Source Tracking of *Escherichia coli* Using the 16S-23S Intergenic Spacer Region of the *rrnB* Ribosomal Operon

by

Tom D'Elia

Submitted in Partial Fulfillment of the Requirements
For the Degree of

Masters of Science

In the

Biological Sciences

Program

YOUNGSTOWN STATE UNIVERSITY

August, 2004

Source Tracking of *Escherichia coli* using the 16S-23S Intergenic Spacer Region of the *rrnb* Ribosomal Operon

Tom D'Elia

I hereby release this thesis to the public. I understand that this thesis will be made available from the OhioLINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis for scholarly research.

Signature:

Tom D'Elia — 7/30/04
Tom D'Elia — Date

Approvals:

Carl Johnston — 7/30/04
Dr. Carl G. Johnston, Thesis Advisor — Date

Dr. Chester Cooper — 7/30/04
Dr. Chester Cooper — Date

David Asch — 7/30/04
Dr. David Asch — Date

Peter J. Kasvinsky — 8/2/04
Peter J. Kasvinsky, Dean of Graduate Studies — Date

Abstract

119 *Escherichia coli* isolates from nine different animal sources were subjected to denaturing gradient gel electrophoresis (DGGE) to determine sequence variations within the 16S-23S intergenic spacer region (ISR) of the *rrnB* ribosomal operon. The ISR was analyzed to determine if *E. coli* isolates from various animal sources could be differentiated from each other. In *E. coli,* the 16S-23S ISR has been demonstrated to consist of non-essential sequences that are subject to frequent insertion or deletion events that may allow for differences between different isolates. DNA isolated from the *E. coli* animal sources was PCR amplified to isolate the *rrnB* operons. To prevent PCR amplification of all seven *E. coli* ribosomal operons by PCR amplification by using universal primers, sequence specific primers were utilized for the *rrnB* operon. An additional primer set was then used on these amplimers to prepare samples of the 16S-23S ISR for DGGE. DGGE results show the presence of 40 unique ISR sequences from all of the samples. The highest rate of unique banding patterns, 60%, was observed for humans. The genetic profiles established by the PCR-DGGE method revealed a high genetic diversity for the *E. coli* isolates tested. There was also very little correlation between the ISR profiles created by the DGGE bands between the host sources. These findings suggest that the 16S-23S ISR may contain some host specificity, and that the high diversity of the *E. coli* isolates may allow for the assessment of environmental samples to distinguish similarities.

Table of Contents

List of Figures

List of Tables

List of Appendices

Acknowledgments

United States Geological Survey
I would like to thank Dr. Don Stoeckel of the United States Geological Survey for allowing me to use *Eshcerichia coli* isolates from his collection. Also, he was very helpful through the course of my research in offering important advice.

Youngstown State University
This research was funded by the Youngstown State University PACER grant.

I would like to thank Dr. Carl Johnston, my research advisor, for all of his help and support. He has been very helpful in all parts of my research, including offering lab advice, and help with writing and presentations. I am very fortunate to have conducted my research in his lab, and feel that I will be better prepared for future work because of Dr. Johnston.

I would like to thank Dr. Cooper for his technical support, and use of his lab equipment. As a member of my committee, Dr. Cooper was very important in many aspects of my research, including denaturing gradient gel electrophoresis, PCR, cloning and DNA sequencing. Additionally, he is a very good professor, and I learned a lot from his help and from his classes.

I would also like to thank another committee member, Dr. Asch. He was very helpful in giving advice, and the knowledge I gained from his classes was vital to my research success. Additionally, I would like to thank Dr. Asch for the use of his lab equipment.

I would like to extend a special thank you to Ms. Diana Arnett for all of the help with DNA sequencing.

For helping with data analysis, I would like to thank Dr. Diggins.

A special thank you goes to all the members of my lab.

Additional thanks goes to Jessica Gordon for all of her support, and Donald Brown and James Shevchuck.

A final thanks goes to my parents, family and friends.

We should preserve every scrap of biodiversity as priceless while we learn to use it and come to understand what it means to humanity.
-Edward O. Wilson

Chapter 1: Introduction

I. Microbial Source Tracking

The concept that the origin of fecal pollution can be traced using microbiological, genotypic, phenotypic and chemical methods has been given the general term of microbial source tracking (MST) (Scott et al 2002). This new technology has come about in order to track potentially pathogenic microbes to a particular source to prevent further contamination. Known sources of fecal contamination include combined sewage overflows (CSOs), septic systems, agricultural runoff and wildlife (Dombek et al 2000). In general, there are two main types classification for sources of fecal contamination: point and nonpoint sources. Point sources considered the major contributor to fecal pollution include raw sewage, storm water, CSOs and effluents from waste water pollution. Nonpoint sources are more dispersed and include wildlife, agricultural runoff and pleasure boats (Seurinck et al., 2003). The importance of microbial source tracking is to determine if the source is human, livestock or wildlife, since the microorganisms of human origin are regarded as having greater potential to cause disease in humans and contain human specific enteric pathogens (Scott et al 2002, Guan et al 2002). Furthermore, bacteria from humans found in the environment may indicate the presence of *Salmonella* spp., *Shigella* spp., hepatitis A virus and Norwalk group viruses which are known human pathogens that do not colonize nonhuman species (Parveen et al 1999).

All MST technologies rely on information received by investigating a particular indicator organism. Indicator organisms are used to determine the presence of fecal

pollution in water and are essential for MST. For many years, fecal coliforms have been widely used as indicators of human enteric pathogens, since they are naturally occurring in the gastrointestinal tracts of humans and warm blooded animals (Parveen et al 1999). Additionally, pathogens present in the environment are often in low numbers and more difficult to culture relative to indicator organism (McLellan et al., 2003). In particular, *Escherichia coli* has been extensively used as an indicator organism of fecal pollution. The reason *E. coli* has become a predominant indicator organism is due to the availability of the complete genome sequence, and that the organism is easily cultured in the lab. Furthermore, *E. coli* is not normally pathogenic to humans and is present in much higher concentrations than the other environmental pathogens it predicts, and thus reveals the presence of human enteric pathogens (Scott et al 2002).

Some strains of *E. coli* are primary pathogens with an enhanced potential to cause disease, and have been linked to worldwide outbreaks of severe disease (Kuhnert et al 2000). Therefore, it is important to monitor the input of *E. coli* into waterways, which is a widespread problem in the U.S. and is correlated with increase risk of several diseases (Dombek et al., 2000). When fecal coliforms, including *E. coli*, are found in high levels they impair the water quality in lakes and rivers, and bring a threat to those that use the water as evident by water borne outbreaks of *E. coli* 0157:h7 (Guan et al 2002). The effects of an *E. coli* infection in humans can be very serious and life threatening. *Eshcherichia. coli* infections are often enteric, and cause severe nausea and diarrhea, while extraintesintal infections are possible that are related to urinary tract infections, sepsis and meningitis (Kuhnert et al., 2000). The fact that *E. coli* is known to exist in the natural flora of the intestine, and is also a potentially severe pathogen, makes this

bacterium a very good indicator organism and a microbe that itself should be closely monitored itself in the environment.

Most of the various MST methods work by comparing environmental samples to a data bank of *E. coli* isolates from known sources. Therefore, MST is based on the assumption that specific genetic markers or strains of bacteria are associated with specific animal sources (Hartel et al., 2003). The presence of a predominant strain of *E. coli* in a particular host is most likely due to genetic drift. There are several variables that influence the potential success of MST by impacting strain selection and enrichment. These factors include the microenvironment of the particular host, intestine temperature, pH and diet (Carson et al 2001). Recently, the diet of confined deer, compared to wild deer, was shown to significantly affect ribotypes of *E. coli* isolates (Hartel et al., 2003). Other assumptions associated with MST are that particular *E. coli* clones are more likely to be isolated from one particular host species than another (host specificity) and that the clonal composition of the species isolated from soil or water represent the clonal composition of the species in the host population responsible for the fecal inputs to the environment (Gordon, 2001).

Based on previous MST investigations, there is a substantial amount of information about the genetic diversity, clonality and spatial and temporal distribution of *E. coli* strains in different hosts from different environments currently available (Farleiter et al 2000). This information suggests that it is possible to assign a host source to an environmental sample of *E. coli.* Rep-PCR utilizes PCR amplification of DNA between repetitive extragenic elements to obtain strain specific fingerprints. When analyzing 154 *E. coli* isolates, Rep-PCR using BOX primers has been able to correctly classify 94.7%

human, 100% chicken and 100% cow isolates (Dombek et al., 2000). However, Rep-PCR using ERIC primers was only capable of correctly classifying 28.6, 0 and 76% human, bovine and pig isolates from a total of 62 *E. coli* isolates (Leung et al., 2004). Other MST methods, such as amplified fragment length polymorphisms (AFLP) and multiple antibiotic resistance have given classification rates up to 97% when analyzing *E. coli* isolates from humans, wildlife and livestock (Guan et al., 2002). Additionally, when *E. coli* isolates were grouped as from either human or non-human sources, ribotyping was able to accurately identify 245 of 247 nonhuman and 38 of 40 human *E. coli* sources (Carson et al., 2001) and 67% human and 100% non-human from a total of 238 *E. coli* isolates (Dombek et al., 2000). Denaturing gradient gel electrophoresis using the 16S-23S intergenic spacer region of *E. coli* was also capable of correctly classifying 72% human and 94% animal isolates (Seurinck et al., 2003). While these different MST methods prove promising, the influence of the various variables on *E. coli* genetic diversity still needs more investigation, as do the issues of the reproducibility and cost effectiveness of the various methods.

II. *E. coli*: Ribosomal DNA and 16S-23S Intergenic Spacer Regions

The ribosomal genes in bacteria are part of a multigene family consisting of a various number of ribosomal (*rrn*) operons depending on the particular species (Cilia et al., 1996). *Escherichia. coli* has 7 *rrn* operons which consist of the 16S rRNA gene, an intergenic spacer region, the 23S rRNA genes, another intergenic spacer region, and the 5S rRNA gene (figure 1) (Anton et al., 1998, Fukushima et al., 2002, Garcia-Martinez et

al., 1999). The ISR are short regions that contain tRNA genes, target sequences for RNase III and other recognition signals for processing the transcript (Garcia-Martinez et al., 1996a), including a well-known consensus antiterminator (Anton et al., 1998). The 7 *rrn* operons in *E. coli* consist of two main types of 16S-23S ISR based on the number and specific tRNA genes present. *Escherichia coli* has 4 ISR type 1 (ISR1) regions that have only one tRNA gene encoding $tRNA^{glu-2}$ and are located in *rrnB, rrnC, rrnE,* and *rrnG* (figure 2) (Brosius et al., 1981). There are also three ISR2 regions that have two tRNA genes, $tRNA^{ile}$ and $tRNA^{ala}$, and are located in *rrnA, rrnD,* and *rrnH* (Anton et al., 1998, Garcia-Martinez et al., 1996a).

After understanding the distribution and arrangement of the ISR in *E. coli* it becomes easier to understand why this region has the potential to be used as a method of differentiating bacteria at the sub-species level. The rRNA operon sequences have become useful to differentiate strains because they are relatively easy to sequence (Cilia et al., 1996). Resent research has also utilized the characterization of the 16S-23S ISR for the comparison of closely related organisms when 16S rRNA has been inadequate for discriminating (Nagpal et al., 1998). The sequences of the rRNA genes undergo much slower divergences than their flanking non-genic sequences (Lia, 2000). This results in more variability and less homogeneity between ISR sequences compared to the ribosomal genes. The main reason the ISR has the potential to discriminate between closely related organisms is because of the presence of nucleotide sequences that apparently neither transcribe for genes or play a vital role in the secondary structure of the ribosomal RNA operon transcript. This is an important aspect to consider, since the secondary structure is essential for the cleavage and release of the rRNA molecules from the primary transcript

Figure 1: Ribosomal operon location in the *E. coli* genome.
Genomic distribution of the rRNA genes. The size (kb) of the entire *E. coli* genome is indicated in parenthesis. The rRNA genes are depicted as arrows, and the distances (kb) between each gene are indicated (adapted form Cilia et al., 1996)

***E. coli* (4639 kb)**

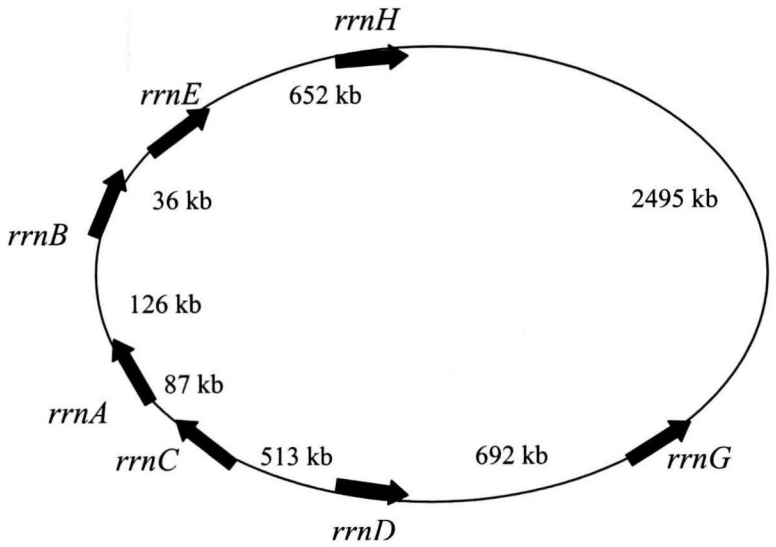
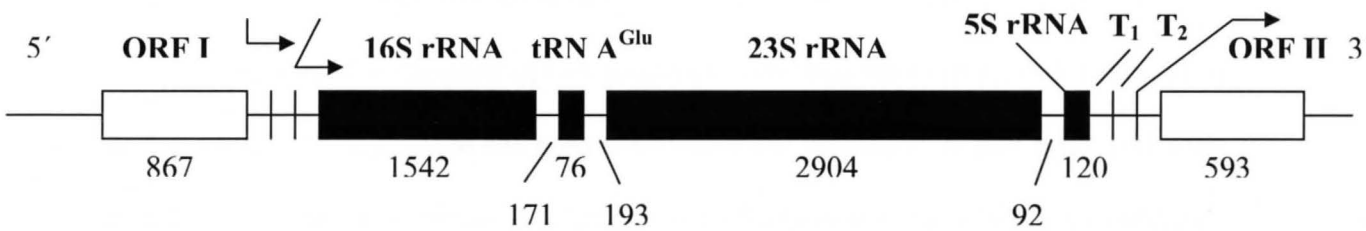


Figure 2: Gene organization of the *rrnB* operon of *E. coli.*
Horizontal arrows at the promoters specify direction of transcription. The RNA genes are indicated by filled bars and the open reading frames are indicated by open bars. The numbers under the bars indicate the length (bp) of the genes and spacers (adapted from Brosius et al., 1981).


