2013

Development of a 16S rRNA gene sequence as a biomarker specific to the gastrointestinal tract and feces of swine based on suppressive subtractive hybridization

Autumn Sayre
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Sayre, Autumn, "Development of a 16S rRNA gene sequence as a biomarker specific to the gastrointestinal tract and feces of swine based on suppressive subtractive hybridization" (2013). *Graduate Theses, Dissertations, and Problem Reports*. 360.
<https://researchrepository.wvu.edu/etd/360>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**Development of a 16S rRNA gene sequence as a biomarker specific
to the gastrointestinal tract and feces of swine based on suppressive
subtractive hybridization**

Autumn Sayre

**Thesis submitted to the
Davis College of Agriculture, Forestry, and Consumer Sciences
at West Virginia University**

in partial fulfillment of the requirements for the degree of

**Master of Science in
Applied and Environmental Microbiology**

**Alan J. Sexstone, Ph.D., Chair
Jennifer L. Weidhaas, Ph.D.
Gary K. Bissonnette, Ph.D.**

Division of Plant and Soil Sciences

**Morgantown, West Virginia
2013**

**Keywords: swine feces, microbial source tracking, biomarker, suppressive subtractive
hybridization, quantitative polymerase chain reaction**

ABSTRACT

Development of a 16S rRNA gene sequence as a biomarker specific to the gastrointestinal tract and feces of swine based on suppressive subtractive hybridization

Autumn Sayre

Fecal contamination of water bodies causes environmental issues and human health risks. Swine are a major contributor of fecal inputs to environmental waters due to their importance in American agriculture. Water quality currently is regulated using fecal coliforms, which have natural reservoirs in the environment and may not correlate with pathogens. In recent years there has been a shift towards microbial source tracking (MST). MST employs microbiological, genotypic, phenotypic and chemical methods to determine the source of fecal pollution in water bodies. Previous studies aimed at developing swine fecal biomarkers have had limited success. The overall objective of this research was to identify a microbial marker for swine fecal contamination. Swine (n=8) and non-target fecal samples (n=23) were collected in Morgantown, WV and the surrounding area. Additional soil (n=2) and lagoon (n=3) samples impacted with swine manure and soil (n=3) samples from areas not expected to be impacted by swine manure were collected. DNA extraction was performed followed by 16S rRNA amplification. The amplified DNA was taken into suppressive subtractive hybridization (SSH) to enrich for swine specific sequences. The final products obtained in SSH were cloned and sequenced. Swine-specific primers were designed and tested in PCR and qPCR assays against swine and non-target DNA. Sequence A5, closely related to Prevotellaceae, was detected in all swine samples tested, as well as the environmental samples impacted by swine manure. The relationship between closely related 16S rRNA sequences of Prevotellaceae and the swine marker were compared in a phylogenetic tree. The SYBR green qPCR assay did show amplification of sequences in cattle, human, and two non-target soil samples; however, apparent positive detection of the swine marker was rejected based on dissociation profiles. Therefore, the SYBR green qPCR assay was determined to be specific to detect the swine marker in target samples and environmental samples impacted by swine manure and to discriminate between closely related genes in non-target fecal material. Furthermore, the results of this study provide evidence that SSH is an effective tool to generate source-specific markers. Future efforts will increase the sensitivity of the qPCR assay.

ACKNOWLEDGMENTS

This thesis would not have been completed without the help of many individuals. I would first like to thank Dr. Barton Baker and the Division of Plant and Soil Sciences for their continual support throughout my graduate level work. Research was made possible through funding by the National Science Foundation's ADVANCE IT Program under Award HRD-1007978.

I would like to extend my sincerest gratitude to my committee chair, Dr. Alan Sexstone. His encouragement and knowledge has helped mold me into the scientist I am today. I would also like to thank my other two committee members: Dr. Jennifer Weidhaas and Dr. Gary Bissonnette. Dr. Weidhaas was gracious and allowed me into her lab. She has provided me with more guidance and motivation than I will ever be able to thank her for. She was always available and provided advice when I was faced with a difficult question or obstacle. Dr. Gary Bissonnette was my undergraduate advisor and is responsible for igniting my interest in microbiology. He was always constructive when giving suggestions or correcting mistakes.

Additionally, I would like to thank the members of the Weidhaas lab: Pujya Wagle, Thomas Richard, Xiang Li, Sirisha Mantha, Emily Lipscomb, and Elliott Hair. Their help during my research and throughout the writing process was invaluable. Furthermore, I would like to thank all of the professors and graduate students in the Division of Plant and Soil Sciences. Each and every one of them has had an impact on me.

Last, but certainly not least, I want to express my love and appreciation for my family and friends. My parents, Cam and Leann, have been incredibly supportive, emotionally and financially, throughout my life and particularly in the past two years. My sisters, Nikki and Missy, are my best friends and were always available when I needed to complain. They are the best role models I could have ever been given. Anthony and his words of encouragement were imperative to my success. He has never once given up on me or the goals I set for myself.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS	iii
1. INTRODUCTION	1
1.1 Environmental impact of swine operations	1
1.2 Current water quality regulations and standards.....	2
1.3 Microbial source tracking.....	3
1.4 Existing swine markers	4
1.5 Suppressive subtractive hybridization	5
1.6 Specific study objectives.....	6
2. MATERIALS AND METHODS	8
2.1 Sample collection	8
2.2 Genomic DNA extraction.....	12
2.3 16S rRNA amplification	13
2.4 Suppressive Subtractive Hybridization	14
2.5 Clone Libraries	15
2.6 Primer Design.....	16
2.7 PCR assay.....	16
2.8 qPCR assay.....	17
2.9 Phylogeny.....	18
3. RESULTS	19
3.1 First round of suppressive subtractive hybridization (SSH).....	19
Primers designed for swine specific sequences	19
PCR optimization	20
PCR assay on swine and non-targets	21
3.2 Second round of suppressive subtractive hybridization (SSH)	22
Primers designed for swine specific sequences	22
PCR optimization	24
qPCR assay testing A5 primer set	24
Phylogeny of A5 Prevotellaceae bacterium marker.....	33
4. DISCUSSION	34
REFERENCES.....	39

APPENDIX.....	43
First round of SSH clone sequences	43
Second round of SSH clone sequences.....	49

LIST OF FIGURES AND TABLES

Figure 2.1 Simplified schematic diagram of methods performed in the study	9
Figure 3.1 Dissociation curve for plasmid A5, rooster and all other non-targets	26
Figure 3.2 Standard curve	28
Figure 3.3 Dissociation curve for plasmid A5 and Soil 2 ESB	30
Figure 3.4 Dissociation curve for plasmid A5 and Soil 3 SAS	30
Figure 3.5 Dissociation curve for plasmid A5 and cattle	31
Figure 3.6 Dissociation curve for plasmid A5 and human	31
Figure 3.7 Dissociation curve for plasmid A5 and swine	32
Figure 3.8 Phylogenetic tree	33
Table 2.1 Samples used to isolate swine fecal marker using FastDNA SPIN Kit for Soil	10
Table 2.2 Samples used to test swine fecal marker using DNA/RNA coextraction method	11
Table 3.1 Primers specific to first round SSH clones	19
Table 3.2 PCR optimization using first round SSH specific primers	20
Table 3.3 Results of first round primers in PCR with swine and non-target fecal sample	21
Table 3.4 Primers specific to second round SSH clones	23
Table 3.5 PCR optimization using second round SSH specific primers	24
Table 3.6 Results of A5 primers in SYBR qPCR with swine and non-target fecal samples	26
Table 3.7 Results of A5 primers in SYBR qPCR testing new coextracted samples	29

1. INTRODUCTION

Water quality has become a top priority in the United States since the passage of the Clean Water Act in 1972; however, many lakes, rivers and streams do not meet the CWA criteria based on pollutant levels (USEPA 2000b). Water can be impaired in numerous ways, but fecal contamination is of great importance, especially in waters used for human recreation, drinking, and aquaculture (Simpson et al., 2002). Many animals, including swine, are known to be reservoirs for human pathogens, including *E. coli* O157:H7, *Salmonella* spp., *Mycobacterium* spp. and *Listeria* spp. (Cox et al., 2005). These pathogens may persist in the soil or surface waters (Brown et al 2004) posing a threat to public health.

1.1 Environmental impact of swine operations

Pork production is of great importance in American agriculture. In 2012, there were approximately 60,200 hog and pig operations in the United States (USEPA 2007). Most of the manure produced at these swine operations is applied to land as fertilizer (USEPA 2007). Unfortunately, it has been shown that pathogens can survive for several weeks during storage of manure and in the soil after manure is applied to land as fertilizer (Nicholson et al., 2005). Overall, sources of fecal contamination can be classified into two major groups: point and nonpoint. Examples of point sources are raw or treated sewage and sewer overflow; nonpoint sources are agriculture, wild-life, and urban runoff (Okabe et al., 2007). The USEPA's 1998 *National Water Quality Inventory* indicates that agricultural operations are a significant source of water pollution in the U.S. It is estimated that agriculture contributes in part to the impairment of at least 170,750 river miles, 2,417,801 lake acres, and 1,827 estuary square miles (USEPA 2007).

The manure that is produced during pork production can be treated in different ways. Historically, swine manure has been handled as a solid (USEPA 2007). The manure is either applied directly to land or kept in storage for composting. Storage is important to maintain nutrient quality and also decrease amount of runoff losses. Manure is rich in nutrients, such as nitrogen, phosphorus and potassium, which make it a suitable way of fertilizing croplands. Manure also has the ability to improve soil quality and increase biological activity.

1.2 Current water quality regulations and standards

Current regulatory strategies require culturing of fecal indicator bacteria (FIB) such as *E. coli*, enterococci and fecal coliforms. An ideal indicator is “nonpathogenic, rapidly detected, easily enumerated, have survival characteristics that are similar to those pathogens of concern, and can be strongly associated with the presence of pathogenic microorganisms” (USEPA 2005). These indicator organisms are present in the intestinal tracts of both humans and warm-blooded animals and are, therefore, unable to distinguish fecal input from swine and other species (Marti et al., 2010). The other drawback associated with the culture-dependent FIB method is the requirement for a long analytical turn-around times (>24hr) (Weidhaas et al., 2011). Some effort has been made to circumvent the weaknesses associated with the traditional FIB culture methods. The U.S. EPA has described improved methods for the enumeration of the two recreational water quality indicators: enterococci and *Escherichia coli*. The new method for *E. coli* does decrease the time required to analyze water quality, but not significantly (USEPA 2000a). A qPCR assay has been designed to detect DNA from enterococci in water. Results can be obtained from this assay in 3-4 hours allowing for rapid water quality analysis (USEA 2010).

1.3 Microbial source tracking

Microbial source tracking (MST) methods targeting the genes of microorganisms that are specific to the gastrointestinal tract and feces of a particular host can be used to supplement traditional FIB methods. The term microbial source tracking or bacterial source tracking, describes the concept that “fecal pollution can be traced using microbiological, genotypic, phenotypic, and chemical method” (Scott et al., 2002). MST methods can be grouped into two major types: library-dependent and library-independent. Library-dependent methods rely on culturing, while library-independent methods are based on detection of a specific host-associated genetic marker or DNA extract by PCR (Stoeckel et al., 2007). Quantitative PCR (qPCR) is also commonly used to quantify the marker in environmental samples. The phenotypic methods used in MST are used for discriminating among different groups of bacteria. The two methods that fall under this category are multiple antibiotic resistance (MAR) analysis and immunological procedures. There are also many genotypic methods that can be used in MST, and they include pulsed-field electrophoresis (PFGE), repetitive element PCR, ribotyping, and host specific molecular markers. Other microbiological and chemical methods are used in microbial source tracking.

MST methods have already been employed to help identify nonpoint sources responsible for fecal pollution of water (USEPA 2005). There are key characteristics that are ideal for a fecal source marker. The following list is specified in the 2005 MST guide document:

1. Host specificity
2. Distribution in all host species members or waste
3. Marker is rarely subject to mutation or methodological variability
4. Temporal marker stability in the host

5. No geographic variability in marker/host association
6. Diversity of the marker in the host and in water is represented by a small sample size
7. Limited or a predictable rate of decay in environment
8. Abundance in primary and secondary habitat is related
9. Marker can be used to regulate water quality
10. The marker itself constitutes a health risk

1.4 Existing swine markers

A large amount of research has been aimed at distinguishing human and animal sources of contamination (Bernhard et al., 2003, Carson et al., 2005, King et al., 2007, and Lynch et al., 2002). During the past ten years, there has been effort placed in finding markers that are appropriate for swine fecal contamination. Of all the previously reported targets and methods used for swine fecal pollution, none are fool-proof. Problems experienced involve the inability to perform quantitative analysis or the limited persistence of the marker in the environment.

Molecular swine markers have been developed to target the 16S rRNA gene sequences of dominant Eubacteria, including *Bacteroides-Prevotella* (Okabe et al., 2007), *Lactobacillus-Streptococcus* (Marti et al., 2010), and *Bifidobacterium* (Marti et al., 2009). Methanogenic *Archaeobacteria*, such as the *mcrA* gene (Ufnar et al., 2007), also have been studied as swine targets. Mitochondrial DNA (Kortbaoui et al., 2009 and Martellini et al., 2005) and adenoviruses (Maluquer de Motes et al., 2004) also have been used as targets for fecal contamination specific to that of swine.

Khatib et al. (2003) developed a PCR method using a portion of the STII toxin gene from enterotoxigenic *E. coli* as the target sequence. The swine biomarker in this study did not show any cross-reactivity against any of the 274 DNA extracts from animal feces, human sewage, and

pure cultures, proving that the occurrence of STII outside of swine is rare. While the STII toxin gene portion did show specificity to swine, the measurement was strictly qualitative. Mieszkina et al., 2009 did perform quantitative PCR (qPCR) targeting a swine-specific *Bacteroidales* 16S rRNA gene marker. The biomarker appears to be suitable for quantifying pig contamination; however, the *Bacteroidales* group employs strictly anaerobic metabolism. This may have a negative impact on their persistence in well-oxygenated water (Marti et al., 2010).

