

PULSENET STANDARD OPERATING PROCEDURE FOR ANALYSIS OF MLVA DATA OF <i>SALMONELLA ENTERICA</i> SEROTYPE ENTERITIDIS IN BIONUMERICS-BECKMAN COULTER CEQ 8000/8800/GeXP DATA	CODE: PND19		
	Effective Date:		
	06	08	10

1. **PURPOSE:** to describe the standardized protocol for analysis of MLVA data of *Salmonella enterica* serotype Enteritidis in BioNumerics.
2. **SCOPE:** to provide PulseNet participants with a single protocol for analyzing MLVA data of *S. enterica* serotype Enteritidis, thus ensuring inter-laboratory comparability of the generated results.
3. **DEFINITIONS:**
 - 3.1. **MLVA:** Multiple-locus variable-number tandem repeat analysis
 - 3.2. **VNTR:** Variable-number tandem repeat
 - 3.3. **CDC:** Centers for Disease Control and Prevention
 - 3.4. **SOP:** Standard Operating Procedure
4. **RESPONSIBILITIES/PROCEDURE:**
 - 4.1 **Software needed for data analysis**
 - 4.1.1. BioNumerics version 4.5 or higher
 - 4.1.2. Customized scripts for data import (*VNTRImport_v4.bns*), copy number calculation (= allele assignment) (*VNTRCalc_v4.bns*), troubleshooting (*VNTRReport.bns*), and support functions (*VNTRDetails_v2.bns*)
 - 4.1.3 Look-up table (*BeckmanSE.txt*) for allele size ranges
 - 4.2 **General overview**
 - 4.2.1 The import process consists of three major steps:
 - 4.2.1.1 Exporting the appropriate data file from the Beckman system (refer to SOP PNL27 step 4.10.1.6)
 - 4.2.1.2 Importing this file into BioNumerics (using the script *VNTRImport_v4.bns*)
 - 4.2.1.3 Determining the copy numbers (assigning alleles) for each VNTR (using the script *VNTRCalc_v4.bns* and the look-up table *BeckmanSE.txt*)
 - 4.2.1.3.1 *VNTRCalc_v4.bns* script facilitates the allele assignment in two different ways:
 - 4.2.1.3.1.1 “BeckmanSE” – assigns the copy number based on the look-up table (Appendix PND19-1). This method should be used in all routine analysis.
 - 4.2.1.3.1.2 “Predicted” – calculates the copy number based on the mathematical formula: (Observed fragment size – offset)/repeat size. This method should only be used when a new allele (not specified in the look-up table) is identified
 - 4.3 **Required import format**
 - 4.3.1 The data should be exported from the Beckman software as a CSV (comma-delimited) text file containing the fragment length information. Multiple columns will be present in the file but only four columns are of interest: “RN”, “dye”, “est size (nt)” and “pk height (rfu)”. All other

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 1 of 12
-----------------	---------------------	-----------------------	---------------------

columns will be ignored by the script during import. The field “RN” should contain information designating the BioNumerics key number (isolate identifier) loaded in each lane of the sequencer, as well as which VNTR mastermix or reaction was loaded. The import script will ignore any text that appears beyond the dot “.” in the RN field.

4.3.1.1 **NOTE:** no spaces or underscores are allowed between the strain ID and the reaction ID or between the reaction ID and the dot “.” in the RN field.

4.4 Setting up a new database

- 4.4.1 Click “Create New Database” icon to set up a new database
- 4.4.2 Type in the name (for example “SEMLVA”) of your new database and click “Next”
- 4.4.3 Select the directory (default on CDC training laptops is [HOMEDIR]\SEMLVA; this will save the data on the G-drive) in which you want to set up the new database. Choose the default “Yes(recommended)” when asked “Do you want to automatically create the required directories?”. Click “Next”.
- 4.4.4 Choose “Yes” to enable creation of log files and click “Finish”
- 4.4.5 In the appearing “Set up new database” pop-up window, choose “Local database (single user only)” and click “Proceed”
- 4.4.6 Click “Yes” in the next pop-up confirmation window (“Choosing ‘local database’ will restrict some functionality of the software. Are you sure you want to continue?”) and “Proceed” in the following “Plug-in” pop-up window
- 4.4.7 Close the newly created database.
- 4.4.8 Go to the directory in which you placed your database at step 4.4.3 (on CDC training laptops: C:\Program files → BioNumerics → Data), open folder “SEMLVA” and create two subfolders named “Scripts” and “VNTRtables”.
- 4.4.9 Save the four MLVA specific scripts ([VNTRImport_v4.bns](#), [VNTRCalc_v4.bns](#), [VNTRReport.bns](#), [VNTRDetails_v2.bns](#)) in the “Scripts” folder and the look-up table ([BeckmanSE.txt](#)) in the “VNTRtables” folder
- 4.4.10 Open the database. The scripts (except the VNTRDetails script) should now appear under the “Scripts” drop-down menu

