

Validating Bionumerics 7.6: A strategic approach from Oregon

Karim Morey, MS, M(ASCP)

Oregon State Public Health Laboratory

PulseNet West Coast Regional Meeting

February 2019

Outline

- Compliance requirements
- Strategy development
- Strategy description
- Summary and discussion

WGS Analysis Validation/Verification

- What type of approach should be applied if lab is CLIA certified and CAP accredited?
- Guidelines available?
- Is it Validation or Verification?
- What needs to be verified or validated?

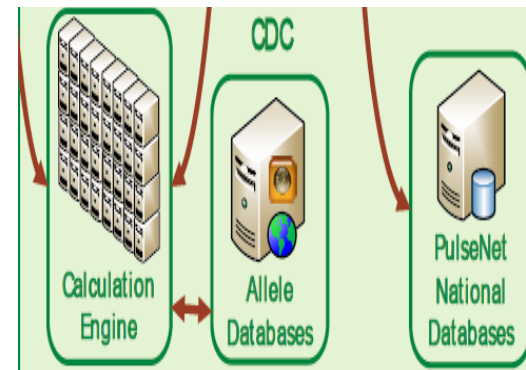
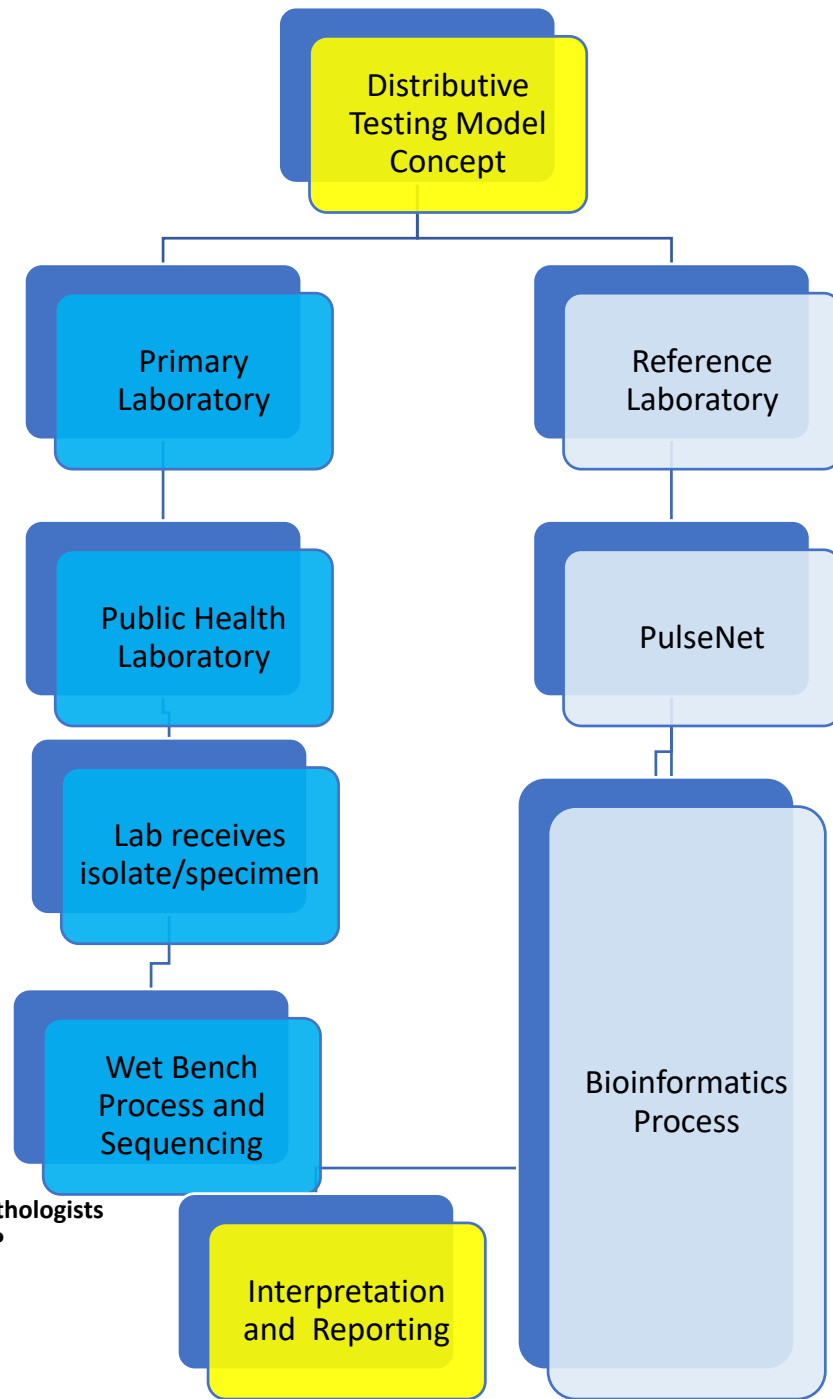
Definitions

