PulseNet Standard Operating Procedure for Analysis of MLVA Data of Shiga-Toxin Producing *Escherichia coli* O157 (STEC O157) In BioNumerics – Applied BioSystems Genetic Analyzer 3130/3500 Data

1. PURPOSE:

To describe the standardized protocol for analysis of MLVA data of Shiga toxin-producing *Escherichia coli* O157 (STEC O157) in BioNumerics.

2. SCOPE

To provide PulseNet participants with a single protocol for analyzing MLVA data of STEC O157, 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***) and troubleshooting (***VNTRReport.bns***) and support functions (***VNTRDetails_v2.bns***)
4.1.3. Look-up table (***ABIEcoli.txt***) for allele size ranges

4.2. General Overview

4.2.1. The analysis process consists of three major steps:
4.2.1.1. Exporting the appropriate data file from the Genetic Analyzer system (Refer to SOP PNL23, PNL28, 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 ***ABIEcoli.txt***)
4.2.1.3.1. ***VNTRCalc_v4.bns*** script facilitates the allele assignment in two different ways:
4.2.1.3.1.1. "ABIEcoli" – assigns the copy number based on the look-up table (PND16-1). 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. (not an option for the ABI data)

4.3. Required Import Format

4.3.1 The data should be exported from the Gene Mapper software as a txt file (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 number (isolate identifier) loaded in each well of the sequencer plate run, as well as which VNTR master mix 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 "Create New Database" icon to set up a new database

4.4.2. Type in the name (for example "O157MLVA") of your new database and click "Next"

4.4.3. Select the directory (default on CDC training laptops is [HOMEDIR]\O157MLVA; 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 "O157MLVA" 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 (***ABIEcoli.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 size for the ROX labeled 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 "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 first 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.

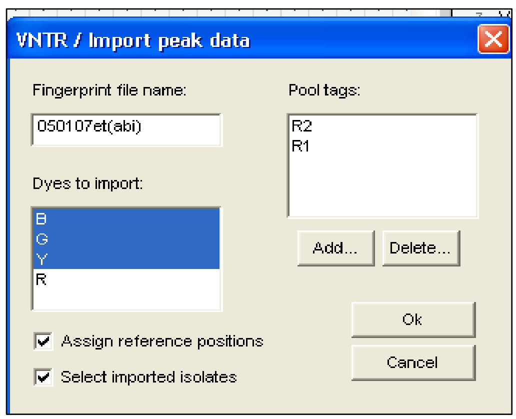
4.5.2.3. *Assign reference positions*. Leave this option checked.

4.5.2.4. *Select imported isolates*. Leave this option checked.

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_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 "RN" field.

4.5.3.4. Created new database entries for all isolates found

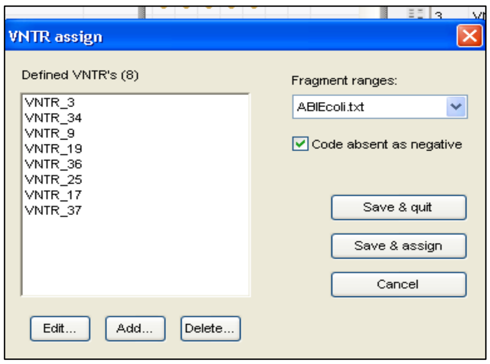
4.5.3.5. Linked the fingerprints to their corresponding entries

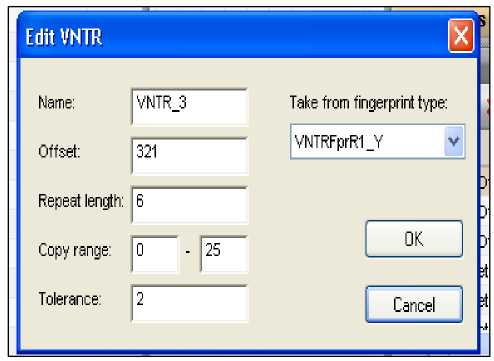
4.5.3.6. Selected all isolates for which fingerprint data was imported

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_frags*" which holds the fragment size with the 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, prompting for all properties of this first VNTR. Refer to the appendix PND16-2 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 VNTR_3, VNTR_34, 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 "***ABIEcoli***" 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 calculating 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 "***ABIEcoli***" look-up table (attachment PND16-1). 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 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 (VNTR_34, VNTR_25, VNTR_17, VNTR_19).

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.