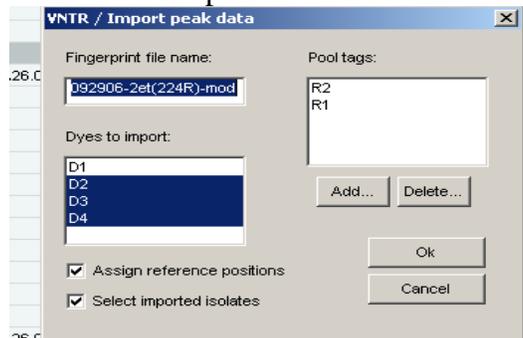
4.5 Importing a peak file for the first time

- 4.5.1 Review the peak file: make sure that the observed sizes for the D1 labelled molecular size standard peaks are within ± 1 bp from the expected size. Remove any data (failed reactions, controls, internal ladder) that you don’t want to import in the BioNumerics from the CSV file. Re-name and re-save the CSV file either on your hard drive or on the flash drive.
- 4.5.2 In BioNumerics, run the script [VNTRImport_v4](#) from the “Scripts” drop-down menu. The “Import VNTR peak data” dialog box will appear. From the “Peak file format” drop-down menu, select “Beckman peak file”. The

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 2 of 12
-----------------	---------------------	-----------------------	---------------------

script pops up a file dialog box, prompting for the name of the file to import. Select the appropriate file and click “OK”. A second dialog box pops up, prompting for a variety of other information:

- 4.5.2.1 *Fingerprint file name.* This is the name of the fingerprint file that will be used in BioNumerics. Leave this unchanged, unless a file has already been imported with the same name. In this case, change the file name into a new, unique and informative name.
- 4.5.2.2 *Dyes to import.* The first dye (D1) contains only reference markers and does not need to be imported. Use this list to select what dyes will be imported; they are D2, D3 and D4. Each dye will be stored in a different file, resulting from appending the dye name to the fingerprint file name.
- 4.5.2.3 *Assign reference positions.* Leave this option checked.
- 4.5.2.4 *Select imported isolates.* Leave this option checked.
- 4.5.2.5 *Pool tags.* For this protocol, there are two pool tags representing the two PCR reactions for each isolate (all loaded on different wells) and tagged with names “R1” and “R2”. The two tags should appear in the list. If they do not, they can be added by selecting “Add” to update.
- 4.5.2.6 If everything is filled appropriately, click “OK” to start the import.
- 4.5.2.7 **NOTE:** These settings are automatically saved and reloaded the next time this script is run.



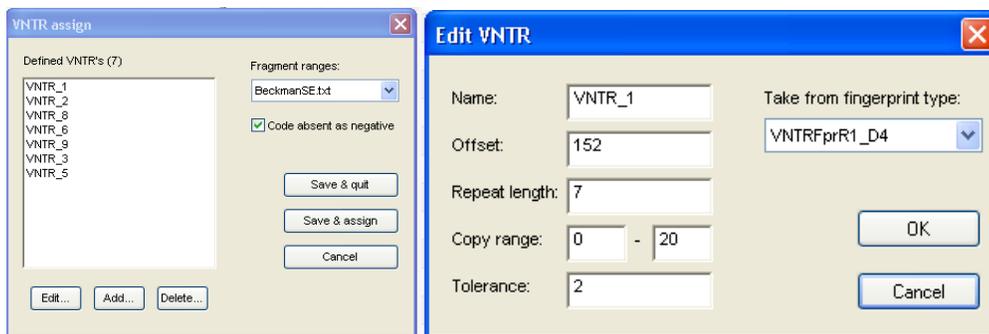
- 4.5.3 When the script is finished, it has:
 - 4.5.3.1 Created a fingerprint type “VNTRFpr” with appropriate settings
 - 4.5.3.2 Created fingerprint types with appropriate settings for each dye and each pool used. The names of these fingerprint types are “VNTRFpr”+“pool name”+“dye” (for example: VNTRFprR1_D2).
 - 4.5.3.3 Created fingerprint files for each dye in the fingerprint type “VNTRFpr,” and imported the band information in these lanes. The individual lanes are assigned to their appropriate pool fingerprint type, according to the known pool tags found in the “RN” field.
 - 4.5.3.4 Created new database entries for all isolates found
 - 4.5.3.5 Linked the fingerprints to their corresponding entries
 - 4.5.3.6 Selected all isolates for which fingerprint data was imported

