THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF VETERINARY AND BIOMEDICAL SCIENCES

COMPARATIVE SEQUENCE ANALYSIS OF NON-TYPABLE H ANTIGENS OF PATHOGENIC SHIGA TOXIN-PRODUCING *ESCHERICHIA COLI*

SAMANTHA FANELLI
SPRING 2013

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Veterinary and Biomedical Sciences
with honors in Veterinary and Biomedical Sciences

Reviewed and approved* by the following:

Chitrita DebRoy
Director of E. Coli Reference Center
Thesis Supervisor

Lester C. Griel Jr.
Professor of Veterinary and Biomedical Sciences
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Escherichia coli bacteria populate a vast proportion of the gut, thus studying and appreciating the vast differences between strains would be to the benefit of the field of human medicine. Potentially pathogenic Shiga toxin-producing *E. coli* (n = 22) isolated from leafy green vegetables that could not be classified using standard H typing protocol, were analyzed in an attempt to determine if these strains were simply mutants or could be grouped together and classified as a new H type. The *fliC* gene, that encodes for the flagellar antigen, was amplified, sequenced, and comparative analyses were conducted to determine the similarities and differences in the sequences of this gene. Of these samples, 5 were almost identical in their *fliC* sequences. Further comparative analysis of the DNA sequences confirmed they were variants of the H7 serotype group that exhibited 26 base pair substitutions that resulted in 21 amino acid substitutions. This is an important finding, as it is a key to their mobility and potential pathogenicity and an indicator of the significance just a few base pair substitutions can have. This also opens doors for further research into the specific impact of amino acid substitutions on mobility and potential pathogenicity, in host species such as leafy vegetables.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to formally thank and acknowledge Dr. Chobi DebRoy and Dr. Lester Griel for editing and approving this thesis. Without your help, I would have been lost. I would also like to thank my friends and family for their continued support of my academic pursuits and most importantly, the Schreyer Honors College for endowing me with the skills and knowledge base to achieve my goals beyond all expectation. Beth “Chief Elf” Roberts, thank you is not enough.

Chapter 1

Introduction

Gram-negative *Escherichia coli* are one of the most prevalent bacteria in the gut flora, and the numerous varieties can be classified based on two major antigens: O polysaccharide (O-antigen) and flagellin (H-antigen). *E. coli* strains have been classified into 181 O types and 53 H types since the 1940's.[1, 2] H-antigen typing relies on difference in genes encoding the proteins that make up *E. coli's* motility mechanism, the flagellum. The flagellum is an organelle projecting from the cell that allows for movement of the bacterium by rotation. The complete assembly, configuration, and operation of the flagella is dictated by over 40 genes, the most important of which being the *fliC* gene.

The *fliC* gene encodes flagellin, the repeated protein subunit of the filament of the flagella, responsible for H antigen specificity. While the flagellin is fairly conserved in the terminal regions, variability exists in the central region. These variations are in regions that encode for the portion of the protein that is exposed to the surface and specific to a particular H-type. Agglutination tests can be used to distinguish these different H-types using H-specific antibodies. The *fliC* gene can also be easily amplified through polymerase chain reaction (PCR) due to the tremendous sequence conservation of the distal portions of the *fliC* alleles. Subsequent restriction digestion of amplicons or DNA sequencing can reveal the H types.[3] Frequently, however, strains do not fall easily into a known H-type and are labeled "untypeable." These untypeable strains exacerbate that difficulty to classify pathogenic strains and to study trends in microbial diseases.

Classification of untypeable strains into new H-types can make distinction and prediction of pathogenic properties of strains more conclusive. Research into untypeable strain classification can offer insight into the prevalence or rarity of these potential new H-types or possibly reveal unique mutations that affect flagellar structure. Information gathered from this research could be applied world-wide to microbiology research laboratories as well as epidemiological study centers, especially in the case of new H-types. The global implications of H-typing research make it extremely important in the field of microbiology and a serious endeavor to pursue.

Chapter 2

Materials and Methods

Sample Preparation

Shiga toxin-producing *Escherichia coli* (n = 22) that exhibited the presence of the *fliC* gene but could not be classified into any of the 53 H groups were taken from The Pennsylvania State University *E. coli* Reference Center's repository (see Table 2-1). These samples were grown individually on tryptic soy agar plates, and the DNA was isolated and cellular debris removed to prepare samples for further analyses.

Table 2-1. Reference Strains Studied. *Escherichia coli* samples deemed untypeable were further analyzed by *fliC* gene sequencing to determine the H group.

Sample Reference Number	O-Type	Species	Sample Reference Number	O-Type	Species
11.1701	1	Spinach	12.3163	74	Spinach
11.1819	1	Spinach	12.3164	74	Spinach
11.1924	76	Spinach	12.3165	113	Spinach
12.2366	145	Cow	12.3166	74	Spinach
12.2788	Untypeable	Spinach	12.3167	74	Spinach
12.2855	Untypeable	Spring Mix	12.3235	Untypeable	Spinach
12.2856	Untypeable	Spring Mix	12.3236	Untypeable	Spinach
12.2857	Untypeable	Spring Mix	12.3237	Untypeable	Spinach
12.2858	Untypeable	Spring Mix	12.3551	76	Food
12.2859	Untypeable	Spring Mix	12.3552	76	Food
12.2860	Untypeable	Spring Mix	12.2649	73	Spinach

Polymerase Chain Reaction

Each isolate underwent Polymerase Chain Reaction (PCR) to amplify the DNA product in preparation for sequencing. Amplification was achieved by several cycles of denaturing the

DNA strands, adding in priming template strands to build new ones, annealing them together, and elongating the ends.[5] Amplification of the DNA allows for a larger quantity of viable DNA strands for further analyses. Specific conditions can be found in Appendix A. The amplified DNA samples were resolved on 1% agarose gel and visualized by UV photography using a gel imaging system from Kodak. *E. coli* with H type H2 was used as a positive control sample, and *Salmonella* DT104 was used as a negative control.

Ion Torrent Sequencing

Amplified DNA samples were used for sequencing and purified using a QIAquick ® PCR purification kit. Specific conditions can be found in Appendix A. Samples were then sent to the Genomics Core Facility at Penn State, University Park for Sanger method sequencing to determine the precise sequence of each sample's *fliC* gene. This is achieved by selectively incorporating fluorescently-labeled dideoxynucleotides into a DNA strand as it undergoes replication. These fluorescent labels are then assembled into the correct order of the DNA strand based on length.[6]

BLAST Analysis

Each sample's *fliC* sequence was assembled using SeqMan of the DNASTAR Lasergene 9 Core Suite, and all of the sequences were aligned using the Clustal-W method in MegAlign to identify similarities and differences. Once similar strains were established, their sequences were compared to other *fliC* gene sequences using the National Center for Biotechnology Information's Basic Local Alignment Search Tool (BLAST). Significantly similar sequences found using BLAST were then translated into their corresponding amino acid sequences using SeqBuilder and compared.

Chapter 3

Results

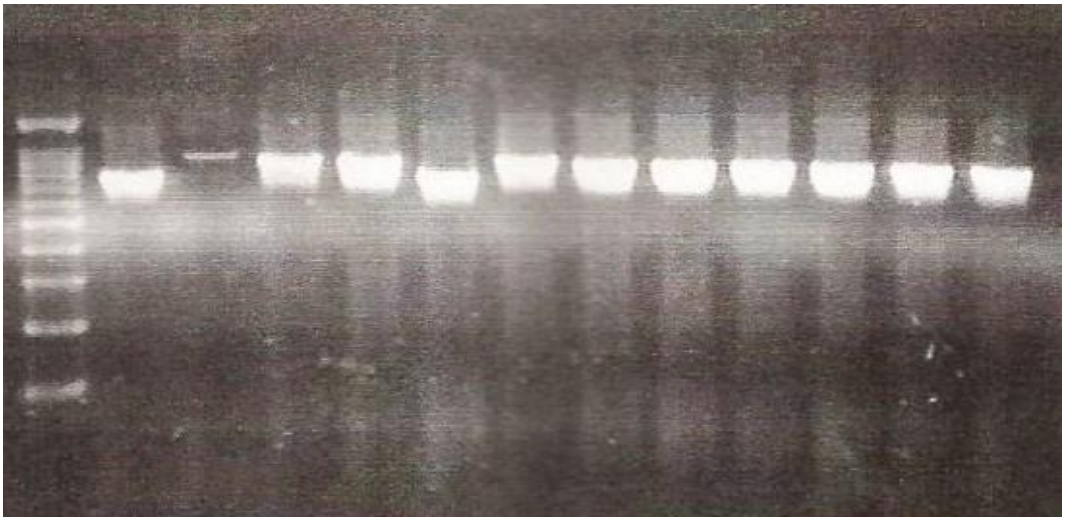

Figure 3-1. Gel electrophoresis results of *fliC* positive *E. coli* isolates. Lane 1 – DNA ladder, Lane 2 – *fliC* positive control (H2), Lane 3 – *fliC* negative control (DT104), Lane 4 – Sample 11.1701, Lane 5 – Sample 11.1819, Lane 6 – Sample 11.1924, Lane 7 – Sample 12.2649, Lane 8 – Sample 12.2366, Lane 9 – Sample 12.2788, Lane 10 – Sample 12.2855, Lane 11 – Sample 12.2856, Lane 12 – Sample 12.2857, Lane 13 – Sample 12.2858.

