ABSTRACT

Title of Document: ALTERNATIVE APPROACHES IN MOLECULAR CHARACTERIZATION OF FOODBORNE PATHOGENS: SHIGA TOXIN-PRODUCING *Escherichia coli* AND *Salmonella* SEROTYPES.

Magaly Alejandra Toro Ibaceta, Doctor of Philosophy, 2014

Directed By: Professor Jianghong Meng, Department of Nutrition and Food Science

Shiga toxin-producing *E. coli* (STEC) and *Salmonella enterica* subspecies *enterica (S. enterica)* are two major foodborne pathogens. They cause almost 1.5 million of cases of disease each year in the US. Due to their public health impact, development of new methods for their detection and identification are top priority. This research focused on identifying alternative molecular methods and markers for the identification of STEC and *Salmonella*.

First, a suspension array was developed to simultaneously identify the seven most prevalent STEC (O26, O45, O103, O111, O121, O145, and O157) in the US. The panel targeted genes *wzx* or *wzy* and Shigatoxin genes. Testing and optimization employed four to eleven isolates of each serotype in the panel. STEC fluorescence values were 30 to >270 times greater than those of negative controls, demonstrating the method's effectiveness for the molecular serotyping of STEC.

STEC strains (n=194) of 43 serotypes were examined for clustered regularly interspaced short palindromic repeats (CRISPR) arrays to study relatedness among serotypes. A subset of strains (n=81) was analyzed for *cas* and virulence genes to determine a possible relationship. CRISPR spacer content correlated well with serotypes, although some strains with different serogroup but the same H type shared identical arrays (O26:H11, O103:H11, and O111:H11). *cas* and virulence genes were not associated, but strains with greater probability of causing outbreaks and disease showed fewer spacers than those less likely to cause them ($p<0.05$). Therefore, CRISPR array content correlated well with STEC serotype, and CRISPR-*cas* systems were inversely related to strain virulence potential.

Finally, the CRISPR arrays of 221 *S. enterica* of 53 serotypes were analyzed to define their relationship. CRISPR-cas systems of 50 *S. enterica* serotype Bareilly (*S.* Bareilly) were analyzed to resolve intra-serotype variations. CRISPR arrays correlated well with serotypes, although some serotypes displayed more than one type of array (e.g. *S.* Bareilly). Additionally, CRISPR-cas system elements reflected *S.* Bareilly phylogeny, but the array content was not linked to food vehicle or isolate's geographical origin. In conclusion, CRISPR array are useful for designing molecular serotyping assays, but a range of strains should be included to account for variation in *S. enterica*.

ALTERNATIVE APPROACHES IN MOLECULAR CHARACTERIZATION OF FOODBORNE PATHOGENS: SHIGA TOXIN-PRODUCING *Escherichia coli* AND *Salmonella* SEROTYPES.

By

Magaly Alejandra Toro Ibaceta.

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Jianghong Meng, Chair
Professor Mei-Ling Ting Lee
Professor Robert Buchanan
Dr. Y. Martin Lo
Dr. Xianghe Yan

Dedication

A mis padres, ejemplo de trabajo y esfuerzo, quienes me educaron con recursos limitados en tiempos turbulentos.

Acknowledgements

First, I would like to express my great appreciation to my advisor, Dr. Jianghong Meng, for the guidance, support and trust along this journey. I cannot imagine a better advisor, not only in academics, but he also has been a mentor for life events; something really important when family is thousands of miles away.

Also, I would like to thank the members of my doctoral committee for their time and commitment with this work. Dr. Lee and Dr. Yan were always willing to take some extra time to listen to me; thank you for your kindness! I also need to thank Dr. Robert Buchanan for his useful and constructive recommendations on this project, and for our enlightening discussion sessions. I learned so much! Dr. Martin Lo became a source of hope and positive energy in moments of tiredness. I will always appreciate his hard work not only in this committee, but also his dedication to undergraduate and graduate students in our department. You have been such an inspiration!

This research was possible thanks to the JIFSAN/UM partnership with the FDA. I want to specially thank Dr. Eric Brown and Dr. Marc Allard from the Center for Food Safety and Applied Nutrition (CFSAN) for their kind support, as well as Dr. Shaohua Zhao and Sherry Ayers from the Center for Veterinary Medicine (CVM) for all of their help.

I also received fundamental help from my lab mates. I am particularly grateful for the assistance given by Dr. Guojie Cao, who helped with phylogenetic analysis,

almost daily weather forecasts, and food provisions in long evenings of work. Dr. Wenting Ju and Dr. Lydia Rump contributed in many ways, including data collection and analysis, edition, and many moments of conversation and laughs. Although our visitors Jinling Shen, Yu Hu, and Likou Zhou were only temporarily on our team, they made our lab feel like a family. I feel very lucky to have found such an incredible team. More importantly, I want to thank them all for their friendship.

The Department of Nutrition and Food Science was very supportive as well. I want to thank specially to Dr. Robert Jackson and Dr. Nadine Sayhoun for their support and to Sara, Nenita, Rita, and Mary for their very important help during all these years. Classmates Dr. Nancy (Tong) Liu, Dr. Margaret Slavin, Ms. Ruth Oni, Dr.Herman Lutterodt and so many others helped with academic and non-academic conversation and support.

This project would have been impossible without the economic support from several institutions: The Fulbright program, Conicyt-Chile, and the University of Maryland through the Ann Wiley Dissertation Fellowship and the Dean Fellowship from the NFSC department and the College of Agriculture and Natural Resources. I also would like to thank to the ELI program at Oregon State University who gave me the tools I needed to succeed in a graduate program in a second language.

The completion of my PhD program would not been possible without great friends. In order of appearance: Paulina Perez, Marcelo Manterola, Danae Azuara, Angelo Gomez and Karina Herrera, among others were friends living this wonderful graduate school experience as well. We shared not only academic accomplishments, but also important life events. We grew together. Also, thanks to the

Fulbright/BIO2008 group who, from the distance, kept a sense of community while living the challenges of a brand new life in a foreign country. My dear friends back in Chile were a great support for keeping in touch and listening in my happy and sad times all these years. I specially thank Alvaro Cabrera and the group he created over 40 years ago…not just a choir, but a group of friends who sing: El coro Antumapu. Many of those friends have been supporting me from a distance.

My parents deserve most of the credit for this achievement. They gave me a great education with very limited resources. Also, thank you for teaching me that responsibility and honesty are the most significant values in life. Thanks to my brother Fabian for all the love and care he has shown to me and to Roman for taking care of my parents. I could not have left my country without knowing my parents were going to be loved and taken care of.

Finally, I want to thank my personal cheerleader, the most important person in my life: My husband Charles. There is no way I could ever have reached this goal without his unconditional support and love. I know all the sacrifices you do every day for us, and I will not forget them. I love you so much!

Table of Contents

List of Tables

Table IV.4. General characteristics of CRISPR arrays from *Salmonella enterica* serotypes (n=224).