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 3 of 12
-----------------	---------------------	-----------------------	---------------------

4.6 Determining the copy numbers for the first time (allele assignment)

4.6.1 Run the script *VNTRCalc_v4* from the scripts menu. The first time this script is run, it will automatically create a new character set called “*VNTR_vals*”. This character set will hold the copy numbers for each VNTR. The script will also create a second character set called “*VNTR_frags*” which holds the fragment size with a highest fluorescence for each VNTR. Click “OK” on the two pop-up windows notifying that these two character types will be created.

4.6.2 Before the script can determine copy numbers from the fragment length information, it should have sufficient information about what VNTRs can be found in which pool and dye, what are the repeat lengths, etc. To this end, the script pops up a dialog box called “VNTR assign”. This dialog box contains a list of defined VNTRs that is initially empty. To enter the parameters of the first VNTR, click on “Add”. This brings up a second dialog box, “Edit VNTR”, prompting for all properties of this first VNTR. Refer to the appendix PND19-2 for all necessary information to load the specifications for each VNTR.



4.6.3 The VNTR information includes:

- 4.6.3.1 *Name*. This name will be used for further reference and for the character name in the character type *VNTR_vals*. Each VNTR should have a unique name (for example VNTR_1, VNTR_2, etc).
- 4.6.3.2 *Offset*. Each fragment consists of a repeat portion and a constant portion, due to the fact that the primers do not locate exactly at the start and the end of the repeat region. This parameter specifies the size (in base pairs) of the constant portion.
- 4.6.3.3 *Repeat length*. This parameter specifies the size of the repeat block unit (in base pairs).
- 4.6.3.4 *Copy range*. These parameters specify the minimum and maximum number of copies that the script will consider during the copy number determination.
- 4.6.3.5 *Tolerance*. This specifies the maximum difference between the expected fragment length (calculated by the software) and the actual length estimated by the sequencer.

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 4 of 12
-----------------	---------------------	-----------------------	---------------------

- 4.6.3.6 *Take from fingerprint type.* This drop-down list should be used to indicate on what fingerprint type this VNTR was run. The fingerprint type is determined by the dye and the pool tag.
- 4.6.4 If all parameters are filled in appropriately, click OK to add the VNTR and return to the previous dialog box (“VNTR assign”). Repeat the same actions to enter the information for all VNTRs used for this protocol. When all VNTRs are entered, make sure the box for “Code absent as negative” is checked. Select “*BeckmanSE*” from the “Fragment ranges” drop-down menu and press the “Save&Assign” button to let the script assign copy numbers for the currently selected entries.
- 4.6.4.1 **NOTE 1:** pressing “Save&Quit” would store the information without assigning copy numbers for the current selection.
- 4.6.4.2 **NOTE 2:** all information regarding the VNTRs is stored with the database. The next time the script is run, the VNTR definitions will be loaded automatically.
- 4.6.4.3 For every fragment with the highest fluorescence level in the fingerprint type (= the combination of the reaction and the dye), the script will assign an allele type (=a copy number) based on the fragment size ranges specified in the “*BeckmanSE*” look-up table (Appendix PND19-1). Note that, within the same fingerprint type, more than one VNTR can be loaded if there is no overlap within the fragment size ranges.
- 4.6.5 When the script has completed, copy numbers for all VNTRs for the currently selected entries are determined.
- 4.6.5.1 **Note:** Two types of problems may arise during the process (if one or more such errors were encountered during the assignment, an error report is displayed listing all the problems):
- 4.6.5.1.1 *None of the peaks present in a fingerprint are compatible with the fragment size ranges in the look-up table.* In this case, the corresponding character value will be scored “-2.0”. In this situation, run the “*VNTRReport*” script from the “Scripts” drop-down menu to verify the reason for the problem:
- 4.6.5.1.1.1 No amplification (a null allele): mutations, insertions or deletions in the primer annealing region, locus located on a plasmid that was lost. PCR needs to be repeated only if the null allele occurred in a locus in which null alleles have not been previously detected (SE5, SE8, SE9; null alleles in locus SE6 are extremely rare and hence always need to be confirmed).
- 4.6.5.1.1.2 Fragment size outside the acceptable range (either slightly outside of range for a previously detected allele or a possible new allele). Rerun the fragment analysis reaction to confirm the accuracy and reproducibility of the sizing. The isolate should be submitted to CDC for confirmation