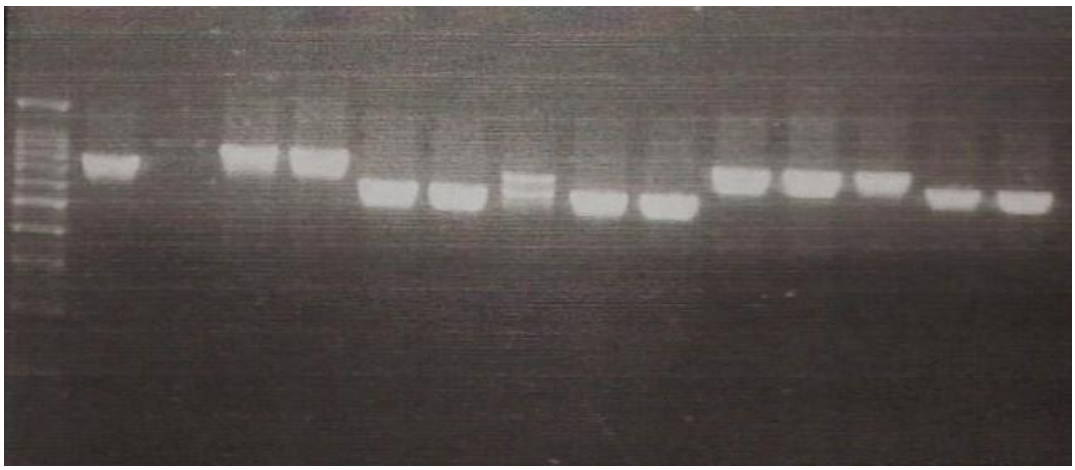

Figure 3-2. Gel electrophoresis results of *fliC* positive *E. coli* isolates. Lane 1 – DNA ladder, Lane 2 – *fliC* positive control (H2), Lane 3 – *fliC* negative control (DT104), Lane 4 – Sample 12.2859, Lane 5 – Sample 12.2860, Lane 6 – Sample 12.3163, Lane 7 – Sample 12.3164, Lane 8 – Sample 12.3165, Lane 9 – Sample 12.3166, Lane 10 – Sample 12.3167, Lane 11 – Sample 12.3235, Lane 12 – Sample 12.3236, Lane 13 – Sample 12.3237, Lane 14 – 12.3551, Lane 15 – 12.3552.

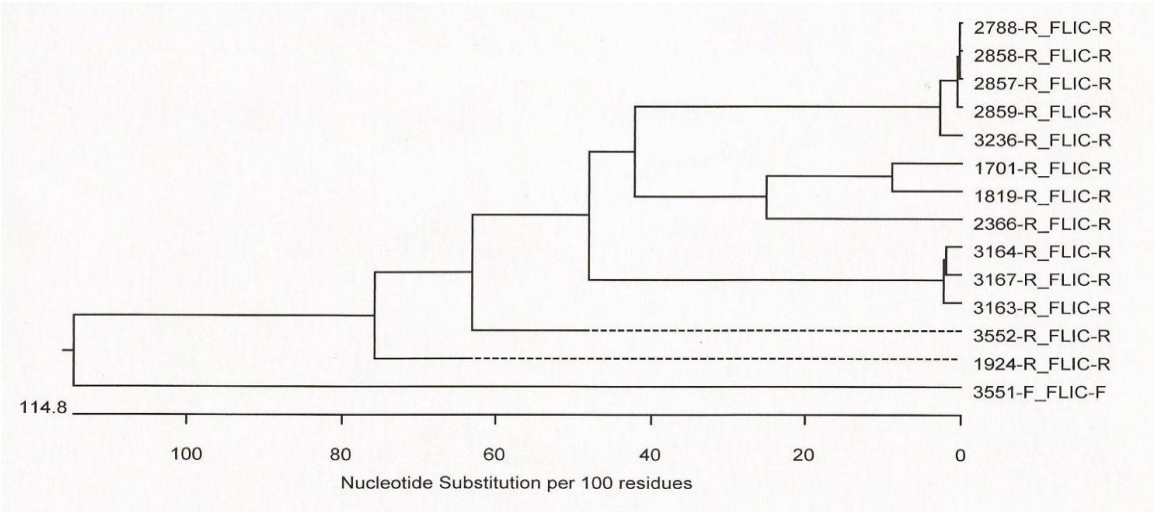

Figure 3-3. Phylogenetic tree of *fliC* sequences. A phylogeny illustrating the relative similarity of each sample's specific *fliC* sequence by nucleotide substitutions. Samples grouped at the top are the closest in similarity sequence with decreasing similarity at the bottom.

Escherichia coli O55:H7 fliC gene for flagellin, complete cds, strain: WC416

Sequence ID: **dbj|AB334574.1|** Length: 1758 Number of Matches: 1

Range 1: 1427 to 1708

Score	Expect	Identities	Gaps	Strand	Frame
377 bits(204)	6e-103()	256/282(91%)	0/282(0%)	Plus/Minus	

Features:

```
Query  3     CCAGCACGGAGTTACCGGCCTGCTGGATAATCTGCGCTTTCGACATGTTGGACACTTCGG  62
             ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Sbjct  1708  CCAGCACGGAGTTACCGGCCTGCTGGATGATCTGCGCTTTCGACATGTTGGACACTTCGG  1649

Query  63    TCGCATAGTCGGCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTGGTAGTGGTGTTGT  122
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1648  TCGCATAGTCGGCGTCCTGAATACGGGACTGCGCTTCGGACAGGTTGGTAGTGGTGTTGT  1589

Query  123   TCAGGTTAGTCACCGCAGAATCCAGACGGTTCTGGATAGCACCCAGGGATGAACGGAATT  182
             ||||||| || ||||| ||||||||||||||||||||||||||||||| ||||||||||
Sbjct  1588  TCAGGTTGGTGACCGCGGAATCCAGACGGTTCTGGATAGCACCCAGGGAAGAACGGAATT  1529

Query  183   TGTCGATGGAGCTGATAGCGTCGTCCAGGGCAGCCAGAGGATCTTTGGTTGCAGTGCCTG  242
             |||||||||||||||| ||||||||||||||| || || |   | |||||||  || |
Sbjct  1528  TGTCGATGGAGCTGATTGCGTCGTCCAGGGCAGCAAGCGGGTTGGTCGTTGCAGAACCAG  1469

Query  243   CACTGGTGGTTTCAGTGGTCAGTTTACCTGTAGAGTTAACAT  284
             ||||||| || ||||| || | |||||| |  ||||| ||||
Sbjct  1468  CACTGGTAGTCTCAGTAGTGATTTTACCCGCGGAGTTCACAT  1427
```

Figure 3-4. BLAST results for significantly similar *fliC* sequences. Samples showing significant similarity based on the phylogenetic tree were compared to known samples in the NCBI database using BLAST.

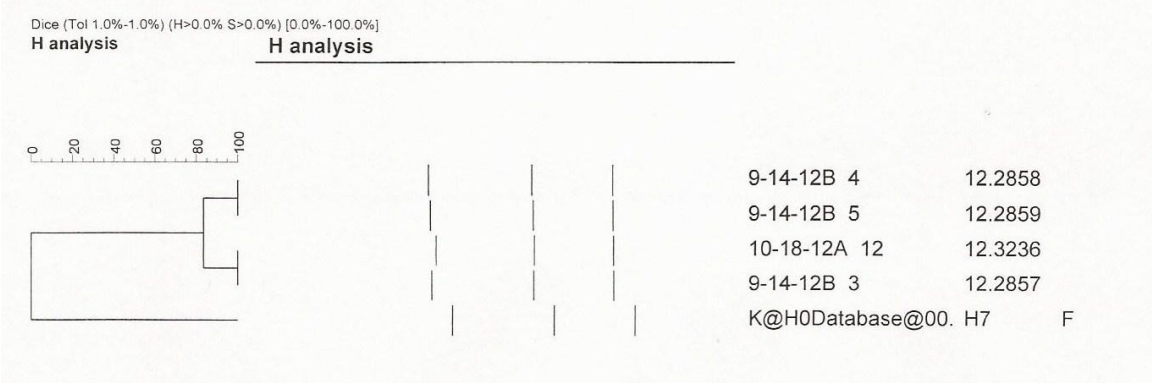

Figure 3-5. H analysis comparison of similar sequences to known H type H7. Samples were compared to the known sequence for the H7 H type and to each other.

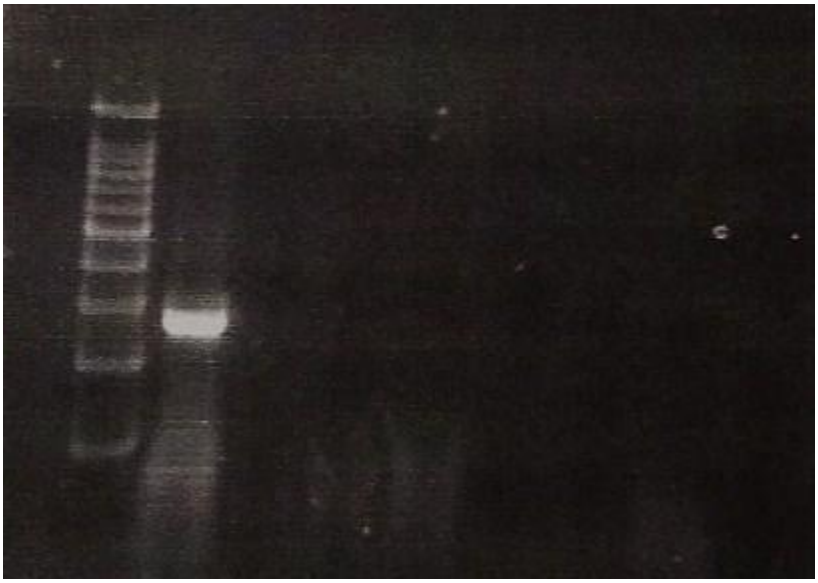

Figure 3-6. Gel electrophoresis results of H7 H type negative *E. coli* isolates. Lane 1 – DNA ladder, Lane 2 – H7 positive control (O157:H7), Lane 3 – H7 negative control (DT104), Lane 4 – Sample 12.2788, Lane 5 – Sample 12.2857, Lane 6 – Sample 12.2858, Lane 7 – Sample 12.2859, Lane 8 – Sample 12.3236