5′
ORF I
16S rRNA
tRN A^Glu
23S rRNA
5S rRNA
T1
T2
ORF II
3′
867
1542
76
171
193
2904
120
92
593

by RNase III at specific recognition sites (Garcia-Martinez et al., 1999). The functional units within the spacer region do not sum up to more than 50% of its whole size, and the rest of the region consists of non-essential sequences submitted to frequent insertion/deletion events as noticed in *E. coli* (Garcia-Martinez et al., 1999). An additional aspect of the ISR that makes it particularly useful for differentiating among closely related species is that the spacer region varies in sequence and in length among species (Fox et al., 1998, Garcia-Martinez et al., 1999, Scheinert et al., 1996).

Currently, the presence of ISR sequence variations between *E. coli* strains and within an individual genome has been observed when looking at *E. coli* K-12 and other members of the *E. coli* reference collection (ECOR) (Anton et al., 1998, Garcia-Martinez et al., 1996a). Three main variations were found among the *E. coli* strains, and include dispersed nucleotide substitutions at certain locations, grouped variable sites of different composition but preserved secondary structure and block substitutions involving putative insertions or deletions that change the secondary structure (Anton et al., 1998). Specifically, *E. coli* K-12 was shown to have an ISR sequence of either 106 bp or 20 bp upstream of the $tRNA^{glu-2}$ (ISR1), a block of 14 bases grouped in a stem loop secondary structure in ISR1 only, and a 17 bp block substitution by a different 8 base sequence (Anton et al., 1998). Other differences observed were single base substitutions differentiating every individual operon and the switching of an ISR1 for an ISR2 and consequently an ISR2 for an ISR1 in the genome of ECOR 40 (Anton et al., 1998). Further analysis also revealed that all sites prone to nucleotide substitutions seem not to be involved in secondary structure, and the presence of a stem-loop with 21 variable positions in ISR1 with no homologous region in ISR2 (Anton et al., 1998). These

variations, among different *E. coli* strains, are valuable when using ISR sequence analysis for strain differentiation.

Ribosomal operons, including the *rrn* of *E. coli,* also exhibit intercistronic heterogeneity. This is a differing of operon sequences within the genome of an individual strain. There has been reportedly almost as much variation among different operons of the same strain as among different strains. Rarely are identical operons found even in the same genome when looking at ECOR samples (Garcia-Martinez et al., 1996a). Having as much variation among different operons in an *E. coli* genome as in other strains could limit the usefulness of ISR sequencing as a differentiating device (Nagpal et al., 1998). This is a very important aspect to consider when using the ISR sequence to compare *E. coli* strains.

Intercistronic heterogeneity in *E. coli* ribosomal operons can be overcome as a hindrance to strain typing and can be exploited as a way to differentiate between *E. coli* samples. PCR amplification of the 16S-23S ISR with general ribosomal operon primers produces a broad mixture of amplicons from each of the 7 *rrn* operons in *E. coli.* Direct sequencing of these PCR products produces a mean sequence in which mutations in the most variable domains become hidden. Cloning a single operon actually results in a sequence that differs from that of other operons, and of the mean sequence, by several point mutations. For this reason a mean sequence should be avoided to identify strains at the species level or below (Cilia et al., 1996). Besides using cloning to overcome the intercistronic heterogeneity, primers have been created that allow amplification of each individual *rrn* operon in *E. coli* (Anton et al., 1998). These primers are positioned on genes or open reading frames upstream of the 5' end of the 16S rRNA genes and use a

universal primer for the 23S rRNA gene. Utilization of these primers allows for the individual amplification of each operon for sequencing, avoiding production of a mean sequence.

The ribosomal operons in *E. coli* are complex in compositions having the sequences for important genes and cleavage sites, along with a complex and important secondary structure. Even though the ribosomal operons, and genes it encodes, are so essential to life, there are still some areas within the operon that are prone to higher levels of variation without effecting the transcription of the genes. Additionally, these variations can be as high, if not higher, between the operons of a single genome as between various strains. The fact that such variations exists, along with the capability to PCR amplify each individual operon, allows for the comparison of individual operons of one *E. coli* isolate to that of another, and shows promise as a useful tool for discriminating among samples.

III. Denaturing Gradient Gel Electrophoresis (DGGE)

DGGE is a gel separating method that can be used to distinguish two DNA molecules that differ by as little as a single base substitution (Sheffield et al., 1989). The ability of DGGE to separate PCR amplified DNA differing by a single base substitution has contributed to the wide use of the method in various fields of microbiology. In fact, DGGE has become routinely used in microbiology labs around the world as a molecular tool to compare the diversity of microbial communities and to monitor population

dynamics (Muyzer, 1999). In parallel DGGE a gradient of denaturants, urea and formamide, are established in a polyacrylamide gel. DNA samples are loaded onto the gel and an electric current is applied in the same direction as the denaturants. As DNA fragments migrate through a denaturing gel they remain double stranded until they reach a concentration of denaturants equivalent to a melting temperature (T_m) that causes the fragments lower melting temperature domains to melt resulting in the reduced mobility of the fragment as the DNA denatures (Sheffield et al., 1989). Sequence variations within the melting domains of various DNA samples will alter their melting temperatures, resulting in fragments that stop migrating at different positions in denaturing gels (Muyzer et al., 1993). Therefore, it is possible to separate PCR amplified DNA samples of target genes from various sources to determine sequence variations.

DGGE DNA separation is even more enhanced with the addition of a GC-clamp by using specially designed primers. GC-clamps are a series of 40 guanine and cytosine nucleotides that are added to one of the PCR primers, and they allow for separation of single base changes in the highest melting domains (Sheffield et al., 1989). DNA fragments up to 1000bp can be separated by DGGE. However, smaller fragments are easier to separate due to their increased mobility and the presence of fewer melting domains. Specifically, single base changes of PCR amplicons up to 500 bp joined to a 40 bp GC-clamp can be separated by DGGE (Farnleitner et al., 2000, Sheffield et al., 1989).

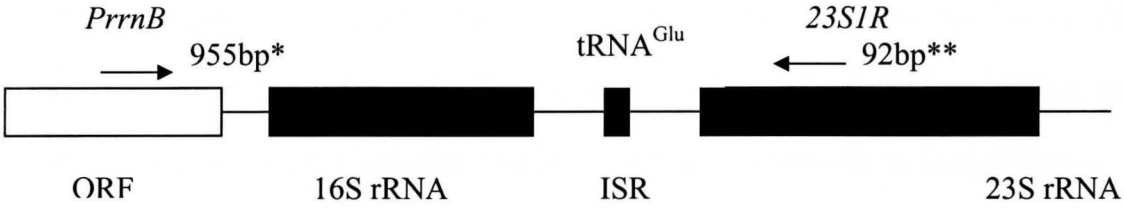
Chapter 2: Method Development and *E. coli* source isolates PCR-DGGE

Specific Aims:

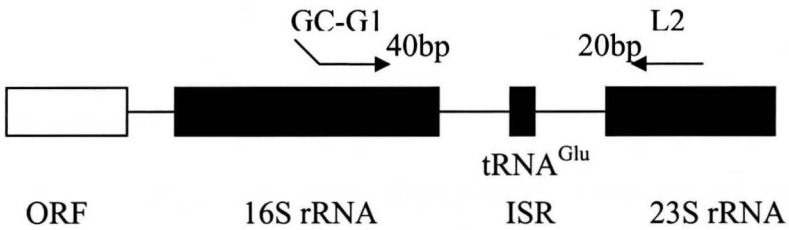
The goal of this study was to evaluate a microbial source tracking method for differentiating between strains of *E. coli* from known sources based on DGGE analysis of the 16S-23S ISR. Specifically, the *rrn*B operon was subjected to PCR amplification using sequence specific primers. Two new primers, one of which had a GC-Clamp, were then used to prepare PCR amplified 16S-23S ISR for DGGE analysis (figure 3). These techniques were performed on a total of 119 *E. coli* from 8 different sources, which were generously donated by the United States Geological Survey (USGS). The resulting gels were compared utilizing an *E. coli* PCR amplified 16S-23S ISR and commercially available standards, which were ran on each DGGE gel. In order to verify DGGE results, and to determine the amount of sequence variation in the ISR, DNA sequencing of cloned ISR was performed. Control experiments were performed to determine the reproducibility of this PCR-DGGE method, and included repeated independent PCR amplification of *E. coli* isolates followed by DGGE, cloning of an amplified 16S-23S ISR for PCR efficiency analysis. Additionally, 3 of the isolates were carried through a series of generations over the course of the experiment to evaluate any changes in the ISR over time.

Figure 3a and b: Primer locations for PCR reactions.
3a: Primer sets PrrnB and 23S1R are isolate the *rrnB* operon specifically in *E. coli* by binding to an open reading frame upstream of the 16S rRNA gene and a conserved region on the 5' end of the 23S rRNA gene. *Distance from the 3' end of the primer to the 5' end of the 16S rRNA gene. **Distance from the 5' end of the primer to the 5' end of the 23S rRNA gene.
3b: Primer sets GC-G1 and L2 isolate the ISR, distance form these primers to the ISR are listed. Open reading frames are designated as open bars, and rRNA genes are indicated by filled bars.

3a:


PrrnB
955bp*
tRNA^Glu
23S1R
92bp**
ORF
16S rRNA
ISR
23S rRNA

3b:


GC-G1
40bp
20bp
L2
tRNA^Glu
ORF
16S rRNA
ISR
23S rRNA

Materials and Methods

E. coli isolates and collection:

E. coli isolates from nine animal sources (Canada goose, chicken, cow [beef and dairy], deer, dog, horse, human, swine) were subcultured from an *E. coli* isolate collection from various farms in Berkely County West Virginia. The samples consisted of one isolate per individual animal source and 15 individual isolates from the 8 animal sources for a total of 120 samples. Isolates were grown overnight in 3 ml of Luria-Bertani broth at 37°C and then frozen in a 20% glycerol solution. The cells were stored at -80° C. Each isolate was labeled by using a host source abbreviation followed by the isolate number (01-15) (appendix A).

DNA Isolation:

E. coli isolates were cultured overnight in Luria-Bertani broth at 37°C. The genomic DNA was then isolated using the GenElute^{tm} Genomic DNA Kit (Sigma-Aldrich, Inc.), according to its protocol. Isolated genomic DNA was subjected to PCR followed by quantitation using agarose gel electrophoresis. The isolated DNA used in the PCR was diluted by a factor of 1:1, 1:10, 1:100 and 1:1000, to optimize future reactions. The optimal concentration was determined and used for subsequent reactions. The isolated DNA was stored at -20°C.

PCR Reactions:

PCR I: Reagents for the Polymerase Chain Reaction.

Reagent	Amount per Reaction (µl)
10x Gold PCR Buffer (Applied Biosystems)	5
25 mM $MgCl_2$ (Applied Biosystems)	5.5
50 pmol/µl Primer 23S1R	1
50 pmol/µl Primer PrrnB	1
10mM dNTPs (Roche)	1
Taq polymerase 5U/µl (Applied Biosystems)	0.4
H_2O	26.1

PCR amplification was carried out in 50 µl volumes using 40 µl of the PCR I mix and 10 µl a 1:10 dilution of isolated genomic DNA. Primer 23S1R (5' GGG TTT CCCC A TT CGG AAA TC 3') hybridizes 96bp from the 3' of the 23S rRNA gene (Garcia Martinez et. al 1996b). Primer PrrnB (5' AAC ACT GCC AGT ACC GTT TC 3') binds to an open reading frame 955 bp from the 5' end of the 16S rRNA gene (Anton et al, 1998). All PCR amplifications were carried out in a MJ Research PTC-200 thermal cycler. An initial 1 min at 95°C was followed by 35 cycles of: 30s at 94°C, 30s at 56.8°C, 2 min at 72°C. The final cycle was followed by an additional 5 min at 72°C,

with a holding temperature of 4°C. PCR products were visualized by agarose gel electrophoresis.

PCR II: Reagents for the Polymerase Chain Reaction.

Reagent	Amount per Reaction (µl)
10x Gold PCR Buffer (Applied Biosystems)	5
25 mM $MgCl_2$ (Applied Biosystems)	5
6 nM Primer GC-G1	2
400 nM Primer L2	2
10mM dNTPs (Roche)	0.2
Taq polymerase 5U/µl (Applied Biosystems)	0.4
H_2O	26.8

PCR amplifications were carried out using 40 µl of the PCR II mix and 1:100 dilutions of PCR I products. Primer GC-G1 (5' CGC CCG CCG CGC CCC GCG CCG T CCC GCC GCC CCC CGC CCC CGA AGT CGT AAC AAG G 3') binds within the 16S rRNA gene, approximately 40 bp upstream of the ISR. The G-C rich region of primer GC-G1 is a GC clamp, and is essential to prevent complete denaturization during DGGE analysis (Buchan et al., 2001). Primer L2 (5' CAA GGC ATC CAC CGT 3') is the reverse primer, and binds to a region of the 23S rRNA gene approximately 20 bp

downstream of the ISR (Jensen et al., 1993). All PCR amplifications were carried out in a MJ Research PTC-200 thermal cycler. An initial 3 min at 94°C was followed by 25 cycles of: 1 min at 94°C, 1 min at 55°C, 2 min at 72°C. The final cycle was followed by an additional 7 min at 72°C, with a holding temperature of 4°C. PCR products were visualized by agarose gel electrophoresis.