VALIDATION

“...the process of assessing the assay and its performance characteristics to determine the optimal conditions that will generate a reproducible and accurate result...”

VERIFICATION

One-time process to determine or confirm a test’s expected performance compared to actual results produced by the lab.



Adapted from College of American Pathologists
 NGS: What does compliance look like?
 2018 Focus in Compliance

How do we demonstrate compliance?

- CAP MOL.3615. Analytical Bioinformatics Process Validation must determine performance characteristics for all microbial targets.
- Apply Distributive Testing Model concept
- Organize a process for ID Validation and BN 7.6 performance
- Develop plan and validation/verification strategy

Validation and Verification strategy

BN 7.6
Software
Verification

Validation
plan for
organism
identification

BN 7.6
pipeline
verification

Validation and Verification strategy



BN 7.6 Software
Verification

BN 7.6 Software verification

Verify performance of software as expected

Perform version upgrade
Verify functionality of PFGE component
Verify functionality of WGS component (certification)

| A | B | C | D | E | F | G | H |
|---|--|--------|-----------|------|------|---|---|
| Process flow verification for upgrading and converting PulseNet databases in Bionumerics v 7.6 | | | | | | | |
| Testers | Michael Bitzer (MB), Peter Arvidson (PA), Karim Morey (KM) | | | | | | |
| Stage | Description | Tester | Date | Pass | Fail | Description of Failed process | Corrective action |
| 1 | Prepare to convert databases | KM/MB | 11/1/2018 | ✓ | | | |
| 2 | Install BioNumerics v 7.6 | PA | 11/1/2018 | ✓ | | | |
| 3 | Install Plugins (WGS too and PulseNetClient) | PA/KM | 1/8/2019 | ✓ | | | |
| 4 | Upgrade Database for Salmonella | | | | | | |
| | 4.1 Annotate information for Salmonella database | PA/KM | 1/8/2019 | ✓ | | | |
| | 4.2 Rename PFGE database | PA/KM | 1/8/2019 | ✓ | | | |
| | 4.3 Save Upgrade plugin | PA/KM | 1/8/2019 | ✓ | | | |
| | 4.4 Upgrade Salmonella database to v7.6 | PA/KM | 1/8/2019 | | * | Existence of two databases, local and connected, named similarly prevented initial upgrade of | Determined the correct database was local and contacted Applied Math (AM) for support. File was corrected and converted by AM allowing us to proceed with the upgrade |
| | 4.5 Install PulseNet upgrade plugin | PA/KM | 1/8/2019 | ✓ | | | |
| | 4.6 Rename field IDs | PA/KM | 1/8/2019 | ✓ | | | |
| | 4.7 Copy Upload Dates | PA/KM | 1/8/2019 | ✓ | | | |
| | 4.8 Uninstall upgrade plugin | PA/KM | 1/8/2019 | ✓ | | | |
| | 4.9 Final upgrade verification of database | PA/KM | 1/8/2019 | ✓ | | | |
| 5 | Upgrade Database for Escherichia | | | | | | |
| | 5.1 Annotate information | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.2 Create wgEscherichia database in v 7.6 and install PulseNet upgrade plugin | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.3 Upgrade E. coli O157 to v 7.6 | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.4 Export from E. coli into wgEscherichia in v 7.6 | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.5 Rename field IDs | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.6 Copy upload dates | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.7 Upgrade E. coli NON O157 to v 7.6 | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.8 Export from NON O157 in v 7.6 | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.9 Import NON O157 into wgEscherichia in v 7.6 | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.10 Rename field IDs | PA/KM | 1/9/2019 | ✓ | | | |
| | 5.11 Copy upload dates | PA/KM | 1/9/2019 | ✓ | | | |

