

PULSENET STANDARD OPERATING PROCEDURE FOR ANALYSIS OF MLVA DATA OF <i>SALMONELLA ENTERICA</i> SEROTYPE TYPHIMURIUM IN BIONUMERICS-APPLIED BIOSYSTEMS GENETIC ANALYZER DATA	CODE: PND17
	Effective Date: 10 22 09

1. **PURPOSE:** to describe the standardized protocol for analysis of MLVA data of *Salmonella enterica* serotype Typhimurium in BioNumerics.
2. **SCOPE:** to provide the 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 ([ABIST.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 Genetic Analyzer system (see laboratory protocol PNL24 step 4.10.1.15)
 - 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 [ABIST.txt](#))
 - 4.2.1.3.1 [VNTRCalc_v4.bns](#) script facilitates the allele assignment in two different ways:
 - 4.2.1.3.1.1 “ABIST” – assigns the copy number based on the look-up table (PND17-1)
 - 4.2.1.3.1.2 “Predicted” – calculates the copy number based on the mathematical formula: (Observed fragment size – offset)/repeat size (not an option for the ABI data)

4.3 Required import format

- 4.3.1 The data should be exported from the GeneMapper software as a .txt (tab-delimited) text file containing the fragment length information. Multiple columns will be present in the file but only four columns are of interest: “Dye/Sample Peak”, “Sample File Name”, “Size” and “Height”. All other columns will be ignored by the script during import. The field “Sample File Name” should contain information designating the BioNumerics key

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

PULSENET STANDARD OPERATING PROCEDURE FOR ANALYSIS OF MLVA DATA OF <i>SALMONELLA ENTERICA</i> SEROTYPE TYPHIMURIUM IN BIONUMERICS-APPLIED BIOSYSTEMS GENETIC ANALYZER DATA	CODE: PND17
	Effective Date: 10 22 09

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 “Sample File Name” 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

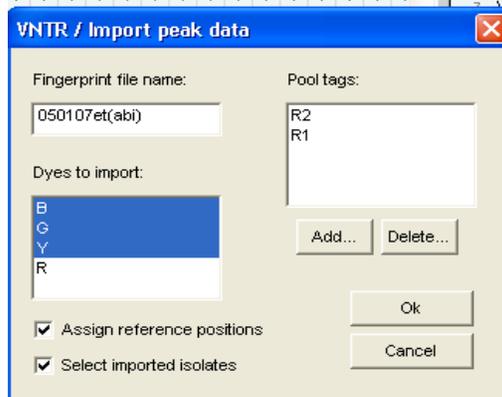
- 4.4.1 Click “New” to set up a new database
- 4.4.2 Type in the name of your new database (for example STMLVA) and click ”Next”
- 4.4.3 Select the default directory to set up a new database and click “Next”
- 4.4.4 Choose “Yes” to enable creation of log files and click “Finish”
- 4.4.5 Choose the default “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 (*ABIST.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 ROX labelled molecular size standard ran properly without skipping any peaks. Remove any data (failed reactions, controls, internal ladder) that you don’t want to import in the BioNumerics from the txt file. Re-name and re-save the txt 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 “ABI GeneMapper 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 dye R (Rox) contains only reference markers and does not need to be imported. Use this list to select what dyes will be imported; they are B (FAM), G (HEX) and Y (Calred 590). 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.

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

- 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 in 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.