List of Figures

List of Abbreviations

ANOVA: analysis of variance

BLAST: Basic Local Alignment Search Tool

bp: base pairs

cas: CRISPR associated genes

CASCADE: CRISPR associated complex for antiviral defense

CDC: Centers for Disease Control and Prevention

CFSAN: Center for Food Safety and Applied Nutrition

CRISPR: Clustered regularly interspaced short palindromic repeats

DNA: Deoxyribonucleic acid

EDC: 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride

EHEC: enterohaemorragic *E. coli*

ERIC: Enterobacterial repetitive intergenic consensus

FDA: US Food and Drug Administration

FWD: Forward

HC: Hemorrhagic Colitis

HIV: Human Immunodeficiency Virus

HPI: High Pathogenicity Island

HUS: Hemolytic Uremic Syndrome

LPS: Lipopolysaccharide

MFI: Median Fluorescence Intensity

MLST: Multi Locus Sequence Typing

NCBI: National Center for Biotechnology Information

NM: Non motile

nr: non redundant

nt: nucleotides

ONT: O antigen non typifiable

OR: O antigen rugose

P/N ratio: Positive MFI /Negative control MFI ratio

PCR: Polymerase Chain Reaction

PFGE: Pulse Field Gel Electrophoresis

REV: reverse

SA-PE: Streptoavidine- Phicoeritrine complex

SD: standard deviation

SNP: Single nucleotide polymorphism

SPT: Seropathotype

STEC: Shiga toxin-producing *Escherichia coli*

Stx: Shiga toxin

TMAC: Tetra Methyl Ammonium Chloride

UN: Un-typifiable

US: United States of America

WGS: Whole genome sequencing

Chapter 1: LITERATURE REVIEW

Introduction

Major foodborne pathogens cause approximately 9.4 million illnesses in the United States each year, including more than 55,000 hospitalizations and 1,351 deaths (1). Shiga toxin-producing *E. coli* (STEC) and *Salmonella* are two of the most important foodborne pathogens; they were implicated in 15 of the 17 multistate outbreaks produced in 2008 (2), and it is estimated that altogether caused almost 1.5 million of diseases a year (1).

Some STEC and *Salmonella* are associated with human diseases more frequently than others; therefore, typing foodborne pathogens is fundamental for epidemiological studies. The present review focuses on Shiga toxin-producing *E. coli* and *Salmonella* as foodborne pathogens and novel approaches for their typing.

Shiga toxin-producing *Escherichia coli*

Overview

Most *E. coli* are normal member of the gastrointestinal flora; however, some subgroups are a threat for human health (3). Shiga toxin-producing *E. coli* (STEC) is among the most important causes of foodborne diseases in the United States and in the world (1, 2, 4, 5). STEC causing severe diseases such as hemorrhagic colitis (HC)

and hemolytic uremic syndrome (HUS) are called enterohaemorragic *E. coli* (EHEC), and *E. coli* O157:H7 is the most known serotype.

In early years, research focused on *E. coli* O157:H7 as it was the first STEC causing severe disease and outbreaks (6-8), but more than 400 serotypes have been involved in human disease (9). In the US, six serogroups –O26, O45, O103, O111, O121 and O145— cause 70% of non O157 STEC diseases (10, 11). As a consequence, regulatory agencies recognized the importance of other serogroups and US agencies required active surveillance of STEC in early 2000's (12), and new regulations increased the awareness of the big six non-O157 STEC serogroups (13).

STEC main virulence factor is Shiga toxin (Stx), a cytotoxin similar to *Shigella dysenteriae* toxin type 1 (14). STEC harbors two main types of Stx: Shiga toxin 1 (Stx_1) and Shiga toxin 2 (Stx_2), each one with several variants (Stx_{1a}, Stx_{1c}, Stx_{1d}, Stx_{2b}, $Stx_{2c,}$ $Stx_{2d,}$ Stx_{2dact}, Stx_{2e}, Stx_{2f}, Stx_{2g},) (15) which damage the intestinal and kidney cells in cases of HC and HUS respectively (16).

STEC detection, isolation and characterization

STEC detection and identification are challenging. Metabolic features make *E. coli* O157:H7 detection easy by growing colonies on sorbitol-McConkey agar. However, other STECs lack unique characteristics useful for detection (17). STEC isolation and detection takes several steps. First, samples are cultured and colonies are biochemically tested (14, 18). Then, other methods measure toxic effects of Stx on Vero cells (19), capture the toxin itself by immunological methods (11), amplify *stx*

genes by PCR (20) or hybridize *stx* genes directly from colonies (21). Finally, isolates are serotyped by a modified Kauffman scheme originally created in 1944 (14).

E. coli serotyping

E. coli serotypes are defined after their O –somatic— and H –flagellar— antigens. The somatic O antigen is a component of the outer membrane in gram negative bacteria, and 179 groups are defined to date (22). Subunits of two to six carbohydrates structure an O unit, and 20 to 30 O units form the lateral polysaccharide chain called O antigen (23) that forms part of the lipopolysaccharide (LPS), the essential element of the outer membrane. Additionally, there are 53 H antigens, proteins that form the flagella (22).

The O- antigen gene cluster carries genetic information to synthesize the O antigen, and it is part of the bacterial chromosome (23). It is generally located between two housekeeping genes –*gnd* and *gal*F –, and the arrangement and number of genes in each O group vary depending on the complexity of the polymer (Fig. I.1). Proteins encoded by this cluster have diverse functions during the O-antigen synthesis process. For example, genes *wzx* and *wzy* codify for the O antigen processing proteins flippase and polymerase. These two genes have conserved DNA sequences among a serotype, but not between serotypes. Due to the multiple combinations of sugar residues, anomeric configurations and posterior linkages with non-sugar residues, the O-antigen is considered exceptionally heterogeneous (23, 24).

Some STEC serogroups are epidemiologically more related to disease. Thus, serogroup identification is fundamental to investigate outbreaks and to prevent

disease (11). STEC serogrouping relies on the use of a large set of antibodies raised in rabbits against each of the known *E. coli* LPS and flagellar antigens. Consequently, serotyping is restricted to laboratories that have the antisera. The technique is time consuming, and cross reactions among different antisera may occur. Moreover, serotyping is not suitable for every strain, resulting in non-typeable strains. These limitations drove the development of molecular methods which provide simplified *E. coli* characterization tools (24, 25).

Several molecular approaches have been developed for STEC serogrouping. Many of these assays target unique genes such as *wzx* (O-antigen flipase) and *wzy* (O-antigen polymerase). Several teams designed simplex and multiplex PCR (22, 25-31) for the detection of one or several serotypes at the same time. Real time PCR takes advantage of probes, improving assay specificity (32-36); however, they allow only a few targets at the same time. Microarrays can detect many targets at the same time (20, 37-40), but they are not easily customizable and require specialized equipment for the microchip printing and detection.

Suspension Arrays

In recent years, suspension arrays emerged as an alternative to traditional, planar microarrays. Advantages are they detect several targets in a single reaction –up to 100 in the case of the Luminex ® and Bioplex ® platforms—and can use different kind of probes –nucleic acids, antibodies, lipids or carbohydrates. Suspension arrays provide numeric data, which is more robust than qualitative information. Each data set is obtained from at least 100 repeats for each target, presenting statistical

superiority over other methodologies. Finally, suspension arrays are flexible allowing homemade design and modifications (41-44).

Suspension arrays have two main components; color codified polystyrene beads and a specialized flow cytromety-like device that detects and classifies beads (Fig. I.2.) (41). The beads are dyed with two flourochromes mixed in different ratios to produce 100 combinations or bead sectors (45). Probes are attached to bead surface and then added to the reaction to hybridize the target. The flow cytometry-like device analyzes the reaction by identifying bead types (color code), counting beads of each type, and measuring and recording the fluorescence emitted by the reporter molecule (42, 43, 45).

Bead-based suspension arrays have identified and detected different targets in molecular biology, immunology, clinical sciences, food microbiology, environmental monitoring, and many others (42-44, 46, 47). In food microbiology, suspension arrays have detected *E. coli, Salmonella, C. jejuni* and *L. monocytogenes*. For instance, nucleic acid and antibodies probes identified *E. coli* at 10^3 cells and 2.5 cells/ml respectively, demonstrating the system's flexibility and sensitivity (45). Also, suspension arrays identified *Campylobacter spp*. and assessed isolate antimicrobial resistance with accuracy equivalent to that obtained by sequencing (48). A suspension array panel serotyped *Salmonella* associated with foodborne illness: A first assay identified the six most common serogroups in the US and serotype Paratyphi A (46), and a second assay identified *Salmonella* H antigens (49).

