

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 1 of 22

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 National Databases and NCBI using PN 2.0.
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 PN 2.0.
 - 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 **Bioproject:** a collection of biological data related to a single initiative, originating from a single organization or a consortium.
 - 3.8 **CDC:** Centers for Disease Control and Prevention
 - 3.9 **cgMLST:** Core Genome Multi-Locus Sequencing Typing
 - 3.10 **Certification file evaluator:** An individual who evaluates certification files.
 - 3.11 **Certification files:** Fastq and/or excel/pdf (completed PN 2.0 Workflow Knowledge Test & Analysis Submission Template) files submitted by PulseNet participants for WGS certification evaluation.
 - 3.12 **CLRW:** Clinical Laboratory Reagent Water
 - 3.13 **Core genome:** Genes shared by a vast majority (e.g., >99% of all *Listeria monocytogenes* strains) of strains of the same species.
 - 3.14 **Coverage:** The average number of reads that include a given nucleotide in the reconstructed sequence.
 - 3.15 **Critical Quality metrics:** Average de novo coverage, average quality (Q score), assembly length, secondary species coverage (contamination determined by MIDAS), percent core present, cgMLST allele difference compared to reference, and average read length. Failing to meet any one of these metrics will result in rejection of the certification submission.
 - 3.16 **De Novo Assembly:** A sequence assembly generated from the short raw reads without the use of a reference genome.
 - 3.17 **DNA:** Deoxyribonucleic Acid
 - 3.18 **FASTQ:** a text-based format for storing both a biological sequence and its corresponding quality scores.
 - 3.19 **Fastq-Certified:** An individual who is certified in laboratory methods for WGS.
 - 3.20 **FTP:** File Transfer Protocol; a standard network protocol used to transfer computer files from one host to another host via Internet.
 - 3.21 **MB:** Mega Base
 - 3.22 **Metadata:** a set of data that describes and gives information about other data.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 2 of 22

- 3.23 MIDAS:** Metagenomic Intra-species Diversity Analysis System. Integrated computational pipeline for quantifying bacterial species abundance and coverage from shotgun metagenomes based on blast alignment against a panel of universal single copy genes.
- 3.24 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.
- 3.25 Organism-Specific Database:** A PN 2.0 database, used for comparing isolate sequences for surveillance. Current organism databases include *Campylobacter*, *Escherichia/Shigella*, *Listeria*, *Salmonella* and *Vibrio*.
- 3.26 PulseNet 2.0 (PN 2.0):** Customized cloud-based (cloud service provider Microsoft Azure) start-to-end automated sequence analysis workflow utilized by PulseNet USA. The workflow is composed of open-source tools that are integrated in a cohesive pipeline within a NextFlow framework.
- 3.27 Production Environment:** PN 2.0 environment that currently includes the organism-specific surveillance databases for *Campylobacter*, *Escherichia/Shigella*, *Listeria*, *Salmonella* and *Vibrio* and the Reference_ID database. Each database includes local views and the national database view and is used to perform the routine surveillance work and submissions to the National Database and NCBI.
- 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.
- 2.29 PulseNet Central:** PulseNet team at CDC comprising of the PulseNet Response and Outbreak Management Team (PulseNet@cdc.gov) and the WGS Core Laboratory Activity (PulseNetNGSlab@cdc.gov).
- 3.30 PulseNet/OutbreakNet SharePoint:** A closed, web-based collaboration application used for communication among PulseNet participants.
- 3.31 QA/QC:** Quality assurance/quality control
- 3.32 Q score:** The sequence quality score for each individual base position in a sequence, indicating the accuracy of the base call. Phred scores are used, where $Q = -10\log(\text{Error Probability})$. The higher the quality score, the more reliable the base call. A Q30 means a 1 in 1000 likelihood of an incorrect base call at that position.
- 3.33 Read:** A unit of continuous DNA sequence (based pairs) derived by sequencing a part of the fragmented target DNA.
- 3.34 Reference_ID Database:** A PN 2.0 database that can be used to initiate the PN 2.0 analysis pipeline. Sequences that are identified as one of the PulseNet surveillance species that have organism-specific databases are automatically transferred to those databases. Species that do not currently have an organism-specific database are housed in the Reference_ID database.
- 3.35 SAMS:** Secure Access Management Services
- 3.36 SOP:** Standard Operating Procedure
- 3.37 STG Environment:** PN 2.0 environment that includes the following organizations: PulseNet Training, PulseNet Certification, PulseNet Proficiency Testing and PulseNet Validation. Each organization further includes the organism-specific databases and the Reference_ID database. All functionalities in STG environment are the same as in the Production Environment. New users are granted access to the STG environment first to familiarize themselves with the PN 2.0 workflow and to perform analyses for the PulseNet

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 3 of 22

certifications. After passing the PulseNet analysis certification, new users are granted access to the Production Environment and the National Databases.