| Function | Tester | Date | Pass | Fail | Description of Failed process | Corrective action | Pass2 | Fail3 | Comments |
|---|--|------|-----------|------|-------------------------------|-------------------|-------|-------|---|
| Main BioNumerics Screen (Overview) | | | | | | | | | |
| 6 | | | | | | | | | |
| 7 | Verify database fields are present | MB | 1/10/2019 | ✓ | | | | | |
| 8 | Verify fingerprint and experiment types are present | MB | 1/10/2019 | ✓ | | | | | |
| 9 | Import a TIFF to analyze (use unique Test tiff name) | MB | 1/10/2019 | ✓ | | | | | |
| 10 | Import demographic information | | | | | | | | Manual entry until harvest is set up to export |
| 11 | Fingerprint Data Window (TIFF) | | | | | | | | |
| 12 | Analyze a TIFF/enter unique Key names | MB | 1/10/2019 | ✓ | | | | | When using auto band assign small circle icons appear marking bands. If you assign a band on top of the circle it makes it a double band. Please delete the small circle and assign a band by selecting the peak and pressing enter. |
| 13 | Make sure the flags on the reference standard in the normalization step are correct (bands marked and Kb values) | MB | 1/10/2019 | ✓ | | | | | |
| 14 | Fingerprint File Window (TIFF) | | | | | | | | |
| 15 | Enter data, change the fingerprint types and make sure the correct standard is listed | MB | 1/10/2019 | ✓ | | | | | When a key that was enter incorrectly was deleted from the database and error was given "can not find key that was deleted". Demographics could not be entered and the database had to be closed and reopened. |
| 16 | Keys and data link correctly | MB | 1/10/2019 | ✓ | | | | | |
| 17 | Verify data entry form and picklists are present (green i) | MB | 1/10/2019 | ✓ | | | | | Base drop down list are present. Custom dropdown needs to be created (dates, counties) |
| 18 | Queries (Daily functions - local) | | | | | | | | |
| 19 | Analysis: Fast band matching | MB | 1/10/2019 | ✓ | | | | | |
| 20 | Charts and Statistics | | | | | | | | |
| 21 | Dendrograms | MB | 1/10/2019 | ✓ | | | | | |
| 22 | Comparisons | MB | 1/10/2019 | ✓ | | | | | Xbsl pattern field has moved. When assigning pattern numbers scroll to almost the end of the data fields |
| 23 | Queries (Daily functions - National) | | | | | | | | |
| 24 | Connect to upload TIFF and data to national database | MB | 1/10/2019 | ✓ | | | | | Use new firewall link to connect to PN https://PulseNet/WGS-USA.cdc.gov/ When connecting to database server you must enter username and password. The old was SQLsmc@cdc.gov and password given by CDC. Now in v7.6 the user name is what you use to access the firewall (mbitzer) and the same password assigned by the CDC you used to use. (xxx-nnn-XX) |
| 25 | Reupload with added/changed serotype/demo data; verify with CDC PN that nothing becomes unlinked, new data is added, PulseNetUploadModified date is filled | MB | 1/11/2019 | ✓ | | | | | |