Salmonella enterica subspecies *enterica*

Overview

Salmonella enterica subspecies *enterica* is a major cause of bacterial foodborne diseases in the US with more than 40,000 laboratory confirmed cases each year. Latest estimations indicate *Salmonella* causes almost 1,300,000 cases a year (1). Most *Salmonella* diseases are self-limited, but death rate reaches 0.5% (1, 50). *Salmonella* also has high prevalence in other countries (51) where less common *Salmonella* serotypes are associated to outbreaks; in Africa, its prevalence increased in children and HIV positive population (52).

Salmonella serotypes have a broad range of hosts, producing different symptoms depending on the host species (50). *Salmonella* can cause persistent infections, so hosts can remain as reservoirs for a long time (51). Finally, *Salmonella* antibiotic resistance is increasing, so treatment of invasive cases of infection is more difficult and risk of death has increased (50, 51, 53).

The main reservoirs of *Salmonella enterica* are food animals (51), and animal derived foods have been frequently associated to disease (54). However, outbreaks have also involved produce, evidencing that *Salmonella* can invade and survive in different types of foods (55). For example, some of the last outbreaks of *Salmonella* in the US involved pine nuts, ground turkey, fresh papayas, dwarf frogs, alfalfa sprouts, cantaloupes, shell eggs, frozen entrees, peanut butter, etc (54, 56).

Salmonella nomenclature and serotyping

Salmonella spp. was first described in the second half of the 19th century, but the name *Salmonella* was adopted in 1900. In 1934, Kauffman and White proposed a

classification scheme based on *Salmonella* antigenic reactions (49). Since then, thousands of *Salmonella* serotypes have been discovered in different hosts, and new ones are described every year. *Salmonella* nomenclature has changed, but thanks to molecular methodologies it is possible to understand their phylogeny (50).

Salmonella enterica is further divided into six subspecies (I, II, IIIa, IIIb, IV, VI, and VII), each one with a variable number of serotypes (49, 50). Similar to *E. coli*, *Salmonella* serotyping is based on the reaction to O (somatic) and H (flagellin) antigens; however, *Salmonella* serotyping is more complicated because they can present two flagellin phases (49). There are 46 *Salmonella* O antigens and 114 H antigens, and different combinations have created 2587 serotypes so far (49). *Salmonella* antigenic formula includes subspecies and the O, H1 and H2 antigens. Traditionally, serotypes of subspecies I (*Salmonella enterica* subspecies *enterica*), are named after the place they were first isolated. Traditionally, serotypes' names are used together with the genera name, omitting species and subspecies. For example, the name *Salmonella* Saintpaul designates *Salmonella enterica* subspecies *enterica* serotype Saintpaul (50).

Salmonella serotyping is fundamental for outbreak investigation (49). Also, some serotypes cause different symptoms depending on the host, and some are more frequently associated with diseases and outbreaks (51). For instance, *Salmonella* serotypes Newport, Typhimurium, Enteritidis, Montevideo and Saintpaul are in the top 10 serotypes list of CDC and FDA at the same time. Traditional *Salmonella* serotyping can take 5 days or more (49), and it has the same disadvantages of *E. coli* serotyping. The 20 most common serotypes from human specimens account for about

70% of all isolates reported in the United States; the top 100 serotypes account for about 98% of all isolates (57).

Molecular serotyping and subtyping of *Salmonella spp*

Molecular genotyping of Nontyphoidal *Salmonella* is used to identify and track strains related more frequently to disease and outbreak (51). The technology uses specific genomic sequences that differentiate *Salmonella* serotypes.

To serotype *Salmonella,* several sequences are commonly used: O antigen gene cluster genes *wzx* and *wzy* genes are O antigen specific. Similarly to *E. coli*, the cluster is located between the genes *gal*F and *gnd*, but it is known as the *rfb* region (23). *Salmonella* H antigen sequences are located in two flagellin loci in the genome; *fli*C and *flj*B. The first one is present in every *Salmonella* strain, and the second one is additional present in diphasic *Salmonella* (58). Both genes have conserved flanks but variable central portions, making them ideal for molecular H typing (49).

To molecular serotype and subtype *Salmonella enterica,* different approaches have been used: Molecular serotyping based on genes coding for O and H antigen based on multiplex PCR (59) and suspension arrays (46, 49), PFGE (60) , Microarrays, and MLST (61) among others were developed (62). New molecular tools for detection and characterization of *Salmonella enterica* are highly important for the control and prevention of Salmonellosis (62).

bbbbbbbbbbbbbbb

Salmonella enterica subspecies *enterica* serotype Bareilly

Salmonella enterica subspecies *enterica* serotype Bareilly (*S.* Bareilly) was first described in 1928 in India and it has been linked not only to gastroenteritis, but also to nosocomial infections (63, 64). In the early 1950's, it was one of the serotypes used for *Salmonella* experimental infection studies; dosages starting at 125,000 cfu caused disease in human volunteers (65). Currently, *S.* Bareilly is one of the 20 most prevalent *Salmonella* serotypes in the US; from 1999 to 2009, confirmed cases rose from 171 to 284 (66%), and its ranking escalated from number 23 to 19. However, there are not many studies in this serotype.

S. Bareilly has been isolated from different host (66). It was found in cobras in Thailand (67), pasture feed chickens environment (68), diverse animals in India (66, 69), etc. In Japan, *S.* Bareilly was the third more frequently isolated serotype from sporadic diarrhea from humans and was also frequently found in lying hens (70), and it was isolated from raw poultry in Ireland (71).

In the past years, this serotype has been associated to several outbreaks (70, 72, 73). Contaminated bean sprouts affected European countries in 2006 and 2010 adding up to over 500 cases (73, 74), and recently, *S.* Bareilly was implicated in a multistate outbreak associated to raw scraped ground tuna product, causing over 400 cases in the United States (72). *S.* Bareilly subtyping strategies were approached in past years: a bacteriophage typing was developed two decades ago (75), and ERIC-PCR was able to differentiate between strains in 2002 (76); however, there are not new studies in the topic.

Other studies focused on antimicrobial resistance of *S.* Bareilly isolated from food. In the US, isolates from pasture raised chicken farms showed high levels of resistance against Sulfisoxazole and Novobiocin and intermediate resistance against Tetracycline, Neomycin, and Streptomicyn (68). Also, *S.* Bareilly was the 9th most found *Salmonella* in imported foods (77), and one of five *S.* Bareilly found (dace fish, Vietnam) displayed resistance to Ampicillin, Chloramphenicol, Sulfamethoxazol, Tetracyclin, and Trimethoprim/sulfamethoxazole. Conversely, 16 *S.* Bareilly isolated from Indian seafood were susceptible to every antimicrobial tested (78).

The CRISPR-cas system

Description

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) are loci described in approximately 40% of bacteria and 90% of archaea (79-81). They represent a heritable immune systems against mobile genetic elements (80, 82-85), with self-regulatory (86, 87) and self DNA repair (88, 89) functions proposed as well.

CRISPR was discovered in *E. coli* K-12 as an arrangement of repetitive sequences located downstream to the gene *iap;* sequences of 29 nucleotides (nt) were separated by variable spacers regions of 32 nt (90). Later, similar structures were found in other bacteria and archaea (80, 91). The loci also included a leader sequence and CRISPR associated proteins (cas).