3.38 WGS Certification Report: A report that contains the evaluation and results of the participant's WGS certification files (Separate Reports for Lab and Analysis Certifications).

3.39 WGS: Whole Genome Sequencing

4. RESPONSIBILITIES:

- 4.1 All individuals performing PulseNet WGS-related work must pass a certification before being able to generate fastq files for PulseNet surveillance or have access to the PN 2.0 databases to perform sequence analyses and upload to the PulseNet National Databases or PulseNet bioprojects at NCBI.
 - 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 PN 2.0.
 - 4.1.2 Individuals can be certified in one of four 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 (full read-write PN 2.0 access):** 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.
 - 4.1.2.4 **Certification for Read-Only Access to PN 2.0:** An individual can create views and comparisons for cluster detection and export line lists and dendrograms. The read-only access does not allow for metadata entry or edit, submission of sequences to pipelines or upload of sequences to the national databases or NCBI.
 - 4.1.3 At least one person from each PulseNet participating laboratory should be fastq-certified and one person should be analysis-certified.
 - 4.1.3.1 Laboratories cannot have analysis-certified personnel without fastq-certified personnel; fastq certification must occur before or concurrent with analysis certification.
 - 4.1.3.2 A single fastq certification set covers all current PulseNet organisms: *Campylobacter*, *Cronobacter*, *Escherichia*, *Listeria*, *Salmonella*, *Shigella*, *Vibrio*, *Yersinia*. See section 5.1.2 for details on what the fastq certification set includes.
 - 4.1.3.2.1 A laboratory can opt out from *Campylobacter* certification if it does not perform *Campylobacter* culture. The remaining organisms **are required** for the fastq certification.
 - 4.1.3.2.2 If a PulseNet Participating Laboratory wishes to outsource its sequencing work to an external laboratory, individuals performing the sequencing in the contract laboratory must also be PulseNet certified before initiating the work.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 4 of 22

4.1.3.3 A single analysis certification grants access to all current PulseNet databases: *Campylobacter*, *Escherichia/Shigella*, *Listeria*, *Salmonella*, *Vibrio* and Reference_ID.

4.1.4 A laboratory is considered fastq- and analysis-certified as long as at least one person in the laboratory has passed each certification.

4.1.4.1 A certified laboratory must participate and pass the annual PulseNet Proficiency Test (PT) in order to maintain its certification status.

5. PROCEDURE:

5.1 PulseNet participants request the WGS laboratory certification strain set (if they do not already have it) and/or analysis certification materials (instructions and Analysis Submission Template) from CDC (PulseNet@cdc.gov).

5.1.1 In the certification request email, include the following information:

5.1.1.1 Name of the individual seeking certification and contact information (email and phone).

5.1.1.2 Position in the laboratory.

5.1.1.3 Copy the laboratory's main PulseNet contact on the certification request email.

5.1.2 **The WGS Fastq (Laboratory) Certification set** includes one strain each of

Escherichia, *Listeria*, *Campylobacter* and *Vibrio*.

NOTE1: *Escherichia* also covers the certifications for *Salmonella* and *Shigella*. *Vibrio* also covers the certifications for *Cronobacter* and *Yersinia*.

NOTE2: A laboratory can opt out from the *Campylobacter* certification if it does not perform *Campylobacter* culture.

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

5.1.2.2 Any DNA extraction method, including automated DNA extracts, are acceptable as long as the DNA quality meets the requirements outlined in the DNA extraction and quality check SOP PNL33.

5.1.2.2.1 The analyst attempting certification needs to perform/set up DNA extractions personally, including automated DNA extractions. Using previously prepared stored DNA extracts is not acceptable.

5.1.2.3 Acceptable library preparation kits are the following: Illumina Nextera XT (PNL34), Illumina DNA Prep (PNL35, PNL44), Qiagen QIAseq FX (PNL36), KAPA HyperPlus (PNL37), NEBNext Ultra II FS (PNL42), and Illumina DNA PCR-Free (PNL43).

NOTE: only manual library preps are accepted for certification submission.

5.1.2.4 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.