Validation and Verification strategy

Validation plan for
organism identification

Organism ID Validation BN 7.6
Pipelines to replace Gold Standard
methods: molecular and traditional
serotyping, biochemical ID

ANI
Validation

SeqSero
Validation

Serotype
Finder
Validation

Validation Process

Performance Specifications

- Selection of validation strains (previously sequenced isolates)
- Accuracy: Comparison with gold standard identification methods and testing performed in different location (e.g. Pulsenet).
- Precision: Reproducibility and Repeatability
- Sensitivity
- Specificity
- Limit of Detection

Validation and Verification strategy

BN 7.6 pipeline verification

BN 7.6 Pipeline
verification/parallel
testing using a set of
PulseNet organisms

BN 7.6 Ref ID and
Genotyping tools

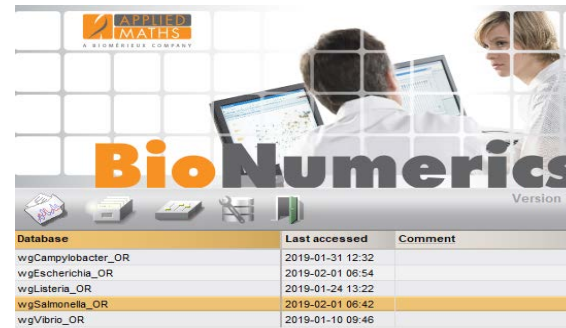
Use of publicly
available and
validated pipelines
by Cloud Computing

Cloud Computing

- Cloud based Virtual Machine (VM) e.g. Google Cloud
- StaPH-B group/CDPHE developed and validated WGS analysis pipelines. Multi-step, multi-software
- Distributable model: Share VMs between institutions from a public repository e.g. Git Hub
- Static, robust workflow, reproducible

Bionumerics 7.6 WGS Analysis Pipeline

Ref ID Database



The screenshot shows the Bionumerics software interface with a table of reference IDs. The table has three columns: Database, Last accessed, and Comment. The rows are:

| Database | Last accessed | Comment |
|--------------------|------------------|---------|
| wgCampylobacter_OR | 2019-01-31 12:32 | |
| wgEscherichia_OR | 2019-02-01 06:54 | |
| wgListeria_OR | 2019-01-24 13:22 | |
| wgSalmonella_OR | 2019-02-01 06:42 | |
| wgVibrio_OR | 2019-01-10 09:46 | |

Genotyping Tools

Quality Control

- Read Quality
- Predicted Coverage
- Contamination

De Novo Assembly

N50
Genome size
Coverage

ANI

Genera and Species for
main PulseNet
organisms

Built in CGE tools

Serotype Finder (Escherichia)
Seqsero (Salmonella)
ResFinder
Virulence Finder
Plasmid Finder
Pathotype