CRISPR-cas system organization

Although CRISPR-cas systems vary between species (80, 91, 92), they have three general components (Fig. I.3):

a) CRISPR array

They are formed by *Repeats*, originated from endogenous DNA sequences, alternated by *Spacers*, acquired from foreign DNA. *Repeats* are of almost indistinguishable size and sequence in a defined locus (81, 85, 93), and due to their palindromic nature, they form stable secondary structures that are fundamental for CRISPR function (79). *Spacers'* length varies from 20 to 72 bp, but they are unique and their length is constant in the same array (80, 81). Array length and number of arrays varies between species and strains (80, 87, 94); documented array length goes from 2 to 375 repeats (85).

b) Leader sequence

The *Leader* sequences are 20 to 534 base pair AT rich sequences located immediately upstream from the array (80). Their sequence is conserved within the same species but not between them (81, 91). Leader sequences lack open reading frames, but they carry the promoter for the array transcription (95); therefore, they are a crucial component of CRISPR systems, and arrays lacking the leader sequence cannot acquire new spacers, express or cause interference (91, 95).

c) Cas proteins

CRISPR associated proteins (Cas proteins) are closely related to CRISPR loci. *cas* genes surround the array, but may be located in different orientation and order (81, 96).

Originally, Cas proteins were classified into 3 families: i) Core cas proteins, *cas* 1 to 6 in different combinations; ii) Eight specific cas proteins subtypes – CRISPR-Cas subtypes Ecoli, Ypest, Nmeni, Dvulg, Tneap, Hmari, Apern, and Mtube— associated with Core cas genes but limited to a narrower group of species; and iii) Modular cas genes that can be distant from the CRISPR array (92). Under a new classification, genes *cas1* and *cas2* are universal markers for CRISPR-cas systems (85, 97), yet genes codifying for nucleases, helicases, and polymerases may be present (85, 96, 98, 99). In *E. coli*, CRISPR 1's genes form the I-E CRISPR-cas system, and it includes genes *cas2, cas1, cas6e, cas5, cas7, cse2, cse1*, and *cas3* (97). Genes *cas6e, cas5, cas7, cse2, cse1* form CASCADE –CRISPR associated complex for antiviral defense—complex with roles in maturation and interference steps. CASCADE transcripts links to one copy of crRNA, and is the basis for recognition and neutralization of alien DNA (99).

CRISPR mechanism of action

CRISPR's mechanism of action is divided in three stages (82, 83, 100, 101).

a) CRISPR adaptation:

Acquisition and incorporation of new spacers that occurs in two phases:

i) Sampling: After bacterial systems recognize foreign DNA from phage, plasmids or other mobile genetic elements inside the cell (102), short sequences, known as protospacers, are removed from the invader (81, 84, 91).

ii) Integration: CRISPR array inserts a protospacer as new spacer, proximal to the Leader sequence (85).

b) CRISPR expression

Assembly of the CRISPR functional structure also happens in two steps:

i) Transcription: Unidirectional transcription of the entire CRISPR, from the leader sequence to the terminal spacer, creating a long pre-crRNA. (103, 104).

ii) Maturation: Small crRNA are cleaved from the long pre-crRNA and linked to CASCADE (84). Each crRNA is formed by an entire spacer and short flanking regions from the adjacent repeats. The 5' crRNA side starts with the last 8 nucleotides of the 5'flanking repeat (103), and it ends with a non-constant number of nucleotides coming from the 3' flanking repeat (91). Handles are conserved sites for the binding of CASCADE (103).

c) CRISPR interference

crRNA/CASCADE complex inactivates the phage by nucleic acid degradation (84). A spacer, homologous to the foreign nucleic acid (82, 101), works as a probe for the CRISPR-mediated interference system (82, 103, 105). The complex binds to a complementary sequence forming a double stranded DNA molecule, and an R loop is formed through Watson and Crick base pairing between crRNA and the protospacer (99). The recognition begins at the 5' region, and further base pairing advances in 3'

direction (98, 106). The R loop works as a marker for the catalytic action of enzymes such as cas3 (99).

Interference depends on the base pairing between the crRNA and the protospacer, but the system accepts up to 4 to 5 mismatches before halting (106).

Use of CRISPR for bacterial subtyping

Spacer acquisition is as a natural process in prokaryotes, and changes in spacers would evidence population changes (91, 107). For example, two neighboring extremophiles populations had similar CRISPR systems, but their spacer content was totally different. Those communities were expected to be clonal (108). CRISPR spacers would show the exposures a bacterial lineage survived (81).

The high variability in spacers could be used for phylogenetic and evolutionary studies (91). Spacers should show extraordinary variability because they are a byproduct of infections by mobile elements present in the host's environment, and they should evidence geographical locations (85). Even closely related strains should display different spacers (81). Moreover, since newer spacers are acquired proximal to the leader sequence, CRISPR arrays represent a chronological record of invasions (109); therefore, distal location indicates an older origin, and those spacers tend to be more frequent among strains (110). However, not all the spacers are kept, and sporadic deletions of internal spacers are described (83).

Some studies have already explored the use of CRISPR systems/arrays to subtype bacterial populations:

Kamerbeek and cols. developed a method –Spoligotyping—to detect and subtype *Mycobacterium tuberculosis* for diagnosis and epidemiology. After PCR and hybridization, the detection of particular spacers defined the specie and strain present (111). The methodology was evaluated later in a clinical setting, helping to select the right treatment and shortening diagnosis time in 70 days when compared to culture and antibiotic resistance tests (112).

CRISPR array was used to study the diversity of bacterial *Streptococcus spp.* communities in the oral cavity: The change in spacers in time evidenced an evolution of *Streptococcus spp*. oral flora (113).

CRISPR was useful for subtyping *Campylobacter jejuni*. A high resolution DNA melt curve analysis of 22 different CRISPR types concluded this methodology, used in combination with other real time techniques, provides an equal or superior method to PFGE subtyping (114).

CRISPR in *Salmonella* and *E. coli* subtyping

Basic research of the CRISPR-cas system has been carried out in *E. coli*; however, a few studies focused on the CRISPR array diversity of this species or in *Salmonella enterica*.

E. coli and *Salmonella* CRISPR arrays were described as poor epidemiological markers because of slow rate of change; strains that diverged in the last 250 thousand years had an almost identical CRISPR array (115). Touchon and cols. later confirmed that close related strains had identical CRISPR arrays (87). Oppositely, other studies found great variation in cas proteins and spacers' identity

(94), and recently, studies demonstrated the relationship between CRISPR array content and serotypes. Delannoy *et al.* reported the presence of specific CRISPR polymorphisms related to O:H serotypes of STEC which were useful to differentiate these serotypes (116, 117). Similarly, Yin *et al.* confirmed a relationship between CRISPR polymorphism and serotypes (118).

Salmonella enterica's CRISPR-cas system would drive *Salmonella* evolution; CRISPR-mediated immunity would prevent strains to evolve in the same way, promoting the development of new subtypes and emergence of new *S. enterica* sublineages. Also, CRISPR systems could be controlling the acquisition of plasmid and phage mediated horizontal gene transfer (HGT), which has been associated to virulence factors in *S. enterica.* (119).

CRISPR-cas system association with bacterial virulence

The function of CRISPR-cas systems would be to protect bacteria and archaea against foreign genetic elements. Since many virulence determinants are acquired through horizontal gene transfer (120), it is possible CRISPR-cas systems interfere with the acquisition of virulence determinants. However, it is still controversial whether there is a relationship between the acquisition of virulence elements and the presence of CRISPR-cas systems.