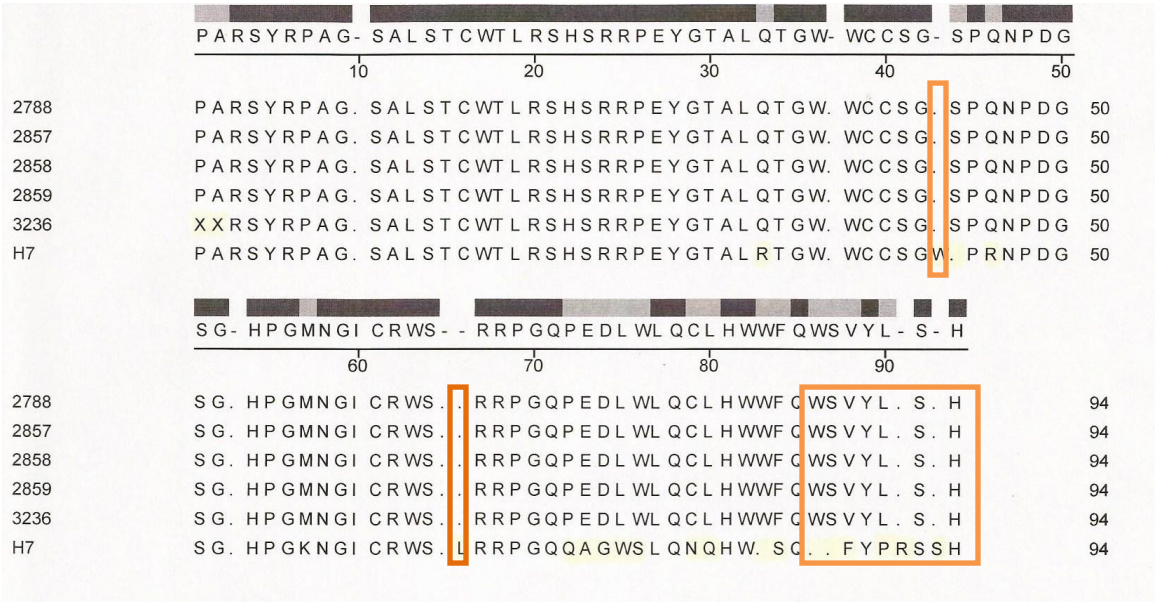

Figure 3-7. Amino acid alignment of similar sequences and H7 H group. Sequences of the samples of interest were converted to their corresponding amino acid sequence and aligned with that of the H7 H type as a basis for comparison of amino acid substitutions in the untypeable strains.

Table 3-1. Amino Acid Abbreviations. The amino acids are abbreviated by the SeqBuilder program using these given symbols.[7]

Symbol	Amino Acid	Symbol	Amino Acid
.	Termination (Stop)	M	Methionine
A	Alanine	N	Asparagine
C	Cysteine	P	Proline
D	Aspartic Acid	Q	Glutamine
E	Glutamic Acid	R	Arginine
F	Phenylalanine	S	Serine
G	Glycine	T	Threonine
H	Histidine	V	Valine
I	Isoleucine	W	Tryptophan
K	Lysine	X	Unknown
L	Leucine	Y	Tyrosine

Chapter 4

Discussion

Verification of the Presence of fliC

The *fliC* gene from Shiga toxin-producing *E. coli* (n=2) was amplified by PCR and visualized by gel electrophoresis using *E. coli* with H type H2 as a positive control and *Salmonella* DT104 as a negative control (Figure 3-1 and 3-2). The DNA fragments produced of all samples were roughly 1000 base pairs and identical to that of the positive control. These strains contain a viable *fliC* gene for comparative analyses.

Analysis of the fliC Gene Sequences

Conclusion of the presence of the *fliC* gene allowed for the continuation of the project and subsequent sequencing. The purified samples were submitted to the Genomics Core Facility for sequencing, and most samples provided successful and utilizable results. Of the 22 samples, 14 proved pure enough for sequencing, while the other 8 samples had sequences with too few base pairs for reliable analysis. This may have resulted from a poorly purified sample or laboratory error in sequencing, but the 14 viable sequences were enough to carry on with analysis. The precise assembled sequences of the *fliC* gene for the successful samples can be found in Appendix B.

Once samples had been sequenced, they could be compared for similarities and differences through alignment using the MegAlign program. Alignment produced a phylogenetic tree (Figure 3-3) suggesting Samples 12.2788, 12.2857, 12.2858, 12.2859, and 12.3236 were significantly similar in their *fliC* sequence and were almost identical at all nucleotides. This is a

notable finding and suggests there has been conservation of a mutation or other alteration to the normal *fliC* sequences found in known H groups. Comparison to samples in the NCBI's database using BLAST produced an almost identical sequence to *Escherichia coli* O55:H7 WC416 (accession number AB334574.1) with only 26 base pair substitutions (Figure 3-4). 256 of 282 base pairs matched with a 91% similar identity. This known sequence was reported with an H group of H7 and was a reference strain used in a comparison of typical and atypical O-antigen sequences.[8] This reference strain was used as the basis for further comparison of these samples to the H7 H type.

Comparison of Similar Sequences to H7 H Type

Once it was established that these strains were significantly similar to H7, further analyses attempted to determine to what degree and the effects these alterations might have. An H analysis comparing four of the five similar sequences to the standard H7 H type can be seen in Figure 3-5 (one of the samples could not be found in the database to include in the analysis). Though these samples produce similar bands to that of H7, they are not the same as H7 but are all very alike. This indicates these strains are an H7 variant. To confirm they were not true H7 H types, PCR analysis of the *fliC* gene was performed to determine if the samples were H7 serotype (Figure 3-6). All samples tested negative, confirming these variants are dissimilar enough from the H7 H type to be a new group.

Amino Acid Sequence Comparison

The final step was to determine precisely how the nucleotide substitutions in the variant strains affected the amino acid sequence. After translating the codons into amino acids using the SeqBuilder tool, the amino acid sequences were aligned using MegAlign and compared to the

amino acid sequence of the reference strain of the H7 H type (Figure 3-7). This alignment showed the five variant strains had identical amino acid sequences (only sample 12.3236 was missing the first two codons). These variant strains also differed from the H7 reference strain at 21 locations in the sequence with a large proportion concentrated at the end. These amino acid substitutions are the likely cause for the change that separates these variant strains from H7 H type and may have an impact on these strains' pathogenicity.

The amino acid sequences that encode the *fliC* gene are often highly conserved in certain regions and highly variable in others. Often, the N- and C-terminal ends of the molecule, responsible for encoding secretion and polymerization of the flagella, are similar across all strains and species, while the middle region is the highly variable portion responsible for difference in flagellar expression, which in turn can impact pathogenicity and H type expression.[9, 10] The variant strains' *fliC* genes match with 1427-1708bp of the reference strain, which is part of one of the sections of the gene generally considered to be conserved.[4] This indicates these strains are highly abnormal and likely have alterations in their secreting or polymerizing abilities. Identification of these differences warrants further research into precisely how these amino acid substitutions alter flagellar expression.

Chapter 5

Conclusion

Pathogenic Shiga toxin-producing *Escherichia coli* (n=22) were studied in an attempt to find new H types or variants of known H types based on sequencing and comparative analyses of the *flic* gene. After these analyses, five samples were determined to be identical strains that are variants of the H7 H type. Though not all samples were successfully sequenced and thus could not be included in analysis, this experiment is the basis for much more research into these specific strains and other untypeable strains. Conclusively determining the effect these variations have on pathogenicity can provide better predictability of virulence which is crucial in managing and understanding Shiga toxin-producing *E. coli*. Further pursuance of comparing *fliC* genes of untypeable strains will also undoubtedly result in the discovery of new H types to additionally improve serological typing and our knowledge of *E. coli.*

Appendix A

Procedure Conditions

Sample Preparation

Shiga toxin-producing *Escherichia coli* (n = 22) that exhibited the presence of the *fliC* gene but could not be classified into any of the 53 H groups were taken from The Pennsylvania State University *E. coli* Reference Center's repository (see Table 2-1). These samples were swabbed from individual freezer vials in the *E. coli* Reference Center's -80^{O}C freezer and grown on tryptic soy agar plates. Samples were allowed to incubate for 24 hours in a 37^{O}C incubator. A small swab of each sample was added to 150µL of distilled water, and the solution was vortexed for roughly 15 seconds at 3000rpm. The solutions were then heated in a Techne Progene Thermocycler at 99^{O}C for 10 minutes and centrifuged in a Beckman Coulter Microfuge at 13000 x g for 3 minutes. Samples were then frozen at -19^{O}C for later use.

Polymerase Chain Reaction

A master mix consisting of 184µl of enzyme diluent, 40µl of deoxyribonucleotide diphosphates (2mM), 16µl of Taq, 147µl of cresol red, and 613µl of distilled water was prepared and mixed with the fliC primers mentioned below to create the PCR working mix. The primer sequences used were:

FLIC-F: 5'-CCGAATTCATGGCACAAGTCATTAATAC-3'

FLIC-R: 5'-CCGAATTCTTAACCCTGCAGTAGAGACA-3'

Working mix (32µl) and template DNA (12µl) from each isolate and the positive and negative controls were placed in a 96 well plate and placed in the Eppendorf thermocycler. Amplification was conducted for 30 cycles with denaturation at 94^{o}C for 30 seconds, annealing at 52^{o}C for 30 seconds, and extension at 72^{o}C for 1 minute.