4.7. Analyzing a Peak File on a Routine Basis

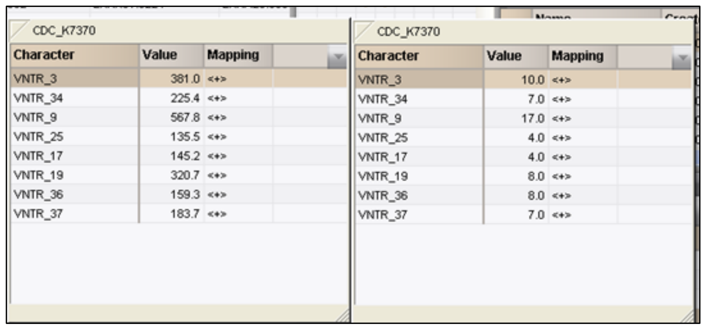
4.7.1. Review the peak file: make sure that the observed size for the D1 labeled molecular size standard peaks is within ± 1 bp from the expected size. Remove any data (failed reactions, controls, internal ladder) that you don't want to import in the BioNumerics from the CSV file. Re-name and re-save 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 "ABI Gene Mapper 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 “***ABIEcoli***” from the “Fragment ranges” drop-down menu and click “Save&Assign” to assign allele numbers.

4.8. Verifying the Allele Assignment

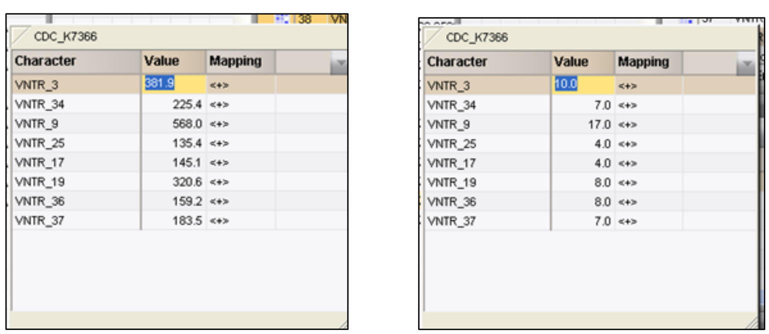
4.8.1.For each isolate in the database that has some VNTR data associated, you can click on the “VNTR frags” and “VNTR vals” entries in the list of yellow buttons to open the “Entry edit” windows.



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 “***ABIEcoli***” 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 **ABIEcoli.txt** file. Once a modification has been made, the modified file will be posted on the PulseNet SharePoint site under 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)



4.9. Visualizing Allele Types

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.

4.9.1.3. *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 (note that 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). This can be shown next to a dendrogram analysis of the data set.

4.10. Performing Comparisons Based on the VNTR Data

4.10.1. The VNTR data contained in the character set *VNTR_vals* can be analyzed 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 (i.e. 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.

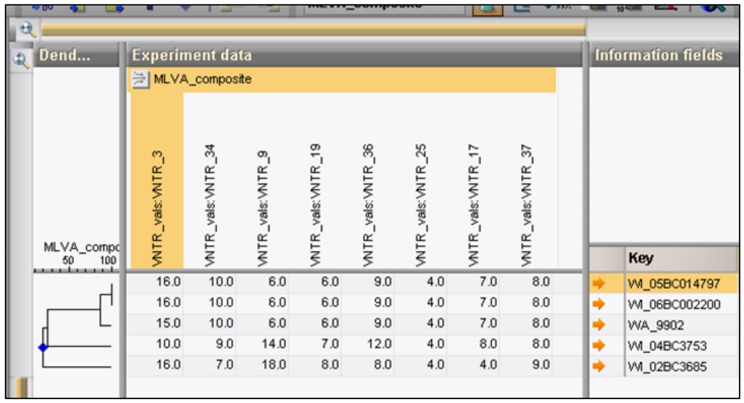
4.10.2.4. From the "Clustering" drop-down menu, select the option "Calculate…Cluster analysis (similarity matrix)" and a "Composite data set comparison" dialog box will appear