Multiples studies evidenced the relationship between CRISPR-cas systems and virulence elements in bacteria. *Streptococcus thermophillus* with longer CRISPR arrays were more resilient to alien DNA (82). The presence of CRISPR-cas systems was inverse correlated to multidrug resistance in *Enterococcus faecalis,* but no

evidence of spacer identity with vectors for antibiotic resistance or known mobile genetic elements was found (121). An artificially incorporated spacer added to *S. thermophillus* CRISPR1 caused the loss of a plasmid carrying antibiotic (122). In *Enterococcus* spp, an inverse correlation was reported between the presence of two virulence genes and the distribution of *cas* genes, and fewer virulence genes were detected when *cas* genes were present (123). In *E. coli*, the acquisition of plasmids carrying antimicrobial genes was not related to the presence of the CRISPR-cas system (124). However, a recent study showed uropathogenic *E. coli* seemed less likely to have CRISPR loci than non-uropathogenic *E. coli* strains from the same patient, suggesting CRISPR-cas may have a role in the acquisition of phage and plasmids and serving as an adaptive advantage for the group (125). These findings are consistent with the documented role of CRISPR-*Cas* immune systems in limiting the uptake of genetic material derived from mobile and invasive elements such as phages and plasmids, yet experiments have failed in proving that wild type *E. coli* CRISPR systems actively function as immune systems (87, 101, 126).

New evidence indicates the relationship between CRISPR-cas systems and bacterial virulence may be indirect. For example, *cas9* from *Franciscella novocida* prevents host recognition by indirect gene regulation (127), and *Legionella pneumophilla*'s *cas2* is required for intracellular infection of amoebae –an amplification step in their lifecycle (128). Non-functional CRISPR are associated to more pathogenic *Campylobacter jejuni* strains; a higher production of gangliosides (linked to Guillain-Barré syndrome) would make strains more resistant to phage, decreasing the evolutionary pressure on CRISPR-cas system (129).

Research Overview

Serotyping is one of the most extensively used tools to characterize foodborne pathogens; information provided is of great use in epidemiological investigations (46). Conversely, traditional serotyping methods are restricted to few laboratories due to its technical difficulties. In recent years, genomic sequences are broadly available which has driven the development of different molecular techniques (36, 39, 40, 46). The application of new techniques and the identification of novel markers for molecular serogrouping, serotyping, and subtyping are two main lines of research in the food microbiology field. The aim of this research was to explore the use of different approaches in molecular serotyping of food borne pathogens, and the identification of molecular markers.

1. **To develop a simultaneous molecular serogrouping methodology for Shiga toxin-producing *Escherichia coli*.**

Molecular techniques identifying STEC serotypes have been previously developed, however, most of those assays require several runs to identify whether a strains belongs to one of the most common STEC serotypes. Thus, the first objective of this research was to develop a suspension array assay that identifies the 7 most common STEC isolates in a single reaction. The assay simultaneously indicates the presence of shiga toxin. This is the first time all these targets are detected in a single reaction assay. These findings were published in the journal Foodborne Pathogens and disease (130)

2. **To describe CRISPR array in shiga toxin-producing *Escherichia coli* (STEC) and determine the association of CRISPR-cas system elements with virulence profiles in STEC.**

CRISPR (Clustered regularly interspaced short palindromic repeats) are bacterial immune systems protecting the cell against the invasion of foreign elements. CRISPR systems are dynamics, and array section would change depending on the environment a bacterial lineage evolved. In this study, CRISRP arrays of STEC were analyzed to establish the relationships between CRISPR arrays and STEC serotypes. Additionally, since many virulence determinants are acquired through horizontal gene transfer, the relationship between different CRISPR-cas system elements and some virulence markers in STEC was investigated. These findings were recently published in the journal Applied and Environmental Microbiology (131)

3. **To describe CRISPR arrays in *Salmonella* serotypes and study CRISPR-cas system in *Salmonella* Bareilly (*S.* Bareilly).**

In this study, the relationship between CRISPR array content and *Salmonella* serotypes was analyzed. The use of CRISPR-cas system for subtyping was analyzed by studying the variation among strains of *S.* Bareilly, a non-well studied serotype. The relationships between CRISPR-cas system components and strain characteristics such as food of isolation, and geographical origin were investigated. Additionally, whether CRISPR –cas system behaves similarly to phylogenies in *S.* Bareilly was also explored.

References

1. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.
2. [**CDC] Centers for Disease Control and Prevention.** 2011. Surveillance for foodborne disease outbreaks --- United States, 2008. MMWR Morb Mortal Wkly Rep **60:**1197-1202.
3. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2:**123-140.
4. **Erickson MC, Doyle MP.** 2007. Food as a vehicle for transmission of Shiga toxin-producing *Escherichia coli*. J Food Prot **70:**2426-2449.
5. **Eurosurveillance.** 2011. EFSA publishes scientific report on the public health risk of Shiga-toxin producing *Escherichia coli* (STEC) in fresh vegetables. Euro Surveill **16** (23). Available at: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19888. Accessed January 7, 2014.
6. **Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N Engl J Med **308:**681-685.

7. **Johnson KE, Thorpe CM, Sears CL.** 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin Infect Dis **43:**1587-1595.

8. **Bettelheim KA.** 2007. The non-O157 shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. Crit Rev Microbiol **33:**67-87.

9. **Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J.** 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J Clin Microbiol **42:**311-319.

10. **Griffin P.** 2007. CDC perspective on non-O157 Shiga toxin-producing *E. coli* (STEC) in the United States. Available at: http://www.fsis.usda.gov/wps/wcm/connect/6200a224-5afa-48c8-a8c5-e87cb9c391f1/Non-0157_STEC_Griffin.pdf?MOD=AJPERES, accessed on January 2014.

11. **Gould LH, Bopp C, Strockbine N, Atkinson R, Baselski V, Body B, Carey R, Crandall C, Hurd S, Kaplan R, Neill M, Shea S, Somsel P, Tobin-D'Angelo M, Griffin PM, Gerner-Smidt P.** 2009. Recommendations for diagnosis of shiga toxin--producing *Escherichia coli* infections by clinical laboratories. MMWR Recomm Rep **58:**1-14.

12. **Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, Strockbine NA.** 2005. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983-2002. J Infect Dis **192:**1422-1429.

13. **[USDA] United States Department of Agriculture, Food Safety and Inspection Service.** 2012. Shiga Toxin Producing *Escherichia coli* in certain raw beef products. Available at: http://www.fsis.usda.gov/OPPDE/rdad/FRPubs/2010-0023FRN.pdf, accessed May2012.

14. **Nataro JP, Kaper JB.** 1998. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev **11:**142-201.

15. **Johannes L, Römer W.** 2010. Shiga toxins--from cell biology to biomedical applications. Nat Rev Microbiol **8:**105-116.

16. **Karmali MA, Petric M, Lim C, Fleming PC, Steele BT.** 1983. *Escherichia coli* cytotoxin, haemolytic-uraemic syndrome, and haemorrhagic colitis. Lancet **2:**1299-1300.

17. **Hussein HS, Bollinger LM.** 2008. Influence of selective media on successful detection of Shiga toxin-producing *Escherichia coli* in food, fecal, and environmental samples. Foodborne Pathog Dis **5:**227-244.

18. **Gyles CL.** 2007. Shiga toxin-producing *Escherichia coli*: an overview. J Anim Sci **85:**E45-62.

19. **Karmali MA, Steele BT, Petric M, Lim C.** 1983. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. Lancet **1:**619-620.

20. **Call DR, Brockman FJ, Chandler DP.** 2001. Detecting and genotyping *Escherichia coli* O157:H7 using multiplexed PCR and nucleic acid microarrays. Int J Food Microbiol **67:**71-80.

21. **Auvray F, Lecureuil C, Taché J, Leclerc V, Deperrois V, Lombard B.** 2007. Detection, isolation and characterization of Shiga toxin-producing *Escherichia coli* in retail-minced beef using PCR-based techniques, immunoassays and colony hybridization. Lett Appl Microbiol **45:**646-651.