1.5 Suppressive subtractive hybridization

Suppressive subtractive hybridization (SSH) was first introduced in 1996 by Clontech laboratories. The method is based on a suppression PCR, which was introduced a year earlier by the same company. Originally, SSH was developed to generate subtracted cDNA libraries that are differentially regulated or tissue specific (Diachenko et al., 1996). Scientist did note that this method could be applied to many other molecular genetic studies. Simple subtractive hybridization is a popular technique that has been used to compare bacterial genomes for identification of virulence factors (Calia et al., 1998, Choi et al., 2002 and Perrin et al., 1999) and to define regions present in a sequenced genome but absent in an unsequenced genome (Agron et al., 2002). In 2008, the first suppressive subtractive hybridization methods were described for prokaryotic DNA, using *Pseudomonas putida* as the model organism (De Long et al., 2008).

In the PCR-Select Bacterial Genome Subtraction Kit the DNA sample that contains the sequences of interest is referred to as the “tester” sample, while the reference sample is the “driver”. In the present study, swine fecal DNA serves as the tester sample and pooled non-target DNA as the driver. The DNA is first digested with a restriction enzyme then adaptors are ligated. These steps are vital for downstream procedures, such as hybridization and PCR. In the next step, sequences that are the same between the tester and driver samples will hybridize and

then be removed. The last two steps are PCR reactions. The primers anneal to sequences on the adaptors. The second reaction is a nested PCR in which the goal is to further enrich to tester specific sequences. The sequences obtained in this particular kit can be taken directly into T/A cloning.

In 2005 a microplate subtractive hybridization protocol was used to enrich for the host-specific *Bacteroidales* rRNA gene fragments that were different from those of closely related reference sources (Dick et al., 2005). The study was able to successfully distinguish between similar fecal sources (e.g. cow and elk) by using host-specific primers designed from the sequences obtained in subtractive hybridization. In 2006, a method known as genome fragment enrichment (GFE) was applied to identify host-specific markers (Shanks et al., 2006). Overall, the data obtained in the study demonstrated that competitive solution hybridization is an efficient method to identify fecal genetic markers. SSH has also been used to develop goose- and duck-specific DNA markers (Hamilton et al., 2006). *E.coli* from host and non-target samples was pooled and used as the tester and driver DNA (respectively) in SSH. The markers developed in this study did show cross reactivity with human *E.coli* strains and also were subject to regional specificity.

1.6 Specific study objectives

A. Determine a 16S rRNA sequence specific to swine feces by suppressive subtractive hybridization. The CLONTECH PCR-Select™ Bacterial Genome Subtraction kit (BD Biosciences CLONTECH, Mountain View, CA) was used in order to complete the first objective. The final PCR products obtained from the kit were taken directly into cloning. The clones were sent to the WVU Genomics core facility for sequencing.

B. Design primers for the candidate 16S rRNA sequence and test the specificity by PCR and/or qPCR on swine and non-target fecal samples. To accomplish the second objective, ABI Primer Express v.3 program (Applied Biosystems, Foster City, CA) and NCBI primer design was utilized to design primers against the sequences obtained in the SSH kit. The primers were tested against swine and non-target fecal samples in PCR and qPCR assays.

C. Establish the environmental persistence of the 16SrRNA sequence by confirming the presence in soils that have historically had swine manure applied. To complete the third objective, genomic DNA was extracted from soil samples obtained from Moorefield, WV and used as template DNA in PCR and qPCR assays with swine specific primers. These soil samples have had swine manure applied to them historically. Additionally environmental sampled collected from WVU campus and animal science farm were tested.

The next section of this paper (Chapter 2) will discuss the materials and methods utilized in the study. The results obtained from each experiment are presented in Chapter 3. The results will then be discussed further and compared to previously developed MST markers in Chapter 4. Finally, conclusions and goals for future research are stated. Additional results are presented in the Appendices.

2. MATERIALS AND METHODS

2.1 Sample collection

A complete overview of the methods used in this study is shown in Figure 1. Fecal samples were collected from six different sites: WVU Animal Sciences Farm (Morgantown, WV), Burn's Family farm (Grafton, WV), Lazy J. Stables (Uniontown, PA), waste water treatment plants (Star City and Cheat Lake, WV) and a swine facility (Moorefield, WV). Fecal samples were collected from the following sources: swine, beef cow, dairy cow, turkey, poultry, rooster, geese, sheep, goat, horse and human, dog and duck. Lagoon samples and two different agricultural soils impacted with swine manure were also collected. A list of the samples collected is shown in Tables 2.1 and 2.2.

Approximately 1 gram of a composited fecal material was collected in sterile, nonpyrogenic 15mL centrifuge tubes (Corning Incorporated, Corning, NY). Composite fecal samples were taken for swine and non-target sources when possible. Composite samples were made by collecting scoops of fecal material from different animals (e.g. pats, scats or droppings) in a sterile, stainless steel bowl and then mixing thoroughly. A subsample was then taken from the bowl, and served as a composite sample for that specific animal species. Composite samples ensure a more accurate representation of the bacteria present in the gut and feces. All samples were held on ice for no longer than 3 hours prior to DNA extraction. Most DNA extractions were done the day of collection. Samples that were not extracted for DNA the day of collection had 20% glycerol added and were placed in a -80°C freezer.

Approximately 500 milliliters of lagoon sample was collected in a sterile 1 liter bottle. Once in the lab, three subsamples were made and used for DNA extraction. The soil samples

collected from on the WVU campus and at the animal science farm were taken from 3 inches below the soil surface. Each soil sample was made of aa composite from three scoops.

Figure 2.1 Simplified schematic diagram of methods performed in the study

Table 2.1 List of samples used to isolate swine fecal marker using FastDNA SPIN Kit for Soil

Fecal Source	Type of sample	n	Location
Swine	Composite of 6 pats	4	WVU ASF*, Morgantown, WV
	Washout trough	2	Swine facility, Moorefield, WV
Beef cattle	Composite of 3 pats	1	WVU ASF, Morgantown, WV
Dairy cattle	Composite of 8 pats	1	WVU ASF, Morgantown, WV
Turkey	One scat	1	WVU ASF, Morgantown, WV
Poultry liter	Composite of 5 scoops	1	WVU ASF, Morgantown, WV
Rooster	Composite of 4 scats	1	WVU ASF, Morgantown, WV
Goose	Composite of 9 scats	1	WVU ASF, Morgantown, WV
Sheep	Composite of 3 scats	1	WVU ASF, Morgantown, WV
Goat	Composite of 2 scats	1	Burn's Farm, Grafton, WV
Horse	Composite of 2 pats	1	Burn's Farm, Grafton, WV
	Composite from manure pile	1	Lazy J Stables, Uniontown, PA
	Composite of 6 pats	1	WVU Reedsville Farm, Reedsville, WV
Human	Septic Tank (3 X 15mL)	1	Morgantown, WV
	WWTP* influent (3 X15 mL)	1	Morgantown WV
Dog	Composite of 3 scats	1	Fairmont and Morgantown, WV
Duck	Composite of 3 scats	1	Morgantown, WV
Soil	Composite of 20 samples	10	Swine facility, Moorefield, WV

n, number of individual samples

ASF*, Animal Science Farm

WWTP*, Wastewater treatment plant

Table 2.2 List of samples used to test swine fecal marker using manual coextraction method

Fecal Source	Type of sample	n	Location
Swine	Composite of 4 pats	1	ASF*, Morgantown
	Composite of 5 pats	1	
Lagoon	3 samples taken	3	ASF
Soil A	Composite of 3 scoops	1	Area directly outside lagoon, ASF
Soil B	Composite of 3 scoops	1	10 ft. from lagoon, ASF
Soil 1 Pine	Composite of 3 scoops	1	Under pine trees, Morgantown
Soil 2 ESB	Composite of 3 scoops	1	Outside SAS*, Morgantown
Soil 3 SAS	Composite of 3 scoops	1	Outside ESB*, Morgantown
Poultry liter	Composite of 8 scoops	1	ASF
	Composite of 5 scoops	1	
Duck/goose	Composite of 5 scats	1	Star City, Morgantown
Cattle	Composite of 5 pats	1	ASF
Human	WWTP* Influent	1	Star City, Morgantown

n, number of individual samples

ASF*, Animal Science Farm

SAS*, South Agricultural Sciences Building

ESB*, Engineering Sciences Building

WWTP*, Waste water treatment plant

2.2 Genomic DNA extraction

Genomic DNA was extracted from fecal and soil samples using two different methods: FastDNA SPIN Kit for Soil (MP Biomedicals, Solon, OH) and a DNA/RNA coextraction method. The coextraction method was only used in the final experiment of this study. Separate samples were collected for each type of extraction (see Tables 2.1 and 2.2).

The FastDNA SPIN Kit for Soil was used following the manufacturer's instructions. For solid fecal and soil samples, DNA was extracted from approximately 500 mg (wet weight) of material placed directly into the extraction kit tube. For liquid samples, approximately 15mL was filtered through a 0.45- μ M filter (Fisher Scientific, Fair Lawn, NJ). The filtrate was then placed into a sterile 50mL tube (Corning Incorporated, Corning, NY). Sterile glass beads and 1mL of water was added. The tubes were then vortexed for 20 seconds to pull the cells from the membrane. The resulting liquid was added directly to the extraction tube provided in the kit. Purification of the DNA was carried out using ethanol precipitation. Purified DNA was quantified using a Nanodrop ND-1000 UV Spectrometer (Nanodrop Technologies, Wilmington, DE).

The manual DNA/RNA coextraction method was done according to a method proposed by Griffiths et al. (2000). Approximately 0.25 grams of feces or soil was measured and placed into a FP 120 bead system tube (MP Biomedicals, Solon, OH). 0.5 mL of CTAB (cetyltrimethyl ammonium bromide) mixture and 0.5 mL of phenol: chloroform: isoamyl alcohol (25:24:1) (Fisher Biotech, Wembley, WA) was added to the tube. The CTAB mixture was 10% weight per volume CTAB (Sigma Life Sciences, St. Louis, MO) with 0.7M NaCl (Fisher Scientific, Fair Lawn, NJ) and equal volume of 0.24M potassium phosphate (Fisher Scientific, Fair Lawn, NJ) at pH 8. The tube was vortexed for ten minutes at a setting of 8 in a horizontal position, then spun

at 4°C, 16,000 x g for five minutes in a Accuspin Micro 17R centrifuge (Fisher Scientific, Asheville, NC). The aqueous phase was transferred to a new, sterile 1.5 mL tube () and the same volume of chloroform: isoamyl alcohol (24:1) (Sigma Life Sciences, St. Louis, MO) was then added. The tube was spun again for five minutes at 4°C, 16,000 x g. The aqueous layer was again taken into a new, sterile 1.5 mL tube and 2 volumes of PEG (Polyethylene glycol) solution was added. The PEG solution was made with 30% weight per volume PEG-6000 (Alfa Aesar, Ward Hill, MA) and 1.6M NaCl. The tube was then incubated in the dark at room temperature for two hours. Following incubation, the tube was spun at 4°C, 17,000 x g for fifteen minutes. The supernatant was removed by pipette, then 100uL of ice cold 70 % molecular grade ethanol (Sigma Life Sciences, St. Louis, MO) was added. The tube was spun for a final time at 4°C, 17,000 x g for fifteen minutes. The ethanol was removed with a pipette and then placed in a hood to evaporate any remaining ethanol. The DNA was eluded in 10 mM Tris (Ambion, Grand Island, NY) with a final volume of 30uL. To rid the sample of RNA, 1uL of 2mg/mL RNase A (Thermo Scientific, Waltham, MA) was added to the tube. The tube was then incubated at 37°C for ten minutes. The extracted DNA was held at -20°C until downstream use.

2.3 16S rRNA amplification

Polymerase chain reaction (PCR) was performed to amplify the 16s rRNA gene from the collected fecal and soil DNA. 8F (5' AGA GTT TGA TCC TGG CTC AG 3') and 1492R (5' ACG GCT ACC TTG TTA CGA CTT 3') (Integrated DNA Technologies) primers were utilized. The final primer concentration in the master mix was 0.2µM. The reaction volume was 25 µL, using 1 µL of template DNA. A Mastercycler nexus family thermocycler (Eppendorf, Hauppauge, NY) was used to carry out the reaction. The following thermocycler conditions

were kept for all samples: 95°C for 15 minutes, 35 cycles of 95°C for 60 seconds, 53.5°C for 90 seconds, 72°C for 60 seconds, and a final extension at 72°C for 7 minutes. The PCR products obtained were purified using QIAquick® PCR Purification Kit (QIAGEN, Valencia, CA). The DNA products were then quantified using a Nanodrop ND-1000 UV Spectrometer (Nanodrop Technologies, Wilmington, DE).

2.4 Suppressive Subtractive Hybridization

Suppressive subtractive hybridization (SSH) was done using CLONTECH PCR-Select™ Bacterial Genome Subtraction kit (BD Biosciences CLONTECH, Mountain View, CA) according to the manufacturer's instructions. The 16S rRNA gene products obtained from the methods described above were pooled together making two separate mixtures. The tester was made up of swine fecal 16S rRNA, and the driver was made of the non-target fecal 16S rRNA (e.g. human, cow, dog...etc.). The tester and driver samples were equalized in the amount of DNA added from each individual PCR product.