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

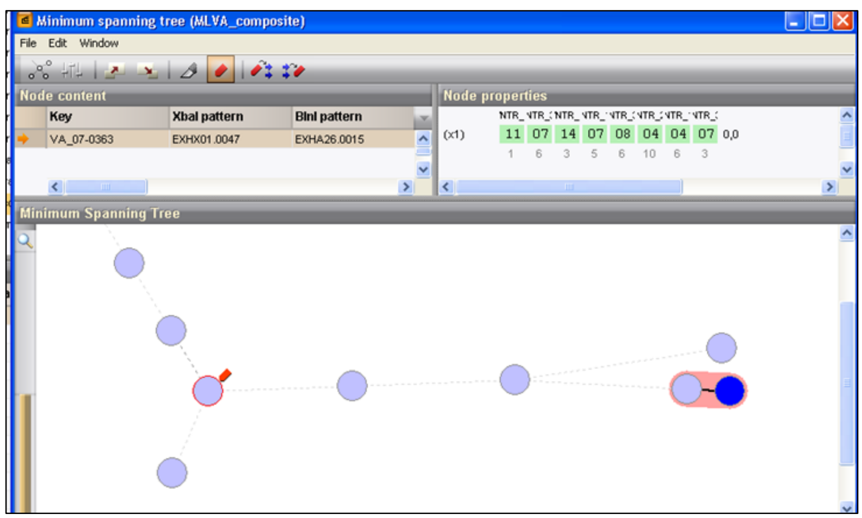
4.10.3.2. From the "Comparison" drop-down menu, select the option "Create new comparison" and a "Comparison" window will appear

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. CONTACTS: PULSENET NEXT GENERATION SUBTYPING UNIT, EDLB, DFWED, CDC

5.1. Eija Trees, D.V.M., Ph.D (404) 639-3672 EHyytia-Trees@cdc.gov

5.2. Patti Lafon (404) 639-2828 PLafon@cdc.gov

5.3. Ashley Sabol (404) 639-2947 ASabol@cdc.gov

Appendix PND16-1

STEC O157 VNTR Allele List and Corresponding Observed Fragment Sizes in the Applied Biosystems Genetic Analyzer 3130/3500 (*ABIEcoli* look-up table)

NOTE: This is posted on the PulseNet SharePoint site under the 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, the latest version is posted on SharePoint site.

Count	VNTR_3	VNTR_34	VNTR_9	VNTR_25	VNTR_17	VNTR_19	VNTR_36	VNTR_37
1			474-475	122-124				
2	339-340			127-128	135-136	283-284		
3	343-346		485-486	132-134	140-142	291-292	123-124	157-158
4	349-352	169-171	491-492	138-140	146-148	297-298	129-131	163-165
5	355-358	187-188	496-498	144-146	153-155	302-304	135-137	169-171
6	361-364	205-206	502-504	150-152	159-161	308-310	141-144	175-176
7	367-370	222-224	508-510	156-158	165-167	314-316	150-151	181-183
8	373-376	242-244	514-516	162-164	172-173	320-322	157-158	187-189
9	379-382	260-262	520-522	168-169	177-179	326-328	163-165	193-194
10	385-388	278-280	526-528	174-175	184-185	332-334	170-172	199-200
11	391-394	295-298	531-534		190-191	338-340	177-179	205-206
12	397-401	315-316	536-540		196-197	344-346	184-185	210-212
13	404-406		543-545	191-192	201-203	350-351	191-192	217-218
14	409-412		549-551			356-357	197-199	223-224
15	415-418		555-557		208-209		204-205	229-231
16	421-424		560-563				211-213	236-237
17	428-430		566-569				218-219	240-244
18	433-435		572-575		225-228		225-226	249-250
19	440-442		578-581				233-234	255-256
20	447-448		584-585				240-242	262-263
21	453-455		590-591		247-248			
22	459-461		596-597					273-274
23	466-467		601-603					
24	472-473		607-608					
25	478-479		613-614					

Appendix PND16-2

BioNumerics specifications of STEC O157 VNTR loci

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

Reaction	***R1***	***R2***
Locus	**VNTR_3**	**VNTR_17**
Dye	**CalRed590=Y**	**CalRed590=Y**
Offset	321	120
Repeat Length	6	6
Copy Range	0-25	0-30
Tolerance	2	2
Fragment Size in EDL933	379-382	159-161

Reaction	***R1***	***R2***
Locus	**VNTR_25**	**VNTR_37**
Dye	**HEX=G**	**HEX=G**
Offset	110	142
Repeat Length	6	6
Copy Range	0-20	0-25
Tolerance	2	2
Fragment Size in EDL933	138-140	187-189

Reaction	***R1***	***R2***
Locus	**VNTR_34**	**VNTR_19**
Dye	**FAM=B**	**FAM=B**
Offset	100	272
Repeat Length	18	6
Copy Range	0-20	0-25
Tolerance	3	2
Fragment Size in EDL933	278-280	308-310

Reaction	***R1***	***R2***
Locus	**VNTR_9**	**VNTR_36**
Dye	**FAM=B**	**FAM=B**
Offset	465	102
Repeat Length	6	7
Copy Range	0-50	0-20
Tolerance	2	2
Fragment Size in EDL933	531-534	157-158