Agarose Gel Electrophoresis:

PCR products and restriction digests were run on 1% high resolution agarose (Sigma, MO) using 1X TAE buffer (40 mM Tris, 20 mM Acetic Acid, 1mM EDTA, pH 8.3). All agarose gels were run at 84 V for 45 min. Gels were stained in 1% ethidium bromide solutions for 20 min, followed by a 5 min rinse in water. The resulting DNA bands were visualized using a UV light, and the size and approximate quantity of DNA present were determined by comparison to a molecular weight marker (Biomarker EXT Plus, Invitrogen, CA).

Denaturing Gradient Gel Electrophoresis (DGGE)

Parallel denaturing gradient gel electrophoresis was performed on the amplified ISR samples (PCR II products) using the DCode Universal Mutation Detection System (BioRad Inc., CA) to detect nucleic acid differences among the samples. Samples were loaded onto 8% (wt/vol) polyacrylamide gels. The gels were prepared using 30% and 60% denaturant acrylamide solutions (30% stock: 40% acrylamide/Bis stock solution (BioRad Inc., CA), 12% deionized formamide, 2.1 M urea, 1X TAE buffer; 60% stock: 40% acrylamide/Bis stock solution, 24% deionized formamide, 4.2 M urea, 1X TAE

buffer). The amount of samples loaded was determined from the agarose gel electrophoresis, and typically ranged from 3-5 µl (approximately 20 ng DNA). Electrophoretic charge was applied in the same direction as denaturants. All gels were run in 1X TAE buffer for 3h 45min at 60°C and 150 V. The amount of time to run each gel was determined by a time course experiment, in which samples were added to the gel each hour for three hours and the gel was ran for a total of 6 hours and 45min. After electrophoresis, each gel was stained in 1% ethidium bromide solution for 7 min, followed by a 10 min de-staining in 1 X TAE. Images were captured using Eagle Eye II image capturing (Stratagene, CA) and saved as digital files.

To allow for comparisons between the denaturing gels, relative distances (RD) were calculated by dividing the distance the individual samples migrated by the distance of the *E. coli* standard on the same gel.

Four independent PCR amplifications were carried out on one *E. coli* isolate, Canada goose isolate 11 (Cg11) and then confirmed to have the same DGGE banding pattern. The *E. coli* PCR product was then subsequently used as a relative standard to compare banding patterns between gels. DCode wild type and mutant controls (BioRad, CA) were also run on each gel to show the capability of the gels to differentiate between single nucleotide differences. Additional isolates were also separately PCR amplified and subjected to DGGE to confirm repeatability.

Statistical Analysis of DGGE Results:

The diversity of the RD values for each *E. coli* source was calculated using the Shannon-Weaver index. This index has been shown to be a useful measure of diversity

for microbial communities, and is the most commonly used method for calculating diversity (Mills et al., 1980). DGGE band diversity for each of the sources was calculated as follows: $H' = \frac{n \log n - \sum_{i=1}^{k} fi \log fi}{n}$

Where n = number of samples, k = the number of categories, or different possible DGGE bands, f_i =number or frequency of DGGE bans present for a given category. The higher the value of H', the higher the diversity.

The maximum possible diversity for each *E. coli* source was calculated using $H'_{max} = \log k$. The magnitude of H' is affected by both the distribution and the number of categories. Therefore, by calculating the evenness (J'), the observed diversity is presented as a proportion of the maximum possible diversity. Evenness was calculated using the following equation: $J' = \frac{H'}{H' \max}$

SPSS 11.5 software (SPSS, IL.) was used to for bivariate correlation analysis of the DGGE band distributions. The Pearson correlation coefficient was calculated, and the significance (2-tailed) of correlation was given for each sources compared.

Cloning:

Selected *E. coli* isolates were subjected to cloning and DNA sequencing of the ISR based on DGGE banding patterns. Within 24 hr of amplification, ISR PCR products were ligated into pCR 4-TOPO vector (TOPO TA Cloning Kit for Sequencing, Invitrogen, CA) following the manufacturer's instructions. The ligation reaction was set up as follows: 2.5 µl PCR products, 1.0 µl salt solution, 1.5 µl H_2O, 1.0 µl vector. The ligation reaction was then transformed into One Shot® TOP10 Competent cells following

the manufacturer's instructions. Following transformation, 50 and 100 µl of the transformed cells were spread onto pre-warmed (37°C) LB-ampicillin (100 µg/µl ampicillin) and incubated at 37°C for 24 h. The vector contains the lethal *E. coli ccd*B gene fused to the C-terminus of the LacZά fragment. When the PCR product ligates to the vector the LacZά –*ccd*B gene fusion is disrupted permitting the growth of positive recombinants. Cells that contain non-recombinant vector are killed upon plating on plates containing ampicillin. After each cloning procedure, 5 colonies were selected from each plated isolate to analyses positive insertion by restriction digest with *Eco*R1.

Restriction Digests:

Following cloning, colonies were isolated and grown overnight in LB-ampicllin broth at 37°C for 24 h. Plasmid DNA was then isolated using the Cyclo-Prep Plasmid DNA isolation kit (Amresco, OH) and eluted in 65 µl H_2O. Plasmid DNA was then digested with *Eco*R1 as follows: 9.6 µl H_2O, 3 µl Buffer, and 0.2µl enzyme were combined with 17 µl of plasmid DNA and incubated for 3h at 37°C. Digests were then visualized by loading 5 µl of digest onto a 1% agarose gel and electrophoresed for 45 min at 85V. Positive inserts were noted as having an inserted 500bp band. A molecular size marker (Biomarker EXT Plus, Invitrogen, CA) was used to determine the size of the insert and to allow quantification of the plasmid DNA for DNA sequencing.

DNA Sequencing:

Isolated plasmids that showed successful insertion of the ISR PCR product were used for DNA sequence analysis. The DNA sequencing reaction was prepared using the

CEQ 2000 Dye Terminator Cycle Sequencing Quick Start kit (Beckman Coulter Inc., CA) according to the manufacturer's instructions. PCR dye labeling was amplified following the manufacturer's program kit (Beckman Coulter Inc., CA) using the M13F and M13R primers for the pCR 4-TOPO plasmid (Invitrogen, CA). The DNA sequencing reaction was ethanol precipitated following the manufacturer's instructions (Beckman Coulter Inc., CA) and resuspended in 40 µl sample loading solution and stored at -20°C. Samples were sequenced on a Beckman Coulter CEQ 2000XL Dye Terminator Cycle DNA Sequencer.

Long Term Cultures:

Three isolates (cow, swine and human) were chosen and ran through a series of cultures. Samples were grown on LB plates overnight at 37°C and then transferred to LB broth tubes for another overnight culture at 37°C. This was repeated over the course of the experiment. Genomic DNA was isolated from the cultures before the time course experiment began, and after 11 culture transfers. The ISR was then PCR amplified following the same protocols as all other isolates and subjected to DGGE analyses to see what affects multiple cultures has on the genetic stability of the ISR over multiple generations of *E. coli* growth.

Chapter 3: Results

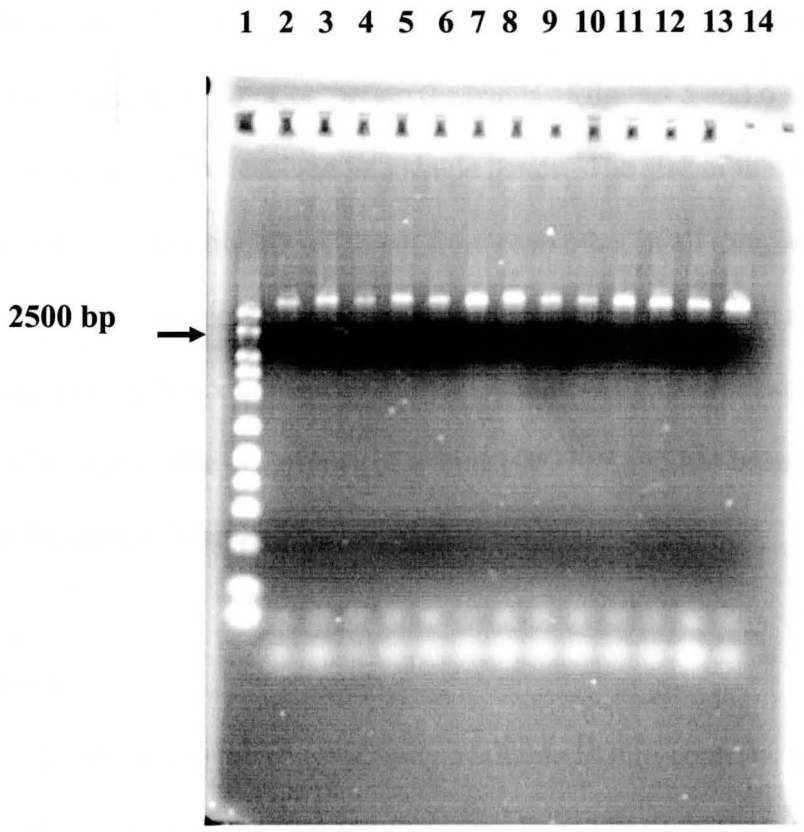
ISR Amplification:

The ISR of 119 *E. coli* isolates from 8 different sources were subjected to PCR amplification to look for DNA sequence variations among the isolates by DGGE analysis. To avoid producing a composite sequence of all 7 *E. coli* ribosomal operons, primers specific to the *rrn*B (Anton et al., 1998) were used for the first PCR (PCR I) amplification. The size of resulting PCR products was then determined by agarose gel electrophoresis and compared to a size standard. Amplification of all isolates with the *rrn*B specific primers yielded a single product of approximately 3000 bp (figure 4), which is consistent with other results using the same primer set (Anton et al., 1998).

After successful amplification of the *rrn*B operon, the ISR was specifically targeted using primers adapted from Jensen et al., 1995. This second PCR (PCR II) resulted in the amplification of the rrnB ISR, which could then be subsequently used for DGGE analysis. PCR II primers incorporated a GC-Clamp into products, which allows for DGGE analysis. Depending on PCR I results, dilutions of 1:100 or 1:1000 of PCR I were used for the PCR II reactions. Resulting PCR products were ran on agarose gels to determine fragment size. Previous studies using the *rrn*B ISR specific primers on isolated *E. coli* genomic DNA have resulted in producing amplicons of 480 and 540 bp corresponding to ISR I and ISR II respectively (Buchan et. al., 2001) or approximately 530 bp for *rrnB, rrnG, rrnD* the hybrid operon *rrnX* (Jensen et al., 1993). In this study, ISR PCR amplification of *rrnB* operons revealed products of approximately 500 bp or

Figure 4: PCR amplification resulting in the isolation of the *rrnB* operon from chicken isolates.

Legend: Chicken isolate numbers are listed
Lane 1-BioMarker Ext PlUS 50-2500bp ladder
Lane 2-01
Lane 3-03
Lane 4-04
Lane 5-06
Lane 6-07
Lane 7-08
Lane 8-09
Lane 9-10
Lane 10-11
Lane 11-12
Lane 12-13
Lane 13-14
Lane 14-15
Lane 15-blank


1 2 3 4 5 6 7 8 9 10 11 12 13 14
2500 bp

580 bp (figure 5). Each *E. coli* isolate that had the ISR amplified produced a single dominate band of either of these two listed sizes. However, on certain occasions, the PCR II products would result in two discrete double bands. These double bands were resolved into single products by increasing the amount of PCR I product used for the ISR amplification (figure 6). Comparing the bands of figures 5 and 6 shows that the double bands were able to be resolved into single bands. The size of the double bands present seemed to be proportionally different from each other in all samples in which they occurred. The cause of the double bands in unknown, and would require additional investigation by DNA sequence analysis. Additionally, a faint band was often observed around 1000 bp. The occurrence of a faint 1000 bp band has been noted elsewhere when using the same primer set (Buchan et al., 2001).

Controls:

In addition to the commercially available DGGE controls, an *E. coli* control was created to use for between gel comparisons. Four 100 µl PCR reaction were performed to amplify the ISR of one *E. coli* isolate, Cg11. The resulting products were run on a DGGE gel to confirm that they all would show the same gel migration patterns (figure 7). After confirmation of similar migrations, the *E. coli* standard, along with the commercial controls (BioRad Inc., CA) was run on each DGGE gel to allow comparison between gels.

In order to determine if sequence variations apparent in different DGGE bands were due solely on naturally occurring nucleotide differences and not from Taq

Figure 5: PCR amplification resulting in the isolation of the *rrnB* 16S-23S ISR from chicken isolates.

Legend: Chicken isolate numbers are listed
Lane 1-BioMarker Ext PlUS 50-2500bp ladder
Lane 2-01
Lane 3-03
Lane 4-04
Lane 5-05
Lane 6-06
Lane 7-07
Lane 8-08
Lane 9-09
Lane 10-10
Lane 11-11
Lane 12-12
Lane 13-13
Lane 14-14
Lane 15-15

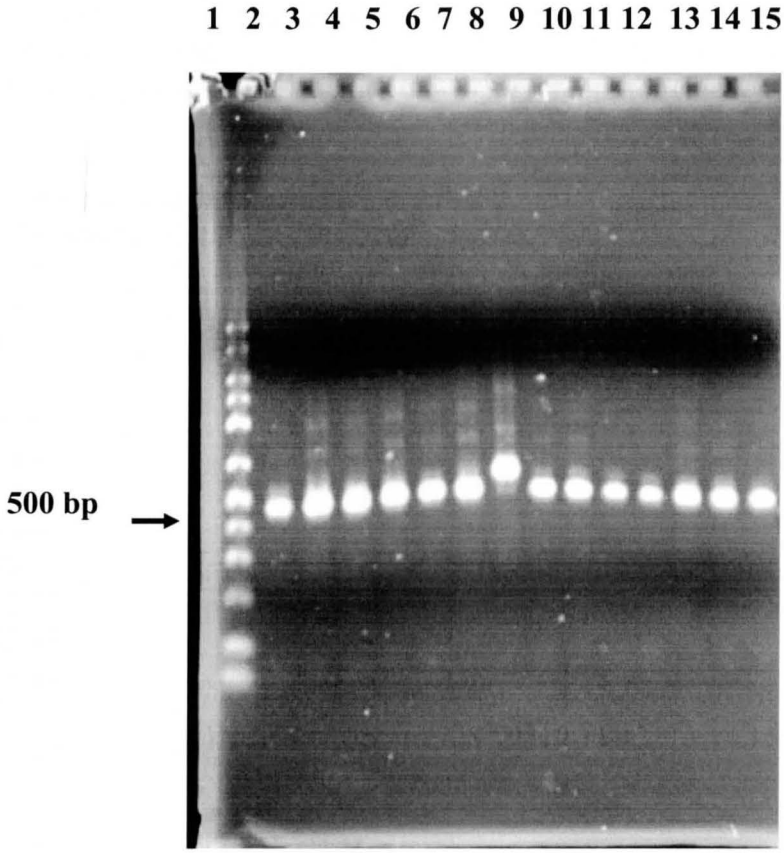

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
500 bp

Figure 6: Resolving of double bands that were present after PCR amplification of the *rrnB* 16S-23S ISR.
The gel pictured shows double bands that were present for chicken isolates 01-15.