The tester and driver DNA was first digested with *RsaI* (provided in the kit) for 5 hours at 37°C. *RsaI* is a four-base cutting restriction enzyme that will yield short, blunt ended molecules that are ideal for hybridization reactions. The next two steps in the kit are hybridization reactions between the tester and driver. The goal of the primary hybridization is to equalize and enrich differentially expressed sequences. In the second hybridization, PCR templates specific to the tester (swine feces) are generated. These templates are then amplified by two rounds of PCR, using primers that are supplied in the kit. The second PCR is performed using nested primers to further reduce any background PCR products and enrich for tester-specific sequences.

Two rounds of SSH were conducted. In the first round, all available non-target DNA (see Table 2.1) was pooled and used as the driver mixture. In the second round, only five non-target fecal samples (cattle, sheep, horse, rooster, and human) were pooled. These particular non-target samples were chosen because they represent the major sources for fecal contamination in national waters.

2.5 Clone Libraries

The tester-specific sequences obtained from SSH were cloned into a plasmid vector (pCRII-TOPO) using the TOPO TA Cloning® kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The pCRII-TOPO construct, made by introducing the swine-specific sequences, was then transformed into competent *E. coli* provided in the kit. Transformed *E. coli* cells were plated onto Difco™ LB agar plates (BD, Dickinson and Company, Sparks, MD, USA) containing 50 µg/ml kanamycin and X-gal. Plates were incubated at 37°C overnight. White or light blue colonies were chosen and cultured overnight in SOC broth media (Teknova, Hollister, CA) containing 50 µg/ml kanamycin. Plasmids were extracted using GeneJET Plasmid Miniprep Kit (Thermo Scientific). The purified plasmids were sequenced at the WVU Genomics Core Facility using M13F (5' GTA AAA CGA GGG CCA G 3') and M13R (5' GAG GAA ACA GCT ATG AC 3') primers. The sequences obtained from the Genomics Core were analyzed using Biological Sequence Alignment Editor (BioEdit Software, Carlsbad, CA). Sequences were trimmed by finding where the nested primers, used in the last step of SSH kit, were located. The sequences were searched in NCBI BLAST and RDP11 database to ensure true unique sequences; the closest related organism was noted. Bellerophon software was also used to check for chimeric sequences (Huber et al 2004).

2.6 Primer Design

PCR primers were designed to amplify candidate tester-specific sequences using the ABI Primer Express v.3 program (Applied Biosystems, Foster City, CA) and NCBI primer design (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). The Oligo Analyzer available through Integrated DNA Technologies was utilized to analyze the primer sets. The following parameters were used in order to design the most effective primers: length between 18 and 22; GC content 40 to 60 percent; delta G value for secondary structures below set value; and annealing temperature for forward and reverse no more than 5 degrees Celsius different. The designed primers were checked in NCBI BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and the ribosomal database (<http://rdp.cme.msu.edu/>) for specificity.

2.7 PCR assay

The primers were first optimized for PCR. Swine fecal genomic DNA was used as the template and increasing melting temperatures (ranging from 57°C-73°C) were tested. The highest melting temperature in which amplification could be observed was used in all subsequent PCR analysis. The sensitivity and specificity of the primer sets were then evaluated by PCR using original swine and non-target genomic DNA.

In all PCR reactions a Mastercycler nexus family thermocycler (Eppendorf, Hauppauge, NY) was used. The final primer concentration in the master mix was 0.2µM. The reaction volume was 25 µL, and 1 µL of template DNA was used. The following thermocycler conditions were kept for all samples, except for the determined annealing temperature: 95°C for 15 minutes, 35 cycles of 95°C for 60 seconds, X°C for 90 seconds, 72°C for 60 seconds, and a final extension at 72°C for 7 minutes.

2.8 qPCR assay

Quantitative PCR (qPCR) was used due to the increased sensitivity compared to gel electrophoresis detection of PCR products. This method was only used for the second round of designed primers. Only the most successful primers observed in PCR were tested. Nested qPCR was also conducted on certain samples. This was done by using 16S rRNA amplified DNA as template DNA, rather than genomic DNA. All reactions were carried out using SYBR Select Master Mix (Applied Biosystems, Austin, TX) on a 7300 Real Time PCR System (Life Technologies, Grand Island, NY). The final concentration of the forward and reverse primers in the reaction mix was 0.5 μ M. The plasmid from which the primer set was designed was used as the positive control in all reactions; PCR grade water was used as negative control. The following thermocycler conditions were used for all qPCR reactions: 50°C for 2 minutes, 95°C for 10 minutes, 35 cycles of 95°C for 1 minute and 67°C for one minute. Primers were also tested against USDA soil samples. The soil was collected from a site in Moorefield to which swine manure had been applied historically. This assay determined if the biomarker is resilient in the environment.

A dissociation curve was run at the end of every qPCR run. This is important to do with SYBR green amplicon detection. SYBR Green will detect any double stranded DNA, including mis-annealed PCR products or primer-dimers. Melting curve analysis was done on the data to determine if the desired amplicon was detected. This analysis can verify even a single base pair difference between samples. The following temperature ramp conditions were used for all dissociation curves: 95°C for 15 seconds, 20°C for 30 seconds, then 60°C to 95°C increasing 2°C per second. A standard curve was developed from serial dilutions of plasmid A5 DNA. Each standard concentration was measured in triplicate.

2.9 Phylogeny

Mega 5 software was used to analyze the evolutionary relationship between the proposed biomarkers. The sequences of the biomarkers were first aligned by Clustal W. The alignment was then used to generate a Maximum Likelihood tree. The bootstrap value was set to 1000. The tree is rooted with the methanogen *Methanosarcina mazei* (AF411468).

3. RESULTS

3.1 First round of suppressive subtractive hybridization (SSH)

Primers designed for swine specific sequences

Two rounds of SSH were conducted. In both rounds the 16S rRNA genes of all samples were first amplified then taken into the SSH protocol. In the first round of suppressive subtractive hybridization (SSH), 15 different non-target DNA samples were combined to serve as the driver DNA in the kit. Seven clones were obtained. The clones were sequenced, and it was determined by use of BioEdit that only five sequences were present. Bellerophon software was utilized to determine whether chimeras were present in the sequences (Huber et al 2004). The closest related culturable organism for each sequence was determined by NCBI BLAST. Primer design gave rise to five primer sets: *Acholeplasma brassiaca*, Rumen bacterium, Rickenellaceae bacterium, *Treponema pectinovorum* and Bacteroidetes bacterium. All primer sequences, the clones from which they were designed, the primer length and the amplicon size can be seen in Table 3.1.

Table 3.1 Primers specific to first round SSH clones

Clone	Closest related organism	Primer sequence	Primer length (bp)	Amplicon length (bp)
5	<i>Acholeplasma brassiaca</i>	5' CTCTAGCGAGACTGCCGTTGA 3' 5' TAGCGATTCCGACTTCATGGA 3'	21 21	200
1	Rumen bacterium	5' CAGGTGGCTTCCCTCTGTAG 3' 5' AGCTCGTGTCGTGAGATGTT 3'	20 20	200
2	Rickenellaceae bacterium	5' GCAGACTCCGATCCGAAGTGA 3' 5' CACATAGTTACCAGCGCGTTGA 3'	20 22	400
3	<i>Treponema pectinovorum</i>	5' GGAACCTTACCTGGGTTTGACA 3' 5' CCACCTTCCTCCGTTTGT 3'	22 19	273
4	Bacteroidetes bacterium	5' CTGCAGAATTCGCCCTTAGC 3' 5' CGAAGGCATCTCACGAATCC 3'	20 20	250

Seven clones obtained from the first round of SSH. Only five unique sequences existed and were used to develop specific primers.

PCR optimization

The primers designed for the clones obtained from the first round of SSH were first tested against genomic swine DNA to optimize conditions. Multiple PCR reactions were carried out with increasing annealing temperatures ranging between 57°C and 66°C (Table 3.2). It was determined that 62°C was the appropriate annealing temperature to use for all primer sets in all subsequent PCR reactions. This high annealing temperature will help to decrease non-specific amplification. At low annealing temperatures, primers may anneal to sequences other than the true target, leading to amplification in samples other than swine.

Table 3.2 PCR optimization using first round SSH specific primers

Primer	Annealing temperature				
	57°C	60°C	62°C	64°C	66°C
<i>Acholeplasma brassiaca</i>	Black	Black	Black	White	White
Rumen bacterium	Black	Black	Black	Grey	White
Rickenellaceae bacterium	Black	Black	Black	White	White
<i>Treponema pectinovorum</i>	Black	Black	Black	Grey	White
Bacteroidetes bacterium	Black	Black	Black	White	White

The PCR results shown were obtained testing only swine genomic DNA. Black = amplification, White = No amplification, Grey = weak amplification.

PCR assay on swine and non-targets

Polymerase chain reaction (PCR) was performed to determine if amplification occurred using the newly designed primers. The five primers were tested against genomic swine and non-target DNA (Table 3.3). The annealing temperature used in the PCR presented was 62°C. This temperature was determined by testing genomic swine DNA in PCR with increasing melting temperatures (Table 3.2). Unfortunately, the primers were amplifying non-target DNA. The Rikenellaceae bacterium primer set was the most successful of all five tested, but still showed cross-reactivity in beef and dairy cow, sheep and goose. At this point, the SSH was re-done. In the second round of SSH, the number of non-target DNA pooled in the driver sample was decreased, with the goal of obtaining sequences truly specific to swine feces.

Table 3.3 Results of first round primers in PCR with swine and non-target fecal sample

Primer	Sample												
	1	2	3	4	5	6	7	8	9	10	11	12	13
5 <i>Acholeplasma brassiaca</i> e	Black	Black	White	Black	Black	White	Black	Black	White	White	Black	White	
1 Rumen bacterium	Black	Black	Black	Black	Black	Black	Black	Black	White	Black	White	Black	White
2 Rickenellaceae bacterium	Black	White	White	Black	Black	White	White	Black	White	White	White	Black	White
3 <i>Treponema pectinovorum</i>	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black
4 Bacteroidetes bacterium	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black

The PCR results shown were obtained using an annealing temperature of 62°C. Black = amplification. White = No amplification. The following genomic DNAs were used in the analysis: 1-Swine; 2-Human (septic tank); 3-Dog; 4-Beef cow; 5-Dairy cow; 6-Horse; 7-Goat; 8-Sheep; 9-Poultry; 10-Rooster; 11-Turkey; 12-Goose; 13-Duck

3.2 Second round of suppressive subtractive hybridization (SSH)

Primers designed for swine specific sequences

In the second round of SSH, only five non-target DNA samples were combined to serve as the driver DNA. The five non-target samples combined were cattle (combining dairy and beef DNA), sheep, horse, rooster, and human (septic tank), each representing major fecal sources in U.S. waters. Twenty clones were obtained from this round of SSH. The clones were submitted for sequencing. It was determined that only twelve unique sequences were present. Bellerophon software was utilized to determine whether chimeras were present in the sequences (Huber et al 2004). Primer design gave rise to the following twelve sets: *Clostridium tepidiprofundum*, *Clostridium sp.*, *Prevotella sp.*, Prevotellaceae bacterium, Bacterium WH2-11, *Gemmiger formicilis*, Butyrate producing bacterium, *Prevotella copri*, Clostridiales bacterium, *Clostridium chauvoei*, *Clostridium butyricum* and *Clostridium sp.* All primer sequences, the clones from which they were designed, the primer lengths and the size of the amplicon can be seen in Table 3.4.

Table 3.4 Primers specific to second round SSH clones

Clone	Closest related organism	Primer sequence	Primer length (bp)	Amplicon length (bp)
A1	<i>Clostridium tepidiprofundum</i>	5' TTGTAGCACGTGTGTAGCCC 3' 5' CCGCACAAGCAGTGGAGTAT 3'	20 20	310
A2	<i>Clostridium sp.</i>	5' AGGGGGAGCAAACTGGAAAA 3' 5' CAAGGCCCGGGAACGTATT 3'	21 19	123
A3	<i>Prevotella sp.</i>	5' ATCCCCACCTTCCTCCGTT 3' 5' CGTGTGTGTGAAATACTGTTCGC 3'	19 21	580
A5	Prevotellaceae bacterium	5' TGGTGTAGCGGTGAAATGCT 3' 5' CATGGCTGATGCGCGATTAC 3'	20 20	682
A8	Bacterium WH2-11	5' GTTAACTGCGGCACTGAAGG 3' 5' ACGTAGGGGGCTAGCGTTAT 3'	20 20	331
A10	<i>Gemmiger formicilis</i>	5' ACTCCCCAGGTGGATTACTT 3' 5' CCGGCAAGTTGGAAGTGAAA 3'	20 20	302
A13	Butyrate bacterium	5' GCCAGAGTCCTCTTGCGTAG 3' 5' CGACCGCAAGGTTGAAACTC 3'	20 20	253
A15	<i>Prevotella copri</i>	5' TGGTGTAGCGGTGAAATGCT 3' 5' TGTAACACGTGTGTAGCCCC 3'	20 20	555
A16	Clostridiales bacterium	5' AGCACGTGTGTAGCCCTAAG 3' 5' AAGTAGCGGAGCATGTGGTT 3'	20 20	300
A18	<i>Clostridium chauvoei</i>	5' CAGGATGACAGGTGGTGCAT 3' 5' CCAGCTTCATGTAGGCGAGT 3'	20 20	300
A19	<i>Clostridium butyricum</i>	5' GCGGTAATACGTAGGTGGCA 3' 5' ACGCATTTCACCGCTACACT 3'	20 20	175
A20	<i>Clostridium sp.</i>	5' GTACTTGAGGAGGAAGCCCC 3' 5' AATCCTGTTTGCTCCCCACG 3'	20 20	304

Twenty clones obtained from the second round of SSH. Only twelve unique sequences existed and were used to develop specific primers.

PCR optimization

The primers designed for the clones obtained from the second round of SSH were first tested against genomic swine DNA to optimize conditions. Multiple PCR reactions were carried out with increasing annealing temperatures ranging between 57°C and 73°C (Table 3.5). Different annealing temperatures were determined for the 12 primer sets. The A5 primer was able to show amplification with an annealing temperature as high as 73°C. The highest annealing temperature in which amplification still occurred was used for each primer set in subsequent PCR reactions.

