DEVELOPMENT OF MULTI-LOCUS VARIABLE NUMBER TANDEM REPEAT ANALYSIS FOR OUTBREAK DETECTION OF *NEISSERIA MENINGITIDIS*

by

Alicia Anne Price

B.S. Biology/Chemistry, Lock Haven University of Pennsylvania, 2004

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master in Science

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Alicia Anne Price

It was defended on

April 7, 2006

and approved by

Lee H. Harrison, M.D.
Thesis Advisor, Professor
Departments of Infectious Diseases and Microbiology, Epidemiology, and Medicine
Graduate School of Public Health and School of Medicine
University of Pittsburgh

M. Catherine McEllistrem, M.D., M.S.
Committee Member, Assistant Professor
Departments of Epidemiology and Medicine
Graduate School of Public Health and Medicine
University of Pittsburgh

Jeremy Martinson, Ph.D.
Committee Member, Assistant Professor
Department of Infectious Diseases and Microbiology
Graduate School of Public Health
University of Pittsburgh

Saleem A. Khan, Ph.D.
Committee Member, Professor
Department of Molecular Genetics and Biochemistry
School of Medicine
University of Pittsburgh

Copyright © by Alicia Anne Price

2006

DEVELOPMENT OF MULTI-LOCUS VARIABLE NUMBER TANDEM REPEAT ANALYSIS FOR OUTBREAK DETECTION OF *NEISSERIA MENINGITIDIS*

Alicia Anne Price, MS

University of Pittsburgh, 2006

Neisseria meningitidis is a major cause of septicemia and meningitis worldwide. Traditional typing methods like pulsed-field gel electrophoresis (PFGE) for identifying outbreaks are subjective and time consuming. Multi-locus variable number tandem repeats analysis (MLVA) is an objective typing method amenable to automation that has been used to type other bacterial pathogens. This report describes the development of MLVA for outbreak detection of *N. meningitidis.* Tandem Repeats Finder software was used to identify variable number tandem repeats (VNTRs) from 3 sequenced *N. meningitidis* genomes. PCR amplification of identified VNTRs was performed on DNA from 7 serogroup representative isolates. PCR products were sequenced and repeats were manually counted. VNTR loci identified by this screen were evaluated on a collection of 46 outbreak and sporadic serogroup C isolates. Alleles at each locus were concatenated to define the MLVA type for each isolate. Minimum spanning tree (MST) analysis was performed to determine the genetic relationships among the isolates. The genetic distance was defined as the summed tandem repeat difference (STRD) between isolates MLVA types. Outbreak clusters were defined by a STRD $\leq$3. These data was compared to PFGE data to determine the utility of MLVA for outbreak detection. Twenty-one VNTR loci with variable copy numbers among the sequenced genomes were identified that met the established criteria of short repeat length and consensus sequence > 85%. Seven VNTR loci were reliably amplified among the 7 serogroups tested. These loci had repeat lengths between 4 and 20 nucleotides and exhibited between 10 and 26 alleles among 61 isolates belonging to 7 different serogroups. MST analysis with 7 loci differentiated serogroups, discriminated sporadic isolates and identified 7 out of 8 serogroup C outbreaks. In summary, MLVA with 5 VNTR loci distinguished *N. meningitidis* isolates from 7 different serogroups and sporadic isolates within each serogroup. In addition, MLVA identified 88% of PFGE-defined serogroup C outbreaks. Further investigation of these and additional outbreak-associated isolates is necessary to define the optimal

combination of VNTR loci and to evaluate MST analysis criteria in order to determine the utility of MLVA for *N. meningitidis* outbreak detection.

TABLE OF CONTENTS

ACKNOWLEDGMENTS IX
1.0 SPECIFIC AIM 1
2.0 INTRODUCTION 2
3.0 MATERIALS AND METHODS 5
3.1 *NEISSERIA MENINGITIDIS* ISOLATES 5
3.2 IDENTIFICATION OF VNTR LOCI 6
3.3 VNTR PCR AND SEQUENCE ANALYSIS 6
3.4 PULSED FIELD GEL ELECTROPHORESIS (PFGE) 8
3.5 DATA ANALYSIS 9
4.0 RESULTS 11
4.1 VNTR LOCUS IDENTIFICATION AND CHARACTERIZATION 11
4.2 MINIMUM SPANNING TREE (MST) ANALYSIS 17
4.3 COMPARISON OF MLVA WITH PFGE 27
5.0 DISCUSSION 31
BIBLIOGRAPHY 34

LIST OF TABLES

Table 1. *N. meningitidis* isolate collection – serogroup (SG) distribution.................................... 5

Table 2. Characteristics of *N. meningitidis* serogroup C outbreak and sporadic isolates............. 6

Table 3. Primer sequences for identified VNTR loci. 8

Table 4. Characteristics of 21 identified VNTR loci. 13

Table 5. Number of VNTR loci by serogroup. 14

Table 6. Characteristics of 7 *N. meningitidis* MLVA loci. 15

Table 7. STRD ranges for outbreaks using different combinations of VNTR loci. 17

Table 8. Categorical coefficient ranges for outbreaks using different combinations of VNTR loci. 23

LIST OF FIGURES

Figure 1. Gel electrophoresis of PCR products demonstrating TR variation at – VNTR01, AP8, AP5, AP6. 15
Figure 2. Distribution of loci on *N. meningitidis* genomes (A) Z2491 (B) MC58 (C) FAM18 ... 16
Figure 3. MST of MLVA data with all 7 VNTR loci using a STRD <8*. 18
Figure 4. MST of MLVA data without AP18 using a STRD <6*. 19
Figure 5. MST of MLVA data without AP10 and AP18 using a STRD ≤5*. 20
Figure 6. MST of MLVA data without AP14 and AP18 using a STRD ≤3*. 22
Figure 7. MST of MLVA data without AP14 using a categorical coefficient ≤1*. 25
Figure 8. MST of MLVA data without AP14 and AP18 using a categorical coefficient ≤1*...... 26
Figure 9. Dendrogram based on PFGE data with MLVA results included. 29
Figure 10. Comparing MLVA and PFGE data from two outbreaks (Maryland and Arizona). (A) MLVA types for Md and Arizona outbreak isolates (B) PFGE results for Md outbreak isolates (C) PFGE results for Arizona outbreak isolates. 30

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Dr. Lee Harrison for his support and guidance on my project, as well as my other committee members Drs. Jeremy Martinson, Catherine McEllistrem, and Saleem Khan. Special thanks go to Dr. Jane Marsh who served as another unofficial advisor and committee member and provided a great deal of guidance, advice, and support every step of the way. I would also like to thank Mary O'Leary, for training me in the ways of the lab, and all of the members of the Public Health Infectious Disease Laboratory for their support. Finally, I want to thank all of my family and a friends, with a special thanks to a great friend and my "IDM sidekick" Leah Kostelnik, who have provided an enormous amount of support along the way.

1.0 SPECIFIC AIM

To develop multi-locus variable number tandem repeat analysis (MLVA) as a method for public health surveillance and outbreak detection of the most prevalent *Neisseria meningitidis* serogroups.

2.0 INTRODUCTION

Neisseria meningitidis is a major cause of bacterial meningitis and septicemia worldwide. It is a gram negative diplococcus and a commensal bacterium that resides in the human nasopharynx [1]. Pathogenic strains of *N. meningitidis* have a polysaccharide capsule while unencapsulated strains are often found in the nasopharynx of asymptomatic carriers [2]. Infection is acquired through inhalation of infectious droplets following close contact with a case or carrier. The case fatality rate for those infected tends to be around 12% depending on clinical presentation [3]. Prompt and accurate diagnosis and treatment helps to increase survival rates and prevent secondary cases of infection [4, 5].

There are 13 serogroups based on the meningococcal capsular polysaccharides with the five most important serogroups associated with disease being A, B, C, W-135, and Y [5, 6]. Periodically, large epidemics of *N. meningitidis* serogroup A occur in the meningitis belt of Africa while the United States is most often affected by outbreaks of serogroup C. Serogroup B is common throughout the world while serogroup C is often present throughout North America and Europe. Serogroup Y disease in common in the United States but relatively rare in Europe. Serogroup W-135 is relatively uncommon in the United States but recently caused a large outbreak in Saudi Arabia with subsequent worldwide spread [3, 5, 7-10].

Molecular epidemiologic studies have improved meningococcal surveillance and allowed a better understanding of the dissemination of this organism. There are a number of molecular subtyping methods that are currently used for characterization of *N. meningitidis*, which include pulsed-field gel electrophoresis (PFGE), multi-locus sequence typing (MLST), and outer membrane protein (OMP) typing [5]. PFGE is the current "gold standard" for outbreak detection and local epidemiologic analysis of *N. meningitidis*. This method involves the separation of large DNA restriction fragments by periodic alteration of non-uniform electric currents [11]. The advantage of this method is that it is highly discriminatory. However, it is also very

subjective, time-consuming, and the data is not easily portable between laboratories [12, 13]. MLST involves DNA sequencing seven housekeeping gene segments. These genes were chosen because they are not under selective pressure and therefore allow for the assessment of the organism's genetic background. Unique housekeeping gene sequences are assigned different numeric alleles and the combination of these alleles results in a sequence type (ST). Some of the advantages of MLST as compared with other methods like PFGE are that it is objective, portable, and automatable. MLST is a useful tool for global epidemiologic analysis and although it has many advantages, it is not as discriminatory as PFGE for outbreak detection of *N. meningitidis* [12, 14]. For example, most serogroup C strains causing invasive disease in the United States below to a single ST complex (ST-11). OMP typing utilizes the inherent variability of the PorA and PorB porin proteins and the iron regulated FetA protein as a molecular tool for the characterization of *N. meningitidis*. These outer membrane proteins are under selective pressure to change because they are surface exposed and subject to immune surveillance. These immunogenic variable regions form the basis for molecular genotyping at the 3 *N. meningitidis* OMP gene loci. Together with MLST, OMP typing can provide enhanced discriminatory power for meningococcal outbreak detection [5].

Multi-locus variable-number tandem repeat analysis (MLVA), as seen with other bacterial pathogens, may provide a higher throughput and automatable subtyping method with more discriminatory power than current subtyping methods for the molecular surveillance of *N. meningitidis*. MLVA utilizes short DNA tandem repeats which are areas of the bacterial genome that evolve rapidly and vary in number among different strains of the same species. MLVA also utilizes repeats at multiple loci for the assessment of relationships between bacterial strains of the same species and has been used to genotype several other important bacterial pathogens including *Escherichia coli* O157:H7 and *Bacillus anthracis* [13, 15]. Some of the advantages of MLVA are that it provides objective data, it is portable between laboratories and a large number of samples can be evaluated simultaneously and automatically. One of the drawbacks of MLVA is that VNTR loci can be highly variable causing them to be over-discriminatory [15]. Therefore, VNTR loci for MLVA must be chosen carefully.

In this study, we developed MLVA for *N. meningitidis* on a subset of isolates from the seven major serogroups that had been previously characterized by MLST and PFGE. The goal of the study was to determine the utility of MLVA as a molecular subtyping tool for public

health surveillance and outbreak detection of the most prevalent global serogroups of *N. meningitidis*. We focused on serogroup C isolates in this study because it is the most common cause of outbreaks in the United States.