Legend: Chicken isolate numbers are listed
Lane 1-BioMarker Ext PLUS 50-2500bp ladder
Lane 2-01
Lane 3-03
Lane 4-04
Lane 5-05
Lane 6-06
Lane 7-07
Lane 8-08
Lane 9-09
Lane 10-10
Lane 11-11
Lane 12-12
Lane 13-13
Lane 14-14
Lane 15-15

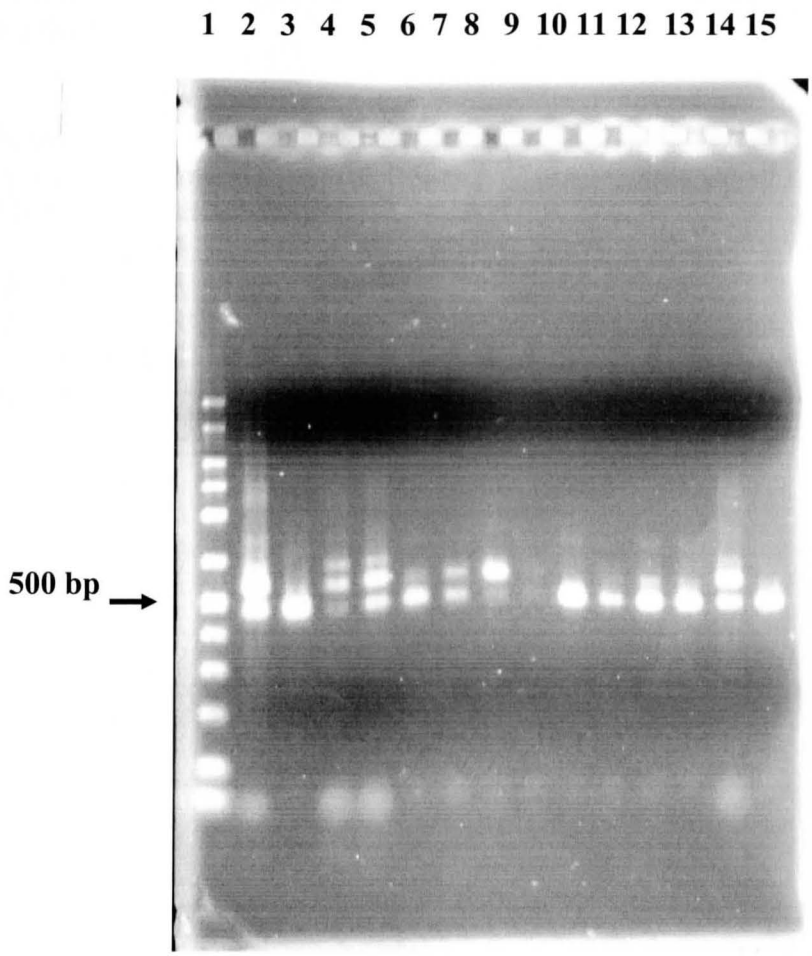
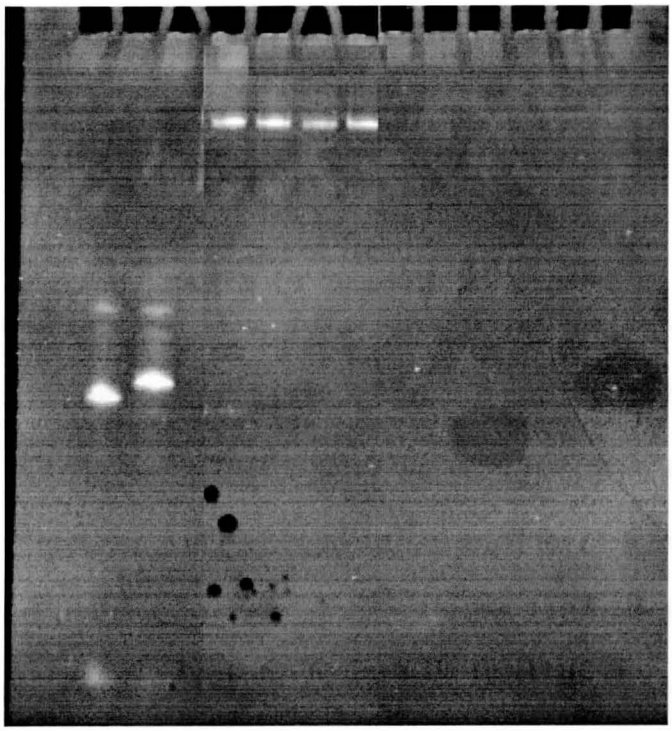

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
500 bp

Figure 7: DGGE gel of *E. coli* standard that was used to compare gel migrations between gels.

Legend:
Lane 1-Blank
Lane 2-Wildtype commercial control
Lane 3-Mutant commercial control
Lane 4- *E. coli* control
Lane 5- *E. coli* control
Lane 6- *E. coli* control
Lane 7- *E. coli* control
Lane 8-13-Blank

1 2 3 4 5 6 7 8 9 10 11 12 13



polymerase inefficiencies, *E. coli* isolate Cg11 PCR amplified *rrnB* ISR was cloned and subjected to ten independent PCR reactions. Resulting products were ran on DGGE gels to confirm consistency in PCR amplification between samples. The resulting DGGE gels showed that each independent PCR of the same clone gave the same gel migration distance (figure 8).

DGGE Analysis of the *rrnB* ISR of 119 *E. coli* Isolates:

DGGE analysis of the *rrnB* ISR from 119 *E. coli* isolates revealed the presence of a total of 40 different possible gel migrations of the single band ISR product based on relative distance (RD) values. The 40 possible DGGE bands that were present were labeled A-OO and the frequency of occurrence of each band was recorded for each animal source (table 1). The distribution of RD values for each animal source overlapped quite a bit towards the middle range. For several of the sources, there were unique bands present, and when there were similar bands present, they were often present in different frequencies. Consequently, each animal source has different frequency of bands and distributions, which may allow for differentiation between sources. Table 2 shows the variations in DGGE patterns between the human and animal sources. Of the 15 Canada goose *E. coli* isolates, a total of nine different DGGE bands were present. Each Canada goose sample produces a single band representing the amplified *rrnB* ISR. Therefore, the ISR bands present from the 15 Canada goose samples were distributed between 9 band possibilities on the denaturing gels. The number of the 15 total Canada goose samples that were unique compared to all other sources was 7. As a result, 47 % of all Canada goose *E. coli* ISR sequences were unique.

Figure 8: DGGE gel of independently cloned Canada goose isolated number 11.

Legend:
Lane 1- *E. coli* control
Lane 2- Cg 11 clone 1
Lane 3- Cg 11 clone 2
Lane 4- Cg 11 clone 3
Lane 5- Cg 11 clone 4
Lane 6- Cg 11 clone 5
Lane 7- Cg 11 clone 6
Lane 8- Cg 11 clone 7
Lane 9- Cg 11 clone 8
Lane 10- Cg 11 clone 9
Lane 11- Cg 11 clone 10
Lane 12- *E. coli* control

1 2 3 4 5 6 7 8 9 10 11 12

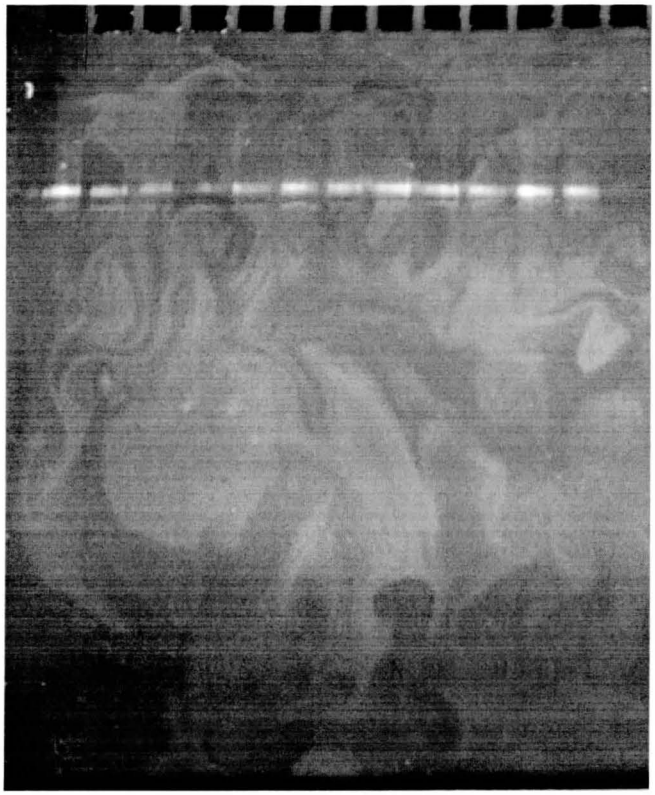


Table 1: Frequency and distribution of gel migrations of the *E. coli* isolates from the various sources.

The table shows how the DGGE bands establish a genetic profile for each animal source. There are unique bands present for many of the sources. There are overlaps in the middle range, but the frequencies are often different between the sources. All nonhuman sources were pooled together for comparison with human source isolates.

	Sources								
	Canadian Goose	Swine	Cattle	Human	Deer	Dog	Horse	Chicken	Nonhuman
A	3	0	0	0	0	0	0	0	3
B	1	0	0	0	0	0	0	0	1
C	0	0	0	2	0	0	0	0	0
D	0	0	0	0	0	1	0	0	1
E	0	0	0	1	0	0	0	0	0
F	0	0	1	0	0	0	0	0	1
G	1	0	0	0	0	0	0	0	1
H	3	0	1	0	0	0	0	0	4
I	0	0	0	0	0	0	0	1	1
J	1	0	0	0	0	0	0	0	1
K	0	0	0	0	1	0	0	0	1
L	0	0	0	0	1	0	0	0	1
M	0	1	0	0	0	0	1	0	2
N	0	0	0	2	0	0	0	0	0
O	0	0	0	0	0	0	0	1	1
P	0	0	2	0	2	3	1	1	9
Q	0	0	1	0	0	0	0	0	1
R	0	0	0	1	0	0	2	0	2
S	0	3	0	0	0	0	1	2	6
T	0	1	0	1	1	0	0	0	2
U	0	0	2	0	3	2	0	1	8
V	0	0	1	0	2	0	0	3	6
W	0	0	1	0	0	0	2	0	3
X	0	0	0	0	0	0	1	0	1
Y	0	2	0	1	1	4	0	1	8
Z	3	0	1	0	0	0	1	3	8
AA	1	4	1	0	0	0	1	0	7
BB	0	0	0	2	0	0	0	0	0
CC	0	0	0	0	0	1	1	0	2
DD	1	1	1	0	1	0	3	1	8
EE	0	2	1	1	1	0	0	0	4
FF	0	0	1	0	1	0	0	1	3
GG	0	0	0	0	0	1	1	0	2
HH	0	1	0	1	0	1	0	0	2
II	0	0	0	0	0	1	0	0	1
JJ	0	0	1	1	0	0	0	0	1
KK	0	1	0	0	0	0	0	0	1
LL	0	0	0	1	0	0	0	0	1
MM	1	0	0	0	0	0	0	0	1
NN	0	0	0	0	0	1	0	0	1
OO	0	0	0	1	0	0	0	0	0

Table 2: Distribution and frequency of DGGE results compared among the isolates.

The different DGGE profiles established for each source are compared. The number and percentage of unique bands present are recorded, and the diversity and evenness are given in the table.

	Source								
	Canadian Goose	Swine	Cattle	Human	Deer	Dog	Horse	Chicken	Total
Number of Isolates	15	15	15	15	14	15	15	15	119
Number of Different DGGE Bands Present	9	8	13	12	10	9	11	10	40
Number of Unique RDs	7	1	2	9	2	2	1	2	26
Percentage of Unique RD Values per Number of Isolates	47	7	13	60	14	13	7	13	21
Shannon Weaver Diversity H'	0.89	0.84	1.14	1.06	0.96	0.88	1	0.95	1.46
H' Max=	0.95	0.95	1.11	1.08	1	0.95	1.04	1	1.6
Evenness J'	0.93	0.88	1	0.98	0.96	0.92	0.96	0.95	0.91

For all of the animal sources, the number of possible DGGE bands ranged from 8 to 13. The number of unique DGGE bands ranged from 1 for swine and horse to 9 for human *E. coli* isolates. Each source produce unique RD values, with the highest proportion of unique DGGE bands, 60%, found for human isolates. Cattle, dog and chicken isolates each had 13% unique DGGE bands for their respective *E. coli* isolates. The Shannon-Weaver Diversities (*H'*) of all samples were moderately high, ranging from *H'*=0.84-1.14. All evenness values (*J'*) were also high, ranging from *J'*=0.92-1. These results show that there is high diversity among the possible DGGE bands present for each animal source, and that the bands identified were evenly distributed among the possible RD values for each human or animal source. Therefore, it is likely that any additional *E. coli* isolates tested for a given source will remain unique to that particular source.