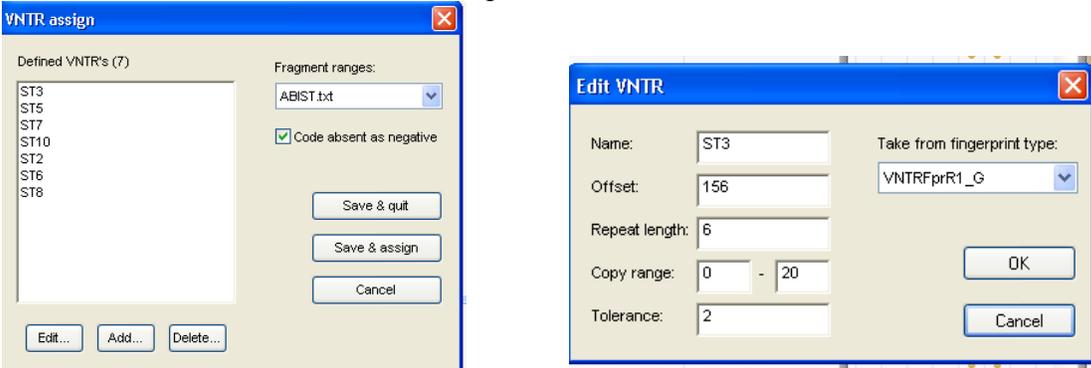


- 4.5.3 When the script is finished, it has:
 - 4.5.3.1 Created a fingerprint type “VNTRFpr” with appropriate settings.
 - 4.5.3.2 Created fingerprint types with appropriate settings for each dye and each pool used. The names of these fingerprint types are “VNTRFpr”+”pool name”+”dye” (for example: VNTRFprR1_Y).
 - 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 “Sample File Name” 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 sets 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 attachment PNL24-4 in the laboratory protocol 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 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 “*ABIST*” from the “Fragment ranges” dropdown menu and press the “Save&Assign” button to let the script assign copy numbers for the currently selected entries.

PULSENET STANDARD OPERATING PROCEDURE FOR ANALYSIS OF MLVA DATA OF <i>SALMONELLA ENTERICA</i> SEROTYPE TYPHIMURIUM IN BIONUMERIC-APPLIED BIOSYSTEMS GENETIC ANALYZER DATA	CODE: PND17		
	Effective Date: <table border="1" style="display: inline-table;"> <tr> <td style="width: 20px; text-align: center;">10</td> <td style="width: 20px; text-align: center;">22</td> <td style="width: 20px; text-align: center;">09</td> </tr> </table>	10	22
10	22	09	

- 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. New incoming VNTR data can be processed 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 “*ABIST*” look-up table (attachment PND17-1). 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 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 calculations, 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 the whole VNTR array and flanking sequences missing) 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 ROX labelled molecular size standard ran properly without skipping any peaks. Remove any data (failed reactions, controls, internal ladder) that you don’t want to import in the

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 5 of 11
-----------------	---------------------	-----------------------	---------------------

BioNumerics from the txt file. Re-name and re-save the txt 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 “ABI GeneMapper 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 “R” 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 “*ABIST*” 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	177.4	<+>	ST3	4.0	<+>
ST5	162.7	<+>	ST5	4.0	<+>
ST7	134.4	<+>	ST7	2.0	<+>
ST10	366.8	<+>	ST10	9.0	<+>
ST2	387.9	<+>	ST2	25.0	<+>
ST6	255.5	<+>	ST6	10.0	<+>
ST8	589.6	<+>	ST8	36.0	<+>

- 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 “*ABIST*” 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 PulseNet database managers are allowed to modify the *ABIST.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		
Character	Value	Mapping
ST3	177.4	<+>
ST5	162.7	<+>
ST7	134.4	<+>
ST10	366.8	<+>
ST2	387.9	<+>
ST6	255.5	<+>
ST8	589.6	<+>

CDC_2009K0740		
Character	Value	Mapping
ST3	4.0	<+>
ST5	4.0	<+>
ST7	2.0	<+>
ST10	9.0	<+>
ST2	25.0	<+>
ST6	10.0	<+>
ST8	36.0	<+>

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

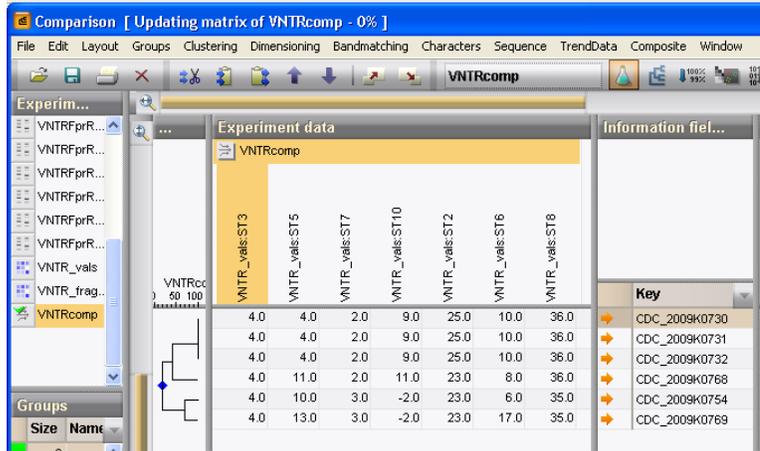
4.10.1. The VNTR data contained in the character set *VNTR_vals* can be analysed in BioNumerics with all the tools that are available to character data. That includes cluster analysis with a variety of methods and similarity coefficients. For VNTR data, the coefficients that make most sense are:

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.