Table 3.5 PCR optimization using second round SSH specific primers

Primer	Annealing temperature						
	57°C	63°C	65°C	67°C	69°C	71°C	73°C
A1 <i>Clostridium tepidiprofundum</i>	Black	Black	Black	Black	Black	Black	Black
A2 <i>Clostridium sp.</i>	Black	Black	White	White	White	White	White
A3 <i>Prevotella sp.</i>	Black	Black	Black	Black	Black	Black	Black
A5 <i>Prevotellaceae</i> bacterium	Black	Black	Black	Black	Black	Black	Black
A8 Bacterium WH2-11	Black	Black	Black	Black	Black	Black	Black
A10 <i>Gemmiger formicilis</i>	Black	Black	Black	Black	White	White	White
A13 Butyrate bacterium	Black	Black	Black	Black	Black	White	White
A15 <i>Prevotella copri</i>	Black	Black	White	White	White	White	White
A16 <i>Clostridiales</i> bacterium	Black	Black	White	White	White	White	White
A18 <i>Clostridium chauvoei</i>	Black	Black	Black	White	White	White	White
A19 <i>Clostridium butyricum</i>	Black	White	White	White	White	White	White
A20 <i>Clostridium sp.</i>	Black	Black	White	White	White	White	White

The PCR results shown were obtained testing only swine genomic DNA. Black = amplification. White = No amplification

qPCR assay testing A5 primer set

The A5 primers were taken into qPCR and were tested against swine and non-target genomic DNA. This assay has greater sensitivity compared to gel electrophoresis and also gives quantitative data. The non-targets tested in this assay were as follows: swine, human, rooster,

sheep, cattle, horse, duck, poultry liter, goose, and environmental soil samples. The soil samples were taken from a site where swine manure had been applied historically. Nested qPCR was also conducted in which 16S rRNA from those same samples as listed above was used as the template. No amplification was observed in all samples tested with regular qPCR. Nested qPCR did, however, amplify the four swine samples as well as the rooster sample (Table 3.7). This cross-reactivity with rooster DNA is not favorable, but the marker was not found in the other non-target samples tested, including human and cow cattle which are common sources of fecal inputs into waterways.

The final step in all qPCR reactions was a dissociation (melting) curve. This was done to determine if the desired amplicon was detected. The final PCR product is exposed to a temperature gradient, in this case between 60°C-95°C, while the fluorescence is continually measured. At low temperatures the PCR product remains double stranded but as temperatures increase the DNA begins to separate, releasing the SYBR green. The melt curves are converted to distinct melting peaks by plotting the first negative derivative of the fluorescence as a function of time. The peak or inflection point on the graph represents an individual PCR product. The dissociation curve generated from the nested qPCR run can be seen in Figure 3.1. The melting temperature of the rooster sample is 1°C lower than that of the plasmid A5 and the swine sample, confirming a difference in the sequence.

Table 3.6 Results of A5 primers in SYBR qPCR with swine and non-target fecal DNA extracted using FastDNA SPIN Kit for Soil

Fecal Source	<i>n</i> *	qPCR	Nested qPCR
Swine	4	ND	+
Human	2	ND	ND
Rooster	2	ND	+
Sheep	2	ND	ND
Cattle	2	ND	ND
Horse	2	ND	ND
Duck	1	ND	ND
Poultry liter	1	ND	ND
Goose	1	ND	ND
Soil	4	ND	ND

+ positive detection

ND Not detected

*n** number of samples tested

Figure 3.1 Dissociation curve for plasmid A5, swine, rooster, and all other non-targets. The x-axis is the negative derivative of the fluorescence over the change in temperature (-dF/dT). The melting temperature for each sample is shown.

The nested qPCR results demonstrate that the A5 marker may be present in low abundance in swine feces, which is not ideal for a MST marker. In order to increase the concentration of DNA tested in qPCR, new fecal and environmental samples were collected and subjected to DNA/RNA coextraction. This particular method results in much more DNA compared to that produced from commercial DNA extraction kits. The newly extracted DNA was then taken into qPCR with the same conditions as used in previous runs. The standard curve generated to quantify the marker is shown in Figure 3.2. The minimum detection limit for this assay was 4,000 gene copies per μL . The swine, lagoon and agricultural soil samples all amplified using the A5 primer set (Table 3.7). This shows that the marker can be detected without performing nested qPCR, if using DNA in high enough concentrations. It also confirms the presence of the marker in the environment. The two soils collected from the WVU campus (Soil 2 ESB and Soil 3 SAS) as well as the cattle and human samples showed amplification; however, the concentration of the marker was significantly lower compared to that of swine fecal samples and lagoon samples (Table 3.7). It is also important to note that the melting temperatures of soil 2 ESB (Figure 3.3), soil 3 SAS (Figure 3.4), cattle (Figure 3.5), and human (Figure 3.6) were all different than the melting temperature of the A5 plasmid, suggesting differences in the nucleotide sequence. The non-target samples also had multiple peaks present on their dissociation curve, verifying that the A5 sequence was not amplified. The swine sample, along with the lagoon and soils samples (Soil A and Soil B), shows an exact match to the melting temperature of the A5 plasmid (Figure 3.7).

Figure 3.2 A representative standard curve of Ct values verses the plasmid A5 concentrations as measured by SYBR green qPCR assay. Error bars depict the standard error between triplicate samples. The minimum detection limit for this assay is 4,000 gene copies / μ L.

Table 3.7 Results of A5 primers in SYBR qPCR testing new coextracted samples

Fecal Source	<i>n</i> *	Amplification of DNA by qPCR	Multiple peaks in dissociation profile	Swine maker present	Gene copies g ⁻¹ feces or g ⁻¹ soil
Swine	2	Yes	No	Yes	6.90 x 10 ⁸ 9.50 x 10 ⁸
Lagoon	3	Yes	No	Yes	3.20 x 10 ⁷ 4.20 x 10 ⁷ 2.50 x 10 ⁷
Soil A (against lagoon)	1	Yes	No	Yes	2.10 x 10 ⁸
Soil B (10ft. from lagoon)	1	Yes	No	Yes	3.30 x 10 ⁷
Soil 1 Pine Soil	1	No	NA	NA	NA
Soil 2 ESB	1	Yes	Yes	No	NA
Soil 3 SAS	1	Yes	Yes	No	NA
Poultry liter	2	No	NA	NA	NA
Duck/ goose	1	No	NA	NA	NA
Cattle	1	Yes	Yes	No	NA
Human	1	Yes	Yes	No	NA

*n**, number of samples tested

NA, not applicable

Figure 3.3 Dissociation curve for plasmid A5 and Soil 2 ESB. The x-axis is the negative derivative of the fluorescence over the change in temperature ($-dF/dT$). The melting temperature for each sample is shown.

Figure 3.4 Dissociation curve for plasmid A5 and Soil 3 SAS. The x-axis is the negative derivative of the fluorescence over the change in temperature ($-dF/dT$). The melting temperature for each sample is shown.

Figure 3.5 Dissociation curve for plasmid A5 and cattle. The x-axis is the negative derivative of the fluorescence over the change in temperature ($-dF/dT$). The melting temperature for each sample is shown.

Figure 3.6 Dissociation curve for plasmid A5 and human. The x-axis is the negative derivative of the fluorescence over the change in temperature ($-dF/dT$). The melting temperature for each sample is shown.

Figure 3.7 Dissociation curve for plasmid A5 and swine. The x-axis is the negative derivative of the fluorescence over the change in temperature ($-dF/dT$). The melting temperature for each sample is shown.

Phylogeny of A5 Prevotellaceae bacterium marker

The A5 clone sequence obtained from the second round of SSH were used to construct a Maximum Likelihood tree (Figure 3.8). The sequence of the 16S rRNA gene for A5 isolate is shown in relationship to the closest related organisms according to NCBI BLAST. The 16S sequence of *Methanocarcina mazei* was used as the root in tree construction. This tree indicates that the partial 16S rRNA sequences obtained from the SSH kit is related to *Prevotella* spp.

Figure 3.8 Phylogenetic tree of 16S rRNA sequences showing the relationship of clone A5 to that of other *Prevotella* spp. 16S rRNA sequences. The tree was generated in using the maximum likelihood method using 1,000 pseudo-replicates. The scale bar represents a 5% estimated sequence divergence.

4. DISCUSSION

Determining the source of fecal contamination in water is necessary for proper remediation and risk assessment. Fecal contamination of water can lead to many detrimental outcomes for public health and the environment. The feces of warm blooded animals, including that of swine, can harbor many human pathogens, including *E. coli* O157:H7 and *Salmonella* spp. Manure contains high levels of phosphorus, nitrogen and potassium and therefore is used to fertilize many croplands. This can eventually lead to eutrophication especially in the event of heavy rainfall. Current methods employed to examine water quality rely on culturing of fecal indicator bacteria (FIB). This approach can conclude whether water is contaminated but has no way of determining the pollution source.

In the United States pork production contributes to profuse amounts of manure. For example, five hundred finishing pigs produce over 265,000 gallons of manure a year (USEPA 2007). In 2012, there were 60,200 hog and pig operations (USEPA 2007). In most cases the manure produced from these operations is used for land application do to its nutrient rich properties and ability to improve soil structure. Manure is, therefore, a cost effective and readily available mode to fertilize croplands. If not managed properly, however, polluting surrounding waterways will most likely occur.

Microbial source tracking techniques attempt to identify the source of contamination by targeting microorganisms or their genes that are specific to a particular animal's feces. In this study, a new 16S rRNA gene sequence was reported as a potential biomarker for swine fecal contamination. Suppressive subtractive hybridization (SSH) was successfully used to identify sequences with high specificity to swine feces.

In the first round of SSH many of the sequences obtained were not specific to swine feces. In this round of SSH, all available non-target DNA was pooled together to use as the driver. This led to a reduced “subtraction efficiency due to increased complexity introduced into the reaction” (Hamilton et al., 2006). It is also important to consider ineffective primer design as a cause for non-specific amplification. There was most success with isolating sequences specific to swine in the second round of SSH when the driver DNA was limited to 5 non-target samples. This is not the first study to use the SSH method. Goose- and duck-specific DNA markers have been developed using the same CLONETECH PCR Select SSH kit (Hamilton et al., 2006).

The most successful marker, the A5 Prevotellaceae bacterium sequence, was first tested in SYBR green qPCR against DNA isolated using the FAST DNA Spin Kit for Soil. The A5 marker was detected in rooster and swine DNA and was absent in all the non-targets tested (Table 3.6). Based on the dissociation curve generated, the sequence being amplified from the rooster DNA was different to that of the marker (Figure 3.1). The sequence did not amplify in some of the most common sources of fecal contamination, including that of human and cattle. The main setback of this data was the requirement of nested qPCR in order to detect the marker and no amplification in the soil samples from Moorefield impacted with swine manure. In general nested PCR is a modified version of PCR used to decrease the detection limit. Two successive PCR reactions are run involving two different primer sets. The second primer set amplifies a target within the primary PCR product. Due to amplification observed in nested PCR only, the sequence may be present in swine feces in limited abundance. This finding explains why the sequence did not amplify in the soil from Moorefield treated with swine manure. It is

also important to note these soils have had swine manure historically applied to them; however, the last date of application is unknown.

Although the nested qPCR with the A5 primers showed some success, it is not considered favorable for the marker to be in low abundance in swine feces. The marker also was not detected in the environmental samples tested. To increase the amount of DNA taken into qPCR reactions, a coextraction method (Griffiths et al. 2000) was introduced. This method extracts DNA in concentrations as high as $3,000 \text{ ng } \mu\text{L}^{-1}$. With this quantity of DNA, amplification with the A5 primers was achieved without having to perform nested qPCR (Table 3.7).

Unfortunately, apparent cross reactivity was observed in cattle and human samples as well as two different soils collected around the WVU campus (Soil 2 ESB and Soil 3 SAS). The dissociation curves produced from these samples show a melting temperature that is different from the plasmid A5 melting temperature, representing a difference in the nucleotide sequence (Weidhaas et al., 2009). There were also multiple peaks seen in the melting curve (Figure 3.3-3.6), showing non-specific amplification.

The concentration of the marker in swine feces was calculated to be between 6.9 and 9.5×10^8 gene copies per gram of feces. This was higher than what was observed in the lagoon and the soil samples outside the lagoon (Table 3.7). A decrease in the marker concentration is expected to occur between the swine feces and the lagoon and/or environmental samples due to conditions such as aerobic or UV degradation. The A5 marker most likely employs anaerobic metabolism because it is closely related to organisms that are obligate anaerobes. The soil samples were collected below the soil surface and the lagoon had a thick film, likely creating conditions for anaerobic life.

The A5 marker was most related to organisms in the family Prevotellaceae, which is composed of four genera, *Prevotella*, *Alloprevotella*, *Hallella*, and *Paraprevotella*. *Prevotella* can be isolated from a variety of environments including the rumen and hind gut of sheep and cattle, anaerobic infections, and human feces. Twenty different species of *Prevotella* have been isolated from anaerobic infections such as gingivitis, periodontitis, and abscesses (Tanaka et al., 2008). *Prevotella* was found to make up 53% of the gut bacteria in African children (De Filippo et al., 2010).

The A5 marker sequence was most closely related to *Prevotella copri* with an approximate 1.3% sequence divergence (Figure 3.8). *P. copri* has been isolated from human feces (Hayashi et al., 2007); however, no work has been done to isolate this organism from environmental samples. Most research focuses on species that can be opportunistic pathogens (e.g. *P. melaninogenica*). *Bacteroides*, a close relative to *Prevotella*, are obligate anaerobic, gram negative bacteria known to compose a large portion of the gut microbiota in humans and warm blooded animals. *Bacteroides* are commonly used as MST markers due to their specificity for an individual host's gut or feces. The anaerobic metabolism of *Bacteroides* may lead to die off in environmental water or wastewater. Therefore, if the *Bacteroides* marker is detected in water, this suggests a recent fecal pollution (Balleste and Blanch 2010).