Effective MST methods should be capable of distinguishing human sources of *E. coli* from nonhuman sources, as the presence of microorganisms from human origin are regarded as having a greater potential to cause disease in humans (Guan et al., 2002). Therefore, all of the nonhuman sources DGGE results were combined for comparison with the human results (table 3). The samples sizes are quite different (human n=15, nonhuman=104). However, the results give an insight to the amount of variation between the two groups. Human source isolates had 60% unique RD values, while nonhumans had 80% unique DGGE bands. This is substantially higher than any other individual nonhuman source. Combining the nonhuman sources also considerably increased the diversity (*H'*=1.39) compared to any individual nonhuman source, and the evenness remained relatively high (*J'*=0.91).

Table 3: Distribution and frequency of DGGE results compared between human and nonhuman isolates.

The different DGGE profiles established for human and nonhuman sources are compared. The number and percentage of unique bands present are recorded, and the diversity and evenness are given in the table.

	Source	
	Human	Non-Human
Number of Isolates	15	105
Number of Different DGGE Bands Present	12	34
Number of Unique RDs	9	83
% Unique RD Values per Number of Isolates	60	80
Shannon Weaver Diversity *H'*	1.1	1.39
H' Max=	1.08	1.53
Evenness *J'*	1	0.91

Pearson correlation coefficients were calculated for the RD values of all the isolates tested. Table 4 shows the results of the correlation test between the DGGE bands from the 8 animal sources. The correlation calculations take into consideration both the frequency and distribution of all RD values for each isolate source, establishing a genetic profile or fingerprint for each source. These profiles are then compared to each other for any correlation between the RD values. The only sources that showed any correlation between isolate profiles were cattle/deer, cattle/chicken, deer/dog, and deer/chicken. All other source profiles showed no correlation. Therefore, based on the isolates tested, the majority of the sources could be differentiated based on their *rrnB* 16S-23S ISR DGGE results. These results suggest that even though the diversity of RD values is high among all the sources tested, the bands that were present for each source were still unique, and that further testing would likely reveal unique RD values for each source. It is important to note that the sample sizes for each source was low (15 isolates/source). Nonetheless, the results reveal the potential of DGGE analysis of the *rrnB* 16S-23S ISR for differentiating between *E. coli* isolates.

When the human and nonhuman sources were compared using Person correlation coefficients, the two sources were nearly negatively correlated (-0.290, P=0.065 2-tailed). Regardless, the two sources were still not correlated by RD values, and the unique profiles established for DGGE RD values were capable of distinguishing each source. Further analysis with a larger set of human isolates would reveal more information about exactly how the two sources genetic profiles are related. The data is also useful for establishing field testing studies. It may not be necessary to determine specific sources

Table 4: Correlation of DGGE Bands from 8 Animal Sources.

The table shows the correlation between the DGGE bands from the 8 animal sources. Only Cattle/Deer, Cattle/Chicken, Deer/Dog, and Deer/Chicken showed any correlation in the DGGE bands produced from amplified ribosomal B ISR. All other comparisons have no correlation between DGGE bands. Number values listed are Pearson correlation values.

*Correlation is significant at the 0.05 level (two-tailed).

**Correlation is significant at the 0.01 level (two-tailed).

Source	Goose	Swine	Cattle	Human	Deer	Dog	Horse	Chicken
Goose	-							
Swine		-						
Cattle			-					
Human				-				
Deer			0.613**		-			
Dog					0.457**	-		
Horse							-	
Chicken			0.365*		0.418**			-

for all nonhuman sources if human and nonhuman sources can be confidently distinguished.

To confirm the ability of denaturing gels to separate PCR amplified 16S-23S ISR, specific isolates that gave the same, and slightly different, RD values for DGGE analysis were cloned and sequenced. Positive clones were analyzed by restriction digest of the isolated plasmid with *Eco*R1 endonuclease. *Eco*R1 restriction sites flank the insert on the plasmid, so resulting fragments of approximately 550 bp were observed by agarose gel electrophoresis (figure 9). The Isolates Cg11 (*E. coli* standard for DGGE gels) and Cd1 both gave RD values of 1.0. Sequence analysis of the 16S-23S ISR from these isolates confirmed that the sequence were in fact identical (figure 10), and that the size of the PCR fragment was 480 bp, as expected (Buchan et al.,2001, Jensen et al.,1993). Isolates Sw15 had an RD value of 1.11, and isolate Do1 had an RD values of 1.05, and subsequent DNA sequence analysis confirmed the sequences did actually differ in 7 locations (figure 11). Additionally, an RD of 0.96 was recorded for isolate Cg4, and isolate Sw5 had an RD of 0.93. The difference in RD values was confirmed by the presence of nucleotide differences determined from DNA sequence analysis (fig 12). The two sequences differed in 6 locations. When the DNA sequences from all of the isolates were aligned there were individual nucleotide differences between samples with different RD values (figure 13). DGGE analysis of sample Sw4 *E. coli* isolate gave a RD value of 0.81. This isolate did not align with any of the other isolates. There was an 80 nt block insert that was not present in any of the other isolates that were analyzed by DNA sequencing (fig of all). The results of the DNA sequence analysis confirm that

Figure 9: Restriction digest of isolated plasmids from positive clones.

Legend:
Lane 1- Blank
Lane 2- BioMarker Ext PLUS 50-2500bp ladder
Lane 3- Canada goose clone plasmid showing insert
Lane 4- Blank
Lane 5- Canada goose clone plasmid showing insert
Lane 6-15- Blank

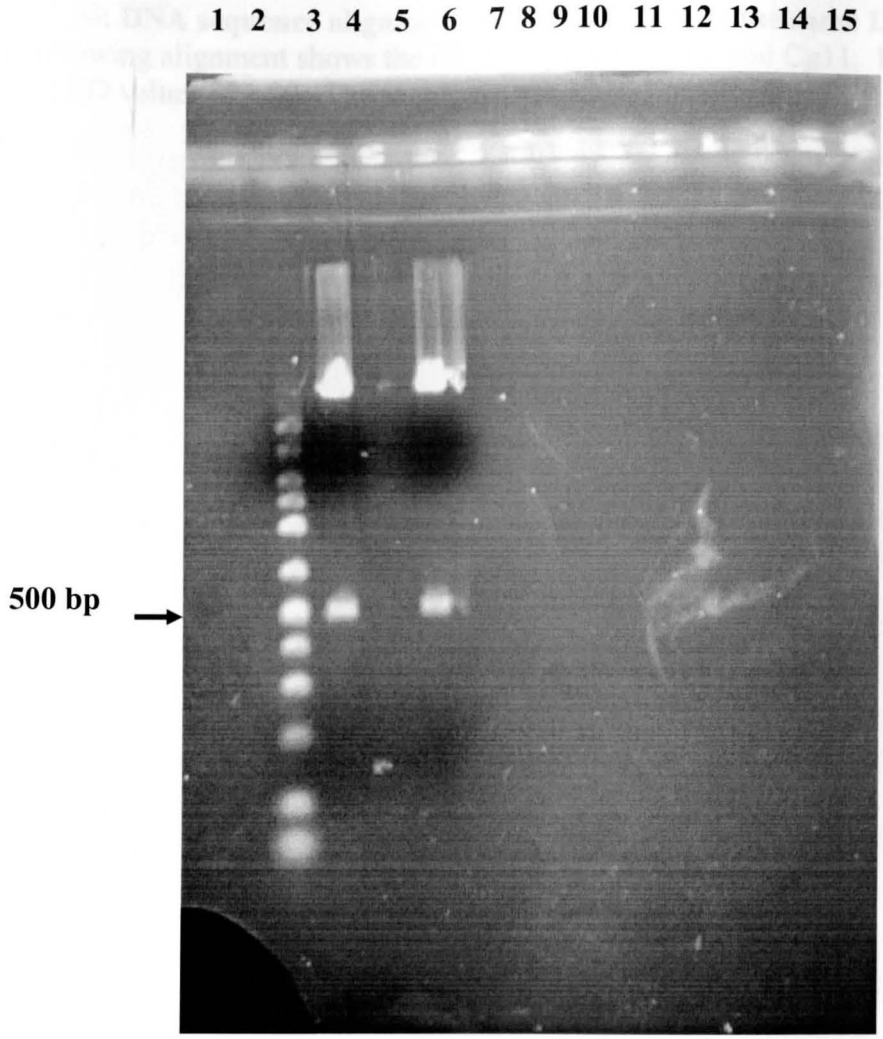

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
500 bp

Figure 10: DNA sequence alignment of two isolates with the same DGGE RD value. The following alignment shows the DNA sequence of Cd1 and Cg11. Both samples had DGGE RD values of 1.00. The sequences are 100% homologous.

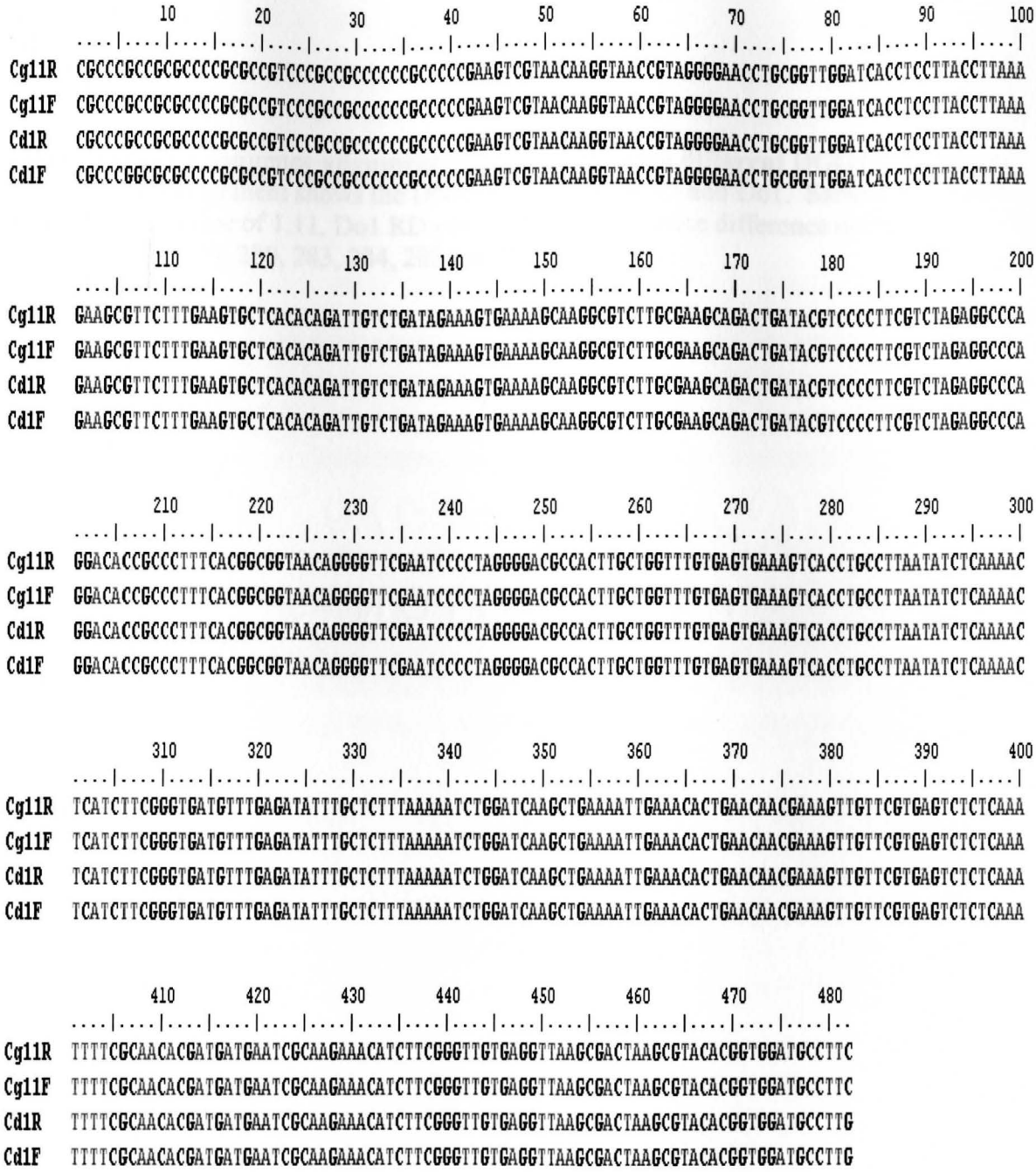

10 20 30 40 50 60 70 80 90 100
Cg11R CGCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAA
Cg11F CGCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAA
Cd1R CGCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAA
Cd1F CGCCCGGCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAA
110 120 130 140 150 160 170 180 190 200
Cg11R GAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCCA
Cg11F GAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCCA
Cd1R GAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCCA
Cd1F GAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCCA
210 220 230 240 250 260 270 280 290 300
Cg11R GGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAAC
Cg11F GGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAAC
Cd1R GGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAAC
Cd1F GGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAAC
310 320 330 340 350 360 370 380 390 400
Cg11R TCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAAA
Cg11F TCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAAA
Cd1R TCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAAA
Cd1F TCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAAA
410 420 430 440 450 460 470 480
Cg11R TTTTCGCAACACGATGATGAATCGCAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCTTC
Cg11F TTTTCGCAACACGATGATGAATCGCAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCTTC
Cd1R TTTTCGCAACACGATGATGAATCGCAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCTTG
Cd1F TTTTCGCAACACGATGATGAATCGCAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCTTG

Figure 11: DNA sequence alignment of two isolates with different DGGE RD values. The following alignment shows the DNA sequence of Sw15 and Do1. Sample Sw15 had an DGGE RD value of 1.11, Do1 RD value of 1.05. Sequence difference occur at positions 177, 220, 280, 283, 284, 288 and 379.

