Standard Operating Procedure for WGS Certification of PulseNet Personnel

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.

Scope

This procedure applies to all PulseNet personnel generating sequence data, checking data quality, and/or performing analysis and uploading to the PulseNet databases using BioNumerics.

Definitions and Terms

1. **Analysis Certified:** An individual who is certified for checking WGS data quality, performing analysis using BioNumerics and uploading to the PulseNet National Database
2. **BaseSpace:** Illumina cloud-based computing environment for next generation sequencing data analysis and management, including data sharing
3. **BioNumerics:** Analysis software used by PulseNet, developed by Applied Maths (Sint-Martens-Latem, Belgium)
4. **Bioproject:** a collection of biological data related to a single initiative, originating from a single organization or a consortium
5. **Bundle file:** A file with a .bdl extension that is produced in BioNumerics and contains the analysis of at least one fastq sequence file
6. **CDC:** Centers for Disease Control and Prevention
7. **Certification file evaluator:** An individual who evaluates certification files
8. **Certification files:** Fastq and/or word document (completed WGS analysis template) files submitted by PulseNet participants for WGS certification evaluation
9. **FASTQ:** a text-based format for storing both a biological sequence and its corresponding quality scores
10. **Fastq Certified:** An individual who is certified in laboratory methods for WGS
11. **FTP:** File Transfer Protocol; a standard network protocol used to transfer computer files from one host to another host via Internet
12. **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
13. **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.
14. **QA/QC:** Quality assurance**/**quality control
15. **SOP:** Standard Operating Procedure
16. **WGS Certification Report:** A report that contains the evaluation and results of the participant's WGS certification files

17. **WGS**: <u>W</u>hole <u>G</u>enome Sequencing

Responsibilities

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.

   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.

   1.2. Individuals can be certified for *Salmonella*, *Escherichia*, *Listeria*, *Campylobacter*, and *Vibrio* in one of three ways:

      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.

      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. Currently, analysis certifications are available for *Escherichia, Salmonella, Listeria* and *Campylobacter*.

      1.2.3 **Both Fastq and Analysis Certified.** An individual can generate fastq files, check the data quality and submit results to the PulseNet bioprojects housed at NCBI and the PulseNet national database. Currently, analysis certifications are available for *Escherichia, Salmonella, Listeria* and *Campylobacter*.

   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.

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.

   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.

.

1. PulseNet participants request the WGS laboratory certification set (if they do not already have it) and/or analysis fastq files, associated metadata and bundle file from CDC (pulsenet@cdc.gov).

    1.1. The WGS Laboratory Certification set includes STEC*, Salmonella, Listeria, Campylobacter* and *Vibrio*. The STEC certification also covers the certification for *Shigella*.

    1.2.1 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, analyze and submit results (fastq files) to CDC. See Appendices PNQ08-1 and PNQ08-2 for details.

NOTE: PulseNet laboratory SOPs are currently only available for the Illumina MiSeq platform. However, certification files generated using the Ion Torrent PGM and S5 will be accepted and reviewed by CDC.

    1.2. The WGS Analysis Certification currently includes *STEC, Salmonella, Listeria,* and *Campylobacter*. The STEC certification also covers *Shigella*.

    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), along with instructions for accessing fastq files, associated metadata, a bundle file and an analysis submission template via the PulseNetQA FTP site. An example of the analysis submission template is provided in Appendix PNQ08-4.

    1.2.2 Participants are asked to perform quality checks on fastq files provided by CDC, correctly identify which metadata should be submitted to PulseNet National Database, and perform quality checks on the contents of a bundle file (de novo assemblies and allele calls). Using the provided submission template, all results and responses must be emailed to CDC for review.

NOTE: Analysis certification must be performed using BioNumerics version 7.6.

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).

    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. If a person relocates to a different PulseNet laboratory, they must be re-certified.

    2.1.1 Upon becoming WGS Analysis Certified personnel will be given credentials to access the NCBI uploader in BioNumerics.