The persistence of DNA in the environment from non-viable cells is another problem to circumvent when developing a microbial source marker. RNA-based methods can be used; however, they are technically demanding. Ethidium monoazide bromide (EMA) is another alternative to determine if amplified DNA is from viable or non-viable cells. EMA is a DNA intercalating dye that can only penetrate dead cells, leading to PCR inhibition (Nocker and Camper 2006).

Many swine fecal biomarkers have been proposed in recent years. Molecular markers have been developed to target 16S rRNA, swine mitochondrial DNA, and other genes associated with specific bacteria. Research has focused on 16S rRNA gene sequences of dominant Eubacteria, including *Bacteroides-Prevotella*, *Lactobacillus-Streptococcus*, and *Bifidobacterium* (Okabe et al., 2007, Marti et al., 2010, and Marti et al., 2009). The STII toxin gene from enterotoxigenic *E. coli* (Khatib et al., 2003) and the *mcrA* gene from methanogenic *Archaeobacteria* (Ufnar et al., 2007), have also been studied as swine targets. These studies have experienced problems involving the quantitative measurement of their target or the low abundance of the target in the environment. In the current study, the A5 marker was successfully quantified in swine feces and in environmental samples impacted with swine manure.

Future work could be done to increase the sensitivity of the qPCR assay. In this study SYBR green qPCR was employed. A more sensitive Taqman qPCR could be developed. Taqman qPCR utilizes a probe having a fluorescent reporter dye at the 5' end and a quencher dye at the 3' end. New, possibly more specific primers could also be designed to avoid non specific amplification. Non-target PCR products could also be taken into cloning and sequencing in order to determine the location and what nucleotide changes are present compared to plasmid A5. The marker should also be tested in water samples impacted by agricultural practices.

In summary, this study supports the use of SSH as a tool for developing MST markers. The A5 Prevotellaceae bacterium sequence did show specificity to swine feces and is unique in that it differs from all other proposed sequences in literature. The marker could be amplified by qPCR and was found in environmental samples, making it appropriate for use in the field.

REFERENCES

- Agron, P., M. Macht, L. Radnedge, E. Skowronski, W. Miller, and G. Anderson.** 2002. Use of subtractive hybridization for comprehensive surveys of prokaryotic genome differences. *FEMS Microbiol. Lett.* **66**:175-182
- Balleste, E., and A. Blanch.** 2010. Persistence of *Bacteroides* Species Populations in a River as Measured by Molecular and Culture Techniques. *Appl. Environ. Microbiol.* **76**: 7608-7616.
- Bernhard, A., T. Goyard, M. Simonich, and K. Field.** 2003. Application of a rapid methods for identifying fecal pollution sources in a multi-use estuary. *Water Res.* **37**:909-913
- Brown, P., O. Christensen, H. Clough, P. Diggle, C. Hart, S. Hazel, R. Kemp, A. Leatherbarrow, A. Moore, J. Sutherst, J. Turner, N. Williams, E. Wright, and N. French.** 2004. Frequency and special distribution of environmental *Campylobacter* spp. *Appl. Environ. Microbiol.* **70**:6501-6511.
- Calia, K., M. Waldor, and S. Calderwood.** 1998. Use of representational differences analysis to identify genomic differences between pathogenic strains of *Vibrio cholerae*. *Infect. Immun.* **66**:849-852.
- Carson, C., J. Christiansen, H. Yampara-Tquise, V. Benson, C. Baffaut, J. Davis, R. Broz, W. Kurtz, W. Rogers, and W. Fales.** 2005. Specificity of a *Bacteroides thetaiotaomicron* marker for human feces. *Appl. Environ. Microbiol.* **71**:4945-4949.
- Choi, J., C. Sifri, B. Goumnerov, L. Rahme, F. Ausubel, and S. Calderwood.** 2002. Identification of virulence genes in pothogenic strain *Pseudomonas aeruginosa* by representational difference analysis. *J. Bacteriol.* **184**:952-961.
- Cox, P., M. Griffith, M. Angles, D. Deere, and C. Ferguson.** 2005. Concentration of pathogens and indicators in animal feces in the Sydney watershed. *Appl. Environ. Microbiol.* **71**:5929-5934.
- De Filippo, C., D. Cavalieri, M. Di Paola, M. Ramazzotti, J. Poullet, and S. Massart.** 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci.* **107**:14691-14696
- De Long, S., K. Kinney, and M. Kirisits.** 2008. Prokaryotic suppression subtractive hybridization PCR cDNA subtraction, a targeted method to identify differentially expressed genes. *Appl. Environ. Microbiol.* **74**:225-232.

Dick, L., A. Bernhard, T. Brodeur, J. Santo Domingo, J. Simpson, S. Walters, and K. Field. 2005. Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source contamination. *Appl. Environ. Microbiol.* **71**:3184-3191.

Dick, Linda K., Michael T. Simonich, Katherine G. Field. 2005. Microplate subtractive hybridization to enrich for Bacteroidales genetic markers for fecal source identification. *Appl. Environ. Microbiol.* **71**:4179-3183

Griffiths, R., A. Whiteley, A. O'Donnell, and M. Bailey. 2000. Rapid Method for Coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA-and rRNA-Based Microbial Community Composition. *Appl. Environ. Microbiol.* **66**: 5488-5491.

Hamilton, M. J., T. Yan, and M. J. Sadowsky. 2006. Development of goose- and duck-specific DNA markers to determine sources of *Escherichia coli* in waterways. *Appl. Environ. Microbiol.* **72**:4012-4019.

Harwood, V., M. Brownell, S. Wang, J. Lepo, R. Ellender, A. Ajidahun, K. Hellein, E. Kennedy, X. Ye, and C. Flood. 2009. Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico. *Water Res.* **43**:4812-4819.

WHO. 2003. Guidelines for safe recreational water environments. Volume 1: coastal and fresh waters. World Health Organization. Geneva, Switzerland.

Hayashi, H., S. Shibata, M. Sakamoto, S. Tomita, and Y. Benno. 2007. *Prevotella copri* sp. nov. and *Prevotella stercorea* sp. nov., isolated from human feces. *Int. J. Syst. Evol. Microbiol.* **57**: 941-946

Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**:2317-2319.

Khatib, L. A., Y. L. Tsai, and B. H. Olson. 2003. A biomarker for the identification of swine fecal pollution in water, using the STII toxin gene from enterotoxigenic *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **63**:231-238.

King, E., D. Bachoon, and K. Gates. 2007. Rapid detection of human fecal contamination in estuarine environments by PCR targeting of *Bifidobacterium adolescentis*. *J. Microbiol. Methods.* **68**:76-81

Kortbaoui R, Locas A, Imbeau M, Payment P, Villemur R. 2009. Universal mitochondrial PCR combined with species-specific dot-blot assay as a source-tracking method of human, bovine, chicken, ovine, and porcine in fecal-contaminated surface water. *Water Res.* **43**:2002-2010.

- Lynch, P., B. Gilpin, L. Sinton, and M. Savill.** 2002. The detection of *Bifidobacterium adolescentis* by colony hybridization as an indicator of human fecal pollution. J. Appl. Microbiol. **92**:526-533.
- Maluquer de Motes, C., Clemente-Casares P, Hundesa A, Martin M, Girones R.** 2004. Detection of bovine and porcine adenoviruses for tracing the source of fecal contamination. Appl. Environ. Microbiol. **70**:1448-1454.
- Martellini A, Payment P, Villemur R.** 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. Water Res. **39**:541-548.
- Marti, R., P. Dabert, and A. Pourcher.** 2009. Pig manure contamination marker selection based on the influence of biological treatment on the dominant fecal microbial groups. Appl. Environ. Microbiol. **75**:4967-4974
- Marti, R., P. Dabert, C. Ziebal, and A.-M. Pourcher.** 2010. Evaluation of *Lactobacillus sobrius*/L. *amylovorus* as a New Microbial Marker of Pig Manure. Appl. Environ. Microbiol. **76**:1456-1461.
- Mieszkina, S., J.-P. Furet, G. Corthier, and M. Gourmelon.** 2009. Estimation of Pig Fecal Contamination in a River Catchment by Real-Time PCR Using Two Pig-Specific Bacteroidales 16S rRNA Genetic Markers. Appl. Environ. Microbiol. **75**:3045-3054
- Nicholson, Fiona A., Simon J. Groves, and Brian J. Chambers.** 2005. Pathogen survival during livestock manure storage and following land application. Elsevier. **96**:135-143.
- Nocker, A., and A. Camper.** 2006. Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. Appl. Environ. Microbiol. **72**: 1997-2004.
- Okabe, S., N. Okayama, O. Savichtcheva, and T. Ito.** 2007. Quantification of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers for assessment of fecal pollution in freshwater. Appl. Microbiol. Biotechnol. **74**:890-901.
- Perrin, A., X. Nassif, and C. Tinsley.** 1999. Identification of regions of the chromosome *Neisseria meningitidis* and *Neisseria gonorrhoeae* which are specific to pathogenic *Neisseria* species. Infect. Immun. **67**:6119-6129
- Scott, Troy M., Joan B. Rose, Tracie M Jenkins, Samuel R. Farrah, and Jerzy Lukasik.** 2002. Microbial source tracking: current methodology and future applications. Appl. Environ. Microbiol. **68**:5796-5803.
- Shanks, Orin C., Jorge W. Santo Domingo, Regina Lamendella, Catherine A. Kelty, and James E. Graham.** 2006. Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. Appl. Environ. Microbiol. **72**:4054-4060.

Simpson, J., J. Santo Domingo, and D. Reasoner. 2002. Microbial Source Tracking: State of Science: Critical Review. *Environ. Sci. Technol.* **24**:5279-5288

Stoeckel, D. and V. Harwood. 2007. Performance, design and analysis in microbial source tracking studies. *Appl. Environ. Microbiol.* **73**:2405-2415.

Tanaka S., M. Yoshida, Y. Murakami, T. Ogiwara, M. Shoji, S. Kobayashi, S. Watanabe, M. Machino, and S. Fujisawa. 2008. The relationship of *Prevotella intermedia*, *Prevotella nigrescens* and *Prevotella melaninogenica* in the supragingival plaque of children, caries and oral malodor. *J Clin Pediatr Dent.* **32(3)**:195-200.

Ufnar, J., D. Ufnar, S. Wang, and R. Ellender. 2007. Development of a swine-specific fecal pollution marker based on host differences in methanogen *mcrA* genes. *Appl. Environ. Microbiol.* **73**:5209-5217.

USEPA. 2000a. Improved enumeration methods for the recreational water quality indicators: enterococci and *Escherichia coli*. U.S. Environmental Protection Agency.

USEPA. 2000b. The quality of our Nation's waters. U.S. Environmental Protection Agency.

USEPA. 2005. Detecting and Mitigating the Environmental Impact of Fecal Pathogens Originating from Confined Animal Feeding Operations: Review. U.S. Environmental Protection Agency.

USEPA. 2007. Census of Agriculture. U.S. Environmental Protection Agency.

USEPA. 2010. Method A: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) Assay. U.S. Environmental Protection Agency.

Weidhaas, J., T. Macbeth, R. Olsen, and V. Harwood. 2011. Quantitative PCR for a Poultry-specific *Brevibacterium* Marker Gene Correlates with Bacterial and Chemical Indicators of Water Pollution in a Watershed Impacted by Land Application of Poultry Litter. *Appl. Environ. Microbiol.* **77**:2094-2102.

Weidhaas, J., T. Macbeth, R. Olsen, M. Sadowsky, D. Norat, and V. Harwood. 2009. Identification of a *Brevibacterium* marker gene specific to poultry litter and development of a quantitative PCR assay. *J. Appl. Microbiol.* **109**: 334-347

APPENDIX

First round of SSH clone sequences:

The primer sequences and the position at which they anneal are presented. DNA sequences are written in the 5' to 3' direction. The reverse primer sequence listed here is the reverse compliment of the sequence seen in Table 3.1.

Clone 1

```

...|...|...|...|...|...|...|...|
 5      15      25      35      45
Clone 1  TGCAGAATTC GCCCTTAGCG TGGTCGCGGC CGAGGTACAA GGCCCGGGAA
Rum F    -----
Rum R    -----

```

```

...|...|...|...|...|...|...|...|
55      65      75      85      95
Clone 1  CGTATTCACC GCGCCATGGC TGATGCGCGA TTACTAGCGA ATCCAGCTTC
Rum F    -----
Rum R    -----

```

```

...|...|...|...|...|...|...|...|
105     115     125     135     145
Clone 1  ATGGAGTCGA GTTGCAGACT CCAATCCGAA CTGAGATAGC TTTTCGAGAT
Rum F    -----
Rum R    -----

```

```

...|...|...|...|...|...|...|...|
155     165     175     185     195
Clone 1  CTGCTCCCTG TCACCAGGTG GCTTCCCTCT GTAGCTACCA TTGTAACACG
Rum F    ----- ---CAGGTG GCTTCCCTCT GTAG.....
Rum R    -----

```

```

...|...|...|...|...|...|...|...|
205     215     225     235     245
Clone 1  TGTGTCGCCC CGGACGTAAG GGCCGTGCTG ATTTGACGTC ATCCCCGCCT
Rum F    -----
Rum R    -----

```

```

...|...|...|...|...|...|...|...|
255     265     275     285     295
Clone 1  TCCTCACACC TTGCGGTGGC AGTCTCGATA GAGTCCCCAG CTTTATCTGT
Rum F    -----
Rum R    -----

```

```

...|...|...|...|...|...|...|...|
305     315     325     335     345
Clone 1  TGGTAACTAT CGACAAGGGT TGGCTCGTT ATGGCACTTA AGCCAACATC
Rum F    -----
Rum R    ----- ---AACATC

```

```

...|...|...|...|...|...|...|...|
355     365     375     385     395
Clone 1  TCACGACACG AGTGACGAC AACCATGCAG CACCTCGACA GTCGTCCCGA
Rum F    -----
Rum R    TCACGACACG AGCT.....