only if a possible new allele is detected. If the sizing is reproducibly slightly outside the range for an existing allele, the CDC database managers will adjust the look-up table based on the evidence the submitting laboratory provides.

4.6.5.1.2 *More than one peak in a fingerprint is compatible with an acceptable fragment size range.* In this case, the script will use the solution that corresponds to the peak with the largest “peak height” value. Possible causes for multiple peaks:

4.6.5.1.2.1 Primer stutter: multiple peaks with sizes 1-2 bp from each other. No further action required.

4.6.5.1.2.2 Double alleles: two peaks differing by one or more full repeats from each other. Further actions required:

4.6.5.1.2.2.1 Repeat the PCR with a freshly made template. If two peaks still observed and the same peak consistently has the highest fluorescence intensity, report the predominant peak only. If the same peak does not consistently have the highest fluorescence intensity, proceed to the next step.

4.6.5.1.2.2.2 Test ten single colony picks from the culture. Report the peak that has the highest fluorescence intensity in the majority of the picks.

4.6.5.1.2.2.3 **Note:** when the 433-435 bp allele is the predominant allele in locus SE8, in some strains a secondary allele at 344-347 bp may be detected. No further testing is needed in this case because this is a known case of a valid double allele (= two copies of the same locus in the genome).

4.7 Importing a peak file and assigning alleles on a routine basis

4.7.1 Review the peak file: make sure that the observed sizes for the D1 labelled molecular size standard peaks are within ± 1 bp from the expected size. Remove any data (failed reactions, controls, internal ladder) that you don’t want to import in the BioNumerics from the CSV file. Re-name and re-save the CSV file either on your hard drive or on the flash drive.

4.7.2 In BioNumerics, run the script *VNTRImport_v4* from the “Scripts” drop-down menu. The “Import VNTR peak data” dialog box will appear. From the “Peak file format” drop-down menu, select “Beckman peak file”. The script pops up a file dialog box, prompting for the name of the file to import. Select the appropriate file and click “OK”. “VNTR/Import peak data” dialog box will appear. De-select “D1” from the “Dyes to import” and click “OK”.

4.7.3 Run the script *VNTRCalc_v4* from the scripts menu. “VNTR assign” dialog box will appear. Select “*BeckmanSE*” from the “Fragment ranges” drop-down menu and click “Save&Assign” to assign allele numbers.

4.8 Verifying the allele assignment

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 6 of 12
----------	--------------	----------------	--------------

4.8.1 For each isolate in the database that has some VNTR data associated, you can click on the “VNTR frags” and “VNTR vals” entries in the “Experiments” list to open the “Entry edit” windows.