Ion Torrent Sequencing

Amplified DNA samples (30µl) were used for sequencing. A QIAquick ® PCR purification kit was used to purify the samples. Buffer PB (150µl) was added to each DNA sample, and the sample was applied to a QIAquick spin column. Samples were centrifuged for 60 seconds at 13,000 x g, and the supernatant was removed. Samples were washed with Buffer PE (0.75ml) and centrifuged at 13,000 x g for another 60 seconds. The remaining supernatant was discarded, and the sample was centrifuged a final time for 60 seconds at 13,000 x g. To elute the DNA, Buffer EB (50µl, 20mM Tris·Cl, pH 8.5) was added, and the sample was centrifuged for another 60 seconds at 13,000 x g. Samples were then sent to the Genomics Core Facility at University Park for Sanger method sequencing.

Appendix B

Successfully Assembled Sequences

Table 6-1. Assembled *fliC* Sequences. The *fliC* sequences of *E. coli* samples of untypeable H group were successfully assembled.

Sample	Sequence
11.1701	NNTGCCAGCACGGAGTTACCTGCCTGCTGAATGATCTGAGCTTTCGACATATTGGACACTTCGGTCGCA TAGTCGGCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTGGTGACC GCGGAATCCAGACGGTTCTGGATAGCACCCAGGGATGAACGGAATTTGTCGATCTGGCTGATTGCGTC ATCCAGAGCGGCCAGCGGATCGGTGGTGGTTGCGGCTTTGGTTGCAGCGTCGGTAGTGAAATCACCAT CCGCTTCTACATAGACCGCTTTACCTGAACCGTTGGTTACAGAGCCATCTTTCTGCAGATACAGTTTAG TGGTGTTATCAGTAGTCAGTGTACCGTCAGTGTTGGTGTAAGAATCTTTGTTGCCAGAAGTCGCAACAT CATATGCCTGAACACTCCCATCAGTACCTACAGAGATATAACCAGCTGTGCCTGCTGTACCAACAGTG AAACCAGTCGTTTTGCTTGCAGATTGCAGAGCATCTGAACTAATTACAGCGCCTGTAACAGCAATCGT ACCTTCAGTACCAGCATTACCAGTACCATCGCTTGAACCATTCAGTACAATCGAAGTGTtATCTGCTGT AGTAATTGTGGATTTAACAGCCGCTCCACCATTGTTGTTCAGAGCTAAACCACTCAGAGTCGCTTCCGC GAGTGTTCCAACACCGTTATTACCACCATTCTGAGTTAAGTTACCAGTTGTATCAATGTAAAGCTTAGA GCCGTCCTTnAGCAGTGAGTTTGCCATCACTGGACAGCACGACGTCTTGTGTAGTACCACCAATTTCAA CACTGAATGTTGCAGTGTCGCCCGCTTTCGGAGTCAGGTAAGACTGCAGCTCAGCTGTCGTAGCAGTC GTGTCATAAGTAAATGTATTATTATCTTTGTTAAGTTTATAGGCATTGGAAGTGGCTCCTGCAGCAAAG CCATTTTTCACGCCTGTGGCTGTAATAGTCGTACCATCAGCGAGGCTAGACAGTACATCTGCGGCTGTA GTTTTAGTCAGCCCAGCACTTACTGTGTAAGCAATGGCACCTGTAGAATCTGCTGCACCAGGGGTACC AATTGCAGCAGCAGCCAAATCAGCTTTAGTCGCCGCAGTATTCGCCACAGAACCAGAACCATTCACGT TAAAACCAGTCAGTTTCAACGTAGAAGAGTCAATCTTCTTCAGGTCGATAGTGATGGTCTGGCCATCAT TCGCGCCAACCTGAATTTTCATGGAACCGTCTTTTGCCAGTACGTTCACGCCGTTGAACTGAGTCTGAC CAGATACGCGGTCAATTTCGTCCAGACGGGATTTGATTTCGTCCTGGATAGAGTCCAGATCAGACTGG GAGTTGGTACCGGTGGTCGCCTGAACGGTCAGTTCACGCACACGCTGTAAGTTGTTGTTGATTTCGGAC AGCGCGCCTTCAGTGGTCTGTGCAACAGAAATACCGTCGTTGGCGTTACGTGCAGCCTGAGTCAGGCC TTTAATATTAGAAGTAAAACGGTTAGCAATCGCCTGACCTGCGGCGTCATCCTTCGCGCTGTTAATACG CAAGCCAGAAGACAGACGCTCGATAGAACTCGACAGCGCAGACTGGTTCTTGTTGATANN
11.1819	NNGCCAGCACGGAGTTACCTGCCTGCTGAATGATCTGAGCTTTCGACATATTGGACACTTCGGTCGCAT AGTCGGCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTGGTGACC GCGGAATCCAGACGGTTCTGGATAGCACCCAGGGATGAACGGAATTTGTCGATCTGGCTGATTGCGTC ATCCAGAGCGGCCAGCGGATCGGTGGTGGTTGCGGCTTTGGTTGCAGCGTCGGTAGTGAAATCACCAT CCGCTTCTACATAGACCGCTTTACCTGAACCGTTGGTTACAGAGCCATCTTTCTGCAGATACAGTTTAG TGGTGTTATCAGTAGTCAGTGTACCGTCAGTGTTGGTGTAAGAATCTTTGTTGCCAGAAGTCGCAACAT CATATGCCTGAACACTCCCATCAGTACCTACAGAGATATAACCAGCTGTGCCTGCTGTACCAACAGTG AAACCAGTCGTTTTGCTTGCAGATTGCAGAGCATCTGAACTAATTACAGCGCCTGTAACAGCAATCGT ACCTTCAGTACCAGCATTACCAGTACCATCGCTTGaACCATTCAGTACaATCGAAGTGTTATCTGCTGTA GTAATTGTGGATTTAaCAGCCGCTCCACCATTGTTGTTCAGAGCTAAACCACTCAGAGTCGCTTCCGCG AGTGTTCCAACACCGTTATTACCACCATTCTGAGTTAAGTTACCAGTTGTATCAATGTAAAGCTTAGAG CCGTCCTTAGCAGTGAGTTTGCCATCACTGGACAGCACGACGTCTTGTGTAGTACCACCAATTTCAACA CTGAATGTTGCAGTGTCGCCCGCTTTCGGAGTCAGGTAAGACTGCAGCTCAGCTGTCGTAGCAGTCGTG TCATAAGTAAATGTATTATTATCTTTGTTAAGTTTATAGGCATTGGAAGTGGCTCCTGCAGCAAAGCCA TTTTTCACGCCTGTGGCTGTAATAGTCGTACCATCAGCGAGGCTAGACAGTACATCTGCGGCTGTAGTT TTAGTCAGCCCAGCACTTACTGTGTAAGCAATGGCACCTGTAGAATCTGCTGCACCAGGGGTACCAAT TGCAGCAGCAGCCAAATCAGCTTTAGTCGCCGCAGTATTCGCCACAGAACCAGAACCATTCACGTTAA AACCAGTCAGTTTCAACGTAGAAGAGTCAATCTTCTTCAGGTCGATAGTGATGGTCTGGCCATCATTCG CGCCAACCTGAATTTTCATGGAACCGTCTTTTGCCAGTACGTTCACGCCGTTGAACTGAGTCTGACCAG ATACGCGGTCAATTTCGTCCAGACGGGATTTGATTTCGTCCTGGATAGAGTCCAGATCAGACTGGGAG TTGGTACCGGTGGTCGCCTGAACGGTCAGTTCACGCACACGCTGTAAGTTGTTGTTGATTTCGGACAGC GCGCCTTCAGTGGTCTGTGCAACAGAAATACCGTCGTTGGCGTTACGTGCAGCCTGAGTCAGGCCTTTA ATATTAGAAGTAAAACGGTTAGCAATCGCCTGACCTGCGGCGTCATCCTTCGCGCTGTTAATACGCAA GCCAGAAGACAGACGCTCGATAGAACTCGACAGCGCAGACTGGTTCTTGTTGATANN