```

```

...|...|...|...|...|...|...|...|
405     415     425     435     445
Clone 1  AGGACCTTTG CATCTCTGCA TCGTTCAACT GCCGTTTCGAG CCCGGGTAAG
Rum F    -----
Rum R    -----

```

```

...|...|...|...|...|...|...|...|

```

```

      455   465   475   485   495
Clone 1 GTTCCTCGCG TATCATCGAA TTAAACCACA TGTCCTCCG CTTGTGCGGG
Rum F .....
Rum R .....

      ...|...|...|...|...|...|...|...|
      505   515   525   535   545
Clone 1 CCCCCGTCAA TTCCTTTGAG TTCAACCTT GCGGTCGTAC CTCGGCMRCR
Rum F .....
Rum R .....

      ...|...|...|...|...|...|...|...|
      555   565   575   585   595
Clone 1 ACMACGCTAA GGGCGAATTC CAGCAGACTG GCGGCMRTTA CTAGTGATC
Rum F .....
Rum R .....

      ...|...|...|...|...|...|...|...|
      605   615   625   635   645
Clone 1 CNAGCTCKGT ACCAAGCTTG GCGTAATCAT GGTCATAGCT GTTCCCNA
Rum F .....
Rum R .....

.

Clone 1 A
Rum F .
Rum R .

```

Clone 2

```

      ...|...|...|...|...|...|...|...|
      5    15    25    35    45
Clone 2 CGCCAGTGTG ATGGATATCT GCAGAATTCG CCCTTTCGAG CGGCCGCCCCG
Rik F -----
Rik R -----

      ...|...|...|...|...|...|...|...|
      55   65   75   85   95
Clone 2 GGCAGGTACA AGGCCCGGGA ACGTATTCAC CGCAGCATGG CTGATCTGCC
Rik F -----
Rik R -----

      ...|...|...|...|...|...|...|...|
      105  115  125  135  145
Clone 2 ATTACTAGCG AATCCATCTT CACGAAGTCG GGTTGCAGAC TCCGATCCGA
Rik F -----GCAGAC TCCGATCCGA
Rik R -----

      ...|...|...|...|...|...|...|...|
      155  165  175  185  195
Clone 2 ACTGAGACAG ACTTTCGAGA TTCGCATTGC ATCTCTGCAT AGCTGCCCTC
Rik F ACTG.....
Rik R -----

      ...|...|...|...|...|...|...|...|
      205  215  225  235  245
Clone 2 TGTATCTGCC ATTGTAGCAC GTGTGTCGCC CCGAACGTAA GGGCCGTGCT
Rik F .....
Rik R -----

      ...|...|...|...|...|...|...|...|
      255  265  275  285  295
Clone 2 GATTTGACGT CATCCCCACC TTCTCTCAG CTTGCGCCGG CAGTCCACAT
Rik F .....
Rik R -----

      ...|...|...|...|...|...|...|...|
      305  315  325  335  345
Clone 2 AGAGTCCCCA CCTCAACGCG CTGGTAACTA TGTGCAGGGG TTGCGCTCGT

```



```

          905   915   925   935   945
Clone 2   GCGCTCTTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCGTTCGGN
Rik F     .....
Rik R     .....

          .....|.....|.....|.....|.....|.....|.....|
          955   965   975   985   995
Clone 2   TGCGGNNAGN GGTATCAGCT TCACTCAAGT CGTAATACGT ATCACAGATC
Rik F     .....
Rik R     .....

          .....|.....|.....|.....|.....|.....|.....|
          1005  1015  1025  1035  1045
Clone 2   AGGGGATACG NAGAAGACAT GTGACANGCA GCAAGGTCAT GACGTAANGG
Rik F     .....
Rik R     .....

          .....|.....|.....|
          1055  1065
Clone 2   CGGTTGCNGN NTNNTNNNN
Rik F     .....
Rik R     .....

```

Clone 3

```

          .....|.....|.....|.....|.....|.....|.....|
          5    15    25    35    45
Clone 3   GAGGAACCTT ACCTGGGTTT GACATACACA CTGATTATCA GGAGACTGGT
Trep F    --GGAACCTT ACCTGGGTTT GACA.....
Trep R    -----

          .....|.....|.....|.....|.....|.....|.....|
          55   65   75   85   95
Clone 3   AAGCGTAGCA ATACGAGTGT GAACAGGTGC TGCATGGCTG TCGTCAGCTC
Trep F    .....
Trep R    -----

          .....|.....|.....|.....|.....|.....|.....|
          105  115  125  135  145
Clone 3   GTGYGTGAN GTGTTGGGTT AAGTCCC GCA ACGRGCGCAA CCCCTACNGN
Trep F    .....
Trep R    -----

          .....|.....|.....|.....|.....|.....|.....|
          155  165  175  185  195
Clone 3   CAKTACNAA CANGTAAWGC TGAGRACTCT GGCGRARCTG CCGGTGACAA
Trep F    .....
Trep R    -----ACAA

          .....|.....|.....|.....|.....|.....|.....|
          205  215  225  235  245
Clone 3   ACCGGAGGAA GGTGGGGATG ACGTCAAGTC ATCATGGCCC TTAYGWCCAG
Trep F    .....
Trep R    ACCGGAGGAA GGTGG.....

          .....|.....|.....|.....|.....|.....|.....|
          255  265  275  285  295
Clone 3   GGCTACACAC GTGCTACAAT GGTCGGTACC TCGGCMRCRA CMACGCTAAG
Trep F    .....
Trep R    .....

          .....|.....|.....|.....|.....|.....|.....|
          305  315  325  335  345
Clone 3   GGCGAATTCC RSCACACTGG CGGCNNTTNC TRGKGGATCC NAGCTCSKTA
Trep F    .....
Trep R    .....

          .....|.....|.....|.....|.....|.....|.....|
          355  365  375  385

```

Clone 3 CCAAGCTTGG CKTAATCATG GTCRTAGCTG TTTYNNAAAN
Trep F
Trep R

Clone 4

.....|.....|.....|.....|.....|.....|.....|.....|
5 15 25 35 45
Clone 4 CGAGCGGCCG CCAGTGTGAT GGATATCTGC AGAATTCGCC CTTAGCGTGG
Bac F -----CTGC AGAATTCGCC CTTAGC....
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
55 65 75 85 95
Clone 4 TCGCGGCCGA GGTACTCCCC AGGTGGATAA CTTATCGCTT TCGCTTGGCC
Bac F
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
105 115 125 135 145
Clone 4 ACCGACTGTC TATCGCCGAC GGTTAGTTAT CATCGTTTAC TCGGTGGACT
Bac F
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
155 165 175 185 195
Clone 4 ACCAGGGTAT CTAATCCTGT TTGATCCCCA CGCTTTCGTG CCTCAGCGTC
Bac F
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
205 215 225 235 245
Clone 4 AATGACGGAT TCGTGAGATG CCTTCGCAAT CGGTGTTCTG AGTAATATCT
Bac F
Bac R -----GGAT TCGTGAGATG CCTTCG.... ..

.....|.....|.....|.....|.....|.....|.....|.....|
255 265 275 285 295
Clone 4 ATGCATTTC ACGCTACACT ACTATTTCCT CCCACGGCAT CCGTATTCCA
Bac F
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
305 315 325 335 345
Clone 4 GCGCAACAGT ATCAATGGCA GTTCCGGTGT TGGGCACCGG GCTTTCACCA
Bac F
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
355 365 375 385 395
Clone 4 CTGACTTGCT ACACAGCCTA CGCACCCCTT AAACCCAATA AATCCGGATA
Bac F
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
405 415 425 435 445
Clone 4 ACGCTTGCACT CCTCCGTATT ACCGCGGCTG CTGGCACGGA GTTAGCCGAT
Bac F
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
455 465 475 485 495
Clone 4 GCTTATTCGT GATGTACCTC GGCMRCRACM ACGCTAAGGG CGAATTCAG
Bac F
Bac R -----

.....|.....|.....|.....|.....|.....|.....|.....|
505 515 525 535 545

...|...|...|...|...|...
 555 565 575
 Clone 4 TAATCATGGT CATAGCTGTT TCNNAAGN
 Bac F
 Bac R

.....|||||||
5 15 25 35 45
Clone 5 CCGCCAGTGT GATGGATATC TGCAGAATTC GCCCTTAGCG TGGTCGCGGC
Ach F -----
Ach R -----

Clone 5 CAAGCGGTGG ATTATGTGGT TTAATTCGAA GCAACGCGAA GAACCTTACC
Ach F -----
Ach R -----

...|...|...|...|...|...|...|...|...|...|
205 215 225 235 245

Clone 5 GGAAAGTAGA GACAGGTGGT GCATGGTTGT CGTCAGCTCG TGTCGTGAGA
Ach F -----
Ach R -----

.....||||||
305 315 325 335 345
Clone 5 CAAGAGCACT CTAGCGAGAC TGCCGTTGAC AAAACGGAGG AAGGTGGGGA
Ach F -----CT CTAGCGAGAC TGCCGTTGA.
Ach R -----

.....|.....|.....|.....|.....|.....|.....|.....|
 405 415 425 435 445

Clone 5	ATGGCTGGAA CAAAGAGAAG CGAAGCAGTG ATGTGAAGCA AAACTCAGAA
Ach F	
Ach R	-----

48

```

Clone 5   AAACAGTCTC AGTTCGGATT GGAGTCTGCA ACTCGACTCC ATGAAGTCGG
Ach F     .....
Ach R     -----TCC ATGAAGTCGG

      ...|...|...|...|...|...|...|...|
      505   515   525   535   545
Clone 5   AATCGCTAGT AATCGCAGAT CAGCATGCTG CGGTGAATAC GTTCTCGGGG
Ach F     .....
Ach R     AATCGCTA.. .....

      ...|...|...|...|...|...|...|...|
      555   565   575   585   595
Clone 5   TTTGTACCTC GGCMRCNACM ACGCTAAGGG CGAATTCCAG CACACTGGCG
Ach F     .....
Ach R     .....

      ...|...|...|...|...|...|...|...|
      605   615   625   635   645
Clone 5   GCMGTTAYTA GTGGATCCGA GCTCGGTACC AAGCTTGGCG TAATCATGGT
Ach F     .....
Ach R     .....

      ...|...|...|...|...|...|...|...|
      655   665
Clone 5   CATAGCTGTT TCCTGAGAAA A
Ach F     .....
Ach R     .....

```

Second round of SSH clone sequences:

The primer sequences and the position at which they anneal are presented. DNA sequences are written in the 5' to 3' direction. The reverse primer sequence listed here is the reverse complement of the sequence seen in Table 3.4.

Clone A1

```

      ...|...|...|...|...|...|...|...|
      5    15    25    35    45
Clone A1  TGCTCGAGCG GCCGCCAGTG TGATGGATAT CTGCAGAATT CGCCCTTAGC
A1 F     .....
A1 R     -----

      ...|...|...|...|...|...|...|...|
      55   65   75   85   95
Clone A1  GTGGTCGCGG CCGAGGTACC GACCATTGTA GCACGTGTGT AGCCCTAGAC
A1 F     -----TTGTA GCACGTGTGT AGCCC....
A1 R     -----

      ...|...|...|...|...|...|...|...|
      105  115  125  135  145
Clone A1  ATAAGGGGCA TGATGATTG ACGTCATCCC CACCTTCCTC CCGGTTAACC
A1 F     .....
A1 R     -----

      ...|...|...|...|...|...|...|...|
      155  165  175  185  195
Clone A1  CGGGCAGTCT CGTAGAGTG CCCAACTTAA TGATGGCAAC TAACAATAAG
A1 F     .....
A1 R     -----

      ...|...|...|...|...|...|...|...|
      205  215  225  235  245
Clone A1  GGTTGCGCTC GTTGCGGGAC TTAACCCAAC ATCTCACGAC ACGAGCTGAC
A1 F     .....
A1 R     -----

```


...|...|...|...|...|...|...|...|...|...|
 255 265 275 285 295
 Clone A1 GACAAACCATG CACCACCTGT CTCTGAATTC CCCGAAGGGC ACCCCAGCCA
 A1 F
 A1 R -----

...|...|...|...|...|...|...|...|...|...|
 305 315 325 335 345
 Clone A1 TTACGCCGGG TTCTCAGGAT GTCAAGTTCT GGTAAGGTTT TCGCGTTGC
 A1 F
 A1 R -----

...|...|...|...|...|...|...|...|...|...|
 355 365 375 385 395
 Clone A1 GTCGAATTAA ACCACATACT CCACTGCTTG TGCGGGCTCC CGTCAATTCC
 A1 F
 A1 R -----ATACT CCACTGCTTG TCGGG.....