    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.

    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.

Contacts

1. PulseNet Central, CDC
    PulseNet@cdc.gov
    (404) 639-4558

2. NGS Troubleshooting support
    Pulsenetngslab@cdc.gov

3. Eija Hyytia-Trees, D.V.M., Ph.D
    CDC
    Chief, PulseNet Next Generation Subtyping Methods Unit
    (404) 639-3672
    eih9@cdc.gov

4. Kelley Hise, MPH
    CDC
    Chief, PulseNet Outbreak Detection and Surveillance Unit
    (404) 636-0704
    Kpb6@cdc.gov

Amendments

1. 1/26/2016 **:** Updated appendix PNQ08-1 with instructions for reconstituting Campy lyophils, previously live culture stabs.
2. 10/25/2016 : Added instructions for WGS analysis certification and updated the PulseNet email address to PulseNet@cdc.gov.
3. 02/15/2017 : Included reference to "Sequencing Quality Metrics in BN 7" document on SharePoint in Step 8 of Appendix 8-3.
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 PusleNet Biosample Submission Group on NCBI.
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.)

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 Needed:

- Sterile sturdy forceps
- 1 ml pipetman
- 1 ml sterile pipet tips
- 1 µl sterile inoculating loop

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 alchohol

PROCEDURE FOR REVIVING *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 an 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.**

PROCEDURE FOR REVIVING *CAMPYOLBACTER* 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 250 µl of CLRW. Allow to stand for a few minutes and/or mix gently to produce a uniform suspension. Pipet 100µl of *Campylobacter* cell suspension onto a BHI 5% Rabbit Blood agar plate and with a 10µl loop and streak for growth. Incubate microaerobically at 37°C or 42ºC 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.

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: Instructions for PulseNet Certification for WGS laboratory procedure

Dear PulseNet Participant,

This package should contain 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:

CDC03-98	Shiga toxin-producing *Escherichia coli* O157:H7
C6472	*Salmonella enterica* serovar Heidelberg
CDC#H8394	*Listeria monocytogenes*
D5480	*Campylobacter jejuni*
F5828	*Vibrio parahaemolyticus*

Please follow these supplemental instructions for testing the certification isolates by WGS. Refer to the "Laboratory standard operating procedure for PulseNet Nextera XT library prep and run setup for the Illumina MiSeq (PNL32)" for detailed instructions.

1. Perform the DNA extraction, library preparation, normalization, pooling and loading on the cartridge following the instructions of the standardized protocol.
    1.1. Files must be named according to the PulseNet standardized naming scheme outlined in SOP PNL32. For certification, this includes **strain number, initials, LabID, instrument ID and run date, i.e. CDC03-98jk-GA-M0947-180215.**
    1.2. It is acceptable to include routine isolates on the same cartridge with the certification strains in order not to waste reagents. Please do not share the routine isolate sequences with CDC until you know you have passed your certification.

2. Perform local QC on the raw reads. Refer to the "PulseNet standard operating procedure for Illumina MiSeq data quality control (PNQ07)" for detailed instructions.

3. Transfer the fastq-files to CDC for evaluation either using the Illumina BaseSpace or the PulseNet1 ftp-site. In BaseSpace, the sample data should be shared using the same Projects as used for routine surveillance. Refer to the "PulseNet standard operating procedure for transferring data from the Illumina MiSeq (PND19)" for detailed instructions.

4. Send an email to the PulseNet inbox (pulsenet@cdc.gov) notifying PulseNet Central of the certification files' availability.

> 4.1. In the email to CDC, include "**WGS Certification"** in the subject line.

5. 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 will be eligible to send fastq-files to PulseNet Central or to an in-house analysis certified individual for NCBI upload and analysis. If the submitted certification files are <u>not</u> 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.

Please let us know if you have questions or further clarification is needed.