5.1.2.5. The generated fastq files must be shared with CDC for evaluation using the Illumina BaseSpace or the PulseNet ftp site following the instructions in Appendix PNQ08-2.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 5 of 22

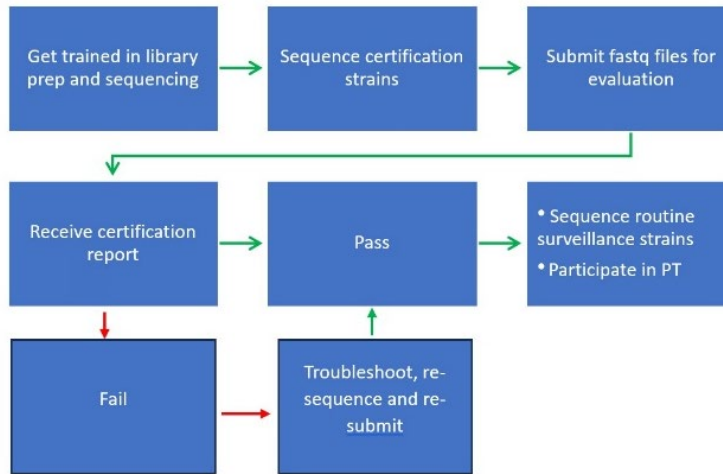
- 5.1.3 **The WGS Analysis Certification** includes a multiple-choice PN 2.0 Workflow Knowledge Test and a set 10 sequences downloadable from NCBI SRA to be analyzed through PulseNet 2.0 Certification Staging Environment.
- 5.1.3.1 Individuals seeking analysis certification must specify in their certification request whether they require **full read-write access or read-only access**. Certification for Read-Only Access will only include the PN 2.0 Workflow Knowledge Test.
- 5.1.3.1.1 Any epidemiologists seeking full read-write access to PN 2.0 must submit their request with the approval of their laboratory's PulseNet program manager/main PulseNet contact.
- 5.1.3.2 Upon receiving a request for WGS Analysis Certification, CDC will submit a request for SAMS access (if requestor does not already have access) as well as access to the PN 2.0 STG Environment and the PulseNet SharePoint site.
- 5.1.3.2.1 New users will receive an email from SAMS requesting to register an account and create a password.
- 5.1.3.2.2 After registering they will receive a second email acknowledging account activation in addition to a separate email from SAMS that includes a link to the STG Environment.
- 5.1.3.3 CDC will email the requestor the WGS Certification SOP (PNQ08) that includes supplemental instructions for performing the analysis certification in PN 2.0 Staging Environment and submitting the results (Appendix PNQ08-3), the PN 2.0 Workflow Knowledge Test, and the Analysis Submission Template. An example of the Analysis Submission Template is provided in Appendix PNQ08-4.
- 5.1.3.4 The analysis certification includes three test modules:
- 5.1.3.4.1 A multiple-choice PN 2.0 Workflow Knowledge Test
- 5.1.3.4.2 Sequence Quality Check (5 sequences)
- 5.1.3.4.3 Cluster Analysis (5 sequences)
- 5.1.3.5 Using the provided PN 2.0 Workflow Knowledge Test and the Analysis Submission Template, all results and responses must be emailed to CDC (PulseNet@cdc.gov) for review with "WGS Analysis Certification" included in the subject line.
- 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 files pass 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).
- 5.2.1.1 All fastq-certified individuals should perform routine sequencing at least every six months.
- 5.2.1.2 Upon becoming WGS Analysis Certified, the person who submitted the files will be granted access to the PN 2.0 Production Environment.
- 5.2.1.2.1 The newly certified analyst will receive an email from SAMS containing a link to the Production Environment.
- 5.2.1.3 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.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 6 of 22

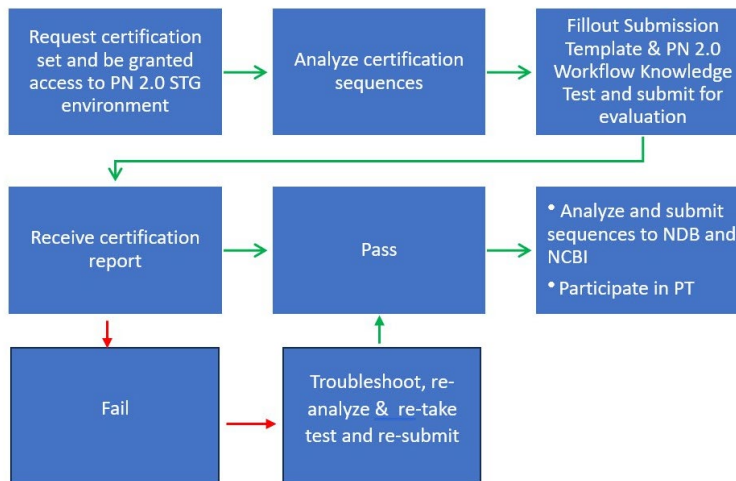
- 5.2.1.4 If a person leaves the PulseNet laboratory and then returns to the same laboratory within six months, no re-certification is required for fastq file generation or analysis. Persons returning after six months of absence are required to get re-certified for both fastq file generation and analysis.
- 5.2.1.5 Supervisors of PulseNet laboratories must notify PulseNet Central about any personnel changes by emailing 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.
 - 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