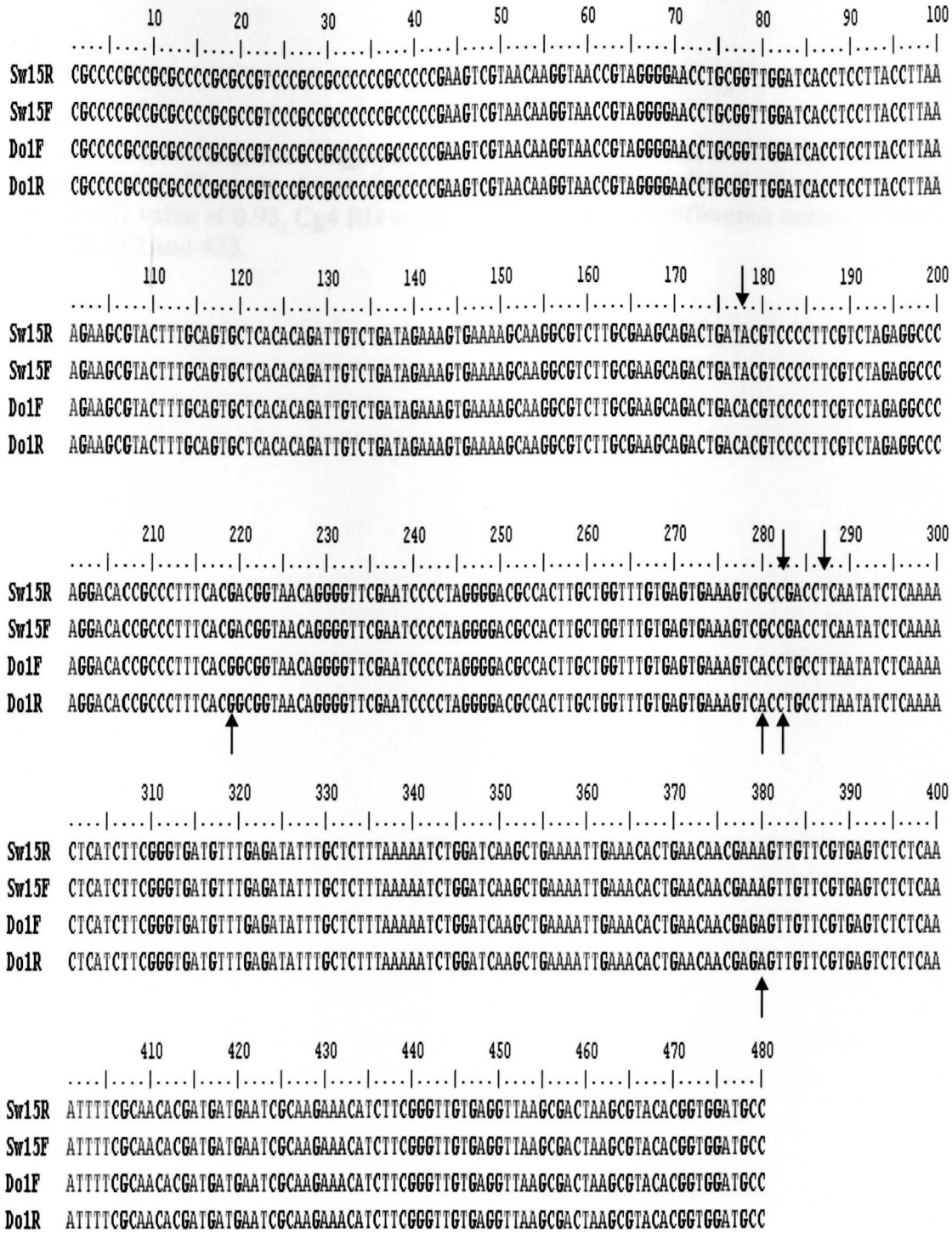

10 20 30 40 50 60 70 80 90 100
Sw15R CGCCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAA
Sw15F CGCCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAA
Do1F CGCCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAA
Do1R CGCCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAA
110 120 130 140 150 160 170 180 190 200
Sw15R AGAAGCGTACTTTGCAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCC
Sw15F AGAAGCGTACTTTGCAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCC
Do1F AGAAGCGTACTTTGCAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGACACGTCCCCTTCGTCTAGAGGCCC
Do1R AGAAGCGTACTTTGCAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGACACGTCCCCTTCGTCTAGAGGCCC
210 220 230 240 250 260 270 280 290 300
Sw15R AGGACACCGCCCTTTCACGACGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCGCCGACCTCAATATCTCAAAA
Sw15F AGGACACCGCCCTTTCACGACGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCGCCGACCTCAATATCTCAAAA
Do1F AGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAA
Do1R AGGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAA
310 320 330 340 350 360 370 380 390 400
Sw15R CTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAA
Sw15F CTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAA
Do1F CTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAGAGTTGTTCGTGAGTCTCTCAA
Do1R CTCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAGAGTTGTTCGTGAGTCTCTCAA
410 420 430 440 450 460 470 480
Sw15R ATTTTCGCAACACGATGATGAATCGCAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCC
Sw15F ATTTTCGCAACACGATGATGAATCGCAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCC
Do1F ATTTTCGCAACACGATGATGAATCGCAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCC
Do1R ATTTTCGCAACACGATGATGAATCGCAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCC

Figure 12: DNA sequence alignment of two isolates with different DGGE RD values. The following alignment shows the DNA sequence of Cg4 and Sw5. Sample Sw5 had a DGGE RD value of 0.93, Cg4 RD value of 1.05. Sequence difference occurs at positions 64, 378, 417 and 425.

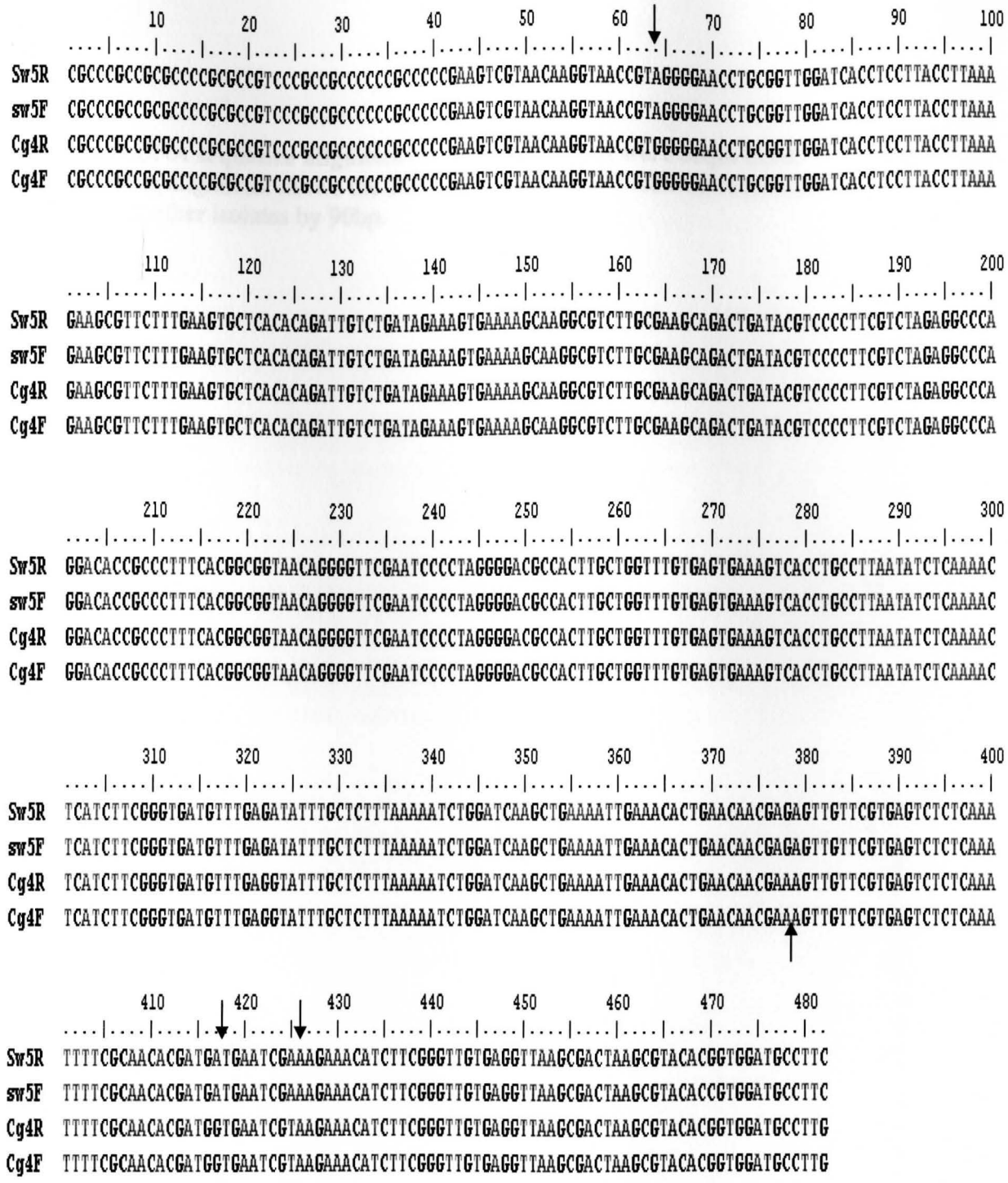

10 20 30 40 50 60 70 80 90 100
Sw5R CGCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAA
sw5F CGCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAA
Cg4R CGCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTGGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAA
Cg4F CGCCCGCCGCGCCCCGCGCCGTCCCGCCGCCCCCCGCCCCCGAAGTCGTAACAAGGTAACCGTGGGGGAACCTGCGGTTGGATCACCTCCTTACCTTAAA
110 120 130 140 150 160 170 180 190 200
Sw5R GAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCCA
sw5F GAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCCA
Cg4R GAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCCA
Cg4F GAAGCGTTCTTTGAAGTGCTCACACAGATTGTCTGATAGAAAGTGAAAAGCAAGGCGTCTTGCGAAGCAGACTGATACGTCCCCTTCGTCTAGAGGCCCA
210 220 230 240 250 260 270 280 290 300
Sw5R GGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAAC
sw5F GGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAAC
Cg4R GGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAAC
Cg4F GGACACCGCCCTTTCACGGCGGTAACAGGGGTTCGAATCCCCTAGGGGACGCCACTTGCTGGTTTGTGAGTGAAAGTCACCTGCCTTAATATCTCAAAAC
310 320 330 340 350 360 370 380 390 400
Sw5R TCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAGAGTTGTTCGTGAGTCTCTCAAA
sw5F TCATCTTCGGGTGATGTTTGAGATATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAGAGTTGTTCGTGAGTCTCTCAAA
Cg4R TCATCTTCGGGTGATGTTTGAGGTATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAAA
Cg4F TCATCTTCGGGTGATGTTTGAGGTATTTGCTCTTTAAAAATCTGGATCAAGCTGAAAATTGAAACACTGAACAACGAAAGTTGTTCGTGAGTCTCTCAAA
410 420 430 440 450 460 470 480
Sw5R TTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCTTC
sw5F TTTTCGCAACACGATGATGAATCGAAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACCGTGGATGCCTTC
Cg4R TTTTCGCAACACGATGGTGAATCGTAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCTTG
Cg4F TTTTCGCAACACGATGGTGAATCGTAAGAAACATCTTCGGGTTGTGAGGTTAAGCGACTAAGCGTACACGGTGGATGCCTTG