Good luck,

PulseNet Central

CDC

Pulsenet@cdc.gov

[(404) 639-4558](tel:4046394558)

Appendix PNQ08-3: Supplemental Instructions for PulseNet WGS Analysis Certification

Dear PulseNet Participant,

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.

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 instructions to access fastq files, associated metadata, a BioNumerics bundle file and a submission template (example template Appendix PNQ08-4) via the PulseNet QA FTP site. There are multiple certification sets available on the FTP site, please use the set assigned to you in the response email sent from CDC.
2. Perform quality checks on the fastq files using FastQC. Please refer to the "PulseNet standard operating procedure for Illumina MiSeq data quality control (PNQ07)" for detailed instructions. Based on the QC, determine if you would import the data to the organism-specific BioNumerics database or upload the file to NCBI. 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. Please use the submission template provided with the files.
3. Provide the name of the correct bioproject to which you would upload the data. For those labs that are members of both PulseNet and GenomeTrakr, appropriate bioprojects would need to be listed depending on the source of the isolate.
4. Metadata for the fastq files are provided in the submission template supplied by CDC. For each isolate (even if the quality metrics did not pass) determine what metadata is required for submission to NCBI. Using the submission template, indicate which metadata you would be required to submit and designate any metadata that would not be required for submission to NCBI as "missing" to meet the PulseNet and/or GenomeTrakr requirements.
5. To open the bundle file, save the file in BioNumerics (organism specific) v7.6. 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.
6. Create a comparison by clicking on the green plus sign in the comparison window (Alt+C)
7. Click on the eye next to the quality experiment in the Experiments panel.
8. Under Experiment Data, click on the first 123 icon (Show values) and change the "Character name" to "Character Description". Now you can see the quality metrics for the isolates selected.
9. Based on the quality metrics outlined in PNQ08-6 the "QC Metrics Summary Table" determine if you would submit the data within the bundle file to the PulseNet National Database. If you determine you would not submit the data to the PulseNet national database, then you must list your reasoning. Please use the template that is provided with the files.

10. In order to be considered analysis certified for WGS you must complete all portions of the submission template and email your submission template to pulsenet@cdc.gov for evaluation. In the email to CDC, include "**WGS Certification"** in the subject line.

After the submission template is emailed to CDC, the PulseNet certification file evaluator will analyze the submission and inform your laboratory 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), 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. These 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.

Please let us know if you have questions or further clarification is needed.

Good luck,

PulseNet Central

CDC

Pulsenet@cdc.gov

(404)639-4558

Appendix PNQ08-4: Example: WGS Analysis Submission Template

Note: This is an example submission template. Please use the appropriate (organism specific) template provided on the PulseNetQA FTP site.

	Listeria WGS Analysis Certification Submission Template A						

I. *Listeria monocytogenes* fastq.gz files data quality check:

WGS_ID	Would you import this data into BioNumerics? Yes/No	If you would NOT import this data, you must explain why by listing any failed quality metrics.
PNUSALCert1		
PNUSALCert2		
PNUSALCert3		
PNUSALCert16		

II. *Listeria monocytogene* s fastq.gz files bioproject selection:

Strain (WGS_ID)	Organism	Collected By	Georgraphic Location (Source Country)	Host (Source Type)	Isolation Source (Source Site)	Collection Date (IsolatDate)	Outbreak Code	PulseNet / GenomeTrackr Bioproject
PNUSALCert1	*Listeria monocytogenes*	VA	USA	Human	blood	2015-10-12	1599VAGX6-1	
PNUSALCert2	*Listeria monocytogenes*	GA	USA	Human	blood	2016-01-14	1699GAGX6-1	
PNUSALCert3	*Listeria monocytogenes*	KY	USA	Environmental	swab	2015-08-03	1599KYGX6-1	
PNUSALCert16	*Listeria monocytogenes*	MN	USA	Food	cantaloupe	2016-09-19	1699MLGX6-1	

III. *Listeria monocytogenes* fastq.gz files metadata check:

In the table below, please remove any metadata that would not be required for submission to NCBI and insert "missing" in that cell:

Strain (WGS_ID)	Organism	Collected By	Georgraphic Location (Source Country)	Host (Source Type)	Isolation Source (Source Site)	Collection Date (IsolatDate)	Outbreak Code	Allele Code
PNUSALCert1	*Listeria monocytogenes*	VA	USA	Human	blood	2015-10-12	1599VAGX6-1	LMO1.0 - 45.1.1.12.34.86
PNUSALCert2	*Listeria monocytogenes*	GA	USA	Human	blood	2016-01-14	1699GAGX6-1	LMO1.0 - 1.4.47.1.4
PNUSALCert3	*Listeria monocytogenes*	KY	USA	Environmental	swab	2015-08-03	1599KYGX6-1	LMO1.0 - 106.1.1.1.20
PNUSALCert16	*Listeria monocytogenes*	MN	USA	Food	cantaloupe	2016-09-19	1699MLGX6-1	LMO1.0 - 106.1.1.1.4

IV. *Listeria monocytogenes* bundle file quality check: CDCListeriaCertA

WGS_ID	Would you upload this entry to the PulseNet National Database?	If you would NOT upload this entry, you must explain why by listing any failed quality metrics.
PNUSALCert4		
PNUSALCert5		
PNUSALCert7		
PNUSALCert9		

LAB ID – WGS

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

Certification Set: 2018 WGS Certification Set
Laboratory Submitting Fastq File:
Submitted by:
Fastq File Name:
Sequencer Type:
Date Fastq Files Received by CDC:
**Date Analysis Certification Submission
Template Received by CDC:**

Escherichia strain 03-98		
QC statistic	**QC value (acceptable value)**	**Comments**
Q-score R1	(min 30)	
Q-score R2	(min 30)	
Average coverage	(min 40x)	
Insert size	(median ≥300 bp)	See E. coli Insert Size Histogram below
No. of hqSNPs compared to reference	(< 1/ MB)	
No. of allele differences compared to reference	NA	
Assembled genome size	(5.5MB, 5% error margin)	

Escherichia WGS Analysis Evaluation		
Quality check	**Pass/Fail**	**Comments**
Fastq.gz file evaluation		
Correct bioproject listed		
Metadata selection		
Bundle file evaluation		

Salmonella serovar Heidelberg strain C6472		
QC statistic	**QC value (acceptable value)**	**Comments**
Q-score R1	(min 30)	
Q-score R2	(min 30)	
Average coverage	(min 30x)	
Insert size	(median ≥300 bp)	See Salmonella Insert Size Histogram below
No. of hqSNPs compared to reference	(< 1/ MB)	
No. of allele differences compared to reference	NA	
Assembled genome size	(5.0MB, 5% error margin)	

Salmonella WGS Analysis Evaluation		
Quality check	**Pass/Fail**	**Comments**
Fastq.gz file evaluation		
Correct bioproject listed		
Metadata selection		
Bundle file evaluation		

Listeria monocytogenes strain H8394		
QC statistic	**QC value (acceptable value)**	**Comments**
Q-score R1	(min 30)	
Q-score R2	(min 30)	
Average coverage	(min 20x)	
Insert size	(median ≥300 bp)	See Listeria Insert Size Histogram below
No. of hqSNPs compared to reference	(< 1/ MB)	
No. of allele differences compared to reference (based on version 4 allele database)	(≤ 3)	
Assembled genome size	(3.0MB, 5% error margin)	

Listeria WGS Analysis Evaluation		
Quality check	**Pass/Fail**	**Comments**
Fastq.gz file evaluation		
Correct bioproject listed		
Metadata selection		
Bundle file evaluation		

Campylobacter jejuni strain D5480		
QC statistic	QC value (acceptable value)	Comments
Q-score R1	(min 30)	
Q-score R2	(min 30)	
Average coverage	(min 20x)	
Insert size	(median ≥300 bp)	See Campylobacter Insert Size Histogram below
No. of hqSNPs compared to reference	(< 1/ MB)	
No. of allele differences compared to reference	NA	
Assembled genome size	(1.7MB, 5% error margin)	