6.1 Fastq certification workflow



6.2 Analysis certification workflow



7. RELATED DOCUMENTS:

Document Number	Title
PNL33	DNA Extraction and QC SOP
PNL34	Illumina Nextera XT Library Prep SOP
PNL35	Illumina DNA Prep Library Prep SOP
PNL36	Qiagen QIAseq FX Library Prep SOP
PNL37	KAPA HyperPlus Library Prep SOP
PNL38	Sequencing on the MiSeq SOP
PNL39	Sequencing on the MiniSeq SOP

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 8 of 22

PNL40	Sequencing on the iSeq SOP
PNL42	NEBNext Ultra II FS Library Prep SOP
PNL43	Illumina DNA PCR-Free Library Prep SOP
PNL44	PulseNet Rapid Library Prep SOP
PNQ07	Illumina Sequence Data QC SOP
PNQ09	PulseNet Proficiency testing SOP
SharePoint: Library of PulseNet Documents/PN 2.0 Training	PN2.0 User Manual
SharePoint: Library of PulseNet Documents/PulseNet 2.0 Reference Documents	PN2.0 Dictionary

8. REFERENCES: N/A

9. CONTACTS:

9.1 PulseNet Outbreak Detection and Surveillance Team

PulseNet@cdc.gov

(404) 639-4558

9.2 PulseNet NGS Core Activity

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.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 9 of 22

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).

10.9 10/25/2024: Switched from organism-specific fastq and analysis certifications to a single fastq and a single analysis certification covering sequencing all PulseNet organisms and access to all PulseNet databases, respectively. Reduced the size of the fastq certification panel from 5 to 4 strains by dropping *Salmonella*; *E. coli* now covers also both *Salmonella* and *Shigella* certifications and *Vibrio* covers also certifications for *Cronobacter* and *Yersinia*. Added PNL43 (Illumina PCR-Free) and PNL44 (PulseNet Rapid Prep) as acceptable library preps. Made sharing the certification fastq files in a separate BaseSpace project with standardized project name **a requirement** (instead of using the surveillance projects). For WGS Analysis Certification, removed BioNumerics and added PN 2.0 and changed the workflow and instructions accordingly. hqSNP difference compared to the reference sequence was removed as a critical quality metric for certifications.

11. APPROVAL SIGNATURES:

Approved By: _____ Date: _____
QA/QC Personnel

Approved By: _____ Date: _____
PulseNet Response and Outbreak Management Team Lead

Approved By: _____ Date: _____
PulseNet WGS Technical Lead

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

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 10 of 22

Appendix PNQ08-1

Laboratory Protocol to Reconstitute Lyophilized (Freeze-Dried) *E. coli* O157:H7, *Listeria monocytogenes*, *Campylobacter jejuni* and *Vibrio parahaemolyticus* Cultures

NOTE1: *If the package does not arrive in satisfactory condition or if cultures are non-viable please contact PulseNet QA/QC Team (PulseNet@cdc.gov).*

NOTE2: *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 the PulseNet certification set for future use, including the PulseNet certification of additional personnel. Avoid completely thawing frozen stock and multiple passes in order to minimize possibility of compromising genetic stability of culture.*

Biological Safety Warning: *E. coli* O157:H7, *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:

***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.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 11 of 22

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 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 the second 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.

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 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 the second 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.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 12 of 22

Appendix PNQ08-2

Supplemental Instructions for PulseNet WGS Fastq Certification

The WGS Fastq Certification Set contains four (4) lyophilized culture vials.

NOTE1: *Please let us know (Pulsenet@cdc.gov) if this package does not arrive in satisfactory condition, or if the cultures are not viable.*

NOTE2: *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 the cultures will ensure the availability of the PulseNet certification set for future use, including WGS certification of additional personnel. Avoid completely thawing frozen stock and multiple passes in order to minimize possibility of compromising genetic stability of culture.*

The strain numbers of the cultures are as follows:

03-98 Shiga toxin-producing *Escherichia coli* O157:H7
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 or PNL44 for Illumina DNA Prep, PNL36 for QIAseq, PNL37 for KAPA HyperPlus, PNL42 for NEBNext Ultra, PNL43 for Illumina PCR-Free Prep) and loading of the cartridge (PNL38 for MiSeq, PNL39 for MiniSeq, PNL40 for iSeq) following the instructions of the standardized protocol.