Surveillance Tools

wgMLST
cgMLST
wgSNP

Google Cloud Computing

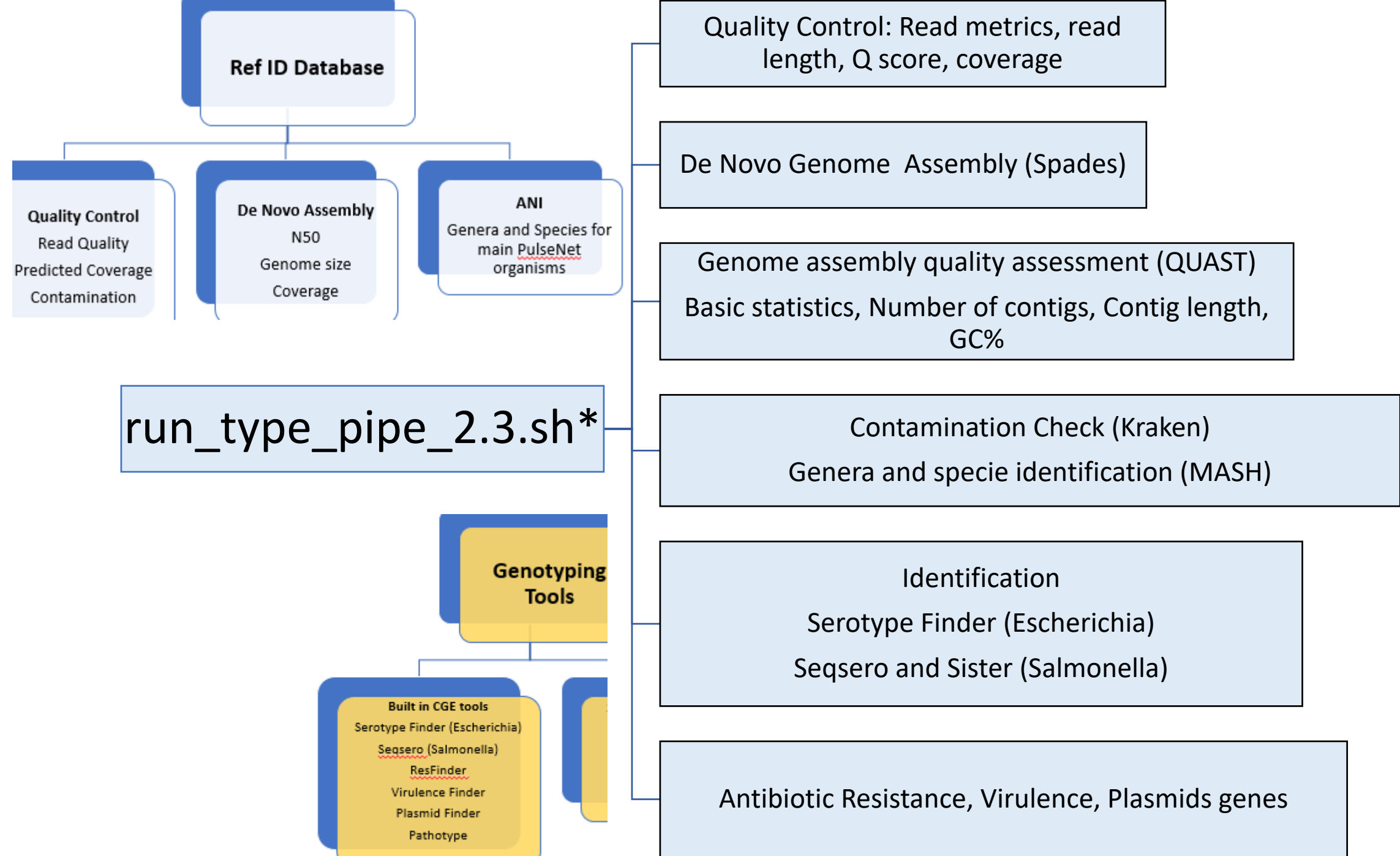
Bioinformatics Pipelines developed by Staph-B group - CDPHE