Campylobacter WGS Analysis Evaluation		
Quality check	Pass/Fail	Comments
Fastq.gz file evaluation		
Correct bioproject listed		
Metadata selection		
Bundle file evaluation		

Vibrio spp. Strain F5828		
QC statistic	QC value (acceptable value)	Comments
Q-score R1	(min 30)	
Q-score R2	(min 30)	
Average coverage	(min 40x)	
Insert size	(median ≥300 bp)	See Vibrio Insert Size Histogram below
No. of hqSNPs compared to reference	(< 1/ MB)	
No. of allele differences compared to reference	NA	
Assembled genome size	(5.0MB, 5% error margin)	

(Organism specific Insert Size Histogram):

QA/QC Certification Result:

Escherichia <u>Fastq files:</u> (Satisfactory) Needs Improvement   STEC analysis <u>files:</u> (Satisfactory) Needs Improvement

Salmonella <u>Fastq files:</u> (Satisfactory) Needs Improvement   *Salmonella* analysis <u>files:</u> (Satisfactory) Needs Improvement

Listeria Fastq files: (Satisfactory)   Needs Improvement   *Listeria* analysis files: (Satisfactory) Needs Improvement

Campy Fastq files: (Satisfactory)   Needs Improvement   *Campy* analysis files: (Satisfactory) Needs Improvement

Vibrio Fastq files: (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:**

Certification Evaluation Performed By: _____   **Date of Evaluation:**

Approved By: _____   **Date of Approval:**

Approved By: _____   **Date of Approval:**

Quality Metric	*Listeria*	*Campylobacter*	*Escherichia*	*Salmonella*	*Vibrio*
ⱡ Median Insert Size:	300bp				
Raw data statistics: R1 & R2 Average Sequence Read Lengths	≥ 138 bp (2 x 150 bp cartridge) ≥ 230 bp (2 x 250 bp cartridge)				
¥ Raw data statistics: R1 & R2 Q-scores	≥ 30				
¥ *De novo* assembly: Average Coverage	≥ 20X	≥ 20X	≥ 40X	≥ 30X	≥ 40X
De novo assembly: Sequence Length (Mbp)	~2.8 – 3.1	~1.4 –2.2	~4.2 – 5.9 ~4.2 – 4.9 (*Shigella,* rare spp.) ~4.9 – 5.9 (most serotypes)	~4.4 – 5.6 (common serotypes usually ~5.0)	~3.8 – 4.3 (Vc) ~4.9 – 5.5 (Vp) ~4.7 – 5.3 (Vv)
Percent Core Present	**≥ 95%**	**≥ 95% (*C. jejuni*)**	**≥ 95%**	**≥ 95%**	**NA**

- ⱡMedian Insert Sizes are determined by mapping sequences back to the reference strain using SMALT. While Median Insert Size cannot be determined in BioNumerics, the Average Sequence Read Lengths, which correlate with insert sizes, can be found in the WGS Experiment card. In this table, the Average Sequence Read Lengths (listed in black) are provided as guidance for projecting the Median Insert size (i.e. sequences with Average Sequence Read Lengths of 230bp or higher would likely have passing Median Insert Sizes of 300bp or greater). This metric can be used as a quality check before submitting files for evaluation.
- ¥ Sequences may pass Certification/ Proficiency Testing, if the Q-scores are between 29.00-29.99 or 28.00-28.99 as long as the coverage is 10X or 20X higher than shown in the table, respectively. If the Q-score falls below 28, then the sequence fails QC regardless of coverage.
- Quality metrics in RED must be met in order to pass WGS Certification and Proficiency Testing.
- Metrics in **bold** are found in the organism-specific databases only; other metrics (in red) can be found in the Reference ID Database.