

1. **PURPOSE:** to describe the standardized protocol for analysis of MLVA data of *Salmonella enterica* serotype Typhimurium in BioNumerics.
2. **SCOPE:** to provide PulseNet participants with a single protocol for analyzing MLVA data of *S. enterica* serotype Typhimurium, 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 3.0 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 (*BeckmanST.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 PNL21)
 - 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 *BeckmanST.txt*)
 - 4.2.1.3.1 *VNTRCalc_v4.bns* script facilitates the allele assignment in two different ways:
 - 4.2.1.3.1.1 “BeckmanST” – assigns the copy number based on the look-up table (PND15-1). 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

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 1 of 11
-----------------	---------------------	-----------------------	---------------------

interest: “RN”, “dye”, “est size (nt)” and “pk height (rfu)”. All other 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

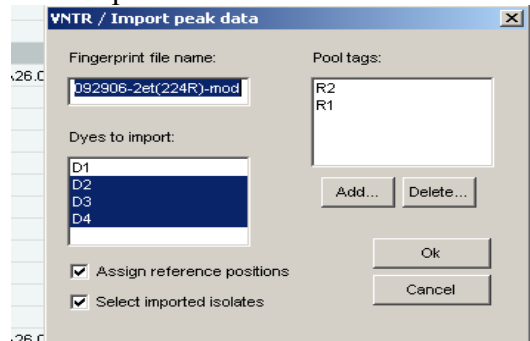
- 4.4.1 Click “New” to set up a new database
- 4.4.2 Type in the name (for example “STMLVA”) of your new database and click ”Next”
- 4.4.3 Select the default directory to set up a new database and click “Next”
- 4.4.4 Choose “Yes” to enable creation of log files and click “Finish”
- 4.4.5 Choose the default “Use the local database” and click “OK”
- 4.4.6 Go to: Program files → BioNumerics → Data → STMLVA and create two folders named “Scripts” and “VNTRtables”.
- 4.4.7 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 ([BeckmanST.txt](#)) in the “VNTRtables” folder

4.5 Importing a peak file for the first time

- 4.5.1 Review the peak file: make sure that the observed size for the D1 labelled molecular size standard peaks is 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 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.

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 2 of 11
-----------------	---------------------	-----------------------	---------------------

- 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

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_frgs” 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 VNTR’s 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 VNTR’s that is initially empty. To enter the parameters of the first VNTR, click on “Add”. This brings up a second

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 3 of 11
-----------------	---------------------	-----------------------	---------------------

dialog box, “Edit VNTR”, prompting for all properties of this first VNTR. Refer to the laboratory protocol PNL21, appendix PNL21-4 for all necessary information to load the specifications for each VNTR.



4.6.3 The VNTR information

- 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 ST3, ST5, 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 occur 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 a unit repeat block (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.
- 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 “*BeckmanST*” from the “Fragment ranges” dropdown 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 VNTR’s is stored with the database. The next time the script is run, the VNTR definitions will be loaded automatically. New incoming VNTR data can be processed

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 4 of 11
-----------------	---------------------	-----------------------	---------------------

simply by selecting the corresponding data entries, launching the script *VNTRCalc_v4* from the “Scripts” drop-down menu, and pressing the button “Save&Assign”.

- 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 “*BeckmanST*” look-up table (attachment PND15-1). 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, all copy numbers for all VNTR’s 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 whether the problem occurred because the VNTR failed to amplify (mutations, insertions or deletions in the primer annealing region) or because the VNTR had a fragment size outside the acceptable range. In the latter case, rerun the reaction to confirm the accuracy and reproducibility of the sizing. The isolate should be submitted to CDC for confirmation if the size upon retesting is still outside the fragment range specified in the look-up table.
 - 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.

4.7 Importing a peak file on a routine basis

- 4.7.1 Review the peak file: make sure that the observed size for the D1 labelled molecular size standard peaks is 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 “*BeckmanST*” from the “Fragment ranges” dropdown menu and click “Save&Assign” to assign allele numbers.

4.8 Verifying the allele assignment

- 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 list to open the “Entry edit” windows.

CDC_2009K0740			CDC_2009K0740		
Character	Value	Mapping	Character	Value	Mapping
ST3	4.0	<+>	ST3	177.4	<+>
ST5	4.0	<+>	ST5	162.7	<+>
ST7	2.0	<+>	ST7	134.4	<+>
ST10	9.0	<+>	ST10	366.8	<+>
ST2	25.0	<+>	ST2	387.9	<+>
ST6	10.0	<+>	ST6	255.5	<+>
ST8	36.0	<+>	ST8	589.6	<+>

- 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 “*BeckmanST*” 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 (**Note:** only the CDC database managers are allowed to modify the *BeckmanST.txt* file. Once a modification has been made, the modified file will be posted on CDCTeamSupport along with a notification that the file has been modified). 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)

CDC_2009K0740			CDC_2009K0740		
Character	Value	Mapping	Character	Value	Mapping
ST3	4.0	<+>	ST3	177.4	<+>
ST5	4.0	<+>	ST5	162.7	<+>
ST7	2.0	<+>	ST7	134.4	<+>
ST10	9.0	<+>	ST10	366.8	<+>
ST2	25.0	<+>	ST2	387.9	<+>
ST6	10.0	<+>	ST6	255.5	<+>
ST8	36.0	<+>	ST8	589.6	<+>