3.0 MATERIALS AND METHODS

3.1 *NEISSERIA MENINGITIDIS* ISOLATES

A total of 61 *N. meningitidis* isolates were used in this study. The Public Health Infectious Disease Laboratory (PHIDL) obtained these isolates from the Center for Disease Control (CDC) and the Active Bacterial Core Surveillance (ABCs) Maryland site, a component of the CDC-funded. Emerging Infectious Diseases Program (EIP). This collection was comprised of 46 serogroup C isolates and 15 isolates that span the other six most commonly occurring *N. meningitidis* serogroups - A, B, W-135, X, Y, and Z (Table 1).

There were 21 isolates from 8 *N. meningitidis* serogroup C outbreaks collected between 1986 and 1999 from various US locations included in this study (Table 2). Each outbreak was indistinguishable by PFGE and confirmed by epidemiologic investigations by the CDC [16]. In addition, there were 25 sporadic serogroup C isolates from the outbreak states that had no known epidemiologic link to each outbreak and were PFGE unique (Figure 9, Table 2). Together, this collection of 46 sporadic and outbreak serogroup C isolates was used for validation of VNTR loci.

Table 1. *N. meningitidis* isolate collection – serogroup (SG) distribution.

SG	# of isolates in each serogroup
A	3
B	5
C	46
W-135	2
X	1
Y	3
Z	1
Total	61

Table 2. Characteristics of *N. meningitidis* serogroup C outbreak and sporadic isolates.

<table>
<tr><th>Outbreak</th><th>Geographic Location</th><th>Year of Isolation</th><th># of Outbreak Isolates</th><th># of Sporadic Isolates</th></tr>
<tr><td>1</td><td>Virginia</td><td>1986</td><td>4</td><td>0</td></tr>
<tr><td>2</td><td>California</td><td>1993</td><td>2</td><td>2</td></tr>
<tr><td>3</td><td>Arizona</td><td>1994</td><td>2</td><td>2</td></tr>
<tr><td>4</td><td>New Mexico</td><td>1995</td><td>3</td><td>2</td></tr>
<tr><td>5</td><td>Texas</td><td>1995</td><td>3</td><td>5</td></tr>
<tr><td>6</td><td>Maryland (Montgomery)</td><td>1995</td><td>2</td><td rowspan="3">14</td></tr>
<tr><td>7</td><td>Maryland (Balt. City)</td><td>1997</td><td>2</td></tr>
<tr><td>8</td><td>Maryland (Prince Charles)</td><td>1999</td><td>3</td></tr>
<tr><td>Total</td><td colspan="2"></td><td>21</td><td>25</td></tr>
</table>

3.2 IDENTIFICATION OF VNTR LOCI

Five variable number of tandem repeats (VNTR) loci originally identified by Yadankhah *et al.* were evaluated on our *N. meningitidis* collection [17]. Additional tandem repeat loci were identified by screening the genomes of *N. meningitidis* strains Z2491 (serogroup A), MC58 (serogroup B), and FAM18 (serogroup C) using Tandem Repeats Finder (TRF) Software, version 3.21 [18]. Tandem repeats identified by TRF were selected for further screening based on the following criteria – a repeat length between 4 and 20 bp and a variable copy number both across and within serogroups.

3.3 VNTR PCR AND SEQUENCE ANALYSIS

PCR primers were designed using DNASTAR Primer Select and MegAlign software, version 6.1 (DNAstar, Madison WI), based on the consensus of sequences flanking the repeats of the 21 loci across all 3 sequenced genomes. PCR amplification was performed in 50µl reactions on a Gene Amp 9700 PCR system (Applied Biosystems, Foster City, CA) with 1.5 U/µl AmpliTaq Gold

polymerase (Applied Biosystems), 0.2 µM primers, and a 1X PCR buffer containing 2.5mM $MgCl_2$,. Initial denaturation was performed at 95°C for 5 minutes followed by 35 cycles at 95°C for 1 minute, annealing for 1 minute and extension at 72°C for 1 minute, followed by a final extension step at 72°C for 7 minutes. Primers and annealing temperatures for the 7 selected VNTR loci retained are shown in Table 3. Gel electrophoresis was performed using 5 µl of each PCR product on a 1% agarose gel containing 0.5 µg/ml ethidium bromide. A 100 bp DNA ladder (Invitrogen, Carlsbad, CA) was used as a size standard. The gel image was digitally photographed using a Gel Doc 2000 (BioRad, Hercules, CA).

For 96 well plates, PCR reactions were prepared for sequencing by precipitation with 60 µl of a 20% polyethylene glycol 8000 (PEG) (Sigma-Aldrich, St. Louis, MO), 2.5 M NaCl solution for 1 h at room temp followed by centrifugation at 2750 x *g* for 1 h at 4°C. The PEG precipitated PCR product was washed twice with 70% ethanol followed by centrifugation at 2750 x *g* for 10 min at 4°C. The cleaned PCR product was resuspended in 15 µl dH_2O and 1 µl of this product was used for sequencing with the ABI Big Dye Terminator Kit v.3.1 (Applied Biosystems) with the same primers employed in the PCR reaction. For single reactions, PCR products were prepared for sequencing using ExoSap-It (USB Corporation, Cleveland, OH), Five µl of the PCR product was mixed with 2 µl of ExoSap-It and this reaction was run on the Gene Amp 9700 PCR system (Applied Biosystems, Foster City, CA). The treatment phase was run for 15 minutes at 37°C followed by inactivation phase which ran for another 15 minutes at 80°C.

Sequencing of PEG precipitated PCR products was performed in 5 µl reactions on a GeneAmp 9700 PCR system (Applied Biosystems) with 1 µl Big Dye Terminator (Applied Biosystems), 0.875 µl 5X Sequencing buffer (Applied Biosystems), and 2.0 µl of either 0.67µM forward or reverse primer. Each PCR product was sequenced on both strands to generate consensus sequence. Sequencing reactions were run for 30 cycles of 96°C denaturation for 10 seconds, followed by 50°C annealing for 5 seconds, and 60°C extension for 2 minutes. Sequencing reactions were then cleaned using a sodium acetate ethanol (NaOAc/ETOH) cleanup. Sequenced products were brought up to 20 µl with dH_2O followed by the addition of 52 µl of a 0.12M sodium acetate and 100% ethanol solution which was kept at room temperature for 45 minutes before centrifugation at 2750 x *g* for 1 h at 4°C. The sequenced product was

washed once with 70% ethanol followed by centrifugation at 2750 x *g* for 10 min at 4°C. One µl of ExoSap-It treated PCR products were sequenced directly in six microliter reactions with 3 µl dH_2O and either 2 µl of 3.2 µM forward or reverse primer. The amount of each product was adjusted based upon the intensity of the PCR product as observed by gel electrophoresis. When sequencing reactions with weaker PCR products, the amounts of both primer and DNA was increased by 1 µl. Automated sequence detection was performed on an ABI 3730 sequence detection system (Applied Biosystems) for 96 well plates and on an ABI 3100 (Applied Biosystems) for single reactions at the University of Pittsburgh Genomic and Proteomic Core Laboratory. Resulting electropherograms were analyzed manually using DNASTAR Seqman software, version 6.1 (DNAstar), by combining both the forward and reverse sequences for each isolate, searching for the 5' and 3' flanking regions, and counting the number of tandem repeats. A MLVA result for an isolate at a locus represents an allele and subsequently alleles at each locus were concatenated to define the MLVA type for each isolate.

Table 3. Primer sequences for identified VNTR loci.

TR ID	Forward primer	Reverse Primer	AT (ºC)
VNTR06*	CCGGCGGCGCGTGATGACTT	GCAGAAACCCCGCAGACAGGATGG	60
AP6	AAAAAGCGGCGCGATACACC	GGGGAAAGGAAAACAGGGAAAAGA	57
AP8	AAGTTATTTGGAAGCGTGTT	TAATAAAATCATCCGAATCAATAA	51
AP9	GGACATCGCCCTTTCACG	GCTTCATCGCCTTGTCCTG	51
AP10	AATTTCTGTCTTCCGCCGCTTCT	AGACCTTTAAACCCCGACCATCCT	57
AP14	CGGCACCCCATATCCTGACAAAAT	CCGCTACAGAAAGTGGCAAGGATG	54
AP16	GCCGATGCATGAGGTTAG	AAGGCGTGAATTTGTATGAA	54
AP18	GAAGTGAATGGTGTGCTGGTGTTT	ATGAGATTTTCGGCGGGTGTG	57

*J Clin Microbiol. 2005 Apr; 43(4):1699-705 [17]

3.4 PULSED FIELD GEL ELECTROPHORESIS (PFGE)

Pulsed field gel electrophoresis was previously reported on the 19 *N. meningitidis* isolates described in this study by the CDC [16]. PFGE was repeated on these isolates and completed for the remaining collection for comparison with MLVA data. The *N. meningitidis* isolates were grown from frozen glycerol stocks at 37°C overnight in the presence of 5% CO_2 on chocolate

agar plates (Becton, Dickinson, and Company, Sparks, Md). The following day isolates were subcultured from an individual colony onto a fresh chocolate agar plate and incubated overnight as described above. A bacterial suspension corresponding to 10% transmission on a colorimeter (Hach Company, Loveland, CO) was prepared in 1 ml of 0.5X TE (1 M Tris-HCl/ 0.1 M EDTA). Then 200 µl of the bacterial suspension was mixed with 200 µl of 2% molten Seaplaque (BioWhittaker Molecular Applications, Rockland, ME) agarose and dispensed into plug casts. The control strain used was MD01329, a serogroup C *N. meningitidis* isolate collected from Maryland. Plugs were digested overnight in ESP buffer (1% Sarkosyl, 1 mg/ml of Proteinase K, and 0.43 M EDTA). The next day, plugs were washed four times in 1X TE at 37ºC for 30 minutes. Plugs were cut into thirds and incubated at 37ºC in 1X NEB2 buffer (New England BioLabs, Beverly, MA) for 1 hour and then restricted using 20 U *Nhe*I (New England BioLabs) overnight at 37°C. The next day, bacterial plugs were loaded onto a 1% SeakemGold agarose (Cambrex Bio Science, Rockland, ME) gel and electrophoresed on a CHEF III system (Biorad) for 14 hours on block 1 with switch times of 1 second and 30 seconds followed by a run of 8 hours on block 2 with switch times of 5 seconds and 9 seconds in 0.5X TE buffer. The gel was stained with 2 µg/ml ethidium bromide for 1 hour followed by a 4 hour wash in dH_2O to destain the gel. The gel was photographed using a Gel Doc 2000 (Biorad).

3.5 DATA ANALYSIS

Bionumerics (Applied Maths, Austin, TX) software v. 4.0 was used for data analysis. Dendrograms and minimum spanning tree (MST) analysis of PFGE and MLVA results were used to determine whether or not isolates were outbreak associated or sporadic.

PFGE results were analyzed by constructing dendrograms in order to identify isolates that were identical or closely related. UPGMA dendrograms were created on the basis of pairwise differences in the allelic profiles of an isolate by using a dice similarity coefficient[19, 20]. The tolerance levels used were an optimization setting of 0.5% and a position tolerance level of 1.2%.