Figure 13: DNA sequence alignment of all isolates that were sequenced.
The following alignment shows the DNA sequence of all 7 isolates sequenced. Sw4 differs by all other isolates by 90bp.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5         15         25         35         45         55
Cg11F           CGCCCGCCGC GCCCCGCGCC GTCCCGCCGC CCCCCGCCCC CGAAGTCGTA ACAAGGTAAC
Cd1F            CGCCCGGCGC GCCCCGCGCC GTCCCGCCGC CCCCCGCCCC CGAAGTCGTA ACAAGGTAAC
Sw15F           CGCCCGCCGC GCCCCGCGCC GTCCCGCCGC CCCCCGCCCC CGAAGTCGTA ACAAGGTAAC
Do1F            CGCCCGCCGC GCCCCGCGCC GTCCCGCCGC CCCCCGCCCC CGAAGTCGTA ACAAGGTAAC
sw5F            CGCCCGCCGC GCCCCGCGCC GTCCCGCCGC CCCCCGCCCC CGAAGTCGTA ACAAGGTAAC
Cg4F            CGCCCGCCGC GCCCCGCGCC GTCCCGCCGC CCCCCGCCCC CGAAGTCGTA ACAAGGTAAC
Sw4F            CGCCCGCCGC GCCCCGCGCC GTCCCGCCGC CCCCCGCCCC CGAAGTCGTA ACAAGGTAAC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     65         75         85         95        105        115
Cg11F           CGTAGGGGAA CCTGCGGTTG GATCACCTCC TTACCTTAAA GAAGCGTTCT TTGAAGTGCT
Cd1F            CGTAGGGGAA CCTGCGGTTG GATCACCTCC TTACCTTAAA GAAGCGTTCT TTGAAGTGCT
Sw15F           CGTAGGGGAA CCTGCGGTTG GATCACCTCC TTACCTTAAA GAAGCGTACT TTGCAGTGCT
Do1F            CGTAGGGGAA CCTGCGGTTG GATCACCTCC TTACCTTAAA GAAGCGTACT TTGCAGTGCT
sw5F            CGTAGGGGAA CCTGCGGTTG GATCACCTCC TTACCTTAAA GAAGCGTTCT TTGAAGTGCT
Cg4F            CGTGGGGGAA CCTGCGGTTG GATCACCTCC TTACCTTAAA GAAGCGTTCT TTGAAGTGCT
Sw4F            CGTAGGGGAA CCTGCGGTTG GATCACCTCC TTACCCTAAA GAAGCGTACT TTGTAGTGCT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    125        135        145        155        165        175
Cg11F           CACACAGATT GTCTGATAGA AAGTGAAAAG CAAGGCGTCT TGCGAAGCAG ACTGATACGT
Cd1F            CACACAGATT GTCTGATAGA AAGTGAAAAG CAAGGCGTCT TGCGAAGCAG ACTGATACGT
Sw15F           CACACAGATT GTCTGATAGA AAGTGAAAAG CAAGGCGTCT TGCGAAGCAG ACTGATACGT
Do1F            CACACAGATT GTCTGATAGA AAGTGAAAAG CAAGGCGTCT TGCGAAGCAG ACTGACACGT
sw5F            CACACAGATT GTCTGATAGA AAGTGAAAAG CAAGGCGTCT TGCGAAGCAG ACTGATACGT
Cg4F            CACACAGATT GTCTGATAGA AAGTGAAAAG CAAGGCGTCT TGCGAAGCAG ACTGATACGT
Sw4F            CACACAGATT GTCTGATAGA AAGTGAAAAG CAAGGCGTTT ACGCGTTGGG AGTGAGGCTG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    185        195        205        215        225        235
Cg11F           CCCCTTCGTC TAGAGGCCCA GGACACCGCC CTTTCACGGC GGTAACAGGG GTTCGAATCC
Cd1F            CCCCTTCGTC TAGAGGCCCA GGACACCGCC CTTTCACGGC GGTAACAGGG GTTCGAATCC
Sw15F           CCCCTTCGTC TAGAGGCCCA GGACACCGCC CTTTCACGAC GGTAACAGGG GTTCGAATCC
Do1F            CCCCTTCGTC TAGAGGCCCA GGACACCGCC CTTTCACGGC GGTAACAGGG GTTCGAATCC
sw5F            CCCCTTCGTC TAGAGGCCCA GGACACCGCC CTTTCACGGC GGTAACAGGG GTTCGAATCC
Cg4F            CCCCTTCGTC TAGAGGCCCA GGACACCGCC CTTTCACGGC GGTAACAGGG GTTCGAATCC
Sw4F            AAGAGAATAA GGCCGTTCGC TTTCTATTAA TGAAAGCTCA CCCTACACGA AAATATCACG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    245        255        265        275        285        295
Cg11F           CCTAGGGGAC GCCACTTGCT GGTTTGTGAG TGAAAGTCAC CTGCCTTAAT ATCTCAAAAC
Cd1F            CCTAGGGGAC GCCACTTGCT GGTTTGTGAG TGAAAGTCAC CTGCCTTAAT ATCTCAAAAC
Sw15F           CCTAGGGGAC GCCACTTGCT GGTTTGTGAG TGAAAGTCGC CGACCTCAAT ATCTCAAAAC
Do1F            CCTAGGGGAC GCCACTTGCT GGTTTGTGAG TGAAAGTCAC CTGCCTTAAT ATCTCAAAAC
sw5F            CCTAGGGGAC GCCACTTGCT GGTTTGTGAG TGAAAGTCAC CTGCCTTAAT ATCTCAAAAC
Cg4F            CCTAGGGGAC GCCACTTGCT GGTTTGTGAG TGAAAGTCAC CTGCCTTAAT ATCTCAAAAC
Sw4F            CAACGCGTGA TAAGCAATTT TCGTGTCCCC TTCGTCTAGA GGCCCAGGAC ACCGCCCTTT
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   305        315        325        335        345        355
Cg11F           TCATCTTCGG GTGATGTTTG AGATATTTGC TCTTTAAAAA TCTGGATCAA GCTGAAAATT
Cd1F            TCATCTTCGG GTGATGTTTG AGATATTTGC TCTTTAAAAA TCTGGATCAA GCTGAAAATT
Sw15F           TCATCTTCGG GTGATGTTTG AGATATTTGC TCTTTAAAAA TCTGGATCAA GCTGAAAATT
Do1F            TCATCTTCGG GTGATGTTTG AGATATTTGC TCTTTAAAAA TCTGGATCAA GCTGAAAATT
sw5F            TCATCTTCGG GTGATGTTTG AGATATTTGC TCTTTAAAAA TCTGGATCAA GCTGAAAATT
Cg4F            TCATCTTCGG GTGATGTTTG AGGTATTTGC TCTTTAAAAA TCTGGATCAA GCTGAAAATT
Sw4F            CACGGCGGTA ACAGGGGTTC GAATCCCCTA GGGGACGCCA CTTGCTGGTT TGTGAGTGAA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   365        375        385        395        405        415
Cg11F           GAAACACTGA ACAACGAAAG TTGTTCGTGA GTCTCTCAAA TTTTCGCAAC ACGATGATGA
Cd1F            GAAACACTGA ACAACGAAAG TTGTTCGTGA GTCTCTCAAA TTTTCGCAAC ACGATGATGA
Sw15F           GAAACACTGA ACAACGAAAG TTGTTCGTGA GTCTCTCAAA TTTTCGCAAC ACGATGATGA
Do1F            GAAACACTGA ACAACGAGAG TTGTTCGTGA GTCTCTCAAA TTTTCGCAAC ACGATGATGA
sw5F            GAAACACTGA ACAACGAGAG TTGTTCGTGA GTCTCTCAAA TTTTCGCAAC ACGATGATGA
Cg4F            GAAACACTGA ACAACGAAAG TTGTTCGTGA GTCTCTCAAA TTTTCGCAAC ACGATGGTGA
Sw4F            AGTCACCTGC CTTAATATCT CAAAACTCAT CTTCGGGTGA TGTTTGAGAT ATTTGCTCTT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   425        435        445        455        465        475
Cg11F           ATCGCAAGAA ACATCTTCGG GTTGTGAGGT TAAGCGACTA AGCGTACACG GTGGATGCCT
Cd1F            ATCGCAAGAA ACATCTTCGG GTTGTGAGGT TAAGCGACTA AGCGTACACG GTGGATGCCT
Sw15F           ATCGCAAGAA ACATCTTCGG GTTGTGAGGT TAAGCGACTA AGCGTACACG GTGGATGCC.
Do1F            ATCGCAAGAA ACATCTTCGG GTTGTGAGGT TAAGCGACTA AGCGTACACG GTGGATGCC.
sw5F            ATCGAAAGAA ACATCTTCGG GTTGTGAGGT TAAGCGACTA AGCGTACACC GTGGATGCCT
Cg4F            ATCGTAAGAA ACATCTTCGG GTTGTGAGGT TAAGCGACTA AGCGTACACG GTGGATGCCT
Sw4F            TAAAAATCTG GATCAAGCTG AAAATTGAAA CACTGAACAA CGAAAGTTGT TCGTGAGTCT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   485        495        505        515        525        535
Cg11F           TC..... ..... ..... ..... ..... .....
Cd1F            TG..... ..... ..... ..... ..... .....
Sw15F           ..... ..... ..... ..... ..... .....
Do1F            ..... ..... ..... ..... ..... .....
sw5F            TC..... ..... ..... ..... ..... .....
Cg4F            TG..... ..... ..... ..... ..... .....
Sw4F            CTCAAATTTT CGCAACACGA TGATGAATCG TAAGAAACAT CTTCGGGTTG TGAGGTTAAG