NOTE1: *NextSeq is also an acceptable sequencing platform.*

NOTE2: *Automated library preps are not accepted for certification. Automated DNA extracts are acceptable if they pass QC metrics outlined in PNL33. The analyst attempting certification needs to perform/set-up the DNA extractions personally. Use of previously prepared stored DNA is not acceptable.*

- 1.1. Files must be named according to the PulseNet standardized naming scheme outlined in SOPs PNL34, PNL35, PNL36, PNL37, PNL42, PNL43, PNL44. For certification, this includes **strain number, initials, LabID, instrument ID and run date, i.e. 03-98jk-GA-M0947-2401024.**
- 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 PNQ08-6. **NOTE:** *Do not*

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 13 of 22

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.
 - 2.1. In BaseSpace, the certification fastq files **must be** shared in a separate project for certification named “LabID_certification”.
 - 2.2. 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 were 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.
 - 4.1. If the certification files are satisfactory, the person who submitted the file(s) will be eligible to send fastq files to an in-house analysis-certified individual for analysis, and upload to the National Database and NCBI.
 - 4.2. 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.
 - 4.3. 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

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 14 of 22

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 seek analysis certification using PN 2.0.

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 be granted access to the PN 2.0 STG Environment where you may familiarize yourself with the PN 2.0 workflow and then perform the analyses required for PulseNet analysis certification.
 - 1.1. New users will receive an email from SAMS requesting to register an account and create a password.
 - 1.2. After registering they will receive a second email acknowledging account activation in addition to a separate email from SAMS that includes a link to the STG Environment.
2. Upon sending a request to PulseNet@cdc.gov for the WGS analysis certification, you will also receive via email the SOP PNQ08 (WGS Certification), the PN 2.0 Workflow Knowledge Test and the Analysis Submission Template (example template Appendix PNQ08-4) for analysis certification containing the SRA accession numbers for the 10 sequences you need to analyze for the certification. There are multiple certification sets, please use the set that is emailed to you.
3. Answer the multiple-choice questions in the **PN 2.0 Workflow Knowledge Test**. Once you have completed the test, please save this file with your initials added to the end of the file name (e.g., PN2.0 workflow test_ET) and return to the PulseNet inbox (pulsenet@cdc.gov) together with the filled out Analysis Submission Template (steps 4 and 5 below).
4. **Sequence QC Module** (Section I in the Submission Template, PNQ08-4):
 - 4.1. In the PN 2.0 STG Environment, select “PulseNet Certification” organization. Go to the “Reference_ID” database, and then “Pipelines”, “Link Files”, and “NCBI SRA”.
 - 4.2. Copy and paste the 5 SRA accession numbers belonging to the Sequence QC Module to the “Search NCI SRA” box. Remove spaces and add the pipe character (|) between the accession numbers. Click on “Search”. **NOTE:** *This is the preferred and recommended method for linking the files for the certification sequences. Importing the SRR numbers to the database first using a spreadsheet would necessitate entering the required minimum metadata. Using the recommended method, no metadata entry is required.*

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 09

Effective Date:

Page 15 of 22

Link Files

1 Select Files 2 Import Rules 3 Link Files 4 Import Metadata

Q SRR7722335|SRR7722358 Search

0 Selected Samples

4.3. Select the sequences returned by the search and click “Next”.

Link Files

1 Select Files 2 Import Rules 3 Link Files 4 Import Metadata

Q SRR7722335|SRR7722358 Search

2 Selected Samples

✓ SRR7722358 2018C-4267:cwgs
Size: 202.48MB Published: Mon Aug 20 2018

✓ SRR7722335 2018C-4267:g:wgss
Size: 217.22MB Published: Mon Aug 20 2018

Displaying 1 to 2 of 2 results.

Back to File Source Select

Next

4.4. Add an underscore and your initials after the SRR ID in the “Key” field for each of the 5 sequences and click “Next”. **NOTE:** *It is critical that each analyst adds unique initials after the SRR numbers, otherwise data gets overwritten in the local certification database view.*

Link Files

1 Select Files 2 Import Rules 3 Link Files 4 Import Metadata

Key	NCBI_ACCESSION	SRR_ID
SRR7722358_et	SAMN09864343	SRR7722358
SRR7722335_et	SAMN09864362	SRR7722335

Reset Next

4.5. Click on “Show details” to verify that the keys are correct, and then verify you have the “PN 2.0 SRA Pipeline” selected and the toggle switch for “Start Pipeline” turned on. Click on “Link Files”. The pipeline should now start after the files are linked.