run_type_pipe_2.3.sh*

run_pipeline_non-ref_tree_build_1.3.sh*

run_lyveset_1.1.sh*

```
Last login: Sun Feb  3 05:18:31 2019 from 159.121.202.174
-bash: /home/staphb/perl5/perlbrew/etc/bashrc: No such file or directory
staphb@staphb-col:~$ cd workspace/
staphb@staphb-col:~/workspace$ cd s_muenster1/
staphb@staphb-col:~/workspace/s_muenster1$ ll
total 5044
drwxrwxr-x 19 staphb staphb  4096 Feb  2 20:15 ./
drwxrwxr-x  6 staphb staphb  4096 Feb  2 18:32 ../
drwxrwxr-x  3 staphb staphb  4096 Feb  1 08:25 abricate/
drwxrwxr-x  2 staphb staphb  4096 Feb  1 14:23 clean/
drwxrwxr-x  2 staphb staphb  4096 Feb  1 15:06 fastq_files/
drwxrwxr-x  2 staphb staphb  4096 Feb  1 14:56 gff_files/
-rw-rw-r--  1 staphb staphb  2018 Feb  1 08:25 isolate_info_.tsv
drwxrwxr-x 18 staphb staphb  4096 Jan 31 22:42 kraken_output/
-rw-rw-r--  1 staphb staphb 1465697 Feb  1 15:06 logfile_non-ref.txt
-rw-rw-r--  1 staphb staphb 3604167 Feb  1 08:25 logfile.txt
drwxrwxr-x 10 staphb staphb  4096 Feb  2 20:15 lyveset/
drwxrwxr-x  3 staphb staphb  4096 Feb  1 07:12 mash/
drwxrwxr-x 18 staphb staphb  4096 Feb  1 14:54 prokka/
drwxrwxr-x 18 staphb staphb  4096 Feb  1 06:53 quast/
drwxrwxr-x  2 staphb staphb  4096 Feb  1 15:06 raxml/
drwxrwxr-x  2 staphb staphb  4096 Feb  1 15:04 roary/
drwxrwxr-x 18 staphb staphb  4096 Feb  1 08:17 SeqSero_output/
drwxrwxr-x  2 staphb staphb  4096 Feb  1 07:12 serotypeFinder_output/
drwxrwxr-x  3 staphb staphb  4096 Feb  3 00:40 .SGELK/
drwxrwxr-x  2 staphb staphb  4096 Feb  1 08:17 sistr/
drwxrwxr-x  2 staphb staphb  4096 Feb  2 20:15 snp_counts/
drwxrwxr-x 18 staphb staphb  4096 Feb  1 06:11 spades_assembly_trim/
-rw-rw-r--  1 staphb staphb   177 Jan 31 20:06 SRR
-rw-rw-r--  1 staphb staphb   352 Feb  1 14:23 tmp1
staphb@staphb-col:~/workspace/s_muenster1$
```



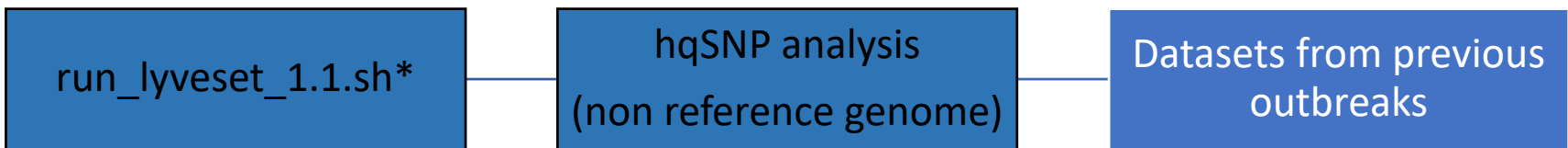
run_pipeline_non-ref_tree_build_1.3.sh*

```
graph LR; A[run_pipeline_non-ref_tree_build_1.3.sh*] --- B[Perform a core genome alignment (non reference genome approach) using evaluated sequences]; A --- C[Genome Annotation (Prokka)]; A --- D[Phylogenetic Analysis (Roary and RAxML)];
```

Perform a core genome alignment
(non reference genome approach)
using evaluated sequences

Genome Annotation (Prokka)

Phylogenetic Analysis (Roary and
RAxML)