Minimum spanning tree (MST) analysis was used to determine genetic relationships of the isolates based on MLVA data. The two main requirements for a valid MST analysis are a short time frame and a complete dataset [21]. The trees display patterns of descent from a

primary founder based on MLVA types [19]. MST analysis connects all of the MLVA types so that the genetic distances between the MLVA types are minimized. In this study, the genetic distance was defined as either the summed tandem repeat difference (STRD) or the categorical coefficient. The STRD accounts for the sum of the tandem repeat differences across all MLVA loci while the categorical coefficient describes the number of locus variants. For both analysis methods, allelic profiles that differ from the founder at only one MLVA locus were called single-locus variants (SLVs). SLVs can diversify further to produce profiles that differ at two loci (double-locus variants [DLVs]) or at three loci (triple-locus variants [TLVs]) [22]. The primary founder was established based upon priority rules that first link the highest number of SLVs followed by the DLVs and so on. As a visual tool, the STRD can be set at a maximum-neighbor distance to generate genetically related clusters [22]. STRD analysis results in a score that represents the total difference in the number of tandem repeats at all loci between two allelic profiles. For example, a DLV differing by 2 at one locus and 1 at another locus would result in an STRD value of 3. MSTs generated using the categorical coefficient result in scores depending on the number VNTR loci that differ between two profiles. SLVs result in a score of 1, while DLVs result in a 2, and TLVs result in a 3. Since this study only examines 7 VNTR loci, the maximum categorical coefficient is 7.

In this study, each circle on the MST represents a unique MLVA type. Open circles represent sporadic isolates while the color-coded circles represent outbreak-associated isolates by geographic location. Genetically related clusters were identified by setting the maximum neighbor STRD to generate colored clouds representative of outbreaks. Larger circles represent two or more isolates with the same MLVA type. Dashed circles represent sporadic isolates that clustered based upon the MLVA data.

4.0 RESULTS

4.1 VNTR LOCUS IDENTIFICATION AND CHARACTERIZATION

Twenty-one tandem repeat loci were identified using TRF software analysis that met the study criteria. These loci underwent both PCR and sequencing analysis on 7 isolates representing each of the following *N. meningitidis* serogroups - A, B, C, W-135, X, Y, Z (Table 4). Of the 21 loci identified, 13 were found to be present in at least two of the three available *N. meningitidis* genomes (Table 5). Of the 6 TR loci common to serogroups A, B and C, only AP6 generated reliable sequence data. Of the 7 TR loci common to either serogroup A and B, serogroups A and C or serogroups B and C, only AP9 and AP10 generated consistent data. In an effort to identify more VNTR for MLVA, 8 TR loci were found in each of the 3 sequenced genomes alone and were screened for variation by PCR. Interestingly, 4 of the 8 VNTR loci that were identified amplified across all 7 serogroups regardless of the fact that the TRF software had not identified these TRs in all 3 available genomes. AP16 and AP18 were identified from the serogroup B genome while VNTR06 and AP14 were identified from the serogroup A and C genomes, respectively. These data suggest that the published serogroup A and B *N. meningitidis* genomes are not accurate with regard to tandem repeat sequences (the serogroup C, FAM18 sequence had not been published and is available through the Sanger Institute website). Loci that produced a strong PCR product, and generated good quality sequence data on forward and reverse strands were deemed reliable. This criterion was used to determine which loci to keep and which to eliminate.

Gel electrophoresis was the first step in determining the reliability of a locus. Figure 1 illustrates how this method was used to show TR variation at VNTR01, AP8, AP5, and AP6. Three of the four loci represented in this figure were eliminated. TR variation can be seen for VNTR01, AP8, and AP6, however, AP6 was the only locus that was not eliminated because it

could be reliably amplified and sequenced. VNTR01 generated multiple PCR bands and unreliably sequenced despite multiple attempts using various primers. Some of the problems with AP8 included that it both unreliably PCR amplified and sequenced. Although AP5 did reliable amplify and sequence, it was invariant and was subsequently eliminated. AP1 and AP11 were also reliable but like AP5 they only produced a single amplicon with little or no polymorphism which was confirmed through sequence analysis. VNTR02 was the only locus found to be serogroup specific for A, C, and Z. Although TRF identified VNTR02 in the genomes of both serogroup A and B, this locus did not amplify any of the serogroup B isolate examined. Six loci were found to be reverse complements of each other - VNTR06/VNTR08, AP3/4, and AP10/AP15 (Table 4). In these cases, the more reliable locus was kept. Two pairs of loci -VNTR10/AP2 and AP7/AP14 - are the same loci with different primer sets.

Table 4. Characteristics of 21 identified VNTR loci.

TR ID	TR sequence	TR length (bp)	SG	# repeat units	Location
VNTR01	CAAACAA	7	B	35.9	657231-657481
VNTR02	GGGCTGTAGAGAT	13	A,B	(A) 3.4	(A) 1234098-1234140
				(B) 29.4	(B) 1131155-1131527
VNTR06/VNTR08 *	GGCA*A/G*CTTT	9	A	9.3	2158511-2158594
	VNTR08 (GCCAAAGCT)				
AP1	CCGTCCGCGTT	11	A,B,C	(A) 2.3	(A) 1042108-1042132
				(B) 2.3	(B) 895592-895616
				(C) 2.3	(C) 826510-826534
AP2/VNTR10#	TTGGG	5	A,B,C	(A) 8.8	(A) 1363389-1363432
				(B) 16.8	(B) 1273601-1273684
				(C) 15.8	(C) 1172828-1172906
AP3/AP4*	GAAGA	5	A,B,C	(A) 7.8	(A) 1600293-1600331
	AP4 (TCTTC)			(B) 13.8	(B) 455614-455667
				(C) 11.8	(C) 1404254-1404312
AP5	GGCGAAGGCAAATGC	15	A,B,C	(A) 2.1	(A) 216572-216603
				(B) 2.1	(B) 2254997-2255028
				(C) 2.1	(C) 2163591-2163622
AP6	CAAG	4	A,B,C	(A) 12.8	(A) 1638925-1638975
				(B) 11.8	(B) 1556762-1556808
				(C) 8.8	(C) 144059-1444093
AP8	AAA*C/TAA/GC/T*	7	A,C	(A) 26.7	(A) 814838-815024
				(C) 30.7	(C) 601066-601280
AP9	CATTTCT	7	A,B	(A) 17.4	(A) 920757-920878
				(B) 4.4	(B) 773265-773295
AP10/AP15*	GCTT	4	A,C	(A) 8	(A) 2123413-2123444
	AP15 (AAGC)			(C) 34	(C) 1892701-1892836
AP11	GTTTTC*A/GG/T*CT	10	A,B	(A) 3.6	(A) 1522510-1522545
				(B) 2.6	(B) 2021040-2021065
AP12	CCGTCATTCCCGCCACTTT	19	B,C	(B) 4.6	(B) 1690136-1690222
				(C) 2.6	(C) 1573750-1573798
AP13	TCAACA	6	C	8.5	1484684-1484734
AP14/AP7#	TAAA	4	C	9.3	2008241-2008277
AP16	GATTCA/G	6	B	7.2	386435-386477
AP17	TCTTCA	6	B	5.3	863508-863539
AP18	*C/T*AGC	4	B	20.5	1400278-1400359

* Loci that are reverse complements of each other
# Same locus with different primers

Table 5. Number of VNTR loci by serogroup.

Serogroups	# of loci identified
A	1
B	6
C	1
A,B	3
A,C	2
B,C	2
A,B,C	6
Total	21

Of the 21 VNTR loci screened by PCR on the 7 serogroup specific isolates, only 7 loci were found to be reliable. These VNTR loci produced a single amplicon that was polymorphic across all 7 serogroups as demonstrated by gel electrophoresis (Figure 1) and subsequent sequence analysis. Of the 7 loci chosen, both AP9 and AP10 were found to be intergenic while the other five loci were intragenic. Both AP6 and AP16 are hypothetical proteins. VNTR06 resides within the gene encoding rotamase, an enzyme that is involved in protein transport and secretion [23]. Both AP14 and AP18 reside within genes encoding a putative modification methylase that results in an authentic frameshift that is not the result of a sequencing error.

MLVA data were completed for 59 of the 61 isolates at all 7 loci. Two isolates from the sporadic collection are missing AP18 data due to a large number of repeats that could not be reliably sequenced. These 7 VNTR loci exhibited between 10 and 26 alleles among the 59 isolates completed. Across all of the loci, repeat numbers ranged between 2 and 39 with the most variable locus being AP10 (Table 6). These 7 VNTR loci were also found to be equally distributed throughout the three *N. meningitidis* genomes (Figure 2).

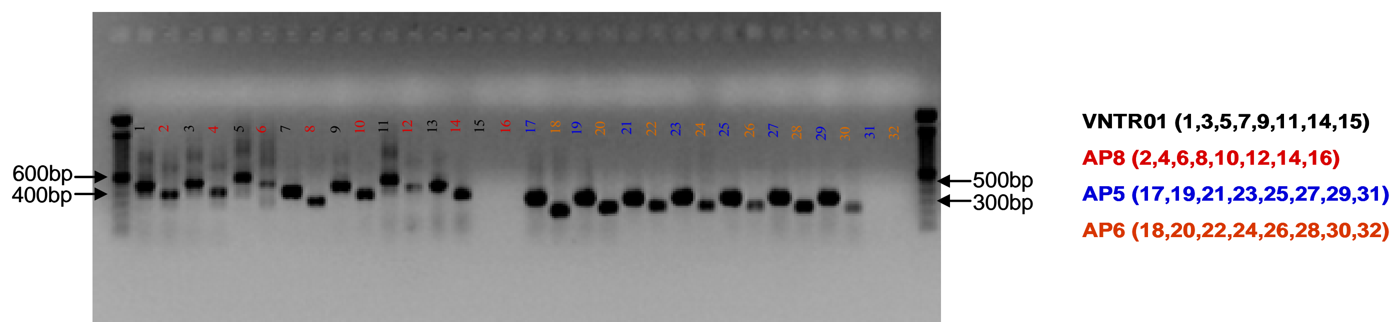

Figure 1. Gel electrophoresis of PCR products demonstrating TR variation at – VNTR01, AP8, AP5, AP6.

Table 6. Characteristics of 7 *N. meningitidis* MLVA loci.

