulseNet@cdc.gov before resequencing.
- ⁶A cgMLST scheme for *Vibrio* is still under development.