Character	Value	Mapping
VNTR_1	194.5 <+>	
VNTR_2	327.7 <+>	
VNTR_8	434.5 <+>	
VNTR_6	479.1 <+>	
VNTR_9	184.9 <+>	
VNTR_3	212.5 <+>	
VNTR_5	202.5 <+>	

Character	Value	Mapping
VNTR_1	6.0 <+>	
VNTR_2	8.0 <+>	
VNTR_8	2.0 <+>	
VNTR_6	10.0 <+>	
VNTR_9	2.0 <+>	
VNTR_3	4.0 <+>	
VNTR_5	8.0 <+>	

4.8.2 If the *VNTRCalc_v4* script did not detect a fragment for a VNTR, refer to the step 4.6.5.1.1 of this protocol for details on how to proceed. If the fragment size is slightly outside the range specified in the “*BeckmanSE*” look-up table it is possible to manually assign a temporary allele size and type for that VNTR until an official confirmation has been performed by CDC

4.8.2.1 **Note:** only the CDC database managers are allowed to modify the *BeckmanSE.txt* file. Once a modification has been made, the modified file will be posted on the PulseNet Sharepoint site under the QA/QC manual. An automatic e-mail notification will be sent about the change in the SOP.

4.8.2.2 To manually assign an allele size and type click on the fragment size in the “VNTRfrags” entry edit window or the allele type in the “VNTR vals” entry edit window. This will open the “Change character value” window in BioNumerics lower than 5.0. In BioNumerics 5.0, the allele size or type can be directly highlighted and changed in the entry edit window (Screen shots below)

Character	Value	Mapping
VNTR_1	6.0 <+>	
VNTR_2	7.0 <+>	
VNTR_8	2.0 <+>	
VNTR_6	10.0 <+>	
VNTR_9	2.0 <+>	
VNTR_3	4.0 <+>	
VNTR_5	8.0 <+>	

Character	Value	Mapping
VNTR_1	194.5 <+>	
VNTR_2	327.7 <+>	
VNTR_8	434.5 <+>	
VNTR_6	479.1 <+>	
VNTR_9	184.9 <+>	
VNTR_3	212.5 <+>	
VNTR_5	202.5 <+>	

4.9 Visualizing the data

4.9.1 In order to display copy numbers next to a dendrogram in a comparison, first create a “*composite data set*” that holds the VNTR data.

4.9.1.1 From the “Experiments” drop-down menu, select the option “Create new composite data set...”, enter a name (e.g. *VNTR_cmp*), and

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 7 of 12
-----------------	---------------------	-----------------------	---------------------

click the “OK” button. The “Composite data set VNTR_cmp” window will appear.

4.9.1.2 Highlight the experiment *VNTR_vals* and from the “Experiment” drop-down menu, select the option “Use in composite data set”. Close the window.

4.9.1.3 **NOTE:** the created composite dataset will be automatically saved in the database, and hence only need to be created once before the first analysis in the database.

4.9.2 Next time a comparison window is opened (see below step 4.10), there will be a new experiment *VNTR_cmp* listed in the bottom of the window (BioNumerics versions lower than 5.0) or in the top left corner of the window (BioNumerics version 5.0). This experiment will facilitate the display of a spreadsheet-like view of the copy numbers that can be shown next to a dendrogram analysis of the data set.

4.9.2.1 **NOTE:** it may be necessary to scroll the experiment list to the right with the arrow button to bring *VNTR_cmp* in display in BioNumerics versions lower than 5.0.

4.10 Performing comparisons based on the VNTR data

4.10.1. The VNTR data contained in the character set *VNTR_vals* can be analysed in BioNumerics with all the tools that are available for character data. That includes cluster analysis with a variety of methods and similarity coefficients. For VNTR data, the coefficients that make most sense are:

4.10.1.1 *Categorical:* preferred if differences in copy numbers should be treated in a qualitative way. This is the only option for creating dendrograms using MLVA data.

4.10.1.2 *Manhattan:* preferred if differences in copy numbers should be treated in a quantitative way (a larger difference means more distantly related organisms). This coefficient can be used to construct minimum spanning trees.

4.10.2 In order to create a dendrogram:

4.10.2.1 Select the isolates to be included in the dendrogram

4.10.2.2 From the “Comparison” drop-down menu, select the option “Create new comparison” and a “Comparison” window will appear