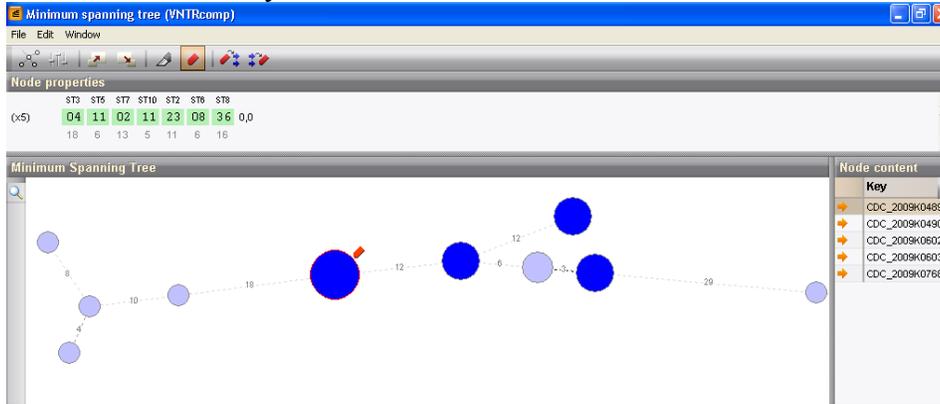
4.10.2 In order to create a dendrogram:

- 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:

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 9 of 11
-----------------	---------------------	-----------------------	---------------------

PULSENET STANDARD OPERATING PROCEDURE FOR ANALYSIS OF MLVA DATA OF <i>SALMONELLA ENTERICA</i> SEROTYPE TYPHIMURIUM IN BIONUMERICS-APPLIED BIOSYSTEMS GENETIC ANALYZER DATA	CODE: PND17		
	Effective Date: <table border="1" style="display: inline-table;"> <tr> <td style="width: 20px;">10</td> <td style="width: 20px;">22</td> <td style="width: 20px;">09</td> </tr> </table>	10	22
10	22	09	

Appendix PND17-1

***S. enterica* serotype Typhimurium VNTR Allele List and Corresponding Observed
Fragment Sizes in the Applied Biosystems Genetic Analyzer 3130xl
(*ABIST* look-up table)**

Note: This is posted on the CDCTeam 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 CDCTeam together with the latest version of the table.

Count	ST3	ST5	ST7	ST10	ST2	ST6	ST8
0			121-122				
1			128-130			205-206	
2	177-180	146-147	137-139	322-323			
3		149-153	147-151				
4	188-192	160-164	157-158	340-341		222-226	
5	200-202	166-168	166-167	345-347		229-230	
6		172-174	175-176	351-354		235-236	
7		178-179	184-185	358-359		241-243	
8		184-185	193-194	364-366		247-250	
9		190-191	202-203	370-372		253-255	298-299
10		195-197		376-378		259-261	
11		201-203		383-384		265-267	331-332
12		207-208	220-221			271-273	
13		213-214	229-230	389-390		277-279	
14		219-220	239-240	394-396	256-257	283-285	
15		225-226		401-402		289-290	364-365
16		231-233	258-259	407-409		295-296	377-378
17		237-239		413-415	286-287	300-302	391-392
18		244-245		420-421	298-299	306-308	
19		249-251	286-287	424-427	324-325	312-314	403-404
20		256-257	295-296	431-433	327-334	318-320	
21		261-263		438-439	351-352	325-326	
22				445-446	362-366	330-332	436-437
23		273-275	321-322	451-453	370-381	336-337	452-454
24		280-281		458-460		342-343	464-466
25		286-287		464-466	399-401	348-349	473-474
26		292-293			409-410	354-355	491-492
27				470-472		360-361	497-499
28				477-478	437-439	366-367	503-504
29		315-316		483-484	448-449	372-373	517-518

VERSION:	REPLACED BY:	AUTHORIZED BY:	Page 10 of 11
-----------------	---------------------	-----------------------	----------------------

PULSENET STANDARD OPERATING PROCEDURE FOR ANALYSIS OF MLVA DATA OF <i>SALMONELLA ENTERICA</i> SEROTYPE TYPHIMURIUM IN BIONUMERICS-APPLIED BIOSYSTEMS GENETIC ANALYZER DATA	CODE: PND17
	Effective Date: 10 22 09

30	489-491	530-532
31	495-496	537-538
32		550-552
33	507-508	556-565
34	514-517	
35	520-521	582-584
36		585-591
37		609-610
38		616-617
39		622-623
40		

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