11.1924	NGCCNGTGCCAGAACAGAGGTACCCGCTTGTTGCAGGATCTGCGCACGAGACATGTTAGACACTTCGG TCGCGTAGTCAGCATCTTCGATACGGCTACGGGCAGAAGACAGGTTGTTTACGGTGTTGCCAAGGTTG GTGATGGCAGAGTCGAAACGGTTTTGTACTGCACCGAGGTCAGAACGCAGATTGTCAACTTTAGCCAA TGCTTTGTCGATAGTTTCGAGCGGGTTGGTGGTAGATTGCAACGATTTTGCTGCATCTTCGTTTACCAG AATCGGGCTACCACCTTCTGATTTGCTCAGATACATGACTTTATTGTTACCAGAAGCAGTCTCCGTtATC GtTTTACCATCTGCACTAACATCGTAAGTTGCACCGTTAACAACTAACGTGCTTCCTGTTTTCTTGGCAG CGTTCAGATCAAGATCAGATAGTGTCGCTGCTTTATTTTCAACTTTTGTTGCAGTCAATTGCCCtGCACT AtTTTTGTATAGGGCTGTCGTAGGGGCAACGGTTGCACTAGTACCTGATGTATCAGTACTCCCCGAAAT AGTGAATTCAACAGCCTTACCATCAACATTGGCGGTTAATTTACCATTACCTGTGGCAGGTATAGCGCC AGTGCCGGTATTTGTAAATTCGATACCTTCGTAGACAATCTTATTACCTTGAGCTAAATCTTTAGCAGC CACTGCAAGCTTAGTTGCATCAATCTTGAATTGAGTATCACTGCTGGTCGTAAGTGAACCATCCGCAGC ACTCACATAAACTTGTTTGCCATCTTTATCCTGTACTACTCCACTATCTACATTAACAGTATAGTTATCA GTACCGTTAATTTGATAATTATCAGTACCTGTCGCTTTAAATTTAGAAATCAGgTCACTGCCGGTTGCTT TCTGCGCGCCATCGATATTAAAACCGTCCAGGCCGAGAGTTTTCGCATCAATTTTTGCCAGATTGATAG TGATGGTTTCACCATCATTAGCACCAACCTGAATTTTCATTTCATTATTTTCAGCAAGGACTTTCACGCC GTTAAACTGAGTTTGCTCAGATACACGGTCAATTTCTTCCAGACGTTGAGTAATTTCAGCCTGGATAGA AGAAAGATCGCTGTCAGAGTTAGTACCGTTAGTTGCCTGAACAGAAAGTTCACGAATACGCTGCAGGT TGTTGTTAATTTCATTCAGCGCACCTTCAGTGGTCTGCGCAACAGAAATACCATCATTCGCGTTACGGG AAGCCTGGGTCAGACCTTTAATATTTGCCGTAAAACGGTTAGCAATCGCCTGACCTGCTGCATCGTCTT TTGCGCTGTTAATACGCAGACCAGAAGACAGACGCTCAATAGCAGAGCTAAGAGAAGACTGAGATTT GTTCANN
12.2366	NNGCCAGCACGGAGTTACCGGCCTGCTGGATGATCTGCGCTTTCGACATGTTGGACACTTCGGTCGCAT AGTCGGCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTAGTAGTGGTGTTGTTCAGGTTGGTGACT GCGGAATCCAGACGGTTTTGCACCGCACCGAGGGAAGAACGGAATTTGTCGATGGAGCTGATCGCTTC GTCCAGAGCTTTCAGGGGATCGGCGGTGGTTGCAGCTTTAGTTTCAGCGTCAGTAGTAAAATTACCATC CGCTGAAACATAGACTGCTTTACCTGAACCATTGGTCACAGTTCCGTCTTTTTGCAAAAACAGATTGGT TGTATTATCGGTAGTCAGTGAACCATCAGCTTTGGTGTAAGCGTCTTCAAAACCTGTACCTGTGTGTTG TTGTACTATACCTTTACTATCAACTTTAATATCGCCTGCGCTGGTACCAGTAGTTCCAACATCAACCGT GAATCCTGTGGACTTAGTTGCCGATTGAAGAGACTCAGCACTCATTCTTGCATTTGCTACAGTGATTGT ACCAGAAGTAGTAGCATCAGTACCGCCAGCCAGAGTGAAGGTTGTTTCATCTTCAGTTGTAATTACCGT AGATACGGCATCCGGTGTGCCTGTTGTATGCCAGTTTTTAGTCAGTGCATCAAGAGTGGCTGCTTTGAC AGTCCCAGCATTGTTTTGAGTGAGGTTACCATCTACGTCAATAAATAGTTCTGAACCATCTTTCGCGGT AATTTTGCCATCGCTCGCCAGAACAACATCAACAGATGTCGTACCGATTTTGACGCTCAGgTTAGCTGT ATCACCTGCTTTTGGTGTCAGGAAGGATTGCAGAGCTGCGCTGTTACTATCACCGACCGCAGCAGCAG CAATGCTGGCATTGAAAGTGAAATCATTTGTAGCTTTATTGTAGgTGTAGTCTGTAGCTGCTGCACCAA AACCGTTGGAAATGCTAGCATTAACAACAGCACCATCTGCAAGGTTTGCAAACACGTTTGCTGCTGTG GCTTTGTTAAGCCCTGCGTCTACGGTATATTTGGTCACGCCGTCAGTACCGACAGCAGGAGTTGTACCC GCAGCAGCAGCAGCAGCAGCAGCCAGTTCATCTTTAGTCGCTGCAGTATTAGCCACAGCCCCGCCACC ATTCACATTAAACCCACTCAGCCCCAGCGTATCTGAGTCAATTTTCTTCAGATCAATAGTGATAGTCTG GCCGTCATTCGCACCAACCTGAATTTTCATTGAACCGTCTTTCGCCAGTACGTTCACGCCGTTGAACTG GGTCTGGCCAGATACGCGGTCAATTTCGTCCAGACGGGATTTGATTTCGTCCTGAATGGAGTCCAGATC CGAATCGGAGTTAGTCCCGGTAGAAGCCTGAACCGTCAGTTCACGGATACGCTGTAAGTTGTTGTTGA TTTCGGACAGCGCGCCTTCGGTGGTCTGTGCAACAGAAATACCGTCGTTGGCGTTACGTGCAGCCTGA GTCAGGCCTTTAATATTAGAGGTGAAACGGTTAGCAATCGCCTGACCTGCGGCGTCATCCTTCGCGCTG TTAATACGCAAGCCAGAAGACAGACGCTCGATAGAACTCGACAGCGCAGACTGGTTCTTGTTGNTN

12.2788	NNCCAGCACGGAGTTACCGGCCTGCTGGATAATCTGCGCTTTCGACATGTTGGACACTTCGGTCGCATA GTCGGCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTAGTCACCGC AGAATCCAGACGGTTCTGGATAGCACCCAGGGATGAACGGAATTTGTCGATGGAGCTGATAGCGTCGT CCAGGGCAGCCAGAGGATCTTTGGTTGCAGTGCCTGCACTGGTGGTTTCAGTGGTCAGTTTACCTGTAG AGTTAACATAAGCCTGTGCCCCAACAGTAGCTGCAAATTTACCAGTTGCACCTGCGCCATTTGAAGCTA CAGTTACTGTACCGGTATCCTTATCGACGTTGTAGTTGATAGTGTAAGATGCGGTTGTGGTGAGCTCTT TGTCTGCATCAGTATAGGTGTCATTCGAATTACCAGCACCATTAGTATTCGCACCACCGAAAGACGCA GTTGCAGACTGAATACCTGNNTTATACTTAATTGTCGCACCAGTAACTGCTGTGTTAGCCGTTGCAGCA CTCTGTACTGTGCTCAGAACAGTATCAGCACTTACCGTTTTGGTGTAGGTTACGCCAGAAGTTGCCGCA GCACCAAAGTTATAAGTTGTGCCAGACAGCTGGATAGAACCAGTTGTCGCTAAAGTACTACCGCTAGC aAACAGATCGGACAATTTCGCTGAAGATGCAATGCCGGCGTTGTTTGTAGATAGGTTACCAGTGGCGTC CAGGTACGCGGCTTTACCACCGATGGTCACATCGCCATTAGCATCTACATCAAACTTCACATCACCAGT ACTACGAGTATAAACACCTGATGCAGTAGTGCCAGCCGCTGGTTTCAGAGTATTTGCAAAGTTAACAA CATCACCATCTGCAACTGTTGCTTGAGTGGTGAAGTTCCCTTTAGCCGCATCATAAGTATAGGTCGCAG CACTCGAGCCAGTAGTAGTAACTGTATCTCCGGTTTTCAGGCGAGACAGTGCATCGCTAGCGCTGAGA GCTGTATTGTTTGTGGTCACAGCATAAGGACCTGTTCCCGTTGCACCAGCAGCGGTCAGATCGCTGACT GTAGCAGCTTTGTTCGCAATAGTACCTTTGCCGTTAACGTTAAACCCATTCAGCCCCAGCGTATCAGAG TCAATTTTCTTCAGATCAATAGTGATAGTCTGGCCGTCATTCGCACCAACCTGAATTTTCATCGAACCG TCTTTTGCCAGTACGTTCACGCCGTTGAACTGGGTCTGGCCGGATACGCGGTCAATTTCGTCGAGACGG GATTTGATTTCGTCCTGGATGGAGTCCAGGTCAGAATCGGAGTTAGTCCCTGTAGTGGCCTGAACCGTC AGTTCACGAATACGCTGTAAGTTGTTGTTGATTTCGGACAGCGCGCCTTCAGTGGTCTGCGCAACGGAG ATACCGTCGTTGGCGTTACGGGCCGCCTGAGTCAGGCCTTTAATATTAGAAGTAAAACGGTTAGCAAT CGCCTGACCCGCTGCGTCATCCTTCGCGCTGTTAATACGCAAGCCAGAAGACAGACGCTCGATAGAAC TCGACAGCGCAGACTGGTTCTTGTTGANTN
12.2857	TTCTNGNGGCCAGCACGGAGTTACCGGCCTGCTGGATAATCTGCGCTTTCGACATGTTGGACACTTCGG TCGCATAGTCGGCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTAG TCACCGCAGAATCCAGACGGTTCTGGATAGCACCCAGGGATGAACGGAATTTGTCGATGGAGCTGATA GCGTCGTCCAGGGCAGCCAGAGGATCTTTGGTTGCAGTGCCTGCACTGGTGGTTTCAGTGGTCAGTTTA CCTGTAGAGTTAACATAAGCCTGTGCCCCAACAGTAGCTGCAAATTTACCAGTTGCACCTGCGCCATTT GAAGCTACAGTTACTGTACCGGTATCCTTATCGACGTTGTAGTTGATAGTGTAAGATGCGGTTGTGGTG AGCTCTTTGTCTGCATCAGTATAGGTGTCATTCGAATTACCAGCACCATTAGTATTCGCACCACCGAAA GACGCAGTTGCAGACTGAATACCTGTATTATACTTAATTGTCGCACCAGTAACTGCTGTGTTAGCCGTT GCAGCACTCTGTACTGTGCTCAGAACAGTATCAGCACTTACCGTTTTGGTGTAGGTTACGCCAGAAGTT GCCGCAGCACCAAAGTTATAAGTTGTGCCAGACAGCTGGATAGAACCAGTtGTCGSTAAAGTACTACCG YTAGCAAACAGATCGGACAATTTCGCTGAAGATGCAATGCCGGCGTTGTTTGTAGATAGGTTACCAGT GGCGTCCAGGTACGCGGCTTTACCACCGATGGTCACATCGCCATTAGCATCTACATCAAACTTCACATC ACCAGTACTACGAGTATAAACACCTGATGCAGTAGTGCCMGCCGCTGGTTTCARAGTATTTGCAAAGT TAACAACATCACCATCTGCAACTGTTGCTTGAGTGGTGAAGTTCCCTTTAGCCGCATCATAAGTATAGG TCGCAGCACTCGAGCCAGTAGTAGTAACTGTATCTCCGGTTTTCAGGCGAGACAGTGCATCGCTAGCG CTGAGAGCTGTATTGTTTGTGGTCACAGCATAAGGACCTGTTCCCGTTGCACCAGCAGCGGTCAGATCG CTGACTGTAGCAGCTTTGTTCGCAATAGTACCTTTGCCGTTAACGTTAAACCCATTCAGCCCCAGCGTA TCAGAGTCAATTTTCTTCAGATCAATAGTGATAGTCTGGCCGTCATTCGCACCAACCTGAATTTTCATC GAACCGTCTTTTGCCAGTACGTTCACGCCGTTGAACTGGGTCTGGCCGGATACGCGGTCAATTTCGTCG AGACGGGATTTGATTTCGTCCTGGATGGAGTCCAGGTCAGAATCGGAGTTAGTCCCTGTAGTGGCCTG AACCGTCAGTTCACGAATACGCTGTAAGTTGTTGTTGATTTCGGACAGCGCGCCTTCAGTGGTCTGCGC AACGGAGATACCGTCGTTGGCGTTACGGGCCGCCTGAGTCAGGCCTTTAATATTAGAAGTAAAACGGT TAGCAATCGCCTGACCCGCTGCGTCATCCTTCGCGCTGTTAATACGCAAGCCAGAAGACAGACGCTCG ATAGAACTCGACAGCGCAGACTGGTTCTTGTNN