22. **DebRoy C, Fratamico PM, Roberts E, Davis MA, Liu Y.** 2005. Development of PCR assays targeting genes in O-antigen gene clusters for detection and identification of *Escherichia coli* O45 and O55 serogroups. Appl Environ Microbiol **71:**4919-4924.

23. **Samuel G, Reeves P.** 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res **338:**2503-2519.

24. **Stenutz R, Weintraub A, Widmalm G.** 2006. The structures of *Escherichia coli* O-polysaccharide antigens. FEMS Microbiol Rev **30:**382-403.

25. **Fratamico PM, Briggs CE, Needle D, Chen CY, DebRoy C.** 2003. Sequence of the *Escherichia coli* O121 O-antigen gene cluster and detection of enterohemorrhagic *E. coli* O121 by PCR amplification of the wzx and wzy genes. J Clin Microbiol **41:**3379-3383.

26. **Wang L, Curd H, Qu W, Reeves PR.** 1998. Sequencing of *Escherichia coli* O111 O-antigen gene cluster and identification of O111-specific genes. J Clin Microbiol **36:**3182-3187.

27. **Fratamico PM, DebRoy C, Miyamoto T, Liu Y.** 2009. PCR detection of enterohemorrhagic *Escherichia coli* O145 in food by targeting genes in the *E.*

coli O145 O-antigen gene cluster and the shiga toxin 1 and shiga toxin 2 genes. Foodborne Pathog Dis **6:**605-611.

28. **DebRoy C, Roberts E, Kundrat J, Davis MA, Briggs CE, Fratamico PM.** 2004. Detection of *Escherichia coli* serogroups O26 and O113 by PCR amplification of the *wzx* and *wzy* genes. Appl Environ Microbiol **70:**1830-1832.

29. **Debroy C, Roberts E, Valadez AM, Dudley EG, Cutter CN.** 2011. Detection of Shiga Toxin-Producing *Escherichia coli* O26, O45, O103, O111, O113, O121, O145, and O157 serogroups by multiplex polymerase chain reaction of the *wzx* gene of the O-Antigen gene cluster. Foodborne Pathog Dis **8:**651-652.

30. **Ren Y, Liu B, Cheng J, Liu F, Feng L, Wang L.** 2008. Characterization of *Escherichia coli* O3 and O21 O antigen gene clusters and development of serogroup-specific PCR assays. J Microbiol Methods **75:**329-334.

31. **Valadez AM, Debroy C, Dudley E, Cutter CN.** 2011. Multiplex PCR detection of Shiga toxin-producing *Escherichia coli* strains belonging to serogroups O157, O103, O91, O113, O145, O111, and O26 experimentally inoculated in beef carcass swabs, beef trim, and ground beef. J Food Prot **74:**228-239.

32. **Hsu CF, Tsai TY, Pan TM.** 2005. Use of the duplex TaqMan PCR system for detection of Shiga-like toxin-producing *Escherichia coli* O157. J Clin Microbiol **43:**2668-2673.

33. **Tsai TY, Lee WJ, Huang YJ, Chen KL, Pan TM.** 2006. Detection of viable enterohemorrhagic *Escherichia coli* O157 using the combination of immunomagnetic separation with the reverse transcription multiplex TaqMan PCR system in food and stool samples. J Food Prot **69:**2320-2328.

34. **Beutin L, Jahn S, Fach P.** 2009. Evaluation of the 'GeneDisc' real-time PCR system for detection of enterohaemorrhagic *Escherichia coli* (EHEC) O26, O103, O111, O145 and O157 strains according to their virulence markers and their O- and H-antigen-associated genes. J Appl Microbiol **106:**1122-1132.

35. **Fratamico PM, Bagi LK, Cray WC, Narang N, Yan X, Medina M, Liu Y.** 2011. Detection by multiplex real-time polymerase chain reaction assays and isolation of shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 in ground beef. Foodborne Pathog Dis **8:**601-607.

36. **Lin A, Sultan O, Lau HK, Wong E, Hartman G, Lauzon CR.** 2011. O serogroup specific real time PCR assays for the detection and identification of nine clinically relevant non-O157 STECs. Food Microbiol **28:**478-483.

37. **Chen S, Zhao S, McDermott PF, Schroeder CM, White DG, Meng J.** 2005. A DNA microarray for identification of virulence and antimicrobial resistance genes in *Salmonella* serovars and *Escherichia coli*. Mol Cell Probes **19:**195-201.

38. **Liu Y, Fratamico P.** 2006. *Escherichia coli* O antigen typing using DNA microarrays. Mol Cell Probes **20:**239-244.

39. **Ballmer K, Korczak BM, Kuhnert P, Slickers P, Ehricht R, Hächler H.** 2007. Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. J Clin Microbiol **45:**370-379.

40. **Bugarel M, Beutin L, Martin A, Gill A, Fach P.** 2010. Micro-array for the identification of Shiga toxin-producing *Escherichia coli* (STEC) seropathotypes associated with hemorrhagic colitis and hemolytic uremic syndrome in humans. Int J Food Microbiol **142:**318-329.

41. **Nolan JP, Sklar LA.** 2002. Suspension array technology: evolution of the flat-array paradigm. Trends Biotechnol **20:**9-12.

42. **Moalic V.** 2004. Technologie Luminex (TM): principe, applications, et perspectives. *Immuno-analyse & Biologie Spécialisée* **19:**7.

43. **Dunbar SA.** 2006. Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. Clin Chim Acta **363:**71-82.

44. **Baums IB, Goodwin KD, Kiesling T, Wanless D, Diaz MR, Fell JW.** 2007. Luminex detection of fecal indicators in river samples, marine recreational water, and beach sand. Mar Pollut Bull **54:**521-536.

45. **Dunbar SA, Vander Zee CA, Oliver KG, Karem KL, Jacobson JW.** 2003. Quantitative, multiplexed detection of bacterial pathogens: DNA and protein applications of the Luminex LabMAP system. J Microbiol Methods **53:**245-252.

46. **Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P.** 2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol **45:**3323-3334.

47. **Griffith JF, Cao Y, McGee CD, Weisberg SB.** 2009. Evaluation of rapid methods and novel indicators for assessing microbiological beach water quality. Water Res **43:**4900-4907.

48. **Barco L, Lettini AA, Pozza MCD, Ramon E, Fasolato M, Ricci A.** 2010. Fluoroquinolone Resistance Detection in *Campylobacter coli* and *Campylobacter jejuni* by Luminex® xMAP™ Technology. Foodborne Pathog Dis 7 (9): 1039-1045.

49. **McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML, Fields PI.** 2011. Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. J Clin Microbiol **49:**565-573.

50. **Su LH, Chiu CH.** 2007. *Salmonella*: clinical importance and evolution of nomenclature. Chang Gung Med J **30:**210-219.

51. **Boyle EC, Bishop JL, Grassl GA, Finlay BB.** 2007. *Salmonella*: from pathogenesis to therapeutics. J Bacteriol **189:**1489-1495.

52. **Casadesús J.** 2011. *Salmonella*: from basic science to clinical issues. Future Microbiol **6:**133-135.

53. **White DG, Zhao S, Sudler R, Ayers S, Friedman S, Chen S, McDermott PF, McDermott S, Wagner DD, Meng J.** 2001. The isolation of antibiotic-resistant *Salmonella* from retail ground meats. N Engl J Med **345:**1147-1154.

54. **Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ.** 2013. Outbreak-associated *Salmonella enterica* serotypes and food Commodities, United States, 1998-2008. Emerg Infect Dis **19:**1239-1244.

55. **Hanning IB, Nutt JD, Ricke SC.** 2009. Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention measures. Foodborne Pathog Dis **6:**635-648.

56. [**CDC] Centers for Disease Control and Precention.** 2013. Outbreaks Involving *Salmonella*. Available at: http://www.cdc.gov/salmonella/outbreaks.html, accessed on January 2014.