TR ID	TR sequence*	TR length (bp)	SG	Amplicon Size (bp)	Min repeats	Max repeats	# alleles	Gene
VNTR06	GGCA*A/G*CTTT	9	A	469	5	33	19	intragenic (rotamase)
AP6	CAAG	4	A,B,C	237	3	20	14	intragenic (hypothetical protein)
AP9	CATTTCT	7	A,B	472	3	28	16	intergenic
AP10	GCTT	4	A,C	462	9	39	22	intergenic
AP14	TAAA	4	C	329	5	17	6	intragenic (modification methylase)
AP16	GATT*CA/G*	6	B	493	2	17	8	intragenic (hypothetical protein)
AP18	*C/T*AGC	4	B	497	3	28	20	intragenic (modification methylase)

* A backslash (/) represents nucleotide variation within TR sequences

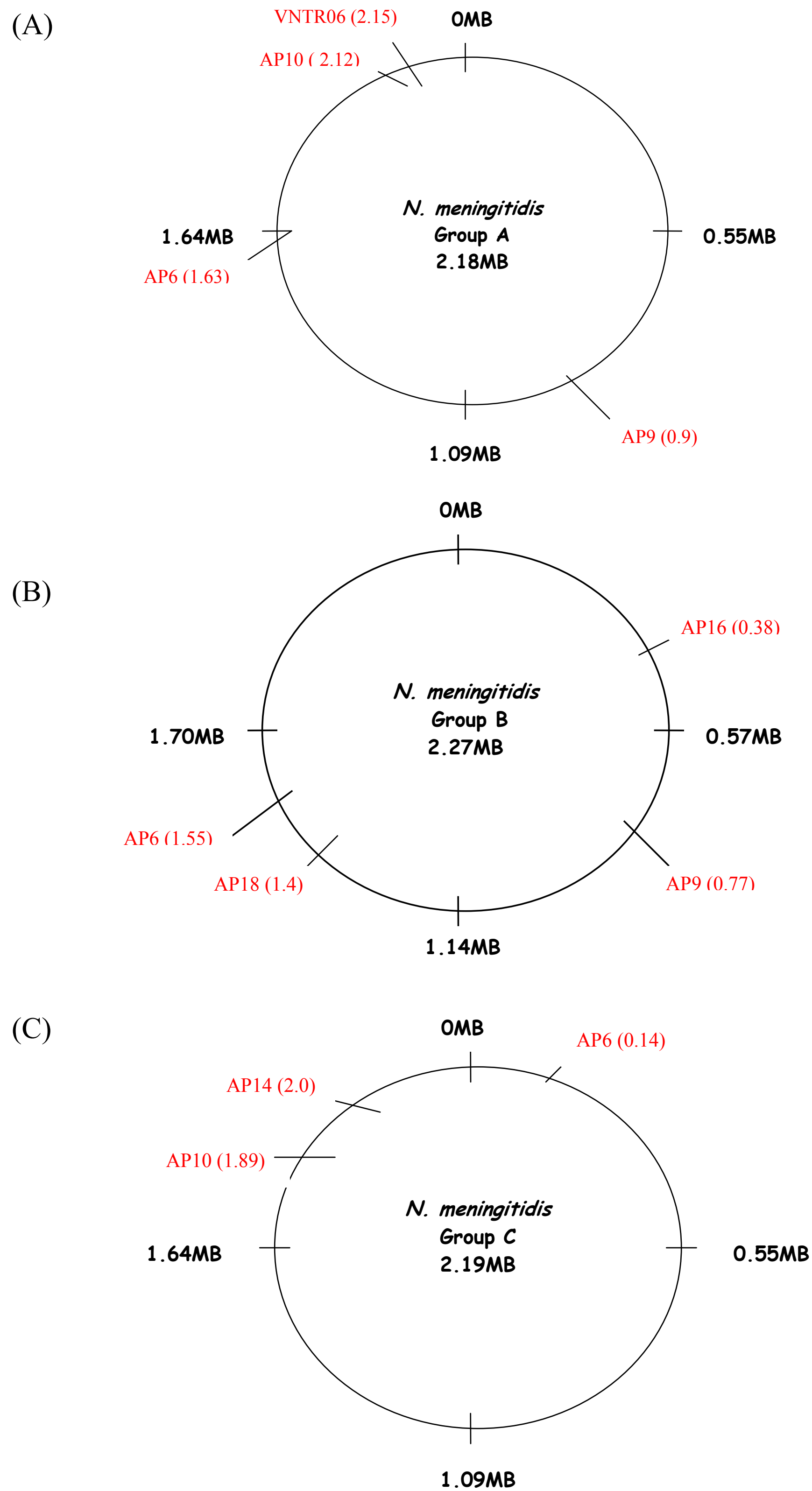

Figure 2. Distribution of loci on *N. meningitidis* genomes (A) Z2491 (B) MC58 (C) FAM18

4.2 MINIMUM SPANNING TREE (MST) ANALYSIS

Minimum spanning tree analysis was performed on MLVA data from the 46 serogroup C isolates with completed datasets to determine their genetic relationships. The MLVA data were analyzed by 2 different methods to define genetic distance: (1) using the sum of the tandem repeat differences (STRD) or (2) using a categorical coefficient (number of locus variants).

To determine the combination of VNTR loci that would best predict the genetic relationships among the 46 *N. meningitidis* strains using the STRD method, MSTs were generated with different combinations of the 7 VNTR loci and the range of the resulting STRD for the isolates within the 8 different outbreaks was determined (Table 7). This analysis illustrates that the combination of 5 VNTR loci without AP14 and AP18 generates the lowest maximum average genetic distance with a STRD ≤ 3 among the 46 isolates tested.

Table 7. STRD ranges for outbreaks using different combinations of VNTR loci.

Outbreak # (# isolates)	All 7 loci	w/o AP18	w/o AP10	w/o VNTR06	w/o AP9	w/o AP6	w/o AP14	w/o AP6 and AP14	w/o AP14 and AP18	w/o AP10 and AP14	w/o AP10 and AP18
VA 1 (4)	2-5	1-2	2-4	2-5	1-5	2-5	1-5	1-5	1	1-4	0-2
CA 2 (2)	5	2	4	4	5	4	5	4	2	4	1
AZ 3 (2)	30	20	25	19	29	27	30	27	20	25	15
NM 4 (3)	3-4	3-4	3-4	3-4	2-3	1-3	1-2	0-1	1-2	1-2	3-4
TX 5 (3)	6-8	5-6	5-8	6-8	6-8	5-6	1-6	1-3	1-3	0-4	5
MD 6 (2)	2	3	2	2	2	1	2	1	1	1	2
MD 7 (2)	3	1	1	3	3	2	2	1	2	2	1
MD 8 (3)	0-1	0-1	0	0-1	0-1	0-1	0-1	0-1	0-1	0	0
Total Range*	0-8	0-6	0-8	0-8	0-8	0-6	0-6	0-5	0-3	0-4	0-5

*Total Range = The minimum and maximum STRD range for all outbreaks except Arizona.
Red columns represent those locus combinations for which MST are shown.

When performing MST analysis using all seven loci and a STRD ≤ 8, MLVA identified 7 of 8 (88%) outbreaks (Figure 3). The only outbreak isolates that did not cluster based on the MLVA data were the 2 isolates from the Arizona outbreak 3. These isolates were found to be genetically distinct by MLVA with a STRD range between 19 and 30 for all combinations of loci examined. Therefore, the STRD ranges for outbreak 3 were omitted in the Total Range calculations for Table 7. As indicated by the dashed circles in Figure 3, there were a number of

sporadic isolates that clustered based on the MLVA results when all 7 loci were used in MST analysis. MDO1353 grouped together with the Maryland outbreak 8 while M143 clustered with the California outbreak 2 isolates. In addition, two sporadic isolates from Maryland, MDO1276 and MDO1296, also clustered together.

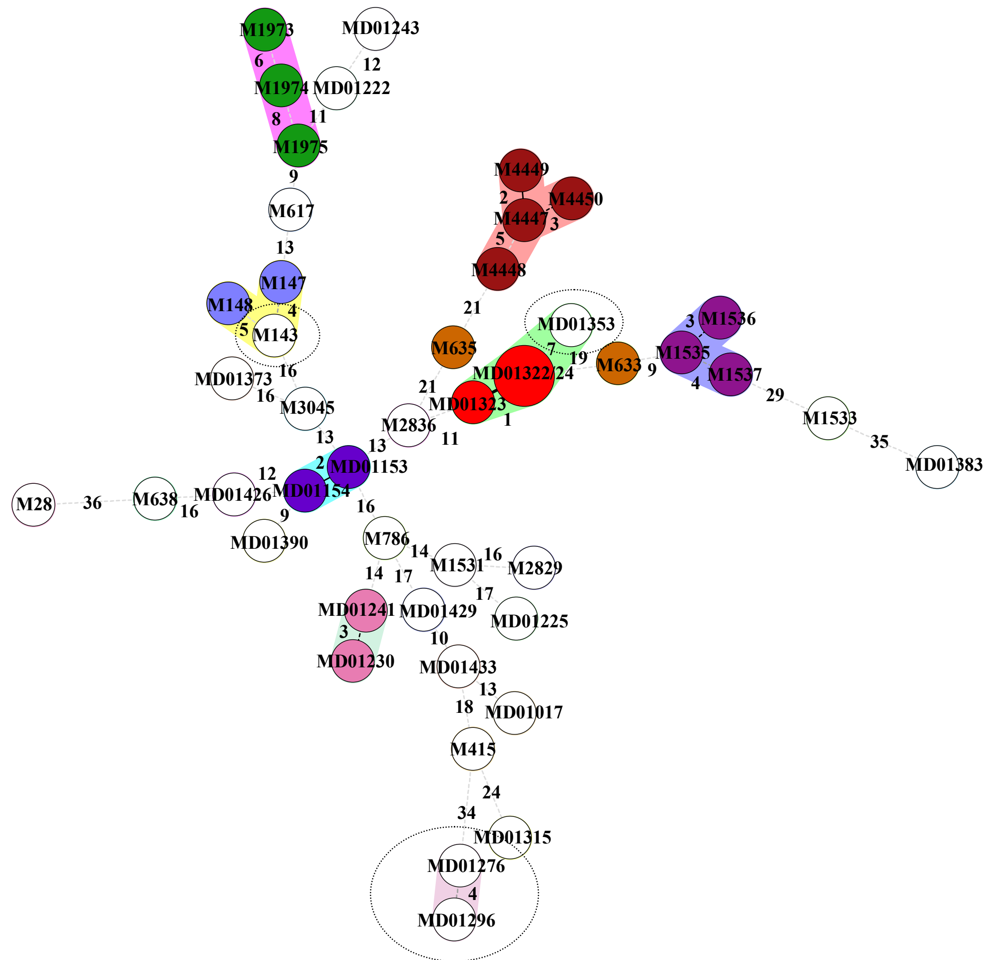

Figure 3. MST of MLVA data with all 7 VNTR loci using a STRD ≤8*.

***Each circle represents a unique MLVA type. Open circles represent sporadic isolates. Outbreak #1- Brown circles, Outbreak #2- light blue circles, Outbreak #3- Orange circles, Outbreak #4- Purple circles, Outbreak #5- Green circles, Outbreak #6- blue circles, Outbreak #7- pink circles, Outbreak #8- red circles. Strain identifiers are represented alphanumerically. Colored clouds denote genetically related clusters based on a STRD ≤8. Larger circles represent two or more isolates with the same MLVA type.**

When MST analysis was done using 6 VNTR loci (without AP18) and a STRD ≤6 was used, MLVA still identified 7 of 8 (88%) known outbreaks (Figure 4).