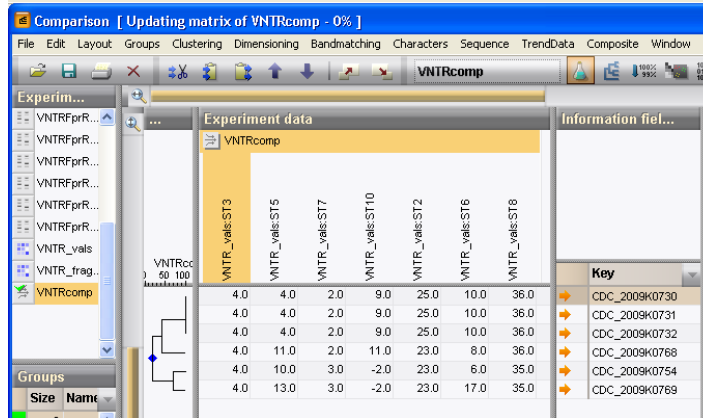
4.9 Visualizing the data

- 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 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.
- 4.9.2 Next time a comparison window is opened (see below “Analysis of the VNTR data”), 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.

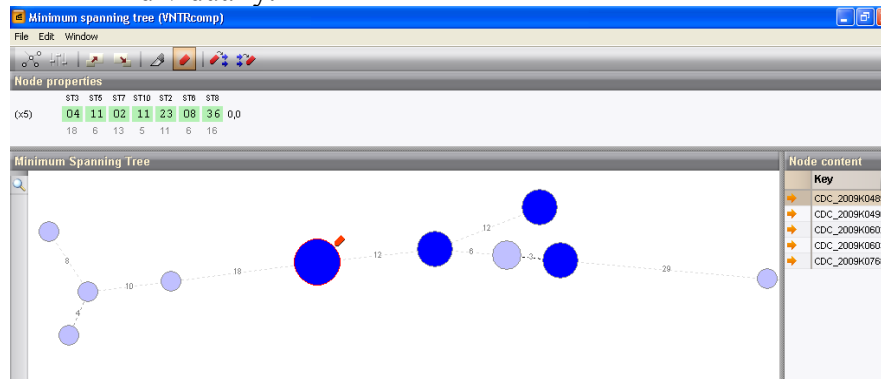
4.10 Analysis of the VNTR data

- 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 to 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 (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 Methods Development and Reference Unit, EDLB, CDC
(404) 639-3672
EHyytia-Trees@cdc.gov
- 7.2.** Patti Lafon
PulseNet Methods Development and Reference Unit, EDLB, CDC
404-639-2828
PLafon@cdc.gov
- 7.3.** Kara Cooper, Ph.D.
PulseNet Methods Development and Reference Unit, EDLB, CDC
404-639-4358
KCooper@cdc.gov

8. AMENDMENTS:

10/22/2009: the allele (= copy number) assignment will be based on a look-up table listing the actual observed fragment size ranges instead of using the mathematical formula that assigns alleles based on predicted fragment sizes. The look-up table approach in allele assignment will facilitate the comparison of fragment size data from the Beckman Coulter CEQ 8000 and the Applied Biosystems Genetic Analyzer 3130xl in the same database.

Appendix PND15-1

***S. enterica* serotype Typhimurium VNTR Allele List and Corresponding Observed Fragment Sizes in the Beckman Coulter CEQ 8000 (*BeckmanST* look-up table)**

Note: This is posted on the CDC Team 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 notification will be posted on CDC Team together with the latest version of the table.

Count	ST3	ST5	ST7	ST10	ST2	ST6	ST8
0			114-115				
1			124-126			202-203	
2	164-171	146-149	130-136	322-323			
3		151-153	139-145				
4	176-181	162-164	151-154	335-336		219-221	
5	187-190	168-170	163-164	341-343		225-227	
6		172-176	172-173	347-349		231-233	
7		179-181	178-182	354-356		234-239	
8		186-188	187-191	360-362		242-245	
9		191-194	200-201	366-368		247-250	294-296
10		197-199		373-375		254-257	
11		203-205		379-381		259-263	327-328
12		209-210	218-219			264-268	
13		214-216	227-229	385-387		272-274	
14		220-222	232-237	392-394	249-252	278-280	
15		226-228		398-400		284-287	359-360
16		232-233	250-256	404-406		290-292	373-375
17		238-239		411-412	279-281	296-298	384-385
18		243-245		417-418	289-292	302-303	
19		249-251	282-283	424-425	312-316	305-309	399-400
20		255-256	291-293	430-431	319-324	314-315	
21		261-262		437-438	341-343	320-322	
22				443-446	351-357	325-327	432-435
23		273-274	321-322	450-452	358-369	331-333	448-449
24		278-279		456-458		336-338	460-464
25		284-286		463-465	387-391	343-344	471-472
26		290-292			397-407	349-350	485-486
27				470-471		355-357	495-497
28				476-478	425-426	361-362	501-508
29		312-314		483-484	433-435	367-368	512-513
30				489-491			526-531

PULSENET STANDARD OPERATING PROCEDURE FOR ANALYSIS OF MLVA DATA OF <i>SALMONELLA ENTERICA</i> SEROTYPE TYPHIMURIUM IN BIONUMERICS-BECKMAN COULTER CEQ 8000 DATA	CODE: PND15
	Effective Date: 05 11 07

31	496-498	532-537
32		547-552
33	509-511	553-565
34	516-517	
35	523-525	580-584
36		587-592
37		599-609
38		610-618
39		622-624
40		631-636

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 11 of 11
-----------------	---------------------	-----------------------	----------------------