12.2858	GNNGCCAGCACGGAGTTACCGGCCTGCTGGATAATCTGCGCTTTCGACATGTTGGACACTTCGGTCGC ATAGTCGGCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTAGTCAC CGCAGAATCCAGACGGTTCTGGATAGCACCCAGGGATGAACGGAATTTGTCGATGGAGCTGATAGCGT CGTCCAGGGCAGCCAGAGGATCTTTGGTTGCAGTGCCTGCACTGGTGGTTTCAGTGGTCAGTTTACCTG TAGAGTTAACATAAGCCTGTGCCCCAACAGTAGCTGCAAATTTACCAGTTGCACCTGCGCCATTTGAA GCTACAGTTACTGTACCGGTATCCTTATCGACGTTGTAGTTGATAGTGTAAGATGCGGTTGTGGTGAGC TCTTTGTCTGCATCAGTATAGGTGTCATTCGAATTACCAGCACCATTAGTATTCGCACCACCGAAAGAC GCAGTTGCAGACTGAATACCTGTATTATACTTAATTGTCGCACCAGTAACTGCTGTGTTAGCCGTTGCA GCACTCTGTACTGTGCTCAGAACAGTATCAGCACTTACCGTTTTGGTGTAGGTTACGCCAGAAGTTGCC GCAGCACCAAAGTTATAAGTTGTGCCAGACAGCTGGATAGAACCAGTTGTCGCTAAAGTACTACCGCT AGCAAACAGATCGGACAATTTCGCTGAAGATGCAATGCCGGCGTTGTTTGTAGATAGGTTACCAGTGG CGTCCAGGTACGCGGCTTTACCACCGATGGTCACATCGCCATTAGCATCTACATCAAACTTCACATCAC CAGTACTACGAGTATAAACACCTGATGCAGTAGTGCCAGCCGCTGGTTTCAGAGTATTTGCAAAGTTA ACAACATCACCATCTGCAACTGTTGCTTGAGTGGTGAAGTTCCCTTTAGCCGCATCATAAgTATAGGTC GCAGCACTCGAGCCAGTAGTAGTAACTGTATCTCCGGTTTTCAGGCGAGACAGTGCATCGCTAGCGCT GAGAGCTGTATTGTTTGTGGTCACAGCATAAGGACCTGTTCCCGTTGCACCAGCAGCGGTCAGATCGCT GACTGTAGCAGCTTTGTTCGCAATAGTACCTTTGCCGTTAACGTTAAACCCATTCAGCCCCAGCGTATC AGAGTCAATTTTCTTCAGATCAATAGTGATAGTCTGGCCGTCATTCGCACCAACCTGAATTTTCATCGA ACCGTCTTTTGCCAGTACGTTCACGCCGTTGAACTGGGTCTGGCCGGATACGCGGTCAATTTCGTCGAG ACGGGATTTGATTTCGTCCTGGATGGAGTCCAGGTCAGAATCGGAGTTAGTCCCTGTAGTGGCCTGAA CCGTCAGTTCACGAATACGCTGTAAGTTGTTGTTGATTTCGGACAGCGCGCCTTCAGTGGTCTGCGCAA CGGAGATACCGTCGTTGGCGTTACGGGCCGCCTGAGTCAGGCCTTTAATATTAGAAGTAAAACGGTTA GCAATCGCCTGACCCGCTGCGTCATCCTTCGCGCTGTTAATACGCAAGCCAGAAGACAGACGCTCGAT AGAACTCGACAGCGCAGACTGGTTCTTGTTGANTN
12.2859	NGNGCCAGCACGGAGTTACCGGCCTGCTGGATAATCTGCGCTTTCGACATGTTGGACACTTCGGTCGC ATAGTCGGCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTAGTCAC CGCAGAATCCAGACGGTTCTGGATAGCACCCAGGGATGAACGGAATTTGTCGATGGAGCTGATAGCGT CGTCCAGGGCAGCCAGAGGATCTTTGGTTGCAGTGCCTGCACTGGTGGTTTCAGTGGTCAGTTTACCTG TAGAGTTAACATAAGCCTGTGCCCCAACAGTAGCTGCAAATTTACCAGTTGCACCTGCGCCATTTGAA GCTACAGTTACTGTACCGGTATCCTTATCGACGTTGTAGTTGATAGTGTAAGATGCGGTTGTGGTGAGC TCTTTGTCTGCATCAGTATAGGTGTCATTCGAATTACCAGCACCATTAGTATTCGCACCACCGAAAGAC GCAGTTGCAGACTGAATACCTGTATTATACTTAATTGTCGCACCAGTAACTGCTGTGTTAGCCGTTGCA GCACTCTGTACTGTGCTCAGAACAGTATCAGCACTTACCGTTTTGGTGTAGGTTACGCCAGAAGTTGCC GCAGCACCAAAGTTATAAGTTGTGCCAGACAGCTGGATAGAACCAGTTGTCGCTAAAGTACTACCGCT AGCAAACAGATCGGACAATTTCGCTGAAGATGCAATGCCGGCGTTGTTTGTAGATAGGTTACCAGTGG CGTCCAGGTACGCGGCTTTACCACCGATGGTCACATCGCCATTAGCATCTACATCAAACTTCACATCAC CAGTACTACGAGTATAAACACCTGATGCAGTAGTGCCAGCCGCTGGTTTCAGAGTATTTGCAAAGTTA ACAACATCACCATCTGCAACTGTTGCTTGAGTGGTGAAGTTCCCTTTAGCCGCATCATAAGTATAGGTC GCAGCACTCGAGCCAGTAGTAGTAACTGTATCTCCGGTTTTCAGGCGAGACAGTGCATCGCTAGCGCT GAGAGCTGTATTGTTTGTGGTCACAGCATAAGGACCTGTTCCCGTTGCACCAGCAGCGGTCAGATCGCT GACTGTAGCAGCTTTGTTCGCAATAGTACCTTTGCCGTTAACGTTAAACCCATTCAGCCCCAGCGTATC AGAGTCAATTTTCTTCAGATCAATAGTGATAGTCTGGCCGTCATTCGCACCAACCTGAATTTTCATCGA ACCGTCTTTTGCCAGTACGTTCACGCCGTTGAACTGGGTCTGGCCGGATACGCGGTCAATTTCGTCGAG ACGGGATTTGATTTCGTCCTGGATGGAGTCCAGGTCAGAATCGGAGTTAGTCCCTGTAGTGGCCTGAA CCGTCAGTTCACGAATACGCTGTAAGTTGTTGTTGATTTCGGACAGCGCGCCTTCAGTGGTCTGCGCAA CGGAGATACCGTCGTTGGCGTTACGGGCCGCCTGAGTCAGGCCTTTAATATTAGAAGTAAAACGGTTA GCAATCGCCTGACCCGCTGCGTCATCCTTCGCGCTGTTAATACGCAAGCCAGAAGACAGACGCTCGAT AGAACTCGACAGCGCAGACTGGTTCTTGTTGACNCN