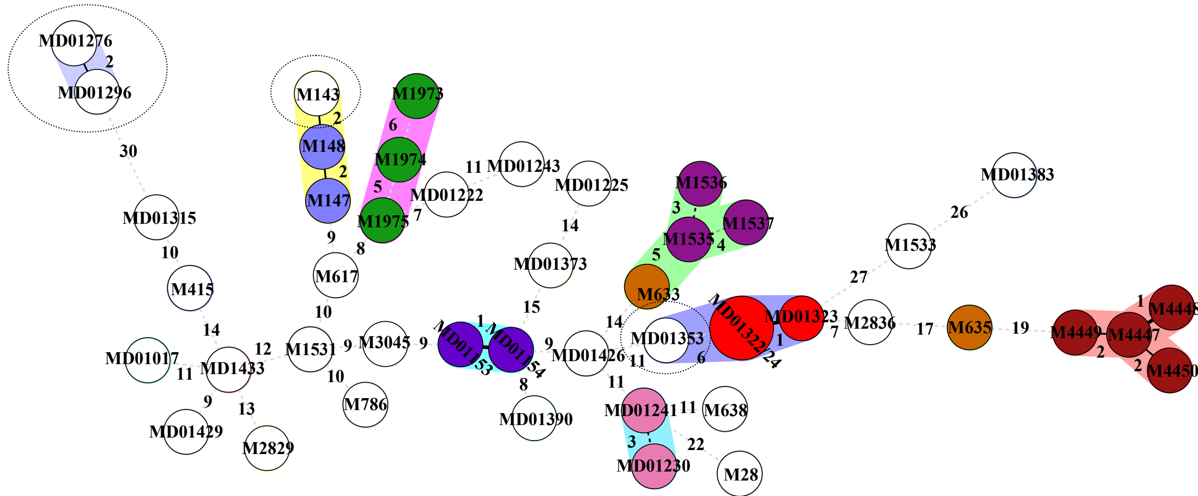

Figure 4. MST of MLVA data without AP18 using a STRD ≤6*.

***Each circle represents a unique MLVA type. Open circles represent sporadic isolates. Outbreak #1- Brown circles, Outbreak #2- light blue circles, Outbreak #3- Orange circles, Outbreak #4- Purple circles, Outbreak #5- Green circles, Outbreak #6- blue circles, Outbreak #7- pink circles, Outbreak #8- red circles. Strain identifiers are represented alphanumerically. Colored clouds denote genetically related clusters based on a STRD ≤6. Larger circles represent two or more isolates with the same MLVA type.**

Again, the only known outbreak that was not identified was Arizona outbreak 3. Using 6 VNTR in the MST analysis, many sporadic isolates clustered with known outbreak isolates. For instance, the Arizona isolate, M633, was grouped with the New Mexico outbreak isolates. There were also two sporadic isolates grouped together with the Texas outbreak and one sporadic

isolate that clustered with the Maryland outbreak 5. The two sporadic Maryland isolates also still grouped together with the removal of AP18.

When performing MST analysis using 5 VNTR loci with the 2 most variable loci removed, AP10 and AP18, a STRD ≤5 was used and MLVA still identified 88% of known outbreaks (Figure 5).

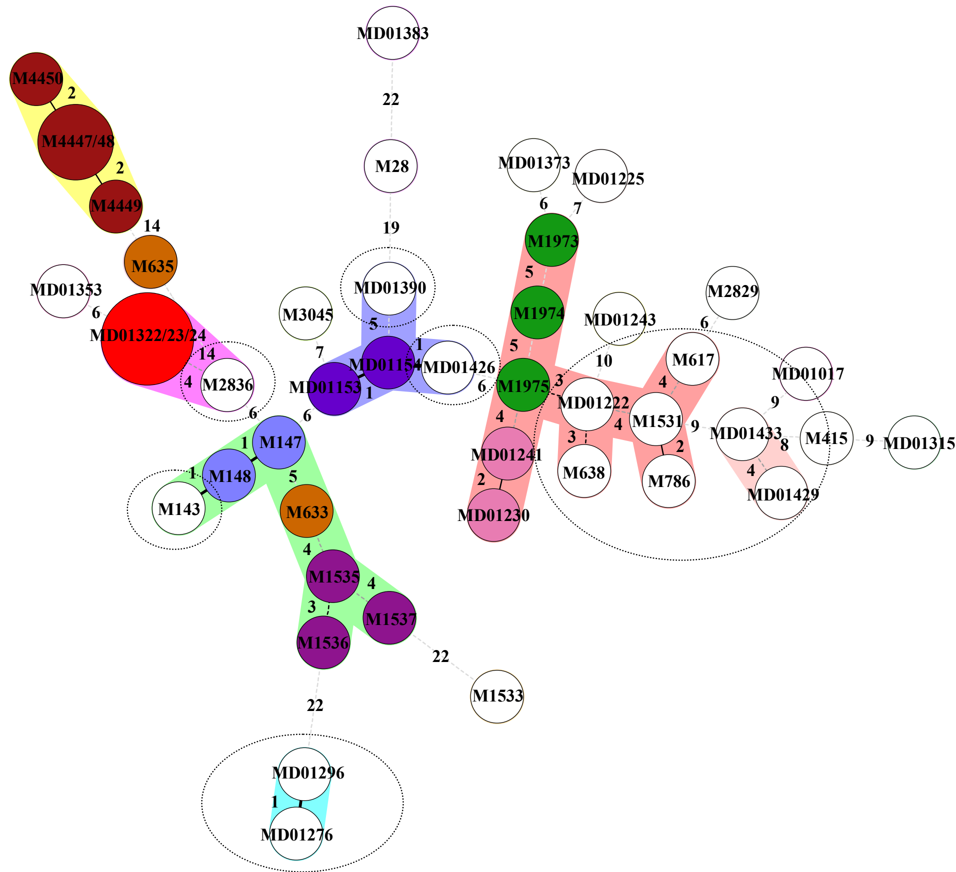

Figure 5. MST of MLVA data without AP10 and AP18 using a STRD ≤5*.

***Each circle represents a unique MLVA type. Open circles represent sporadic isolates. Outbreak #1- Brown circles, Outbreak #2- light blue circles, Outbreak #3- Orange circles, Outbreak #4- Purple circles, Outbreak #5- Green circles, Outbreak #6- blue circles, Outbreak #7- pink circles, Outbreak #8- red circles. Strain identifiers are represented alphanumerically. Colored clouds denote genetically related clusters based on a STRD ≤5. Larger circles represent two or more isolates with the same MLVA type.**

The Arizona outbreak isolates were again the only isolates that did not cluster. In this analysis a STRD $\leq$ 5 was chosen as the maximum neighbor distance for cluster formation based on the STRD ranges for these 5 loci (Table 7). However, while this smaller STRD clustered outbreak isolates, this MST was unable to discriminate sporadic isolates. Many sporadic isolates grouped together with outbreak isolates and isolates from different outbreaks were clustered together. For example, the Maryland outbreak 7 and Texas outbreak 5 isolates clustered together along with 5 sporadic isolates from various locations (Figure 5, pink shaded cluster). In this MST analysis, the New Mexico outbreak 4 isolates and the California outbreak 2 isolates clustered due to their genetic relatedness to 1 of the Arizona outbreak 3 isolates (Figure 5, green shaded cluster).

When MST analysis of the MLVA data was done without AP14 or AP18 a STRD cutoff of $\leq$3 was used and MLVA still identified 88% of known outbreaks and discriminated sporadic isolates (Figure 6). The Arizona outbreak isolates were the only isolates that did not cluster in this analysis. Similar to when all VNTR loci were utilized in the MST analysis (Figure 3), the sporadic Maryland isolates (MD1276 and MD1296) grouped together. However, in this analysis MLVA did not group any other sporadic isolates together or with any known outbreak isolates. Thus, the MST analysis of the MLVA data with different VNTR loci indicates that the best combination for outbreak detection on this collection of *N. meningitidis* isolates is the 5 locus combination of VNTR06, AP6, AP9, AP10 and AP16.

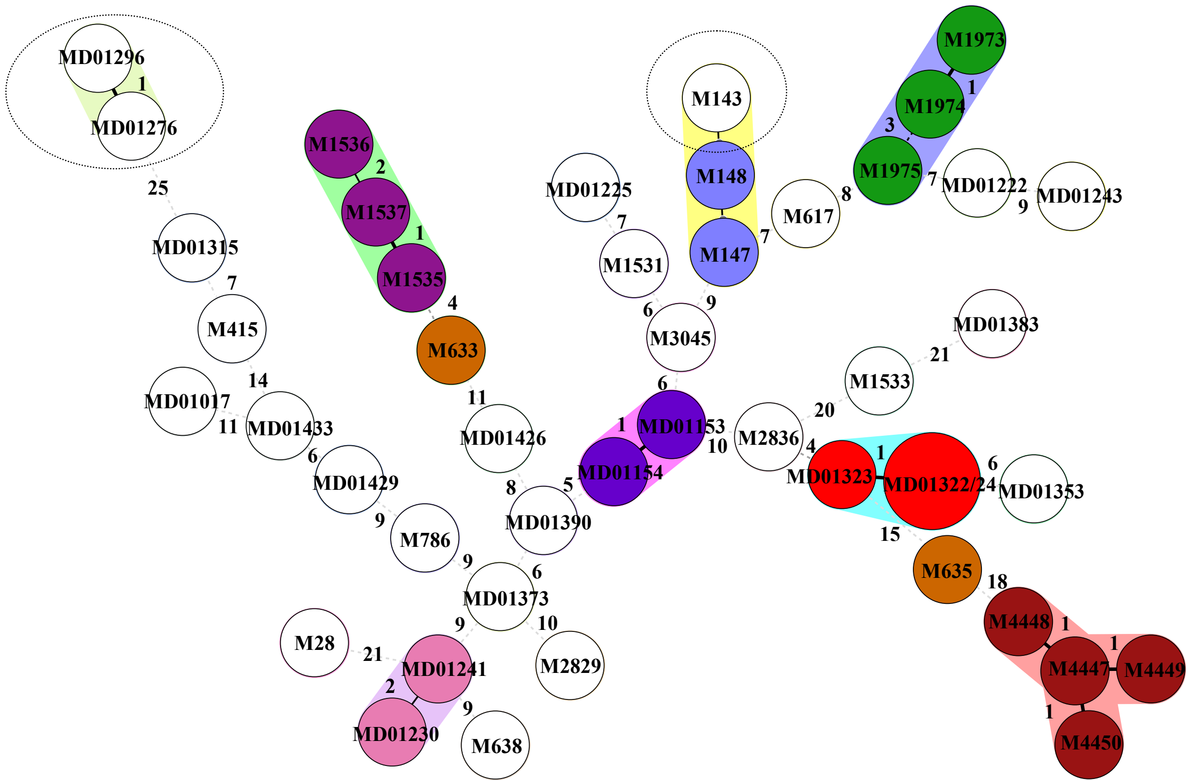

Figure 6. MST of MLVA data without AP14 and AP18 using a STRD ≤3*.

***Each circle represents a unique MLVA type. Open circles represent sporadic isolates. Outbreak #1- Brown circles, Outbreak #2- light blue circles, Outbreak #3- Orange circles, Outbreak #4- Purple circles, Outbreak #5- Green circles, Outbreak #6- blue circles, Outbreak #7- pink circles, Outbreak #8- red circles. Strain identifiers are represented alphanumerically. Colored clouds denote genetically related clusters based on a STRD ≤3. Larger circles represent two or more isolates with the same MLVA type.**

MST analysis was also done using a categorical coefficient. Table 8 describes the different categorical coefficient ranges for outbreak isolates using different combinations of loci. As previously mentioned, the categorical coefficient measures the number of locus variants between two MLVA types. Since 7 VNTR loci were identified in this study, the maximum categorical coefficient is 7. In Table 8, a maximum categorical coefficient of 4 or 5 loci was observed for the Arizona outbreak 3 isolates regardless of the combination of loci examined. Similar to the analysis of STRD ranges, the Arizona isolates were omitted from the total range calculations in Table 8. In this analysis the maximum categorical coefficient was 3 (a triple locus variant) while the smallest maximum categorical coefficient of 2 could be obtained with several different combinations of VNTR loci.