...|...|...|...|...|...|...|...|...|...|
 405 415 425 435 445
 Clone A1 TTTGAGTTTC AACCTTGCGG CCGTACCTCG GCCGCGACCA CGCTAAGGGC
 A1 F
 A1 R

...|...|...|...|...|...|...|...|...|...|
 455 465 475
 Clone A1 GAATTCCAGC AACTGGCGG CCGTTA
 A1 F
 A1 R

Clone A2

...|...|...|...|...|...|...|...|...|...|
 5 15 25 35 45
 Clone A2 GATGGATATC TGCAGAATTC GCCCTTAGCG TGGTCGCGGC CGAGGTACTA
 A2 F
 A2 R -----

...|...|...|...|...|...|...|...|...|...|
 55 65 75 85 95
 Clone A2 TAAGACGCAA GCCC GCGAGG GGGAGCAAAA CTGGAAAACC GATCTCAGTT
 A2 F -----AGG GGGAGCAAAA CTGGAAAA.. ..
 A2 R -----

...|...|...|...|...|...|...|...|...|...|
 105 115 125 135 145
 Clone A2 CGGATTGTAG GCTGAACTC GCCTACATGA GGCTGGAGTT GCTAGTAATC
 A2 F
 A2 R -----

...|...|...|...|...|...|...|...|...|...|
 155 165 175 185 195
 Clone A2 GCGAATCAGA ATGTCGCGGT GAATACGTTT CCGGGCCTTG TACCTCGGCC
 A2 F
 A2 R -----AATACGTTT CCGGGCCTTG

...|...|...|...|...|...|...|...|...|...|
 205 215 225
 Clone A2 GCGACCACGC TAAGGGCGAA TTCCAGCACA C
 A2 F
 A2 R

Clone A3

```
...|...|...|...|...|...|...|...|...|...|
  5      15     25     35     45
Clone A3  CTCGAGCGGC CGCCAGTGTG ATGGATATCT GCAGAATTCG CCCTTAGCGT
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
  55     65     75     85     95
Clone A3  GGTCGCGGCC GAGGTACGTG TGTAGCCCAA GTCATAAAGG GCATGATGAT
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 105     115     125     135     145
Clone A3  TTGACGTCAT CCCCACCTTC CTCCGTTTTC TCAACGGCAG TCCTGCCAGA
A3 F      -----AT CCCCACCTTC CTCCGTT...
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 155     165     175     185     195
Clone A3  GTCCTCTTGC GTAGTAAGT ACAGTAAGGG TTGCGCTCGT TCGGGGACTT
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 205     215     225     235     245
Clone A3  AACCCAACAT CTCACGGCAC GAGCTGACGA CAACCATGCA GCACCTTCAC
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 255     265     275     285     295
Clone A3  AGGAGTCCCG AAGGACCTCA ACATCTCTGT ATCGTTCTCC TGCAATTCAA
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 305     315     325     335     345
Clone A3  GCCCGGGTAA GGTTCCTCGC GTATCATCGA ATTAAACCAC ATGTTCTCTC
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 355     365     375     385     395
Clone A3  GCTTGTGCGG GCCCCCGTCA ATTCCTTTGA GTTTTACCGT TGCCGGCGTA
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 405     415     425     435     445
Clone A3  CTCCCCAGGT GGGATGCTTA ACGCTTTCGC TTGGCCGCGT ACCCAGGGGC
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 455     465     475     485     495
Clone A3  ACACAGCGAG CATCCATCGT TTACCGTGCG GACTACCAGG GTATCTAATC
A3 F      -----
A3 R      -----
```

```
...|...|...|...|...|...|...|...|...|...|
 505     515     525     535     545
Clone A3  CTGTTTCGATA CCCGCACTTT CGAGCTTCAG CGTCAGTTAC GCTACTGCAA
A3 F      -----
A3 R      -----
```

...|...| ...|...| ...|...| ...|...| ...|...|
 555 565 575 585 595
 Clone A3 GCTGCCTTCG CAATCGGAGT TCTTCGTCAT ATCTAAGCAT TTCACCGCTA
 A3 F
 A3 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 605 615 625 635 645
 Clone A3 CACGACGAAT TCCGCCTGCA TTCCGCGCAC TCAAGTCTCC CAGTTCGCGC
 A3 F
 A3 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 655 665 675 685 695
 Clone A3 TGCACCTCTG TCGTTGAGCG ACAGTATTTC ACAACACGCT TAAAAGGCGG
 A3 F
 A3 R -----GCG ACAGTATTTC ACAACACG..

...|...| ...|...| ...|...| ...|...| ...|...|
 705 715 725 735 745
 Clone A3 CCTACGCTCC CTTAAACCC AATAAATCCG GATAACGCCC GGACCTTCCG
 A3 F
 A3 R

...|...| ...|...| ...|...| ...|...| ...|...|
 755 765 775 785 795
 Clone A3 TATTACCGCG GCTGCTGGCA CGGAATTAGC CGGTCCTTAT TCATAAGGTA
 A3 F
 A3 R

...|...| ...|...| ...|...| ...|...|
 805 815 825 835
 Clone A3 CCTCGGCCGC GACCACGCTA AGGGCGAATT CCAGCACAC
 A3 F
 A3 R

Clone A5

...|...| ...|...| ...|...| ...|...| ...|...|
 5 15 25 35 45
 Clone A5 GCATGCTCGA GCGGCCGCCA GTGTGATGGA TATCTGCAGA ATTCGCCCTT
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 55 65 75 85 95
 Clone A5 AGCGTGGTCG CGGCCGAGGT ACGCACAACG TGGGCGGAAT TCGTGGTGTA
 A5 F -----TGGTGTA
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 105 115 125 135 145
 Clone A5 GCGGTGAAAT GCTTAGATAT CACGAAGAAC TCCGATTGCG AAGGCAGCTC
 A5 F GCGGTGAAAT GCT.....
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 155 165 175 185 195
 Clone A5 ACGGGAGCGC AACTGACGCT GAAGCTCGAA AGTGCGGGTA TCGAACAGGA
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 205 215 225 235 245
 Clone A5 TTAGATACCC TGGTAGTCCG CACGGTAAAC GATGGATGCC CGCTGTTGGC
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 255 265 275 285 295
 Clone A5 CTGAATAGGT CAGCGGCCAA GCGAAAGCAT TAAGCATCCC ACCTGGGGAG
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 305 315 325 335 345
 Clone A5 TACGCCGGCA ACGGTGAAAC TCAAAGGAAT TGACGGGGGC CCGCACAAGC
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 355 365 375 385 395
 Clone A5 GGAGGAACAT GTGGTTTAAT TCGATGATAC GCGAGGAACC TTACCCGGGC
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 405 415 425 435 445
 Clone A5 TTGAATTGCA GAGGAAGGAT TTGGAGACAA TGACGCCCTT CGGGGCCTCT
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 455 465 475 485 495
 Clone A5 GTGAAGGTGC TGCATGGTTG TCGTCAGCTC GTGCCGTGAG GTGTCGGCTT
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 505 515 525 535 545
 Clone A5 AAGCGCCATA ACGAGCGCAA CCCCTCTCCT TAGTTGCCAT CAGGTCATGC
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 555 565 575 585 595
 Clone A5 TGGGCACTCT GGGGACACTG CCACCGTAAG GTGTGAGGAA GGTGGGGATG
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 605 615 625 635 645
 Clone A5 ACGTCAAATC AGCACGGCCC TTACGTCCGG GGCTACACAC GTGTTACAAT
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 655 665 675 685 695
 Clone A5 GGCAGGTACA GAGAGATGGT GTTCTGCAAA GCGCATCTAA TCCTTAAAGC
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 705 715 725 735 745
 Clone A5 CTGTCTCAGT TCGACTGGG GTCTGCAACC CGACCCACG AAGCTGGATT
 A5 F
 A5 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 755 765 775 785 795
 Clone A5 CGCTAGTAAT CGCGCATCAG CCATGGCGCG GTGAATACGT TCCCGGGCCT
 A5 F
 A5 R ----GTAAT CGCGCATCAG CCATG.....

...|...| ...|...| ..
 805 815
 Clone A5 TGTACCTCGG CCGCGACCAC GC
 A5 F

A5 R

Clone A8

```

...|...|...|...|...|...|...|...|
 5      15      25      35      45
Clone A8  TGCTCGAGCG GCCGCCAGTG TGATGGATAT CTGCAGAATT CGCCCTTAGC
A8 F      -----
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
55      65      75      85      95
Clone A8  GTGGTCGCGG CCGAGGTACT CCCAGGTGG ATTACTTATT GTGTAACTG
A8 F      -----GTTAACTG
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
105     115     125     135     145
Clone A8  CGGCACTGAA GGGGTCAATC CCCCAACACC TAGTAACCAT CGTTTACGGT
A8 F      CGGCACTGAA GG.....
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
155     165     175     185     195
Clone A8  GTGGACTACC AGGGTATCTA ATCCTGTTTG CTACCCACAC TTTCGTGCCT
A8 F      -----
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
205     215     225     235     245
Clone A8  CAGTGTCAGT TACAGTCCAG AAAGCCGCCT TCGCTACTGG TGTTCTCCTT
A8 F      -----
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
255     265     275     285     295
Clone A8  AATATCTACG CATTTACCG CTACACTAGG AATTCCACTT TCCTCTCCTG
A8 F      -----
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
305     315     325     335     345
Clone A8  CACTCAAGTT TCCCAGTTT AAGAGCTTAC TACGGTTGAG CCGTAGCCTT
A8 F      -----
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
355     365     375     385     395
Clone A8  TCACTCCTGA CTTAAGAAAC CACCTACGCA CCCTTTACGC CCAGTAAATC
A8 F      -----
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
405     415     425     435     445
Clone A8  CGGATAACGC TAGCCCCCTA CGTATTACCG CGGCTGCTGG CACGTAGTTA
A8 F      -----
A8 R      ---ATAACGC TAGCCCCCTA CGT.....
```

```

...|...|...|...|...|...|...|...|
455     465     475     485     495
Clone A8  GCCGGGGCTT CCTCCTCAAG TACCTCGGCC GCGACCACGC TAAGGGCGAA
A8 F      -----
A8 R      -----
```

```

...|...|...|...|...|...|...|...|
505     515
Clone A8  TTCCAGCACA CTGGCGGCCG T
A8 F      -----
A8 R      -----
```

Clone A10

```
...|...|...|...|...|...|...|...|
  5    15   25   35   45
Clone A10  GTGATGGATA TCTGCAGAAT TCGCCCTTTC GAGCGGCCGC CCGGGCAGGT
A10 F      -----
A10 R      -----
```

```
...|...|...|...|...|...|...|...|
  55   65   75   85   95
Clone A10  ACTCCCCAGG TGGATTACTT ATTGTGTAA CTGCGGCACT GAAGGGGTCA
A10 F      ACTCCCCAGG TGGATTACTT .....
A10 R      -----
```

```
...|...|...|...|...|...|...|...|
 105   115  125  135  145
Clone A10  ATCCTCCAAC ACCTAGTAAT CATCGTTTAC GGCATGGACT ACCAGGGTAT
A10 F      .....
A10 R      -----
```

```
...|...|...|...|...|...|...|...|
 155   165  175  185  195
Clone A10  CTAATCCTGT TTGCTACCCA TGCTTTCGAG CCCCAGCGTC AGTTGGTGCC
A10 F      .....
A10 R      -----
```

```
...|...|...|...|...|...|...|...|
 205   215  225  235  245
Clone A10  CAGTAGGTCG CCTTCGCCAC TGGTGTTCTT CCCGATATCT ACGCATTCCA
A10 F      .....
A10 R      -----
```

```
...|...|...|...|...|...|...|...|
 255   265  275  285  295
Clone A10  CCGCTACACC GGAATTCCA CCTACCTCTG CACTACTCAA GGCCAGCAGT
A10 F      .....
A10 R      -----
```

```
...|...|...|...|...|...|...|...|
 305   315  325  335  345
Clone A10  TTTGAAAGCA ATTTATGGGT TGAGCCCATG GTTTTCACTT CCAACTTGCC
A10 F      .....
A10 R      -----TTTCACTT CCAACTTGCC
```

```
...|...|...|...|...|...|...|...|
 355   365  375  385  395
Clone A10  GGTCCGCCTG CGCTCCCTTT ACACCCAGTA ATTCCGGACA ACGCTTGACAC
A10 F      .....
A10 R      GG.....
```

```
...|...|...|...|...|...|...|...|
 405   415  425  435  445
Clone A10  CCTACGTTT ACCGCGGCTG CTGGCACGTA GTTAGCCGGT GCTTTCTGT
A10 F      .....
A10 R      -----
```

```
...|...|...|...|...|...|...|...|
 455   465  475  485  495
Clone A10  TAGGTACCTC GGCCGCGACC ACGCTAAGGG CGAATTCCAG CACTACTG
A10 F      .....
A10 R      -----
```

Clone A13

...|...| ...|...| ...|...| ...|...| ...|...|
 5 15 25 35 45
 Clone A13 CTAGATGCAT GCTCGAGCGG CCGCCAGTGT GATGGATATC TGCAGAATTC
 A13 F -----
 A13 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 55 65 75 85 95
 Clone A13 GCCCTTAGCG TGGTCGCGGC CGAGGTACGT GTGTAGCCCA AGTCATAAAG
 A13 F -----
 A13 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 105 115 125 135 145
 Clone A13 GGCATGATGA TTTGACGTCA TCCCCACCTT CCTCCGTTTT GTCAACGGCA
 A13 F -----
 A13 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 155 165 175 185 195
 Clone A13 GTCCTGCCAG AGTCCTCTTG CGTAGTAACT GACAGTAAGG GTTGCCTCG
 A13 F ----GCCAG AGTCCTCTTG CGTAG.....
 A13 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 205 215 225 235 245
 Clone A13 TTGCGGGACT TAACCCAACA TCTCACGACA CGAGCTGACG ACAACCATGC
 A13 F -----
 A13 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 255 265 275 285 295
 Clone A13 ACCACCTGTC TCTGCGTCCC GAAGGAAACA CATATTTCTA TGTGCGTCGC
 A13 F -----
 A13 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 305 315 325 335 345
 Clone A13 AGGATGTCAA GACTTGGTAA GGTCTTCGC GTTGCCTCGA ATTAAACCAC
 A13 F -----
 A13 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 355 365 375 385 395
 Clone A13 ATACTCCACT GCTTGTGCGG GCCCCGTCA ATTCCTTTGA GTTTCAACCT
 A13 F -----
 A13 R -----GA GTTTCAACCT

...|...| ...|...| ...|...| ...|...| ...|...|
 405 415 425 435
 Clone A13 TGCGGTCGTA CCTCGGCCGC GACCACGCTA AGGGC
 A13 F -----
 A13 R TGCGGTCG..