12.3163	NNGCCAACACGGAGTTACCGGCCTGCTGAATGATCTGTGCTTTCGACATGTTGGACACTTCGGTCGCAT AGTCGGCGTCCTGAATACGGGACTGTGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTGGTGACTG CGGAATCCAGACGGTTTTGAACGGCACCGAGGGAGGAGCGGAAAGTATCAACCTGTGCAATAGCTTTG TCTAAAAGTGCCAGTGGATCAGCAGTGGAAGCCCCAGTAAATTCATCAGCCGTATCCAGATCAATGGT CTGAGCAATATCGGTTGGTnGCAGCAGTGGACTGGTTTTTTTACACCATCTGTAATTTCATAATTCTTACC CTGAACGGTAGCGAAACCAACGGCCTCTCCGTCAGAATTAGCAGCAACTTTAATCAACGTATCATCTTT GGTAGCGCCGGTAATATCACCGCCTGAATAAGTAATATCCGTTTTATTAAGAGTTACAGTCCCGTCATC TGCAACAGATGCAGCGTAATTATCACTACCATAGCTAACAACATAGTTCTCAGTTGCCGCACCATCTTT GTCTAAGATATTGTGCAACGTTAAAGAACTTGCATTAACCTTTTGCCCCAAAGCAGTACCAATATCTGT TGCTGCTGCATCCAGATCCACTTTAACTGGCGCGGCTGAACCATCGCCGACCTGCGTCACTGTATCGCT TAATTTTAGCGCCCCACCGGCAACACTAAAACCACTTAAACCAAGGGCAGAAGAGTCTATTTTCTGCA AATCGATAGAGATGGTCTGCCCATCATTCGCGCCAACCTGAATATTCAGAGAACCGTTTTTAGCCAGC ACGTTCACGCCGTTGAACTGGGTCTGACcAGAGACACGATCGATTTCAGCCAAGCgGGATTTGATTTCG TCCTGGATTGAAGACAGgTCAGAATCAGAGTTAGTACCGGTAGTGGCCTGAACGGTCAGCTCACGAAC ACGCTGCAAGTTGTTGTTGATTTCAGACAGTGCGCCTTCAGTGGTCTGCGCCAGAGAGATACCGTCGTT GGCGTTACGTGCGGCCTGAGTCAGACCTTTGATGTTAGAAGTGAAGCGGTTAGCAATCGCCTGGCCCG CAGCGTCATCTTTAGCGCTGTTAATGCGCAAACCAGAAGAGAGGCGCTCGATAGAAGTCGACAGCGCA GACTGGTTTTTGTTGANN
12.3164	NNGCCAACACGGAGTTACCGGCCTGCTGAATGATCTGTGCTTTCGACATGTTGGACACTTCGGTCGCAT AGTCGGCGTCCTGAATACGGGACTGTGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTGGTGACTG CGGAATCCAGACGGTTTTGAACGGCACCGAGGGAGGAGCGGAAAGTATCAACCTGTGCAATAGCTTTG TCTAAAAGTGCCAGTGGATCAGCAGTGGAAGCCCCAGTAAATTCATCAGCCGTATCCAGATCAATGGT CTGAGCAATATCGGTTGGTGCAGCAGTGGACTGGTTTTTTACACCATCTGTAATTTCATAATTCTTACC CTGAACGGTAGCGAAACCAACGGCCTCTCCGTCAGAATTAGCAGCAACTTTAATCAACGTATCATCTTT GGTAGCGCCGGTAATATCACCGCCTGAATAAGTAATATCCGTTTTATTAAGAGTTACAGTCCCGTCATC TGCAACAGATGCAGCGTAATTATCACTACCATAGCTAACAACATAGTTCTCAGTTGCCGCACCATCTTT GTCTAAGATATTGTGCAACGTTAAAGAACTTGCATTAACCTTTTGCCCCAAAGCAGTACCAATATCTGT TGCTGCTGCATCCAGATCCACTTTAACTGGCGCGGCTGAACCATCGCCGACCTGCGTCACTGTATCGCT TAATTTTAGCGCCCCACCGGCAACACTAAAACCACTTAAACCAAGGGCAGAAGAGTCTATTTTCTGCA AATCGATAGAGATGGTCTGCCCATCATTCGCGCCAACCTGAATATTCAGAGAACCGTTTTTAGCCAGC ACGTTCACGCCGTTGAASTGGKTCTGACCAGAGACACGATCGATTTCAGCCAAGCGGGATTTGATTTCG TCCTGGATTGAAGACAGGTCAGAATCAGAGTTAGTACCGGTAGTGGCCTGAACGGTCAGCTCACGAAC ACGCTGCAAGTTGTTGTTGATTTCAGACAGTGCGCCTTCAGTGGTCTGCGCCAGAGAGATACCGTCGTT GGCGTTACGTGCGGCCTGAGTCAGACCTTTGATGTTAGAAGTGAAGCGGTTAGCAATCGCCTGGCCCG CAGCGTCATCTTTAGCGCTGTTAATGCGCAAACCAGAAGAGAGGCGCTCGATAGAAGTCGACAGCGCA GACTGGTTTTTGTTGATGNTGTN
12.3167	NNCCAACACGGAGTTACCGGCCTGCTGAATGATCTGTGCTTTCGACATGTTGGACACTTCGGTCGCATA GTCGGCGTCCTGAATACGGGACTGTGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTGGTGACTGC GGAATCCAGACGGTTTTGAACGGCACCGAGGGAGGAGCGGAAAGTATCAACCTGTGCAATAGCTTTGT CTAAAAGTGCCAGTGGATCAGCAGTGGAAGCCCCAGTAAATTCATCAGCCGTATCCAGATCAATGGTC TGAGCAATATCGGTTGGTGCAGCAGTGGACTGGTTTTTTACACCATCTGTAATTTCATAATTCTTACCCT GAACGGTAGCGAAACCAACGGCcTYTCCGTCAGAATTAGCAGCAACTTTAATCAAcGTATCATCTTTGg TAGCGCCgGGTAATATCACCGCCTGAATAAGTAATATCCGtTTTTATTAAGAGTTACAGTCCCGTCATCT GCAACAGATGCAGCGTAATTATCACTACCATAGCTAACAACATAGTTCTCAGTTGCCGCACCATCTTTG TCTAAGATATTGTGCAACGTTAAAGAACTTGCATTAACCTTTTGCCCCAAAGCAGTACCAATATCTGTT GCTGCTGCATCCAGATCCACTTTAACTGGCGCGGCTGAACCATCGCCGACCTGCGTCACTGTATCGCTT AATTTTAGCGCCCCACCGGCAACACTAAAACCACTTAAACCAAGGGCAGAAGAGTCTATTTTCTGCAA ATCGATAGAGATGGTCTGCCCATCATTCGCGCCAACCTGAATATTCAGAGAACCGTTTTTAGCCAGCAC GTTCACGCCGTTGAACTGGGTCTGACCAGAGACACGATCGATTTCAGCCAAGCGGGATTTGATTTCGTC CTGGATTGAAGACAGGTCAGAATCAGAGTTAGTACCGGTAGTGGCCTGAACGGTCAGCTCACGAACAC GCTGCAAGTTGTTGTTGATTTCAGACAGTGCGCCTTCAGTGGTCTGCGCCAGAGAGATACCGTCGTTGG CGTTACGTGCGGCCTGAGTCAGACCTTTGATGTTAGAAGTGAAGCGGTTAGCAATCGCCTGGCCCGCA GCGTCATCTTTAGCGCTGTTAATGCGCAAACCAGAAGAGAGGCGCTCGATAGAAGTCGACAGCGCAGA CTGGTTTNGTTGACGCN