The first screenshot shows the 'Link Files' window with the following text: 'Click Link Files to.. Copy 5 files from NCBI SRA. Create 5 new sample entries in database. Show Details'. The second screenshot shows the 'Link Files' window with the following text: 'Click Link Files to.. Copy 2 files from NCBI SRA. Create 2 new sample entries in database. Key: SRR7722358_et, SSR_ID: SRR7722358, Key: SRR7722335_et, SSR_ID: SRR7722335. Hide Details'. The third screenshot shows the 'Link Files' window with the following text: 'Click Link Files to.. Copy 2 files from NCBI SRA. Update 2 sample entries in database. Key: SRR7722358_et, SSR_ID: SRR7722358, NCBI Linked, Key: SRR7722335_et, SSR_ID: SRR7722335, NCBI Linked. Hide Details. Files Linked Successfully, Pipeline Run Started Successfully, View Pipeline Run'.

- 4.6. You will not need to import Metadata for the next step and can click on the X to close out the window. **NOTE:** *Additional files can only be linked by the same user after the on-going file linking is finished.*
- 4.7. After the pipeline has finished, check the QC results for the 5 sequences and answer the questions in the Submission Template Section I (Appendix PNQ08-04). Please refer to Appendix 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.
5. **Cluster Analysis Module** (Section II in the Submission Template, Appendix PNQ08-4):
 - 5.1. Follow steps 4.1-4.5. to start the PN 2.0 SRA analysis pipeline for the 5 sequences belonging to the Cluster Analysis Module.
 - 5.2. After the pipeline finishes, verify the QC results for the 5 sequences.
 - 5.3. Create a comparison of the 5 sequences by selecting them and going to the “X Selected” menu and “Add to New Comparison”.
 - 5.4. Go to “View Analyses” and select “Dendrogram”.
 - 5.5. Go to the “MLST Scheme” and select “cgMLST”.
 - 5.6. 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 **local** cluster detection rules for cgMLST data outlined in the Appendix PNQ08-7. **NOTE:** *your laboratory may have established your own rules for local cluster detection that may differ from those put forward by PulseNet Central. For the*

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 17 of 22

purpose of analysis certification, you are expected to use the local cluster detection guidelines outlined in the Appendix PNQ08-07.

- 5.7. In Section II of the Submission Template, 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.
6. Once you have completed your Submission Template, please save the file as a PDF with your initials at the end (e.g., Submission Template_ET) and submit the template together with the Workflow Knowledge Test for evaluation to the PulseNet Inbox (PulseNet@cdc.gov). In the email, please include “WGS Analysis Certification” in the subject line.
7. After PN 2.0 Workflow Knowledge Test and the Submission Template are 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.
 - 7.1. If the results are satisfactory, the person who submitted the file will be granted access to the PN 2.0 Production Environment. The newly certified analyst will receive a welcome email from SAMS including a link to the PN 2.0 Production Environment.
 - 7.2. 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.
 - 7.3. 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.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 09

Effective Date:

Page 18 of 22

Appendix PNQ08-4

Example: WGS Analysis Submission Template

NOTE: This is an example submission template. Please use the Submission template emailed to you upon certification set request.

WGS Analysis Certification Submission Template 1A		
Section I: Sequence QC Set 1a: PN 2.0 Critical Quality Check		
Analyze the 5 sequences listed below in this section through the PN 2.0 system in the PN 2.0 STG Environment in the PulseNet Certification Organization. It is recommended to start the analysis pipeline in the Reference_ID Database because the sequences represent multiple genera. After the pipeline finishes, assess the quality metrics for each sequence and answer the questions below.		
SRR_ID	Is this sequence of acceptable quality to be uploaded to the PulseNet National Database and NCBI (Y/N)	If you deem this sequence NOT to be of acceptable quality, please state the reasons why
SRR28406699		
SRR28462354		
SRR30687114		
SRR11297423		
SRR11245260		
Section II: Cluster Detection Set 1a: cluster analysis using cgMLST in PN 2.0		
Analyze the 5 sequences listed below in this section through the PN 2.0 system in the PN 2.0 STG Environment in the PulseNet Certification Organization. It is recommended to start the analysis pipeline in the Reference_ID Database. After identification, the sequences will be transferred to the appropriate organism-specific database. After the pipeline finishes, assess the quality metrics for each sequence and then proceed to local cluster detection. Assess the genetic relatedness of the 5 sequences making the assumption that all isolates have been collected within the past 60 (all other organisms except <i>Listeria</i>) or 120 (<i>Listeria</i>) days. Interpret the findings in accordance with the PulseNet <i>local</i> cluster detection rules for cgMLST data. Guidance for PulseNet local cluster detection is outlined in Appendix PNQ08-7 (PulseNet WGS Cluster Coding Guidelines) of the WGS Certification SOP. 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). Paste the dendrogram created in PN2.0 in the designated space below.		
SampleID	Cluster Name	Comments
SRR3291012		
SRR3291014		
SRR3291013		
SRR8144962		
SRR8144970		
Insert Dendrogram Below:		