Table 8. Categorical coefficient ranges for outbreaks using different combinations of VNTR loci.

Outbreak # (isolate #)	All loci	w/o AP18	w/o AP10	w/o VNTR06	w/o AP9	w/o AP6	w/o AP14	w/o AP6 and AP14	w/o AP14 and AP18	w/o AP10 and AP14	w/o AP10 and AP18
Va (4)	2-3	1-2	1-3	2-3	1-2	2	1-2	1-2	1	1-2	0-2
Ca (2)	3	2	2	3	3	2	3	2	2	2	1
Az (2)	5	4	4	4	4	4	5	4	4	4	3
NM (3)	2	2	2	2	1	1	1	0-1	1	1	2
Tx (3)	2-3	2	1-3	2-3	2-3	2	1-2	1	1	0-2	1-2
Md (2)	2	1	2	2	2	1	1	1	1	2	2
Md (2)	3	3	2	3	3	2	2	1	2	1	2
Md (3)	1	1	0	0-1	1	1	0-1	0-1	0-1	0	0
Total Range*	1-3	1-3	0-3	0-3	1-3	1-2	0-3	0-2	0-2	0-2	0-2

*Total Range = The minimum and maximum STRD range for all outbreaks except Arizona.
Red columns represent those locus combinations for which MST are shown.

Figure 7 shows a minimum spanning tree using 6 VNTR loci (without AP14) and cluster formation with a categorical coefficient ≤1. Based on this analysis only 2 of the 8 outbreaks (25%) were correctly identified - the Maryland outbreak 8 and the New Mexico outbreak 4 (Figure 7, yellow and pink shaded clusters, respectively). If the categorical coefficient for cluster formation in this analysis was increased to ≤2, 6 of the 8 outbreaks would be detected

(Figure 7). However, this alteration decreases the clustering tolerance and results in inappropriate cluster formation of sporadic isolates.

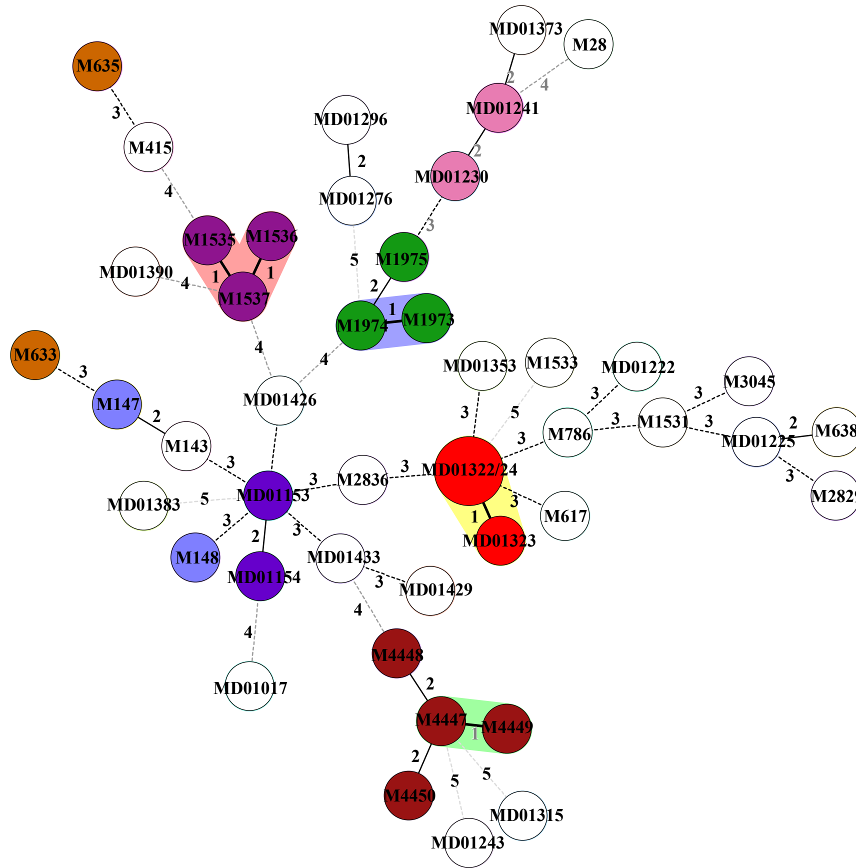

Figure 7. MST of MLVA data without AP14 using a categorical coefficient ≤1*.

***Each circle represents a unique MLVA type. Open circles represent sporadic isolates. Outbreak #1- Brown circles, Outbreak #2- light blue circles, Outbreak #3- Orange circles, Outbreak #4- Purple circles, Outbreak #5- Green circles, Outbreak #6- blue circles, Outbreak #7- pink circles, Outbreak #8- red circles. Strain identifiers are represented alphanumerically. Colored clouds denote genetically related clusters based on a categorical coefficient ≤1. Larger circles represent two or more isolates with the same MLVA type.**

When MST analysis of the MLVA data was performed using the 5 VNTR loci identified as the best by the STRD method (without AP14 and AP18) and a categorical coefficient ≤1 was used for cluster formation, 5 of the 8 of outbreaks (63%) were identified correctly (Figure 8). The 3 outbreaks that did not group were the Arizona outbreak 3, the California outbreak 2, and the Maryland outbreak 7. The Maryland sporadic isolates still clustered together as previously described. Other sporadic isolates also grouped together based on these criteria - MD01373

clustered with Maryland outbreak 7 while both MD01225 and M638 clustered together. If the categorical coefficient for clustering was reduced to ≤ 2 in this analysis, many incorrect clusters would form between sporadic isolates and among outbreaks (Figure 8). Based on these studies, MST analysis using 5 VNTR with STRD ≤ 3 provides sufficient discrimination of outbreaks and sporadic isolates.

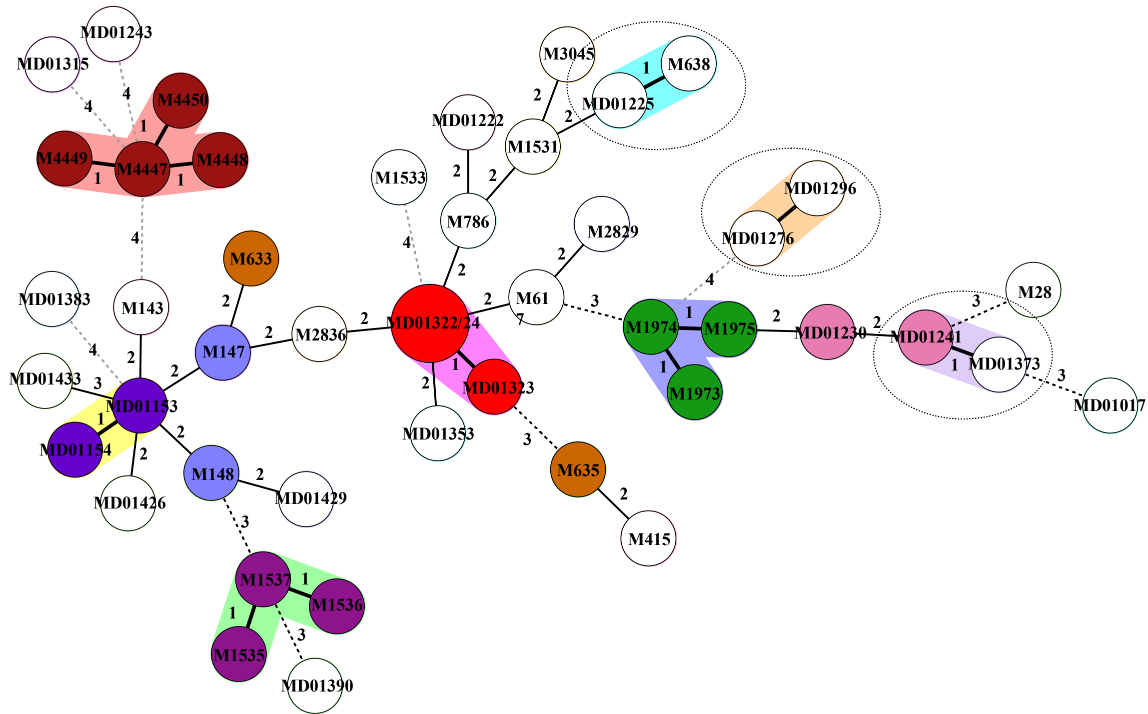

Figure 8. MST of MLVA data without AP14 and AP18 using a categorical coefficient ≤1*.

***Each circle represents a unique MLVA type. Open circles represent sporadic isolates. Outbreak #1- Brown circles, Outbreak #2- light blue circles, Outbreak #3- Orange circles, Outbreak #4- Purple circles, Outbreak #5- Green circles, Outbreak #6- blue circles, Outbreak #7- pink circles, Outbreak #8- red circles. Strain identifiers are represented alphanumerically. Colored clouds denote genetically related clusters based on a categorical coefficient ≤1. Larger circles represent two or more isolates with the same MLVA type.**

4.3 COMPARISON OF MLVA WITH PFGE

PFGE was performed on the 59 *N. meningitidis* isolates for which MLVA data was generated. Dendrograms of the pulsed field profiles from the 46 serogroup C isolates were compared with the MLVA data from the 5 VNTR loci - VNTR06, AP6, AP9, AP10 and AP16 in order to determine if MLVA was as discriminatory as PFGE (Figure 9). All 8 outbreaks were either indistinguishable or highly related based on the PFGE results. Four of the outbreaks had sporadic isolates that were highly related by PFGE. The California outbreak 2 isolates clustered with several New Mexico isolates and 1 Arizona isolate. The MLVA data for these isolates clearly demonstrated that these 3 isolates are unrelated to the California isolates. Similarly, the Maryland outbreak 6 isolates were highly related to a sporadic Maryland isolate which MLVA distinguished as a triple locus variant with a STRD of 3. Maryland outbreak 8 was also highly related to a single sporadic Maryland isolate and MLVA distinguished this isolate as a double locus variant with a STRD equal to 6. The Texas outbreak 5 isolates clustered with 2 sporadic TX cases which were also discriminated by MLVA as TLVs with a STRD of 12 or 4 locus variants with a STRD of 18. This data demonstrates that MLVA is more discriminatory in some cases than PFGE.

The only outbreak that was not identified by MLVA was the Arizona outbreak 3. This data suggest that MLVA may be too discriminatory for some *N. meningitidis* strains. A comparison between the MLVA and PFGE data for the Maryland outbreak 8 and the Arizona outbreak 3 isolates illustrates the similarities and differences between the data for these two molecular subtyping methods (Figure 10). Both the MLVA and PFGE results for Maryland outbreak 8 show that these isolates are either highly related or identical. The MLVA data is identical for all three of these isolates except at AP10 where NM00217 differs by only one tandem repeat (Figure 10A). The PFGE banding pattern for these same isolates is also identical (Figure 10B). Although the Arizona outbreak 3 isolates were selected based on the PFGE data that showed that these isolates were highly related (Figure 10C), the MLVA results show that these isolates are unrelated (Figure 10A). Using the best combination of loci, these isolates differ at four loci (VNTR06, AP6, AP9, and AP10) by a STRD of 20 (Figure 10A). However, the maximum neighbor distance identified in the other 7 outbreaks was a SLV of 3 for Texas isolate M1975 (Figures 6 and 9). Furthermore, epidemiologic information available on the

Arizona outbreak isolates revealed that these isolates were separated in time by 44 days which may explain the differences observed by MLVA. Further investigation of these and other available Arizona outbreak isolates is necessary in order to better understand and explain the differences between the MLVA and PFGE results.