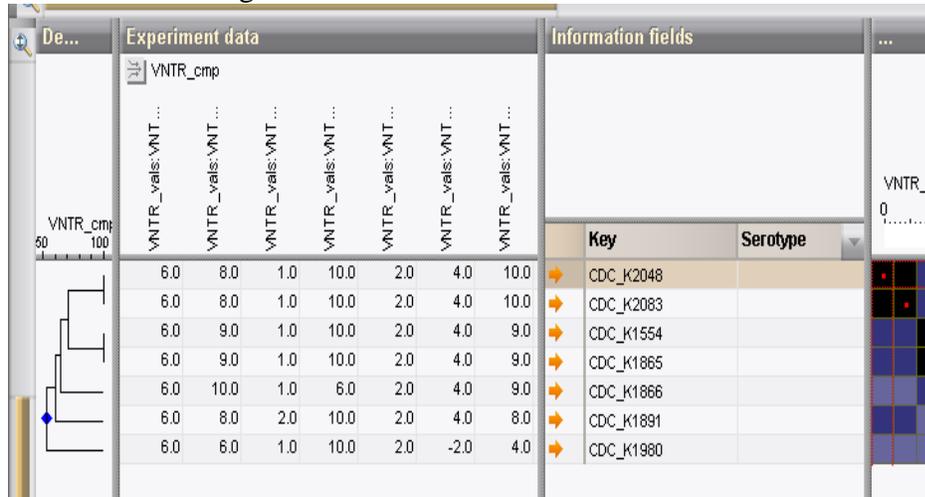
4.10.2.3 Select “*VNTR_cmp*” from the bottom of the window in BioNumerics versions lower than 5.0 or from the top left “Experiments” window in BioNumerics 5.0.

4.10.2.4 From the “Clustering” drop-down menu, select the option “Calculate...Cluster analysis (similarity matrix)” and a “Composite data set comparison” dialog box will appear

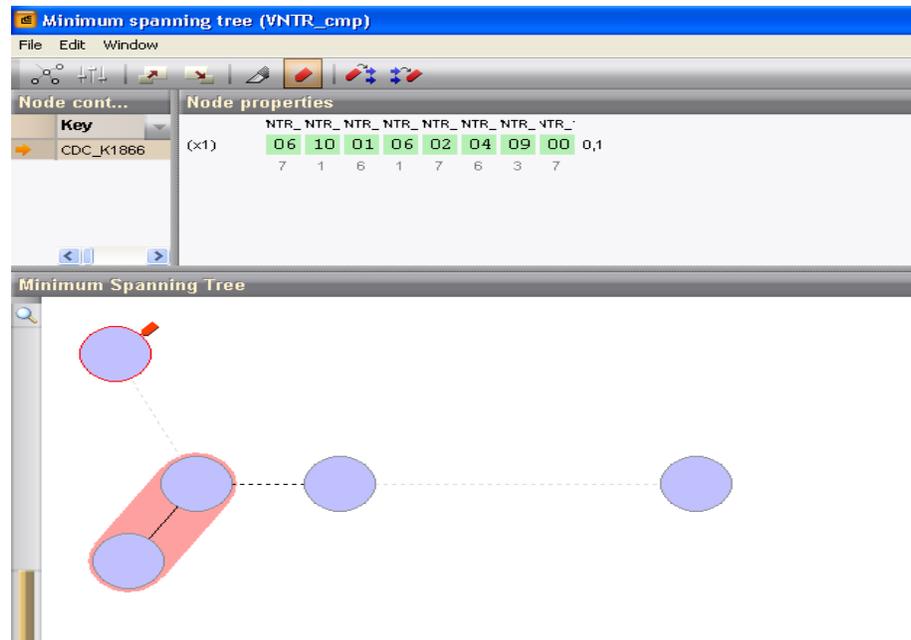
4.10.2.5 Select “Categorical” for “Multi-state coefficient” and “UPGMA” for “Dendrogram type”

4.10.2.6 Click on “OK” button to finish the calculations and the “Comparison” window with the dendrogram will reappear

- 4.10.2.7 From the “Layout” drop-down menu, select the option “Show image”
- 4.10.2.8 From the “Composite” drop-down menu, select the option ”Show quantification (values)” and the copy numbers will appear next to the dendrogram



- 4.10.3 In order to create a minimum spanning tree:
 - 4.10.3.1 Select the isolates to be included in the spanning tree
 - 4.10.3.2 From the “Comparison” drop-down menu, select the option “Create new comparison” and a “Comparison” window will appear
 - 4.10.3.3 Select “VNTR_cmp” from the bottom of the window in BioNumerics versions lower than 5.0 or from the top left “Experiments” window in BioNumerics 5.0.
 - 4.10.3.4 From the “Clustering” drop-down menu, select the option “Calculate... Minimum spanning tree (population modeling)” and a “Minimum spanning tree” dialog box will appear
 - 4.10.3.5 Make sure the default “Manhattan” is checked for “Coefficient” and click “OK” and the “Minimum spanning tree” window with the tree will appear.
 - 4.10.3.6 You can find out the content for each node by clicking on them individually.