                ....|....| ....|....| ....|....
                   545        555        565
Cg11F           ..... ..... .....
Cd1F            ..... ..... .....
Sw15F           ..... ..... .....
Do1F            ..... ..... .....
sw5F            ..... ..... .....
Cg4F            ..... ..... .....
Sw4F            CGACTAAAGC GTACACGGTG GATGCCTTG
```

DGGE analysis is sensitive to single nucleotide alterations in PCR products, and is a sufficient method to compare PCR products.

Long Term Cultures:

Three individual *E. coli* isolates, Hu01, Cd01 and Sw01 were subjected to multiple culture transfers to determine if any natural changes in the *rrnB* 16S-23S ISR DNA sequence occur over time. This is important aspect to examine for future studies using environmental samples. Each of the three isolates DNA was isolated at day 0, and used for PCR-DGGE analysis. The isolates were then transferred between cultures of LB-broth and LB-plates eleven times. The DNA was then isolated from the *E. coli* isolates and examined for any differences from the day 0 isolates by PCR-DGGE. The isolates were then transferred between the two medias 11 more times. Once again the DNA was isolated form the isolates and subjected to PCR-DGGE analysis. Table 5 shows the resulting DGGE gel of the long term cultures compared to the day 0 cultures.

Duplicate runs of several isolates were performed to measure the amount of reproducibility in the method. Each isolate was independently subjected to PCR amplification of the ISR for DGGE analysis. The results show a high degree of reproducibility. Almost all RD values are within 0.02 of each other (table 6).

Table 5: Results of the long term culture studies.
The three long term cultures, human (Hu01), swine (Sw01) and dairy cow (Cd01) are listed showing there day 0 RD values and there RD values after 11 culture changes from LB broth to LB plates.

	RD Value	
Source Isolate	**Day 0**	**5 Culture Changes**
Hu01	0.68	0.76
Cd01	1	0.87
Sw01	0.99	1.04

Table 6: Results of duplicate DGGE results from independently prepared samples compared.

The calculated RD values are given for each of two trials for the isolates that were tested for reproducibility.

Source Isolate	RD Value	
	Trial 1	Trial 2
Sw01	1.01	0.99
Sw03	0.96	0.94
Sw06	1.01	1.01
Sw07	0.89	0.87
Sw08	0.96	0.97
Sw09	0.88	0.88
Sw11	0.88	0.88
Sw12	0.94	0.95
Sw13	1.00	1.01
Cd01	1.00	1.00
De10	0.9	0.9
Ho4	0.87	0.87
Ho5	0.88	0.88
Ho12	0.88	0.87

Chapter 4: Discussion

The results of the PCR-DGGE analysis of the *rrnB* 16S-23S ISR of *E. coli* isolates from eight different animal sources reveals that this method can differentiate between isolates from different sources. Each source *E. coli* isolate tested produced a single band on a denaturing gel corresponding to the DNA sequence of the ISR. The fifteen isolates from each source produced composite genetic profiles of their source based on the unique distribution and abundance of each isolates RD value. There was no correlation between the majority of the sources DGGE genetic profiles. Each source also produced high diversity and evenness of RD values for their isolates tested. This high diversity and evenness, and the fact that there was no correlation between most of the sources, implies that the profiles produced were actually unique to the individual sources, and that there is a high probability that any additional isolates tested would remain unique to a particular source. Additionally, when all nonhuman sources were grouped together, there was no correlation with the DGGE profiles produced compared to those of humans. The diversity of the nonhuman sources actually increased compared to any individual nonhuman source. These results support using this method to differentiate between human and nonhuman *E. coli* isolates.

The DGGE results also revealed a high amount of genetic diversity for the *E. coli* isolates from each of the 8 different sources. Many MST methods have reported finding high diversity in *E. coli* isolates tested (Buchan et. al., 2001, Seurnick et al., 2003, McLellan et al., 2003, Jarvis et al.,2000) using various techniques. Many factors influence the genetic variations found in *E. coli* populations. Host specificity, in

particular, is an important factor influencing *E. coli* populations (McLellan et al., 2003). Buchan et al. found high *E. coli* genetic diversity when performing DGGE analysis of the 16S-23S ISR using primers that amplified all 7 ISR. Of the 132 *E. coli* isolates examined, 84 unique DGGE banding patterns were identified. Similarly, when performing the same ISR-DGGE analysis with primers for all 7 *rrn* operons, Seurnick et. al. identified 87 ISR fingerprints out of a total of 267 isolates examined. The results of our study show that each source had a different *H'* value, and they were all on the high end. Out of the 119 *E. coli* isolates evaluated by the PCR-DGGE of the single *rrnB* ISR, only 40 DGGE bands were observed. Additionally, when our samples were grouped as human or nohuman sources, the nonhuman sources had higher diversity than the human sources. The nonhuman sources also had a higher diversity than any individual nonhuman source. The higher diversity observed in nonhuman sources has been associated with the wide host range of all possible nonhuman sources (Parveen et al., 1999), and strain adaptation to various wild hosts from different regions has been shown to be an important factor in *E. coli* population structure (Souza et al. 1999).

Host specificity refers to the presence of dominant clonal groups of *E. coli* found in a particular host. The Pearson correlation coefficient analysis performed on the eight *E. coli* source ISR DGGE bands showed almost no correlation between the different sources. Many factors are attributed to the occurrence of host specificity for *E. coli.* When examining the ISR in particular, there are stretches of nonfunctional DNA that are present, and these regions should exhibit a considerable degree of variation due to genetic drift (Garcia-Martinez et. al 1996a). Therefore, it is likely that unique ISR sequences may dominate specific animal sources based on genetic drift. The diet of various host

sources also attributes to the host specificity. Recent studies have shown that diet has affected the *E. coli* populations of various hosts (Jarvis et al., 2000, Hartel et al., 2003). Furthermore, the types of sugars that are used by *E. coli* have been shown to be associated with the taxonomic group of the host from which the isolates were obtained (Souza et al., 1999). There are also other various factors between sources that can lead to difference in *E. coli* isolates. The temperature and pH of the host's microenvironment are two important aspects that affect the *E. coli* strains present (Carson et al., 2001). The differences in diets and microenvironments may therefore contribute to host specificity and the genetic drift observed in our *E. coli* samples. Host specificity has only been reported to account for some of the observed diversity of *E. coli* populations, while the the extent to which the host influences the gentic composition of *E. coli* is still unknown (McLellan et al., 2003). Nonetheless, *E. coli* isolates have still been reported to be correctly classified to host sources, and candidate specific genetic fingerprints have been identified using a variety of MST methods.

The main goal of this investigation was to evaluate the ability of a MST method to differentiate between *E. coli* isolates form various sources for future work of testing environmental samples based on the results of the DGGE analysis. The high amount of diversity among the sources tested is a positive result, along with the high evenness of the DGGE band distribution for the sources tested. The low amount of correlation between sources also adds to the potential of this method for evaluating environmental samples. However, the ultimate success of this method is still limited by the inadequate amount of information available on the fate of *E. coli* in the environment. Specifically, the stability of the genetic marker, the ISR, needs to be further investigated. It has been estimated

that a typical *E. coli* bacterium spends half of its life outside of the host in the external environment, and that fate of the clones in the external environment is poorly understood (Gordon, 2001). Furthermore, there appears to be a substantial amount of change in the community composition during the transition from the host to the environment. The conditions in the external environment that differ from the host, and may affect the clonal composition of *E. coli,* include differences in temperature, pH, nutrients, oxygen concentrations and solar irradiation (Buchan et al., 2001). However, isolates from the same source have still been reported to give the same profiles (Buchan et al., 2001). Therefore, the high observed diversity, along with the low correlations between RD values in this investigation, supports using this MST method to analyze environmental samples by comparing collected isolates from the environment with isolates from the presumed source.

Currently, most MST methods require the comparison of environmental samples to a developed host library to determine the source of the contamination. Using a host library has potential drawbacks, which include the possibility that the isolates in the library may overestimate the frequency of a particular strain in the overall population (McLellan et al., 2003). This disadvantage, along with the potential success of our MST method, suggests that comparing different environmental isolates will be adequate to determine the source. However, isolates collected from a stream have been shown to have higher diversity than individual source isolates (Buchan et al., 2001). This may be expected due to the increase in potential sources of *E. coli* in the environment, and the possibility of genetic changes occurring from the transition from host to the environment. Possible geographic and temporal genetic variations that may occur in isolates from the

same host source from different locations makes comparisons to a host library difficult. This problem could be overcome by simply comparing environmental sources from their presumed source in the same geographical area. In fact, it has been stated that given the high amount of *E. coli* strain diversity, isolate characterization may be most feasible within a limited geographical area such as a watershed (McLellan et al., 2003). The ability of our MST method to show no correlation between *E. coli* isolates from various sources with high diversities suggests that the PCR-DGGE method may be able to distinguish environmental samples, even if there diversities are higher than source isolates. Additionally, the repeatability of our method, as shown by the ability to produce the same DGGE bands when individual isolates were independently analyzed by DGGE, demonstrates the reproducibility of this technique.

Summary:

The results of this investigation show promise of 16S-23S ISR DGGE as an effective method for microbial source tracking. DGGE analysis of each *E. coli* source isolate resulted in the production of a genetic profile based on the DNA sequence of their ISR. This finding supports using the ISR as a genetic target for source tracking, and that the sequence variability found in the ISR is sufficient for differentiating isolates from the same bacterial species. Additionally, these results show that *E. coli* demonstrates enough genetic variation to be considered a good candidate for an indicator organism. For each of the sources, the diversity (*H'*) and evenness (*J'*) of banding profiles was relatively high, as would be expected from the reported high genetic diversity of *E. coli*. This makes it difficult to determine were in a specific host group a new isolate would be placed, if further tests were performed. However, the high diversity is important when considering that almost all of the profiles were shown to have no correlation. It can therefore be concluded that each source group produced unique genetic profiles compared to the other sources, excluding the four sets that were shown to be correlated.

There are many conditions that have to be meet in order for a MST method to be successful. The existence of host specificity of the indicator organism is vital to MST methods. This research shows that there appears to be host specificity for most of the sources. Each animal source had unique DGGE bands in their profile. However, the results presented in this study are from a moderately small sample set. Therefore, the results show that further investigation of source isolates by the PCR-DGGE method is merited.

The results of this investigation also shows that it may be possible to overcome some of the problems associated with other MST methods that utilize source reference libraries to assign unknown isolates to a particular source. Reference libraries require vast amounts of isolates from each source to be tested, due to the possible genetic variation in isolates from different geographical locations, and with different diets. Also, reference libraries may over represent the frequency of a particular strain. The apparent ability of our PCR-DGGE method to differentiate isolates from 8 different animal sources shows that the technique can distinguish between two *E. coli* populations. This will strengthen any field results that show similarities, or dissimilarities between DGGE profiles, regardless of the source. Therefore, the use of comparing samples to a reference library is not essential, and any results from analyzing a reference library only support the ability to distinguish possible sources. This supports that using our PCR-DGGE method should be sufficient for concluding if two samples are indeed from the same location.

The stability of the indicator organism and the genetic target in the environment are very important for all MST methods. The conditions an *E. coli* isolate face from the transition from host to environment differ greatly, and may affect the abilities of a MST method. The 16S-23S ISR examined in this study is part of an operon found on the *E. coli* chromosome that is crucial to cell survival. The amount of variation in this region that may come about by genetic recombination in the environment is still not completely understood. However, the ISR as a genetic marker may be a more stable genetic marker than others previously studied. Other genetic targets, such as antibiotic resistance, are present on plasmids and often exchanged in the environment (Guan et al., 2002, Scott et al., 2002).

Appendix A Isolate Name Chart and Individual RD Values

Canadian Goose (Cg01-15)	Relative Distance
Cg01	0.55
Cg02	0.55
Cg03	0.61
Cg04	0.96
Cg05	0.73
Cg06	0.55
Cg07	0.95
Cg08	0.95
Cg09	0.75
Cg10	1.1
Cg11	1
Cg12	0.77
Cg13	0.75
Cg14	0.75
Cg15	0.95

Swine (Sw01-15)	Relative Distance
Sw01	0.99
Sw02	0.93
Sw03	0.94
Sw04	0.84
Sw05	0.93
Sw06	1.01
Sw07	0.87
Sw08	0.97
Sw09	0.88
Sw10	0.86
Sw11	0.88
Sw12	0.95
Sw13	1.01
Sw14	0.88
Sw15	1.11

Human (Hu01-Hu15)	Relative Distance
Hu01	0.68
Hu02	0.98
Hu03	0.82
Hu04	0.68
Hu05	0.98
Hu06	1.2
Hu07	0.89
Hu08	1.09
Hu09	0.94
Hu10	0.87
Hu11	1.01
Hu12	1.05
Hu13	0.82
Hu14	0.7
Hu15	1.07

Cattle Dairy (Cd01-07) Beef (Cb01-08)	Relative Distance
Cd01	1
Cd02	0.89
Cd03	0.95
Cd04	0.82
Cd05	0.9
Cd06	0.91
Cd07	1.07
Cb01	0.72
Cb02	0.75
Cb03	0.85
Cb04	0.85
Cb05	0.86
Cb06	0.87
Cb07	0.96
Cb08	1.02

Deer (De01-15)	Relative Distance
De01	0.91
De02	1.07
De03	0.85
De04	0.9
De05	0.78
De06	1.01
De07	0.89
De08	0.8
De09	1
De10	0.9
De11	0.94
De12	NA
De13	0.91
De14	0.9
De15	0.85

Dog (Do01-Do15)	Relative Distance
Do01	1.05
Do02	0.85
Do03	0.85
Do04	0.94
Do05	0.85
Do06	0.9
Do07	0.99
Do08	0.94
Do09	1.06
Do10	0.9
Do11	0.94
Do12	1.15
Do13	1.03
Do14	0.94
Do15	0.69

Horse (Ho01-Ho15)	Relative Distance
Ho01	0.95
Ho02	0.81
Ho03	1
Ho04	0.87
Ho05	0.88
Ho06	0.85
Ho07	0.92
Ho08	1.03
Ho09	0.92
Ho10	1
Ho11	1
Ho12	0.87
Ho13	0.93
Ho14	0.96
Ho15	0.99

Chicken (Ch01-15)	Relative Distance
Ch01	0.76
Ch02	0.94
Ch03	0.85
Ch04	0.91
Ch05	0.88
Ch06	0.88
Ch07	1
Ch08	0.83
Ch09	0.9
Ch10	0.95
Ch11	0.95
Ch12	0.95
Ch13	1.02
Ch14	0.91
Ch15	0.91

References:

1.) **Antón, A. I., A. J Martínez-Murcia, F. Rodríguez-Valera.** 1998. Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the Escherichia coli ECOR collection. Journal of Molecular Evolution. 47:62-72.

2.) **Brosius, J., T. J. Dull, D. D Sleeter and H. F. Noller.** 1981. Gene organization and primary structure of a ribosomal RNA operon from Escherichia coli. Journal of Molecular Biology. **148**:107-127.

3.) **Buchan, A., M. Alber, and R. E. Hodsom.** 2001. Strain-specific differentiation of environmental *Escherichia coli* isolates via denaturing gradient gel electrophoresis (DGGE) analysis of the 16S-23S intergenic spacer region. FEMS Microbial Ecology. 35:313-321.

4.) **Carson, C. A., B. L. Shear, M. R. Ellersieck and A. Asfaw.** 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. Applied and Environmental Microbiology. **67:**1503-1507.

5.) **Cilia, V., B. Lafay and R. Christen.** 1996. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effects on phylogenetic analyses at the species level. Molecular Biology Evolution. **13:**451-461.

6.) **Dombek, P. E., L. K. Johnson, S. T. Zimmerley and M. J. Sadowsky.** 2000. Use of repetitive DNA sequence and the PCR to Differentiate Escherichia coli isolates from human and animal sources. Applied and Environmental Microbiology. **66:**2572-2577.

7.) **Farleitner, A. H., N. Kreuzinger, G. G. Kavka, S. Grillenberger, J. Rath and R. L. Mach.** 2000. Simultaneous detection and differentiation of Escherichia coli populations from environmental freshwaters by means of sequence variations in a fragment of the ß-D-Glucuronidase gene. Applied and Environmental Microbiology. **66:**1340-1346.

8.) **Fox, K. F., M. Nagpal, P. Steinberg and K. Heroux.** 1998. Identification of *Brucella* by ribosomal-spacer-region PCR and differentiation of *Brucella canis* from other *Brucella* spp. Pathogenic for humans by carbohydrate profiles. Journal of Clinical Microbiology. 36:3217-3222.

9.) **Fukushima, M., K. Kakinuma and R. Kawaguchi**. 2002. Phylogenetic analysis of Salmonella, Shigella, and Escherichia coli strains on the basis of the gyrB gene sequence. Journal of Clinical Microbiology. **40**:2779-2785.

10.) **García-Martínez, J., A. Matínez-Murcia, A. I. Antón and F. Rodríguez-Valera.** 1996a. Comparison of the small 16S to 23S intergenic spacer region (ISR) of the rRNA operons of some Escherichia coli strains of the ECOR collection and E. coli K-12. Journal of Bacteriology. **178:**6374-6377.

11.) **García-Martínez, J., A. Matínez-Murcia, A., F. Rodríguez-Valera and I. Antón.** 1996b. Molecular evidence supporting the existence of two major groups in uropathogenic *Escherichia coli*. FEMS Immunology and Medical Microbiology. **14:**231-244.

12.) **García-Martínez, J., S. G. Acinas, A. I. Antón and F. Rodríguez-Valera.** 1999. Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. Journal of Microbiological Methods. **36:** 55-64.

13.) **Gordon, David M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiology. **147:**1079-1085.

14.) **Guan, S., R. Xu, S. Chen, J. Odumeru and C. Gyles.** 2002. Development of a procedure for discriminating among Escherichia coli isolates form animal and human sources. Applied and Environmental Microbiology. **68:**2690-2698.

15.) **Hartel, G. P., J. D. Summer and W. I. Segars.** 2003. Deer diet affects ribotypes diversity of *Escherichia coli* for bacterial source tracking. Water Research. **37:**3263-3268.

16.) **Jarvis, G.N., M.G. Kizoulis, F. Diez-Gonzalez, and J. B. Russel.** 2000. The genetic diversity of predominant *Escherichia coli* strains isolated from cattle fed various amounts of hay and grain. FEMS Microbiology Ecology. **32:**225-233.

17.) **Jensen, M. A, J. A. Webster, and N. Straus.** 1993. Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. Applied and Environmental Microbiology. 59:945-952.

18.) **Kuhnert, P., P. Boerlin and J. Frey.** 2000. Target genes for virulence assessment of *Escherichia coli* isolates form water, food and the environment. FEMS Microbiology Reviews. **24**:107-117.

19.) **Leung, K. T., R. Mackereth, Y. C. Tien and E. Topp.** 2004. A comparison of AFLP and ERIC-PCR analyses for discriminating *Escherichia coli* from cattle, pig and human sources. FEMS Microbiology Ecology. **47:**111-119.

21.) **Lia, D.** 2000. Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in Bacteria and Archaea. Journal of Molecular Evolution. **51:**305-317.

22.) **McLellan S. L., A. D. Daniels and A. K. Salmore.** Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution by using DNA fingerprinting. Applied and Environmental Microbiology. **69:**2587-2594.

23.) **Mills A. L., and R. A. Wassel.** 1980. Aspects of diversity measurement for microbial communities. Applied and Environmental Microbiology. **40:**578-586.

24.) **Muyzer, Gerard.** 1999. DGGE/TGGE a method for identifying genes from natural ecosystems. Current Opinion in Microbiology. **2:**317-322.

23.) **Muyzer, G., E. C. De Wall and A. G. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology. **59:**695-700.

25.) **Nagpal, M. L., K. F. Fox and A. Fox.** 1998. Utility of 16S-23S rRNA spacer region methodology: how similar are interspace regions within a genome and between strains for closely related organisms? Journal of Microbiological Methods. **33:**211-219.

26.) **Parveen, S., K. M. Portier, K. Robinson, L. Edmiston and M. L. Tamplin.** 1999. Discriminant analysis of ribotype profiles of Escherichia coli for differentiating human and nonhuman sources of fecal pollution. Applied and Environmental Microbiology. **65:**3142-3147.

27.) **Scheinert, P., R. Krausse, U. Ullmann, R. Soller and G. Krupp.** 1996. Molecular differentiation of bacteria by PCR amplification of the 16S-23S rRNA spacer. Journal of Microbiological Methods. **26:**103-117.

28.) **Scott, T. M., J. B. Rose, T. M. Jenkins, S. R. Farrah and J. Lukasik.** 2002. Microbial source tracking: Current methodology and future directions. Applied and Environmental Microbiology. **68:**5796-5803.

28.) **Seurinck, S., W. Verstraete, and S. D. Siciliano.** 2003. Use of the 16S-23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify Nonpoint fecal sources. Applied and Environmental Microbiology. **69:**4942-4950.

30.) **Sheffield, V. C., D. R. Cox, L. S. Lerman and R. M. Myers.** 1989. Attatchment of a 40-base pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-bases changes. Proceedings of the National Academy of Science. **86:**232-236.

31.) **Souza, V., M. Rocha, A. Valera, and L.E. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. Applied and Environmental Microbiology. **65:**3373-3385.