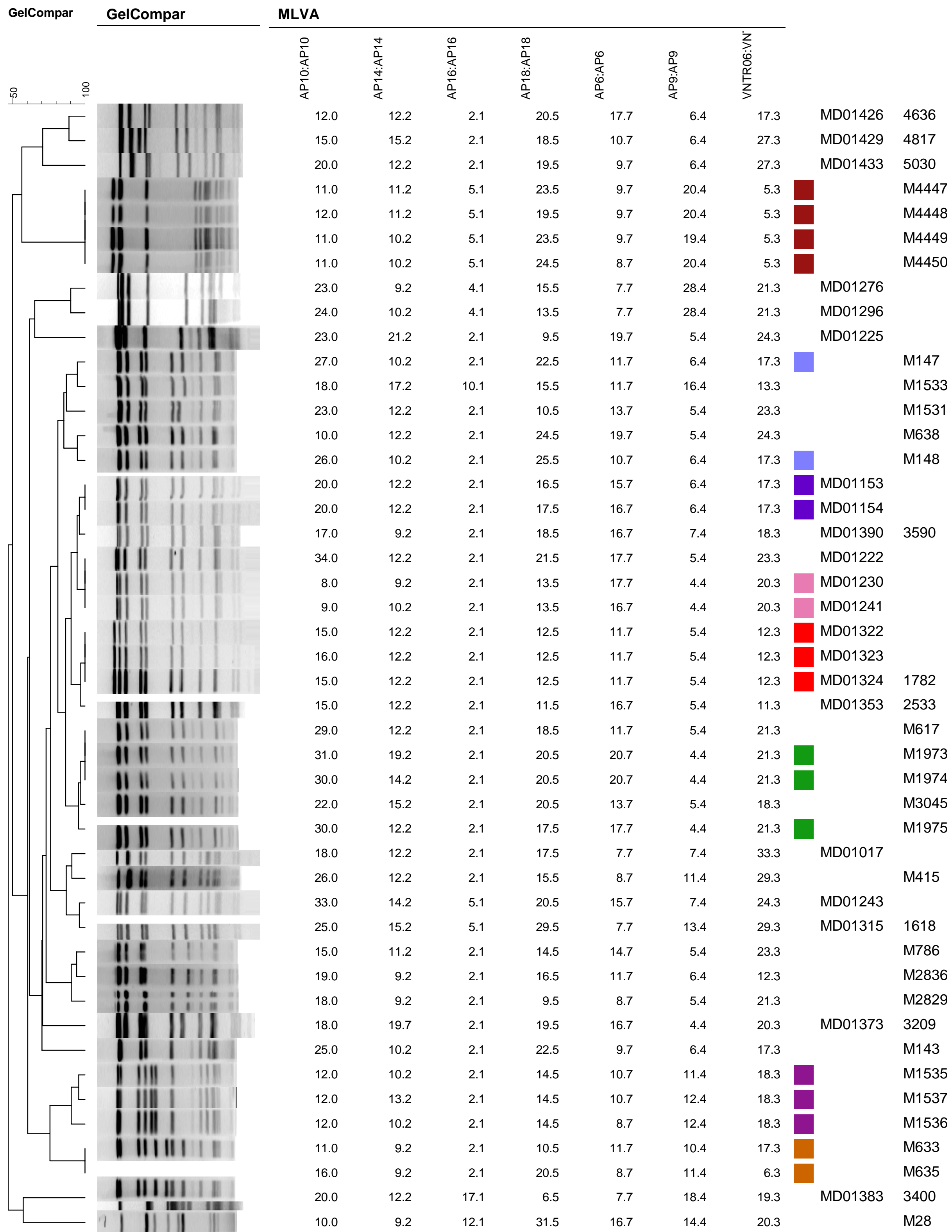

Figure 9. Dendrogram based on PFGE data with MLVA results included.

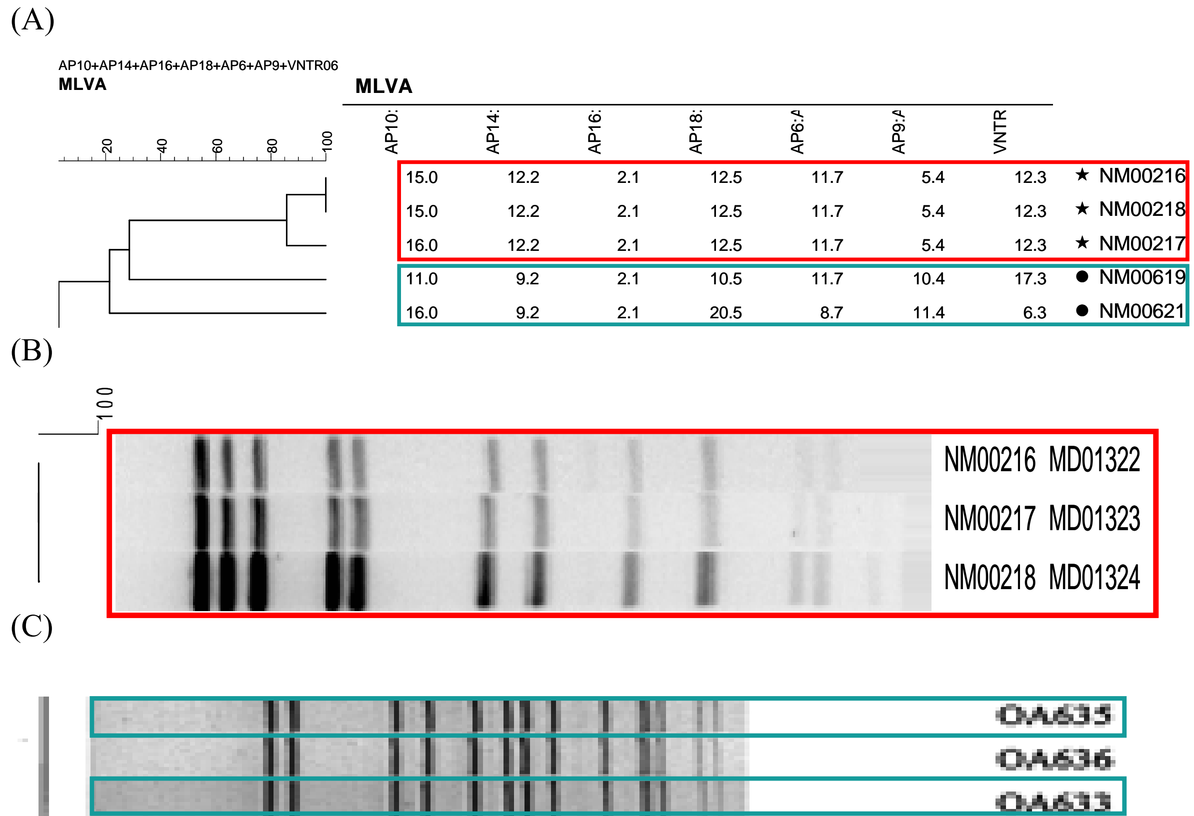

Figure 10. Comparing MLVA and PFGE data from two outbreaks (Maryland and Arizona). (A) MLVA types for Md and Arizona outbreak isolates (B) PFGE results for Md outbreak isolates (C) PFGE results for Arizona outbreak isolates.

5.0 DISCUSSION

The MLVA assay described in this study distinguished *N. meningitidis* isolates from 7 different serogroups and sporadic isolates within each serogroup using the 7 VNTR loci identified. MLVA also identified 88% of PFGE-defined serogroup C outbreaks. Multiple minimum spanning trees were created in order to determine which combination of loci gave the best results based upon PFGE and the known epidemiologic information. Minimum spanning trees were created using both a STRD and a categorical coefficient. The use of STRDs was the primary analysis method for the creation of MST because previous studies in *E. coli* O157:H7 have demonstrated that when mutations occur in these repeat regions, they tend to result in the addition or removal of 1 repeat at a time [24]. Using this type of analysis, no matter what combinations of loci were used, 88% of the known outbreak-associated isolates were identified by MLVA. The main differences between each MST were the STRD cutoffs which were determined by calculating the mean STRD range for each MST (Table 7). When the MLVA data was analyzed by MST using the categorical coefficient, discrimination of outbreaks and sporadic isolates was lost. The categorical coefficient is often used for analysis of MLST data [19]. This method of generating MSTs permits analysis of mutations that arise due to recombination. The mechanism of TR mutation in *N. meningitidis* is not known. In order to settle the debate of STRD versus categorical analysis for *N. meningitidis*, mutational studies are required to determine how repeats are added or removed.

The only outbreak isolates that were not identified by MLVA no matter what locus combination or STRD was used were those from Arizona. These isolates were found to be indistinguishable by PFGE but substantially different by MLVA (Figure 10). Although the exact reason for such a significant difference between the PFGE and MLVA results for these isolates is unknown, *N. meningitidis* is less clonal than other organisms such as *E. coli* O157:H7, which could create the potential for multiple MLVA types to emerge during the course of an outbreak

and the Arizona outbreak occurred over a 25-week time period with these two isolates being cultured 44 days apart [25, 26]. VNTRs are elements that can mutate through slipped-strand mispairing during DNA replication which can influence transcription or translation [27, 28]. The influence of TR on the transcription of promoter or coding regions has also been associated with some human genetic disorders like epilepsy [29]. Some of the factors that influence the frequency and type of TR mutations are the number of repeats as well as the repeat size. As the number of tandem repeats increases, so does the slippage mutation rate because of polymerase instability. Repeats that are shorter in length also have a higher mutational rate [30].

There are two proposed models that explain the rate of mutational changes in bacteria: the stepwise mutation model and the two-phase mutation model. The stepwise mutation model proposes that new alleles are created through the gain or loss of a single repeat, while the two-phase model proposes that a majority of mutations that occur are single repeat changes but there are also a small portion of these mutations that can involve large changes [31, 32].

Some differences seen when using different locus combinations and different STRDs were the ability of different combinations to discriminate sporadic isolates. For instance, when the two most variable loci, AP10 and AP18, were removed, the STRD cutoff was dropped to 5 and all of the known outbreaks were identified except for Arizona. However, using this combination, MLVA was unable to discriminate sporadic isolates which were clustering together. In addition, some of the different outbreak isolates clustered (Figure 5). The best VNTR locus combination was the 5 loci: VNTR06, AP6 AP9, AP10 and AP16. Using this locus combination, with the maximum STRD for cluster formation was ≤ 3, all of the known outbreak isolates clustered except the Arizona outbreak. This VNTR combination also successfully discriminated most of the defined sporadic isolates (Figure 6).