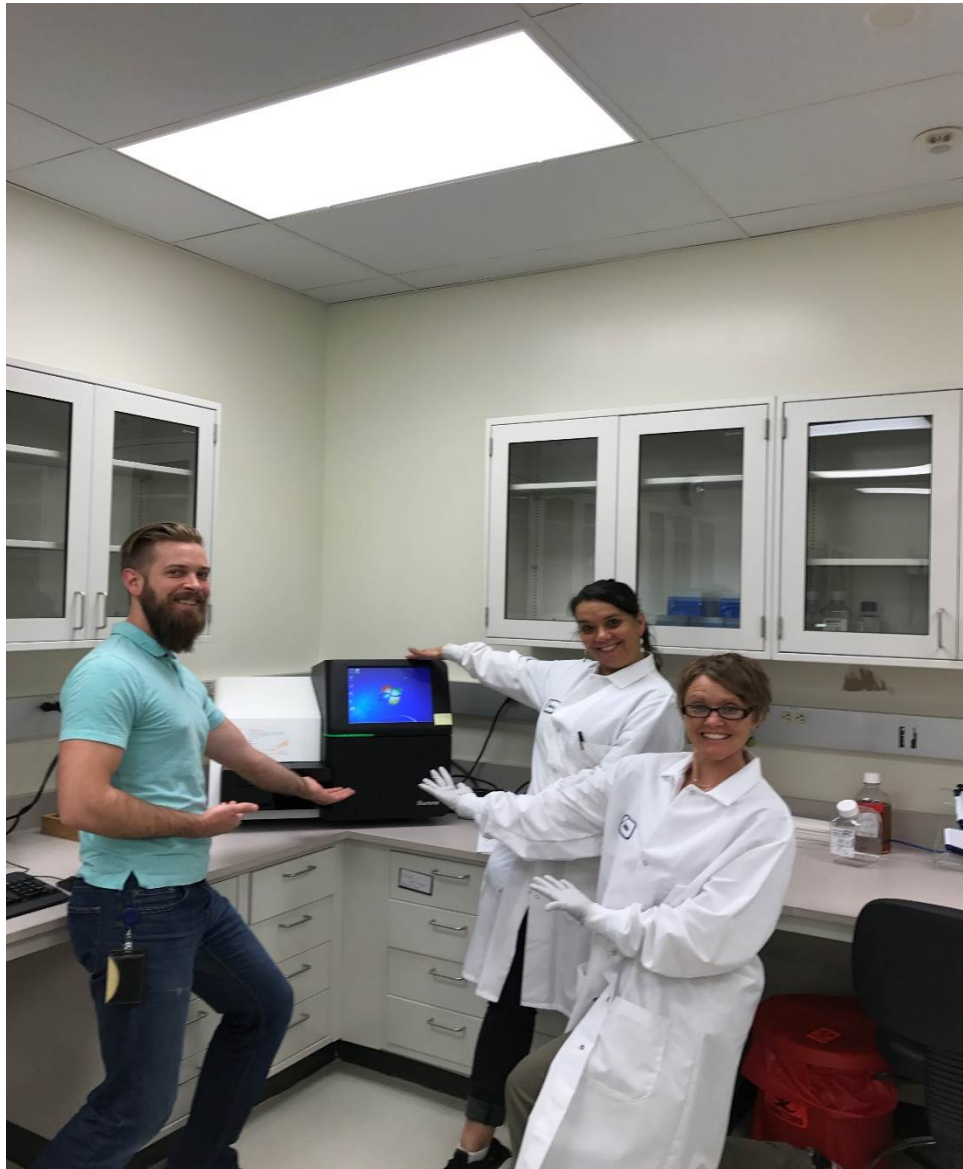
Final remarks

- Plan compliant with CLIA and CAP requirements.
- Plan is doable, however, things can change.
- Challenges: Time! Training, running PFGE and WGS simultaneously.

Acknowledgements

- Joel Sevinsky, Logan Fink, and Curtis Kapsak, CDPHE
- StaPH-B group
- Kelly Hise and Heather Carleton, PulseNet

Thank you!



WGS Microbiology Team at OSPHL

(Right to left):
Kristie Ryder
Veronica Williams
Michael Bitzer