Clone A15

...|...| ...|...| ...|...| ...|...| ...|...|
 5 15 25 35 45
 Clone A15 TGMATGCTCG AGCGGCCGCC AGTGTGATGG ATATCTGCAG AATTGCGCCT
 A15 F -----
 A15 R -----

...|...| ...|...| ...|...| ...|...| ...|...|
 55 65 75 85 95
 Clone A15 TAGCGTGGTC GCGGCCGAGG TACGCACAAA GTGGGCGGAA TTCGTGGTGT
 A15 F -----TGGTGT
 A15 R -----

...|...| ...|...| ...|...| ...|...| ...|...|

```

      105   115   125   135   145
Clone A15  AGCGGTGAAA TGCTTAGATA TCACGAAGAA CTCCGATTGC GAAGGCAGCT
A15 F      AGCGGTGAAA TGCT.....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      155   165   175   185   195
Clone A15  CACTGGAGCG CAACTGACGC TGAAGCTCGA AAGTGCGGGT ATCGAACAGG
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      205   215   225   235   245
Clone A15  ATTAGATACC CTGGTAGTCC GCACGGTAAA CGATGGATGC CCGCTGTTGG
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      255   265   275   285   295
Clone A15  TTCACCTGAA TCAGCGGCCA AGCGAAAGCA TTAAGCATCC CACCTGGGGA
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      305   315   325   335   345
Clone A15  GTACGCCGCG AACGGTGAAA CTCAAAGGAA TTGACGGGGG CCCGCACAAG
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      355   365   375   385   395
Clone A15  CGGAGGAACA TGTGGTTTAA TTCGATGATA CGCGAGGAAC CTTACCCGGG
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      405   415   425   435   445
Clone A15  CTTGAATTGC AGAGGAAGGA TTTGGAGACA ATGACGCCCT TCGGGGCCTC
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      455   465   475   485   495
Clone A15  TGTGAAGGTG CTGCATGGTT GTCGTCAGCT CGTGCCGTGA GGTGTCGGCT
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      505   515   525   535   545
Clone A15  TAAGTGCCAT AACGAGCGCA ACCCCTCTCC CTAGTTGCCA TCAGTTCAAG
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      555   565   575   585   595
Clone A15  CTGGGCACTC TGGGGACACT GCCACCGTAA GGTGTGAGGA AGGTGGGGAT
A15 F      .....
A15 R      -----

      ...|...| ...|...| ...|...| ...|...| ...|...|
      605   615   625   635   645
Clone A15  GACGTCAAAT CAGCACGGCC CTTACGTCCG GGGCTACACA CGTGTTACAA
A15 F      .....
A15 R      -----G GGGCTACACA CGTGTTACA.

      ...|...| ...|...| ...|...| ...|...|
      655   665   675   685
Clone A15  TGGCAGGTAC CTCGGCCGCG ACCACGCTAA GGGCG
A15 F      .....
A15 R      .....

```


Clone A16

```
...|...|...|...|...|...|...|...|
  5    15   25   35   45
Clone A16  TCTAGATGCA TGCTCGAGCG GCCGCCAGTG TGATGGATAT CTGCAGAATT
A16 F      -----
A16 R      -----
```

```
...|...|...|...|...|...|...|...|
  55   65   75   85   95
Clone A16  CGCCCTTAGC GTGGTCGCGG CCGAGGTACC ACCCATTGTA GCACGTGTGT
A16 F      -----A GCACGTGTGT
A16 R      -----
```

```
...|...|...|...|...|...|...|...|
 105   115  125  135  145
Clone A16  AGCCCTAAGC ATAAGGGGCA TGATGATTG ACGTCATCCC CACCTTCCTC
A16 F      AGCCCTAAG. ....
A16 R      -----
```

```
...|...|...|...|...|...|...|...|
 155   165  175  185  195
Clone A16  CAGGTTATCC CTGGCAGTCC CTCTAGAGTG CCCAACTTAA TGATGGCAAC
A16 F      .....
A16 R      -----
```

```
...|...|...|...|...|...|...|...|
 205   215  225  235  245
Clone A16  TAAAGGCAAG GGTGCGCTC GTTGCGGGAC TTAACCCAAC ATCTCACGAC
A16 F      .....
A16 R      -----
```

```
...|...|...|...|...|...|...|...|
 255   265  275  285  295
Clone A16  ACGAGCTGAC GACAACCATG CACCACCTGT CACTTCTGTC CCCGAAGGGA
A16 F      .....
A16 R      -----
```

```
...|...|...|...|...|...|...|...|
 305   315  325  335  345
Clone A16  AAGATGCGAT TAGGCATCGG TCAAAAGGAT GTCAAGCTTA GGTAAGGTTC
A16 F      .....
A16 R      -----
```

```
...|...|...|...|...|...|...|...|
 355   365  375  385  395
Clone A16  TTCGCGTTGC TTCRAATTAA ACCACATGCT CCGCTACTTG TCGGGGTCCC
A16 F      .....
A16 R      -----A ACCACATGCT CCGCTACTT. ....
```

```
...|...|...|...|...|...|...|...|
 405   415  425  435  445
Clone A16  CGTCAATTCC TTTGAGTTTC ACTCTTGCGA GCGTACCTCG GCCGCGACCA
A16 F      .....
A16 R      -----
```

```
...|...|
 455
Clone A16  CGCTAAGGGC
A16 F      .....
A16 R      -----
```

Clone A18

```
...|...|...|...|...|...|...|...|
  5    15   25   35   45
Clone A18  AGATGCATGC TCGAGCGGCC GCCAGTGTGA TGGATATCTG CAGAATTCGC
A18 F      -----
```

A18 R -----

...|...|...|...|...|...|...|...|...|...|
55 65 75 85 95
Clone A18 CCTTAGCGTG GTCGCGGCCG AGGTACGGTC GCAAGATTAA AACTCAAAGG
A18 F -----
A18 R -----

...|...|...|...|...|...|...|...|...|...|
105 115 125 135 145
Clone A18 AATTGACGGG GACCCGCACA AGCAGCGGAG CATGTGGTTT AATTCGAAGC
A18 F -----
A18 R -----

...|...|...|...|...|...|...|...|...|...|
155 165 175 185 195
Clone A18 AACGCGAAGA ACCTTACCTA GACTTGACAT CTCCTGCATT ACTCTTAATC
A18 F -----
A18 R -----

...|...|...|...|...|...|...|...|...|...|
205 215 225 235 245
Clone A18 GAGGAAGTCT CTTCGGAGAC AGGATGACAG GTGGTGCATG GTTGTCGTCA
A18 F -----C AGGATGACAG GTGGTGCAT.
A18 R -----

...|...|...|...|...|...|...|...|...|...|
255 265 275 285 295
Clone A18 GCTCGTGTCTG TGAGATGTTG GGTAAAGTCC CGCAACGAGC GCAACCCTTA
A18 F
A18 R -----

...|...|...|...|...|...|...|...|...|...|
305 315 325 335 345
Clone A18 TTGTAGTTG CCATCATTAA GTTGGGCACT CTAGCGAGAC TGCCCGGGTT
A18 F
A18 R -----

...|...|...|...|...|...|...|...|...|...|
355 365 375 385 395
Clone A18 AACCAGGAGG AAGTGGGGA TGACGTCAAA TCATCATGCC CCTTATGTCT
A18 F
A18 R -----

...|...|...|...|...|...|...|...|...|...|
405 415 425 435 445
Clone A18 AGGGCTACAC ACGTGCTACA ATGGTCGGTA CAATAAGACG CAAGCCCGCG
A18 F
A18 R -----

...|...|...|...|...|...|...|...|...|...|
455 465 475 485 495
Clone A18 AGGGGGAGCA AACTGGAAA ACCGATCTCA GTTCGGATTG TAGGCTGAAA
A18 F
A18 R -----A

...|...|...|...|...|...|...|...|...|...|
505 515 525 535 545
Clone A18 CTCGCCTACA TGAAGCTGGA GTTGCTAGTA ATCGCGAATC AGCATGTGCG
A18 F
A18 R CTCGCCTACA TGAAGCTGG.

...|...|...|...|...|...|...|...|...|...|
555 565 575 585 595
Clone A18 GGTGAATACG TTCCCGGGCC TTGTACACAC CGCCCGTCAC ACCATGAGAG
A18 F
A18 R -----

...|...|...|...|...|...|...|...|...|...|
605 615 625 635 645

Clone A18 TTGGCAATAC CCAAAGTACC TCGGCCGCGA CCACGCTAAG GGCGAATTCC
A18 F
A18 R

...

Clone A18 AGC
A18 F ...
A18 R ...

Clone A19

.....|.....|.....|.....|.....|.....|.....|.....|
5 15 25 35 45
Clone A19 TGCATGCTCG AGCGGCCGCC AGTGTGATGG ATATCTGCAG AATTCGCCCT
A19 F
A19 R

.....|.....|.....|.....|.....|.....|.....|.....|
55 65 75 85 95
Clone A19 TAGCGTGGTC GCGGCCGAGG TACCTAAGGA GGAAGCCACG GCTAACTACG
A19 F
A19 R

.....|.....|.....|.....|.....|.....|.....|.....|
105 115 125 135 145
Clone A19 TGCCAGCAGC CGCGGTAATA CGTAGGTGGC AAGCGTTGTC CGGATTACT
A19 F -----GCGGTAATA CGTAGGTGGC A.....
A19 R

.....|.....|.....|.....|.....|.....|.....|.....|
155 165 175 185 195
Clone A19 GGGCGTAAAG GGAGCGTAGG TGGATATTTA AGTGGGATGT GAAATACCCG
A19 F
A19 R

.....|.....|.....|.....|.....|.....|.....|.....|
205 215 225 235 245
Clone A19 GGCTTAACCT GGGTGCTGCA TTCCAAACTG GATATCTAGA GTGCAGGAGA
A19 F
A19 R

.....|.....|.....|.....|.....|.....|.....|.....|
255 265 275 285 295
Clone A19 GGAAGGGAGA ATTCCTAGTG TAGCGGTGAA ATGCGTAGAG ATTAGGAAGA
A19 F
A19 R -----AGTG TAGCGGTGAA ATGCGT....

.....|.....|.....|.....|.....|.....|.....|.....|
305 315 325 335 345
Clone A19 ATACCAGTGG CGAAGGCGCC TTTCTGGACT GTAAGTGACA CTGAGGCTCG
A19 F
A19 R

.....|.....|.....|.....|.....|.....|.....|.....|
355 365 375 385 395
Clone A19 AAAGCGTGGG GAGCGAACAG GATTAGATAC CCTGGTAGTC CACGCCGTAA
A19 F
A19 R

.....|.....|.....|.....|.....|.....|.....|.....|
405 415 425 435 445
Clone A19 ACGATGAATA CTAGGTGTAG GGGTTGTCAT GACCTCTGTG CCGCCGCTAA
A19 F
A19 R

.....|.....|.....|.....|.....|.....|.....|.....|
455 465 475 485 495
Clone A19 CGCATTAAGT ATTCCGCCTG GGGAGTACCT CGGCCGCGAC CACGCTAAGG
A19 F

A19 R

...

Clone A19 GCG

A19 F ...

A19 R ...

Clone A20

```

...|...|...|...|...|...|...|...|
 5      15     25     35     45
Clone A20  TGCATGCTCG AGCGGCCGCC AGTGTGATGG ATATCTGCAG AATTCGCCCT
A20 F      -----
A20 R      -----
```

```

...|...|...|...|...|...|...|...|
55      65     75     85     95
Clone A20  TAGCGTGGTC GCGGCCGAGG TACTTGAGGA GGAAGCCCCG GCTAACTACG
A20 F      -----G TACTTGAGGA GGAAGCCCC. ....
A20 R      -----
```

```

...|...|...|...|...|...|...|...|
105     115    125    135    145
Clone A20  TGCCAGCAGC CGCGGTAATA CGTAGGGGGC TAGCGTTATC CGGATTACT
A20 F      -----
A20 R      -----
```

```

...|...|...|...|...|...|...|...|
155     165    175    185    195
Clone A20  GGGCGTAAAG GGTGCGTAGG TGGTCTCTTA AGTCAGGAGT GAAAGGCTAC
A20 F      -----
A20 R      -----
```

```

...|...|...|...|...|...|...|...|
205     215    225    235    245
Clone A20  GGCTTAACCG TAGTAAGCTC TTGAAACTGG GAGACTTGAG TGCAGGAGAG
A20 F      -----
A20 R      -----
```

```

...|...|...|...|...|...|...|...|
255     265    275    285    295
Clone A20  GAAAGTGGAA TTCCTAGTGT AGCGGTGAAA TGCGTAGATA TTAGGAGGAA
A20 F      -----
A20 R      -----
```

```

...|...|...|...|...|...|...|...|
305     315    325    335    345
Clone A20  CACCAGTAGC GAAGGCGGCT TCCTGGACTG TAACTGACAC TGAGGCACGA
A20 F      -----
A20 R      -----
```

```

...|...|...|...|...|...|...|...|
355     365    375    385    395
Clone A20  AAGCGTGGGG AGCAAACAGG ATTAGATACC CTGGTAGTCC ACGCCGTAAA
A20 F      -----
A20 R      ---CGTGGGG AGCAAACAGG ATT.....
```

```

...|...|...|...|...|...|...|...|
405     415    425    435    445
Clone A20  CGATGAGTAC CTGCCCGGGC GGCCGCTCGA AAGGGCGAAT TCCAGCACCC
A20 F      -----
A20 R      -----
```

Clone A20

A20 F

A20 R