57. [**CDC] Centers for Disease Control and Prevention.** 2013. National Enteric Disease Surveillance: *Salmonella* annual report, 2011. Available at: http://www.cdc.gov/ncezid/dfwed/pdfs/salmonella-annual-report-2011-508c.pdf, accessed on January 2014.

58. **McQuiston JR, Fields PI, Tauxe RV, Logsdon JM.** 2008. Do *Salmonella* carry spare tyres? Trends Microbiol **16:**142-148.

59. **Hong Y, Liu T, Lee MD, Hofacre CL, Maier M, White DG, Ayers S, Wang L, Berghaus R, Maurer JJ.** 2008. Rapid screening of *Salmonella enterica* serovars Enteritidis, Hadar, Heidelberg and Typhimurium using a serologically-correlative allelotyping PCR targeting the O and H antigen alleles. BMC Microbiol **8:**178.

60. **[CDC] Centers for Disease Control and Prevention.** 2011. PulseNet. Available at: http://www.cdc.gov/pulsenet/, accessed on January 2014.

61. **Liu F, Kariyawasam S, Jayarao BM, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG.** 2011. Subtyping *Salmonella enterica* serovar enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). Appl Environ Microbiol **77:**4520-4526.

62. **Wattiau P, Boland C, Bertrand S.** 2011. Methodologies for *Salmonella enterica* ssp *enterica* subtyping: gold standards and alternatives. Appl Environ Microbiol. 77 (22): 7877-7885.

63. **Aggarwal P, Sarkar R, Singh M, Grover BD, Anand BR, Chowdhuri AN.** 1983. *Salmonella* Bareilly infection in a paediatric hospital of New Delhi. Indian J Med Res **78:**22-25.

64. **Gupta P, Talwar V, Revathi G, Kumar A.** 1997. Nosocomial *Salmonella* Bareilly septicemia: a nursery outbreak. Indian Pediatr **34:**144-146.

65. **Mccullough NB, Eisele CW.** 1951. Experimental human salmonellosis. III. Pathogenicity of strains of *Salmonella* Newport, *Salmonella* Derby, and *Salmonella* Bareilly obtained from spray-dried whole egg. J Infect Dis **89:**209-213.

66. **Singh G, Sharma NC, Jayasheela M, Saxena SN.** 1988. A study of the epidemiology of *Salmonella* Bareilly in India using a new phage-typing system. Epidemiol Infect **100:**221-225.

67. **Prapasarakul N, Pulsrikarn C, Vasaruchapong T, Lekcharoen P, Chanchaithong P, Lugsomya K, Keschumras N, Thanomsuksinchai N, Tanchiangsai K, Tummaruk P.** 2012. *Salmonella* serovar distribution in

cobras (*Naja kaouthia*), snake-food species, and farm workers at Queen Saovabha Snake Park, Thailand. J Vet Diagn Invest **24:**288-294.

68. **Melendez SN, Hanning I, Han J, Nayak R, Clement AR, Wooming A, Hererra P, Jones FT, Foley SL, Ricke SC.** 2010. *Salmonella enterica* isolates from pasture-raised poultry exhibit antimicrobial resistance and class I integrons. J Appl Microbiol **109:**1957-1966.

69. **Rhoades HE.** 1946. *Salmonella* bareilly in quail. North Am Vet **27:**419.

70. **Kudaka J, Itokazu K, Taira K, Iwai A, Kondo M, Susa T, Iwanaga M.** 2006. Characterization of *Salmonella* isolated in Okinawa, Japan. Jpn J Infect Dis **59:**15-19.

71. **Wilson IG.** 2002. *Salmonella* and *Campylobacter* contamination of raw retail chickens from different producers: a six year survey. Epidemiol Infect **129:**635-645.

72. **[CDC] Centers for Disease Prevention and Control.** 2012. Multispate outbreak of *Salmonella* Bareilly and *Salmonella* Nchanga infections associated with a raw scrapped ground tuna product (final update). Available at: http://www.cdc.gov/salmonella/bareilly-04-12/index.html?utm_source=Website+Signup&utm_campaign=f4e381310d-Monday_Morning_Light_April_16_2012_13_2012&utm_medium=email, accessed on January 2014.

73. **de Jong B, Oberg J, Svenungsson B.** 2007. Outbreak of salmonellosis in a restaurant in Stockholm, Sweden, September - October 2006. Euro Surveill **12:**E13-14.

74. **Cleary P, Browning L, Coia J, Cowden J, Fox A, Kearney J, Lane C, Mather H, Quigley C, Syed Q, Tubin-Delic D, team oc.** 2010. A foodborne outbreak of *Salmonella* Bareilly in the United Kingdom, 2010. Euro Surveill **15** (48) pii 19732.

75. **Sharma NC, Bhatia R, Singh S, John PC, Kumar S, Singh H.** 1994. Bacteriophage typing in *Salmonella* Bareilly. Epidemiol Infect **112:**45-49.

76. **Saxena MK, Singh VP, Lakhcharua BD, Taj G, Sharma B.** 2002. Strain differentiation of Indian isolates of *Salmonella* by ERIC-PCR. Res Vet Sci **73:**313-314.

77. **Zhao S, McDermott PF, Friedman S, Qaiyumi S, Abbott J, Kiessling C, Ayers S, Singh R, Hubert S, Sofos J, White DG.** 2006. Characterization of antimicrobial-resistant *Salmonella* isolated from imported foods. J Food Prot **69:**500-507.

78. **Kumar R, Surendran PK, Thampuran N.** 2009. Analysis of antimicrobial resistance and plasmid profiles in *Salmonella* serovars associated with tropical seafood of India. Foodborne Pathog Dis **6:**621-625.

79. **Kunin V, Sorek R, Hugenholtz P.** 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol **8:**R61.

80. **Sorek R, Kunin V, Hugenholtz P.** 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol **6:**181-186.

81. **Karginov FV, Hannon GJ.** 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell **37:**7-19.

82. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

83. **Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S.** 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol **190:**1390-1400.

84. **van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ.** 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci **34:**401-407.

85. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167-170.

86. **Perez-Rodriguez R, Haitjema C, Huang Q, Nam KH, Bernardis S, Ke A, Delisa MP.** 2011. Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. Mol Microbiol **79:**584-599.

87. **Touchon M, Charpentier S, Clermont O, Rocha EP, Denamur E, Branger C.** 2011. CRISPR Distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. J Bacteriol **193:**2460-2467.

88. **Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV.** 2002. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res **30:**482-496.

89. **Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, Gagarinova A, Pogoutse O, Brown G, Binkowski A, Phanse S,**

Joachimiak A, Koonin EV, Savchenko A, Emili A, Greenblatt J, Edwards AM, Yakunin AF. 2011. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol Microbiol **79:**484-502.

90. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A.** 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol **169:**5429-5433.

91. **Al-Attar S, Westra ER, van der Oost J, Brouns SJJ.** 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biological Chemistry **392:**277-289.

92. **Haft DH, Selengut J, Mongodin EF, Nelson KE.** 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol **1:**e60.

93. **Mojica FJ, Díez-Villaseñor C, Soria E, Juez G.** 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol **36:**244-246.

94. **Díez-Villaseñor C, Almendros C, García-Martínez J, Mojica FJ.** 2010. Diversity of CRISPR loci in *Escherichia coli*. Microbiology **156:**1351-1361.

95. **Deveau H, Garneau JE, Moineau S.** 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol **64:**475-493.

96. **Shah SA, Garrett RA.** 2011. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol **162:**27-38.

97. **Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol **9:**467-477.

98. **Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E.** 2011. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature **477:**486-489.