MST analysis always clustered two sporadic Maryland isolates (MD01276, MD01296) based on the MLVA results. These two Maryland isolates were both ST-11 complex, serogroup C isolates from the same city which were highly related by PFGE (Figure 9). These isolates were also cultured nine months apart and were TLV differing by 4 tandem repeats when all 7 VNTR loci were used. Some of the other isolates that were clustering in with a few outbreaks were M143, MD01353, and M633. M143 is a California isolate that grouped with the outbreak-associated isolates (M147, M148) from this region by MLVA but was not considered to be part of the outbreak. M148 was isolated from an inmate in a California jail while M147 was isolated

from a member of a nearby community who had contact with inmates in that jail. Although M143 was collected from a community member who had no contact with the jail, they could have had contact with other infected community members. MDO1353 was also a Maryland isolate that clustered with the Maryland outbreak 8 by MLVA. This isolate was cultured nearly two years after the initial Maryland outbreak 8 began. Previous studies have demonstrated that these isolates belong to the ST-11 complex and have identical *por*A and *fet*A genotypes but were different at *por*B [33]. These isolates had several band differences by PFGE which may be explained by mutation of a single restriction enzyme site. M633, one of the Arizona outbreak isolates, groups closely with the New Mexico outbreak consistently. These outbreaks occurred at least a year apart and although they seem similar by MLVA, there PFGE results were different by four bands. The fact that some of the sporadic isolates are related to some of the outbreak isolates is likely a result of the circulation of outbreak strains in the population, subsequently causing disease in a person with no known epidemiologic link to the persons in the outbreak.

MST analysis using a categorical coefficient proved to be less effective than analysis using a STRD. The categorical coefficient was set to ≤1 for every tree because if the coefficient was set any higher, this method would be unable to distinguish sporadic from outbreak-associated isolates. None of the MST created using a categorical coefficient ≤1 identified all of the outbreak-associated isolates. The most successful combination of loci using this method was with the 5 VNTR loci without AP14 and AP18. Using this combination, 63% of outbreak-associated isolates were identified. In many cases, one or more of the outbreak isolates fell out of their grouping but this method did discriminate sporadic isolates.

A considerable amount of research is necessary to determine the utility of MLVA for *N. meningitidis* outbreak detection. This includes further investigation of these and additional outbreak-associated isolates from across and within different serogroups. In addition, mutational studies need to be performed to determine how tandem repeats are added and removed. MLVA, if validated, has the potential to provide a higher throughput and automatable subtyping method with more discriminatory power than current subtyping methods for the molecular surveillance and outbreak detection of *N. meningitidis*.

.

BIBLIOGRAPHY

1. Nester, E.W., Anderson, D.G., Roberts, C.E., Pearsall, N.N, and Martha T. Nester, *Microbiology: A Human Perspective*. 3 ed. 2001, New York, NY: McGraw-Hill. 291,668-671.
2. Caugant, D.A., E. A. Hoiby, P. Magnus, O. Scheel, T. Hoel, G. Bjune, E. Wedege, J. Eng, and L. O. Froholm, *Asymptomatic carriage of Neisseria meningitidis in a randomly sampled population.* Clinical Microbiology, 1994. **32**: p. 323-330.
3. Harrison, L.H., and C. V. Broome., *The epidemiology of meningococcal meningitis in the U.S. civilian population*, in *Evolution of meningococcal disease*, N.A. Vedros, Editor. 1987, CRC Press: Boca Raton, Fla. p. 27-45.
4. Apicella, M., *Mandell, Douglas, and Bennett's principles and practice of infectious diseases*, in *Neisseria meningitidis*, R.G.D. G. L. Mandell, J. E. Bennett, R. Dolin, and G. V. Kelvin, Editor. 2005, Elsevier/Churchill Livingstone: New York, N.Y. p. 2498-2513.
5. Harrison, L.H., *Prospects for Vaccine Prevention of Meningococcal Infection.* Clinical Microbiology Review, 2006. **19**: p. 142-164.
6. Dolan-Livengood, J.M., Y. K. Miller, L. E. Martin, R. Urwin, and D. S. Stephens, *Genetic basis for nongroupable Neisseria meningitidis.* Infectious Diseases, 2003. **187**: p. 1616-1628.
7. Baker, M.G., D. R. Martin, C. E. Kieft, and D. Lennon., *A 10-year serogroup B meningococcal disease epidemic in New Zealand: descriptive epidemiology, 1991–2000.* Paediatric Child Health, 2001. **37**: p. S13-S19.
8. Caugant, D.A., L. O. Froholm, K. Bovre, E. Holten, C. E. Frasch, L. F. Mocca, W. D. Zollinger, and R. K. Selander., *Intercontinental spread of a genetically distinctive complex of clones of Neisseria meningitidis causing epidemic disease.* PNAS, 1986. **83**: p. 4927–4931.
9. Decosas, J., and J. B. Koama, *Chronicle of an outbreak foretold: meningococcal meningitis W135 in Burkina Faso.* Lancet Infect. Dis., 2002. **2**: p. 763-765.
10. Sacchi, C.T., L. L. Pessoa, S. R. Ramos, L. G. Milagres, M. C. Camargo, N. T. Hidalgo, C. E. Melles, D. A. Caugant, and C. E. Frasch, *Ongoing group B Neisseria meningitidis epidemic in Sao Paulo, Brazil, due to increased prevalence of a single clone of the ET-5 complex.* Clinical Microbiology, 1992. **30**: p. 1734-1738.
11. Achtman, M., and Giovanna Morelli, *Pulsed-Field Gel Electrophoresis*, in *Meningococcal Disease: Methods and Protocols*, A.J. Pollard, and M. C. J. Maiden, Editor, Humana Press: Totowa. p. 147-155.
12. Martin C. J. Maiden, J.A.B., Edward Feil, Giovanna Morelli, Joanne E. Russell, Rachel Urwin, Qing Zhang, Jiaji Zhou, Kerstin Zurth, Dominique A. Caugant, Ian M. Feavers, Mark Achtman, and Brian G. Spratt, *Multilocus sequence typing: A portable approach to*

the identification of clones within populations of pathogenic microorganisms. PNAS, 1998. **95**: p. 3140-3145.

13. Noller, A.C., McEllistrem, M. Catherine, Pacheco, Antonio G. F., Boxrud, David J., Harrison, Lee H., *Multilocus Variable-Number Tandem Repeat Analysis Distinguishes Outbreak and Sporadic Escherichia coli O157:H7 Isolates.* Clinical Microbiology, 2003. **41**: p. 5389-5397.
14. Urwin, R., and Martin C.J. Maiden, *Multi-locus sequence typing: a tool for global epidemiology.* 2003. **11**: p. 479-87.
15. Keim, P., Price, L. B., Klevytska, A. M., Smith, K. L., Schupp, J. M., Okinaka, R., Jackson, P. J., Hugh-Jones, M. E., *Multiple-Locus Variable-Number Tandem Repeat Analysis Reveals Genetic Relationships within Bacillus anthracis.* Bacteriology, 2000. **182**: p. 2928-2936.
16. Popovic, T., Schmink, Susanna, Rosenstein, Nancy A., Ajello, Gloria W., Reeves, Michael W., Plikaytis, Brian, Hunter, Susan B., Ribot, Efrain M., Boxrud, David, Tondella, Maria L., Kim, Chung, Noble, Corie, Mothershed, Elizabeth, Besser, John, Perkins, Bradley A., *Evaluation of Pulsed-Field Gel Electrophoresis in Epidemiological Investigations of Meningococcal Disease Outbreaks Caused by Neisseria meningitidis Serogroup C.* Clinical Microbiology, 2001. **39**: p. 75-85.
17. Yazdankhah, S.P., Lindstedt, Bjorn-Arne, Caugant, Dominique A., *Use of Variable-Number Tandem Repeats To Examine Genetic Diversity of Neisseria meningitidis.* Clinical Microbiology, 2005. **43**: p. 1699-1705.
18. Benson, G., *Tandem repeats finder: a program to analyze DNA sequences.* Nucleic Acids Research, 1999. **27**: p. 573-580.
19. Feil, E.J., Li, Bao C., Aanensen, David M., Hanage, William P., Spratt, Brian G., *eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data.* Bacteriology, 2004. **186**: p. 1518-1530.
20. Bannerman, T., Hancock, GA, Tenover, FC, Miller, JM, *Pulsed-field gel electrophoresis as a replacement for bacteriophage typing of Staphylococcus aureus.* Clinical Microbiology, 1995. **33**: p. 551-555.
21. Maths, A., *Minimum Spanning Trees for population modelling*, in *Bionumerics manual*: Austin, Tx.
22. Schouls, L.M., van der Heide, Han G. J., Vauterin, Luc, Vauterin, Paul, Mooi, Frits R., *Multiple-Locus Variable-Number Tandem Repeat Analysis of Dutch Bordetella pertussis Strains Reveals Rapid Genetic Changes with Clonal Expansion during the Late 1990s.* Bacteriology, 2004. **186**: p. 5496-5505.
23. BA Sampson, a.E.G., *Neisseria meningitidis Encodes an FK506-Inhibitable Rotamase.* PNAS, 1992. **89**: p. 1164-1168.
24. Noller, A.C., McEllistrem, M. Catherine, Shutt, Kathleen A., Harrison, Lee H., *Locus-Specific Mutational Events in a Multilocus Variable-Number Tandem Repeat Analysis of Escherichia coli O157:H7.* Clinical Microbiology, 2006. **44**: p. 374-377.
25. Jackson, L.A., Schuchat, Anne, Reeves, Michael W., Wenger, Jay D., *Serogroup C Meningococcal Outbreaks in the United States: An Emerging Threat.* Journal of the American Medical Association, 1995. **273**(5): p. 383-389.
26. (CDC), N.R., *Evaluation of Meningococcal Disease Outbreaks in Arizona Caused by Neisseria meningitidis Serogroup C*. 2006: Atlanta.

27. Viguera E, C.D., and SD Ehrlich, *Replication slippage involves DNA polymerase pausing and dissociation.* EMBO, 2001. **20**(10): p. 2587-2595.
28. van Belkum, A., Scherer, Stewart, van Alphen, Loek, Verbrugh, Henri, *Short-Sequence DNA Repeats in Prokaryotic Genomes.* Microbiol. Mol, Biol. Rev., 1998. **62**: p. 275-293.
29. Richards RI, a.G.S., *Dynamic mutation: possible mechanisms and significance in human disease.* Trends in Biochemical Science, 1997. **22**(11): p. 432-436.
30. Lai Y, a.F.S., *The relationship between microsatellite slippage mutation rate and the number of repeat units.* Mol Biol. Evol, 2003. **20**: p. 2123-2131.
31. Kimura M, a.T.O., *Stepwise mutation model and distribution of allelic frequencies in a finite population.* PNAS, 1978. **75**: p. 2868-2872.
32. Cozzolino S, C.D., Pellegrino G, Musacchio A, and A Widmer, *Molecular evolution of a plastid tandem repeat locus in an orchid lineage.* Molecular Evolution, 2003. **57**: p. S41-S49.
33. Harrison, L.H., Jolley, Keith A., Shutt, Kathleen A., Marsh, Jane W., O'Leary, Mary, Sanza, Laurie Thomson, and and M.C.J. Maiden, *Antigenic Shift and Increased Incidence of Meningococcal Disease.* Infectious Diseases, 2006. **In Press**: p. In Press.