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 09

Effective Date:

Page 19 of 22

Appendix PNQ08-5Example: WGS Certification Report Template

NOTE: The excerpts below are from both lab and analysis certification reports. *Fastq* (wet lab) certification evaluation reports will only have *fastq* grading criteria on the report form. Analysis certification evaluation reports will only have analysis grading criteria on the report form.

WGS Wet Lab Certification Report:**Lab ID – WGS**

**Enteric Diseases Laboratory Branch
CDC/NCEZID/DFWED
WGS Fastq File Certification Report**



Certification Set:	2024 WGS Fastq Certification Set: <i>Escherichia</i> , <i>Listeria</i> , <i>Campylobacter</i> , and <i>Vibrio</i>
Laboratory Submitting Files:	
Submitted by:	
<u>Fastq</u> File Names:	
Library Prep Kit:	
Sequencer Type:	
Date <u>Fastq</u> Files Received by CDC:	

<i>Escherichia coli</i> O157:H7 strain 03-98		
PN2.0 QC Metric	QC value (acceptable value)	Comments
<u>AverageQuality</u> <u>Qscore</u>	(min 30)	
<u>AvgDeNovoCoverage</u>	(min 40x)	
Genus/species/serotype	<i>Escherichia coli</i> O157:H7	
<u>AverageReadLength</u>	(300 cycle chemistry: ≥ 135 bp; 500 cycle chemistry: ≥ 225 bp; 600 cycle chemistry: ≥ 270 bp)	
<u>MidasSecondaryCoverage</u>	(≤ 1.0)	
No. of allele differences compared to reference (cgMLST)	(≤ 3)	
<u>AllelesCalledCorePercent</u>	($\geq 85\%$)	
Assembled genome size (Length)	(5.27 MB, range 5.01-5.53 MB)	

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 09

Effective Date:

Page 20 of 22

Analysis Certification Report:



Enteric Diseases Laboratory Branch
CDC/NCEZID/DFWED
PN 2.0 WGS Analysis Certification Report



Certification Set:	2024 WGS Analysis Certification Set 1a
Laboratory Submitting Files (Lab ID):	
Submitted by:	
Date Analysis Certification Submission:	

I. Sequence QC check:		
SRR ID	Pass/Fail	Comments
SRR28406699	Pass	
SRR28462354	Pass	
SRR30687114	Pass	
SRR11297423	Pass	
SRR11245260	Pass	

II. Cluster detection: <i>Campylobacter jejuni</i>		
SRR id	Pass/Fail	Comments
SRR3291012	Pass	
SRR3291014	Pass	
SRR3291013	Pass	
SRRB144962	Pass	
SRRB144970	Pass	

III. PN 2.0 WGS Workflow Knowledge Test (Passing =90%)		
Submitter's Score	Pass/Fail	Comments
90%		

IV. Overall PN2.0 WGS Analysis Certification Test Result: Pass/ Fail

Certification is valid as long as laboratory successfully completes annual proficiency tests. This certification allows access to the following National Databases: *Campylobacter*, *Escherichia/Shigella*, *Listeria*, *Salmonella*, *Vibrio*, *Reference_ID*

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL			
Doc. No. PNQ08	Ver. No. 09	Effective Date:	Page 21 of 22