5. FLOW CHART:

6. BIBLIOGRAPHY:

7. CONTACTS:

7.1. Eija Hyytia-Trees, D.V.M., Ph.D.

PulseNet Next Generation Subtyping Methods Unit, DFWED, CDC
(404) 639-3672

EHyytia-Trees@cdc.gov

7.2. Ashley Sabol

PulseNet Next Generation Subtyping Methods Unit, DFWED, CDC
(404) 639-2947

ASabol@cdc.gov

7.3. Patti Lafon

PulseNet Next Generation Subtyping Methods Unit, DFWED, CDC
(404) 639-2828

PLafon@cdc.gov

8. AMENDMENTS:

4/8/2013: appendix PND19-2 (BioNumerics specifications for the *S. enterica* serotype Enteritidis VNTR loci) was added.

4/8/2013: Additions to step 4.6.5: more clarification for reasons for error messages that occur upon allele assignment and guidance for further follow-up for these errors.

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 10 of 12
-----------------	---------------------	-----------------------	----------------------

Appendix PND19-1

***S. enterica* serotype Enteritidis VNTR Allele List and Corresponding
Observed Fragment Sizes in the Beckman Coulter CEQ 8000/8800/GeXP
(*BeckmanSE* look-up table)**

Note: This is posted on the PulseNet Sharepoint site under QA/QC manual as a TXT-file so users may save it locally and use it with the BioNumerics MLVA scripts. Every time the table is updated, a latest version of the look-up table is posted on the Sharepoint.

Count	VNTR_1	VNTR_2	VNTR_8	VNTR_6	VNTR_9	VNTR_3	VNTR_5
1			344-347		175-176		162-164
2	166-168		433-435		183-186		
3	172-174	295-296			192-195	197-201	171-172
4	179-182	299-303			199-200	209-214	177-179
5	185-189	305-309				224-225	183-185
6	193-196	312-316					189-191
7	200-202	318-322					195-197
8	207-209	325-329		410-413		258-259	201-203
9	214-215	332-335		443-445			207-209
10	222-223	338-342		476-481		280-282	213-215
11	229-230	344-349					219-221
12	236-237	353-354		534-535			225-227
13	244-245	360-365		569-570			231-233
14				593-594			237-239
15	256-257	372-374					243-245
16							249-250
17	270-271						255-257
18							260-263
19							266-268
20							272-274
21							
22							284-285
23							
24	317-318						
24							
25							
26							
27							
28							
29							
30	361-362						

Appendix PND19-2
BioNumerics specifications for *S. enterica* serotype Enteritidis VNTR loci

Reaction	R1	R2
Locus	VNTR_1	VNTR_9
Dye	Quas670 (D4)	Quas670 (D4)
Offset	152	166
Repeat length	7	9
Copy range	0-20	0-5
Tolerance	2	2
Fragment size in K1891	193-196	183-186
Copy number in K1891	6	2
Locus	VNTR_2	VNTR_3
Dye	D2	D2
Offset	271	165
Repeat length	7	12
Copy range	0-20	0-10
Tolerance	2	3
Fragment size in K1891	325-329	209-214
Copy number in K1891	8	4
Locus	VNTR_8	VNTR_5
Dye	Quas705 (D3)	Quas705 (D3)
Offset	260	155
Repeat length	86	6
Copy range	0-3	0-25
Tolerance	4	2
Fragment size in K1891	433-435	201-203
Copy number in K1891	2	8
Locus	VNTR_6	
Dye	Quas670 (D4)	
Offset	148	
Repeat length	33	
Copy range	0-15	
Tolerance	3	
Fragment size in K1891	476-481	
Copy number in K1891	10	

Fragment size ranges are based on independent runs on multiple instruments at CDC and in PulseNet Participating Laboratories