12.3236	NNCACGGAGTTACCGGCCTGCTGGATAATCTGCGCTTTCGACATGTTGGACACTTCGGTCGCATAGTCG GCGTCCTGAATACGGGACTGCGCTTCAGACAGGTTGGTAGTGGTGTTGTTCAGGTTAGTCACCGCAGA ATCCAGACGGTTCTGGATAGCACCCAGGGATGAACGGAATTTGTCGATGGAGCTGATAGCGTCGTCCA GGGCAGCCAGAGGATCTTTGGTTGCAGTGCCTGCACTGGTGGTTTCAGTGGTCAGTTTACCTGTAGAGT TAACATAAGCCTGTGCCCCAACAGTAGCTGCAAATTTACCAGTTGCACCTGCGCCATTTGAAGCTACA GTTACTGTACCGGTATCCTTATCGACGTTGTAGTTGATAGTGTAAGATGCGGTTGTGGTGAGCTCTTTG TCTGCATCAGTATAGGTGTCATTCGAATTACCAGCACCATTAGTATTCGCACCACCGAAAGACGCAGTT GCAGACTGAATACCTGTATTATACTTAATTGTCGCACCAGTAACTGCTGTGTTAGCCGTTGCAGCACTC TGTACTGTGCTCAGAACAGTATCAGCACTTACCGTTTTGGTGTANGTTACGCCAGAANTTGCCGCAGCA CCAAAGTTATAAGTTGTGCCAGACAGCTGGATAGAACCAGTtGTCGCTAAAGTACTACCGCTAGCAAA CAGATCGGACAATTTCGCTGAAGATGCAATGCCGGCGTTGTTTGTAGATAGGTTACCMGNGGCGTCCA GGTACGCGGCTTTACCACCGATGNTCACATCGCCATTAGCATCTACATCAAACTTCACATCACCAGTAC TACGAGTATAAACACCTGATGCAGTAGTGCCAGCCGCTGGTTTCAGAGTATTTGCAAAGTTAACAACA TCACCATCTGCAACTGTTGCTTGAGTGGTGAAGTTCCCTTTAGCCGCATCATAAGTATAGGTCGCAGCA CTCGAGCCAGTAGTAGTAACTGTATCTCCGGTTTTCAGGCGAGACAGTGCATCGCTAGCGCTGAGAGC TGTATTGTTTGTGGTCACAGCATAAGGACCTGTTCCCGTTGCACCAGCAGCGGTCAGATCGCTGACTGT AGCAGCTTTGTTCGCAATAGTACCTTTGCCGTTAACGTTAAACCCATTCAGCCCCAGCGTATCAGAGTC AATTTTCTTCAGATCAATAGTGATAGTCTGGCCGTCATTCGCACCAACCTGAATTTTCATCGAACCGTC TTTTGCCAGTACGTTCACGCCGTTGAACTGGGTCTGGCCGGATACGCGGTCAATTTCGTCGAGACGGGA TTTGATTTCGTCCTGGATGGAGTCCAGGTCAGAATCGGAGTTAGTCCCTGTAGTGGCCTGAACCGTCAG TTCACGAATACGCTGTAAGTTGTTGTTGATTTCGGACAGCGCGCCTTCAGTGGTCTGCGCAACGGAGAT ACCGTCGTTGGCGTTACGGGCCGCCTGAGTCAGGCCTTTAATATTAGAAGTAAAACGGTTAGCAATCG CCTGACCCGCTGCGTCATCCTTCGCGCTGTTAATACGCAAGCCAGAAGACAGACGCTCGATAGAACTC GACAGCGCAGACTGGTTCTTGTTNN
12.3551	NNTGAACAAATCTCAGTCTTCTCTTAGCTCTGCTATTGAGCGTCTGTCTTCTGGTCTGCGTATTAACAGC GCAAAAGACGATGCAGCAGGTCAGGCGATTGCTAACCGTTTTACGGCAAATATTAAAGGTCTGACCCA GGCTTCCCGTAACGCGAATGATGGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGAATGAAATTA ACAACAACCTGCAGCGTATTCGTGAACTTTCTGTTCAGGCAACTAACGGTACTAACTCTGACAGCGATC TTTCTTCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACTCAGT TTAACGGCGTGAAAGTCCTTGCTGAAAATAATGAAATGAAAATTCAGGTTGGTGCTAATGATGGTGAA ACCATCACTATCAATCTGGCAAAAATTGATGCGaAAACTCTCGGCCTGGACGGTTTTAATATCGATGGC GCGCAGAAAGCAACCGGCAGTGACCtGATTTCTAAATTTAAAGCGACAGGTACTGATAATTATCAAATT ANCGGTACTGATAACTATACTGTTAATGTAGATAGTGGAGTAGTACAGGATAAAGATGGCAAACAAGT TTATGTGAGTGCTGCGGATGGTTCACTTACGACCAGCAGTGATACTCAATTCAAGATTGATGCAACTAA GCTTGCAGTGGCTGCTAAAGATTTAGCTCAAGGTAATAAGATTGTCTACGAAGGTATCGAATTTACAA ATACCGGCACTGGCGCTATACCTGCCACAGGTAATGGTAAATTAACCGCCAATGTTGATGGTAAGGCT GTTGAATTCACTATTTCGGGGAGTACTGATACATCAGGTACTAGTGCAACCGTTGCCCCTACGACAGCC CTATACAAAAATAGTGCAGGGCAATTGACTGCAACAAAAGTTGAAAATAAAGCAGCGACACTATCTG ATCTTGATCTGAACGCTGCCAAGAAAACAGGAAGCACGTTAGTTGTTAACGGTGCAACTTACGATGTT AGTGCAGATGGTAAAACGATAACGGAGACTGCTTCTGGTAACAATAAAGTCATGTATCTGAGCAAATC AGAAGGTGGTAGCCCGATTCTGGTAAACGAAGATGCAGCAAAATCGTTGCAATCTACCACCAACCCGC TCGAAACTATCGACAAAGCATTGGCTAAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAAC CGTTTCGACTCTGCCATCACCAACCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCGTAGCCGTATC GAAGATGCTGACTACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTAC CTCTGTTCTGGCNN

12.3552	NNGCCAGAACAGAGGTACCCGCTTGTTGCAGGATCTGCGCACGAGACATGTTAGACACTTCGGTCGCG TAGTCAGCATCTTCGATACGGCTACGGGCAGAAGACAGGTTGTTTACGGTGTTGCCAAGGTTGGTGAT GGCAGAGTCGAAACGGTTTTGTACTGCACCGAGGTCAGAACGCAGATTGTCAACTTTAGCCAATGCTT TGTCGATAGTTTCGAGCGGGTTGGTGGTAGATTGCAACGATTTTGCTGCATCTTCGTTTACCAGAATCG GGCTACCACCTTCTGATTTGCTCAGATACATGACTTTATTGTTACCAGAAGCAGTCTCCGTTATCGTTTT ACCATCTGCACTAACATCGTAAGTTGCACCGTTAACAACTAACGTGCTTCCTGTTTTCTTGGCAGCGTT CAGATCAAGATCAGATAGTGTCGCTGCTTTATTTTCAACTTTTGTTGCAGTCAATTGCCCTGCACTATTT TTGTATAGGGCTGTCGTAGGGGCAACGGTTGCACTAGTACCTGATGTATCAGTACTCCCcGAAATAGTG AATTCAaCAGCCTTACCATCAACATTGGCGGTTAATTTACCATTACCTGTGGCAGGTATAGCGCCAGTG CCGGTATTTGTAAATTCGATACCTTCGTAGACAATCTTATTACCTTGAGCTAAATCTTTAGCAGCCACT GCAAGCTTAGTTGCATCAATCTTGAATTGAGTATCACTGCTGGTCGTAAGTGAACCATCCGCAGCACTC ACATAAACTTGTTTGCCATCTTTATCCTGTACTACTCCACTATCTACATTAACAGTATAGTTATCAGTAC CGTTAATTTGATAATTATCAGTACCTGTCGCTTTAAATTTAGAAATCAGGTCACTGCCGGTTGCTTTCTG CGCGCCATCGATATTAAAACCGTCCAGGCCGAGAGTTTTCGCATCAATTTTTGCCAGATTGATAGTGAT GGTTTCACCATCATTAGCACCAACCTGAATTTTCATTTCATTATTTTCAGCAAGGACTTTCACGCCGTTA AACTGAGTTTGCTCAGATACACGGTCAATTTCTTCCAGACGTTGAGTAATTTCAGCCTGGATAGAAGAA AGATCGCTGTCAGAGTTAGTACCGTTAGTTGCCTGAACAGAAAGTTCACGAATACGCTGCAGGTTGTT GTTAATTTCATTCAGCGCACCTTCAGTGGTCTGCGCAACAGAAATACCATCATTCGCGTTACGGGAAGC CTGGGTCAGACCTTTAATATTTGCCGTAAAACGGTTAGCAATCGCCTGACCTGCTGCATCGTCTTTTGC GCTGTTAATACGCAGACCAGAAGACAGACGCTCAATAGCAGAGCTAAGAGAAGACTGAGATTTGTTC ANCN

REFERENCES

1. Ørskov, I., Ørskov, F., Jann, B., Jann, K. 1977. Serology, Chemistry and genetics of O and K antigens of *Escherichia coli*. Bacteriol. Rev. **41**: 667-710.

2. Tiba, M.R., De Moura, C. 2011. Identification of Putative New *Escherichia coli* Flagellar Antigens from Human Origin Using Serology, PCR-RFLP and DNA Sequencing Methods. Braz. J. Infect. Dis. **15**. Print.

3. Machado, J., Grimont, F. 2000. Identification of *Escherichia coli* Flagellar Types by Restriction of the Amplified *fliC* Gene. Res. Microbiol. **151**: 535-546. Print.

4. Wang, L., Rothemund, D., Curd, H. 2003. Species-Wide Variation in the *Escherichia coli* Flagellin (H-Antigen) Gene. J. Bacteriol. **185**: 2936-2943. Print.

5. McPherson, M., Moller, S. 2007. *The Basics: PCR*. New York, NY: Taylor & Francis.

6. Sanger, F., Coulson, A.R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. **94** (3): 441–448.

7. Tropp, B. E. 2012. *Molecular Biology 4E: Genes to Proteins.* The α-Amino Acids. Burlington, MA: Jones and Bartlett Publishing.

8. Iguchi, A., Ooka, T., Ogura, Y. 2008. Genomic comparison of the O-antigen biosynthesis gene clusters of *Escherichia coli* O55 strains belonging to three distinct lineages. Microbiol. **154** (2): 559-570.

9. Reid, S., Selander, R., Whittam, T. 1999. Sequence diversity of flagellin (*fliC*) alleles in pathogenic *Escherichia coli*. J. Bacteriol. **181** (1): 153-160.

10. Abbadi, S.H. and Strockbine, N.A. 2007. Identification of *Escherichia coli* flagellar types by restriction of the amplified *fliC* gene. Egypt. J. Med. Microbiol. **16** (2): 225-232.

ACADEMIC VITA

Samantha Fanelli

SLF5165@psu.edu

Education

B.S., Veterinary and Biomedical Sciences, Expected May 2013, Minor in Equine Science, The Pennsylvania State University, University Park, PA

Honors and Awards

- James A. and Donna E. Bochy Scholarship in Agricultural Sciences, Penn State College of Agricultural Sciences, 2012
- Rosie and Stuart Kahan Scholarship in Animal Health, Penn State College of Agricultural Sciences, 2011
- Richard H. Baker 4-H Scholarship, 4-H of Pennsylvania, 2010
- Penn State Chapter of Gamma Sigma Delta, College of Agriculture Honors Society, Penn State College of Agricultural Sciences, 2010

Memberships/Activities

- Penn State Pre-Vet Club, 2009-Present
- Penn State Lion Ambassador, 2012-Present
- Penn State Equestrian Team, 2010-2012

Professional Experience

- *E. coli* Reference Center, Research Assistant (2010-Present)

Research Interests

I have broad interests in microbiology and pathology. Specifically, my interests lie in virulence factors and the ways in which they can be altered to cause pathogenic effects in a host.

Publications and Papers

Fanelli, S. (2011). Detection of Virulence Factors Shiga-Toxin 1 and 2 and Intimin in *Escherichia coli* O104. 1-5.

Fanelli, S. (2010). CTX-M-15: Good News and Bad News. 1-5.