Appendix PNQ08-6

Critical Quality Metric Thresholds/Acceptable Ranges for PulseNet Certification

Organism	Cleaned AvgDeNovo Coverage ¹	Combined Average Quality Trimmed	Combined AverageRead Length Trimmed (bp)	Cleaned Assembly Length (bp) ²	Secondary Coverage (MIDAS)	Core Percentage	cgMLST Allele Difference Compared to Reference
<i>Listeria monocytogenes</i>	≥ 20x	≥ 30	≥ 135 (300c) ≥ 225 (500c) ≥ 270 (600c)	2800000-3199999	≤ 1	≥ 95	≤ 3
<i>Escherichia coli</i>	≥ 40x	≥ 30	≥ 135 (300c) ≥ 225 (500c) ≥ 270 (600c)	4200000-5999999	≤ 1	≥ 85	≤ 3
<i>Campylobacter</i> spp.	≥ 20x	≥ 30	≥ 135 (300c) ≥ 225 (500c) ≥ 270 (600c)	1400000-2200000	≤ 1	≥ 85	≤ 3
<i>Vibrio parahaemolyticus</i>	≥ 40x	≥ 30	≥ 135 (300c) ≥ 225 (500c) ≥ 270 (600c)	4900000-5500000	≤ 1	≥ 95	≤ 3

¹ The pipeline will not further process the fastq files if the trimmed average read coverage (the coverage after quality-based trimming but before assembly) is <10x.

² The acceptable assembly lengths cited are for the entire genus & species. For certifications, the acceptable length is strain-specific based on the reference sequence generated by the CDC ±5% error of margin.

STANDARD OPERATING PROCEDURE FOR WGS CERTIFICATION OF PULSENET PERSONNEL

Doc. No. PNQ08

Ver. No. 09

Effective Date:

Page 22 of 22

Appendix PNQ08-7

PulseNet WGS Cluster Coding Guidelines

Local Cluster Detection: Things to look for in your local database

Campylobacter, Escherichia, and Salmonella: Look for local clusters of sequences representing three or more human cases within 10 allele differences by cgMLST, with two being no more than 5 alleles different. Isolation dates for all cases should be within the past 60 days of detection. PulseNet upload dates should be within the past 60 days.

Listeria: Using cgMLST, look for local clusters of sequences representing three or more human cases within 10 allele differences, with two being no more than 5 alleles different. You can further expand the cluster definition by using wgMLST, looking for three or more human cases within 25 allele differences with two being no more than 10 alleles different. Isolation dates should be within the past 120 days of detection. PulseNet upload dates should be within the past 120 days.

NOTE: These are general guidelines that should be used for local cluster detection. Each state/jurisdiction may use stricter criteria than what is stated above for the purpose of tracking local disease clusters within their state/jurisdiction. However, only clusters meeting the above criteria should be posted to SharePoint. Once posted to SharePoint, a CDC cluster code will be assigned, or the cluster will be monitored for additional cases. When posting to SharePoint, please include the number of isolates, serotype (if applicable), isolate identifiers (keys), allele differences by cgMLST, isolation/collection dates, and relevant epi data.

National Cluster Detection: PulseNet Central uses the following criteria to code clusters

PulseNet Cluster Code Assignment at the National Level

Organism	Salmonella	Escherichia	Shigella	Listeria	Campylobacter
# days for search, based on Upload Date	60 days	60 days	60 days	120 days	60 days
case threshold	7+ Clinical cases	5+ Clinical cases	10+ Clinical cases	3+ Clinical cases	5+ Clinical cases
methods used	cgMLST	cgMLST	cgMLST	cgMLST/wgMLST	cgMLST
Cluster Coding Criteria*	Within 10 alleles with at least 3 cases within 5 alleles; 3+ from single state	Within 10 alleles with at least 3 cases within 5 alleles	Within 10 alleles with at least 3 cases within 5 alleles	within 5 alleles by cgMLST; further refined by wgMLST	Within 10 alleles
Notes about other dates	30% typically have isolation dates within the past 50 days	30% typically have isolation dates within the past 50 days	30% typically have isolation dates within the past 50 days	Include historical isolates related within 25 alleles by wgMLST	50% typically have isolation dates within the past 30 days
Single and multi-state clusters	Single-state clusters are not coded	Single-state clusters coded for any serotype if there is an epi link	Single-state clusters are not coded	Single-state clusters are not coded	Single-state clusters are not coded
Additional Notes per Organism	Enteritidis, Newport, and Javiana: 10 clinical cases within 5 alleles; Common sequence types are monitored, not coded	-	May include cases within a wider allele range due to person-to-person transmission	-	-

*Rare serotypes, seasonality and matches to nonhuman entries are considered. May go down to 5+ for *Salmonella* and 3+ for *E. coli* for rarer serotypes or slower months.

NOTES: (1) all newly coded clusters are checked against historical outbreaks and NCBI's pathogen detection pipeline
(2) clusters not meeting coding criteria may still be assigned a cluster code based on supporting epi information