99. **Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ.** 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol **18:**529-536.

100. **Marraffini LA, Sontheimer EJ.** 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322:**1843-1845.

101. **Edgar R, Qimron U.** 2010. The *Escherichia coli* CRISPR system protects from λ lysogenization, lysogens, and prophage induction. J Bacteriol **192:**6291-6294.

102. **Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A.** 2010. Transcription profile of Thermus thermophilus CRISPR systems after phage infection. J Mol Biol **395:**270-281.

103. **Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321:**960-964.

104. **Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K.** 2010. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. Molecular Microbiology **77:**1367-1379.

105. **Jore MM, Brouns SJ, van der Oost J.** 2011. RNA in defense: crisprs protect prokaryotes against mobile genetic elements. Cold Spring Harb Perspect Biol. 1; 4 (6).

106. **Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K.** 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A **108:**10098-10103.

107. **Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV, Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R.** 2008. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. PLoS One **3:**e2652.

108. **Tyson GW, Banfield JF.** 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol **10:**200-207.

109. **Vale PF, Little TJ.** 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. Proc Biol Sci **277:**2097-2103.

110. **Pourcel C, Salvignol G, Vergnaud G.** 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology **151:**653-663.

111. **Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J.** 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol **35:**907-914.

112. **Gori A, Bandera A, Marchetti G, Degli Esposti A, Catozzi L, Nardi GP, Gazzola L, Ferrario G, van Embden JD, van Soolingen D, Moroni M, Franzetti F.** 2005. Spoligotyping and *Mycobacterium tuberculosis*. Emerg Infect Dis **11:**1242-1248.

113. **Pride DT, Sun CL, Salzman J, Rao N, Loomer P, Armitage GC, Banfield JF, Relman DA.** 2011. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. Genome Res **21:**126-136.

114. **Price EP, Smith H, Huygens F, Giffard PM.** 2007. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. Appl Environ Microbiol **73:**3431-3436.

115. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.

116. **Delannoy S, Beutin L, Burgos Y, Fach P.** 2012. Specific detection of enteroaggregative hemorrhagic *Escherichia coli* O104:H4 strains by use of the crispr locus as a target for a diagnostic real-time PCR. J Clin Microbiol **50:**3485-3492.

117. **Delannoy S, Beutin L, Fach P.** 2012. Use of CRISPR sequence polymorphisms for specific detection of enterohemorrhagic *E. coli* (EHEC)

strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28 and O157:H7 by real time PCR. J Clin Microbiol.

118. **Yin S, Jensen MA, Bai J, Debroy C, Barrangou R, Dudley EG.** 2013. The evolutionary divergence of shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. Appl Environ Microbiol **79:**5710-5720.

119. **Fricke WF, Mammel MK, McDermott PF, Tartera C, White DG, Leclerc JE, Ravel J, Cebula TA.** 2011. Comparative Genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. J Bacteriol **193:**3556-3568.

120. **Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc Natl Acad Sci U S A **106:**17939-17944.

121. **Palmer KL, Gilmore MS.** 2010. Multidrug-resistant enterococci lack CRISPR-cas. MBio **12; 1** (4).

122. **Garneau JE, Dupuis M, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S.** 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468:**67-71.

123. **Lindenstrauß AG, Pavlovic M, Bringmann A, Behr J, Ehrmann MA, Vogel RF.** 2011. Comparison of genotypic and phenotypic cluster analyses of

virulence determinants and possible role of CRISPR elements towards their incidence in *Enterococcus faecalis* and *Enterococcus faecium*. Syst Appl Microbiol. 34 (8): 553-560.

124. **Touchon M, Charpentier S, Pognard D, Picard B, Arlet G, Rocha EP, Denamur E, Branger C.** 2012. Antibiotic resistance plasmids spread among natural isolates of *Escherichia coli* in spite of CRISPR elements. Microbiology **158:**2997-3004.

125. **Dang TN, Zhang L, Zöllner S, Srinivasan U, Abbas K, Marrs CF, Foxman B.** 2013. Uropathogenic *Escherichia coli* are less likely than paired fecal *E. coli* to have CRISPR loci. Infect Genet Evol **19C:**212-218.

126. **Swarts DC, Mosterd C, van Passel MW, Brouns SJ.** 2012. CRISPR interference directs strand specific spacer acquisition. PLoS One **7:**e35888.

127. **Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS.** 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature **497:**254-257.

128. **Gunderson FF, Cianciotto NP.** 2013. The CRISPR-associated gene *cas2* of *Legionella pneumophila* is required for intracellular infection of amoebae. MBio **4:**e00074-00013.

129. **Louwen R, Horst-Kreft D, de Boer AG, van der Graaf L, de Knegt G, Hamersma M, Heikema AP, Timms AR, Jacobs BC, Wagenaar JA, Endtz HP, van der Oost J, Wells JM, Nieuwenhuis EE, van Vliet AH, Willemsen PT, van Baarlen P, van Belkum A.** 2013. A novel link between

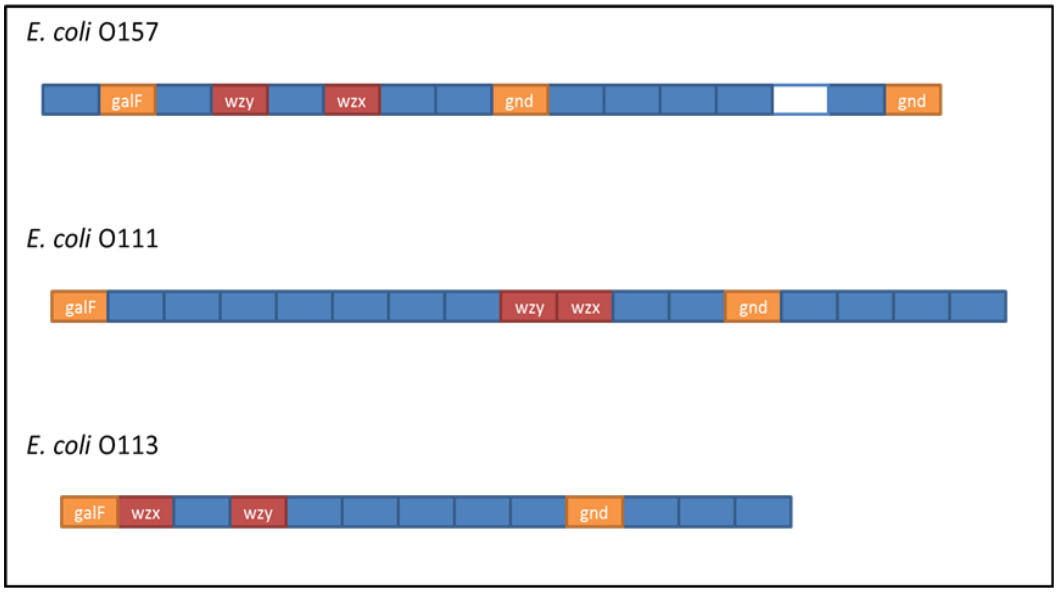
Campylobacter jejuni bacteriophage defence, virulence and Guillain-Barré syndrome. Eur J Clin Microbiol Infect Dis **32:**207-226.

130. **Toro M, Najjar MB, Ju W, Brown E, Zhao S, Meng J.** 2013. Molecular serogrouping of Shiga Toxin-Producing *Escherichia coli* using suspension Array. Foodborne Pathog Dis. **10 (5)**:478-480.

131. **Toro M, Cao G, Ju W, Allard M, Barrangou R, Zhao S, Meng J.** 2014. Association of CRISPR elements with serotypes and virulence potential of Shiga toxin-producing *Escherichia coli*. Appl. Environ Microbiol. **80 (4)**:1411-1420.

1.9 Figures and Tables

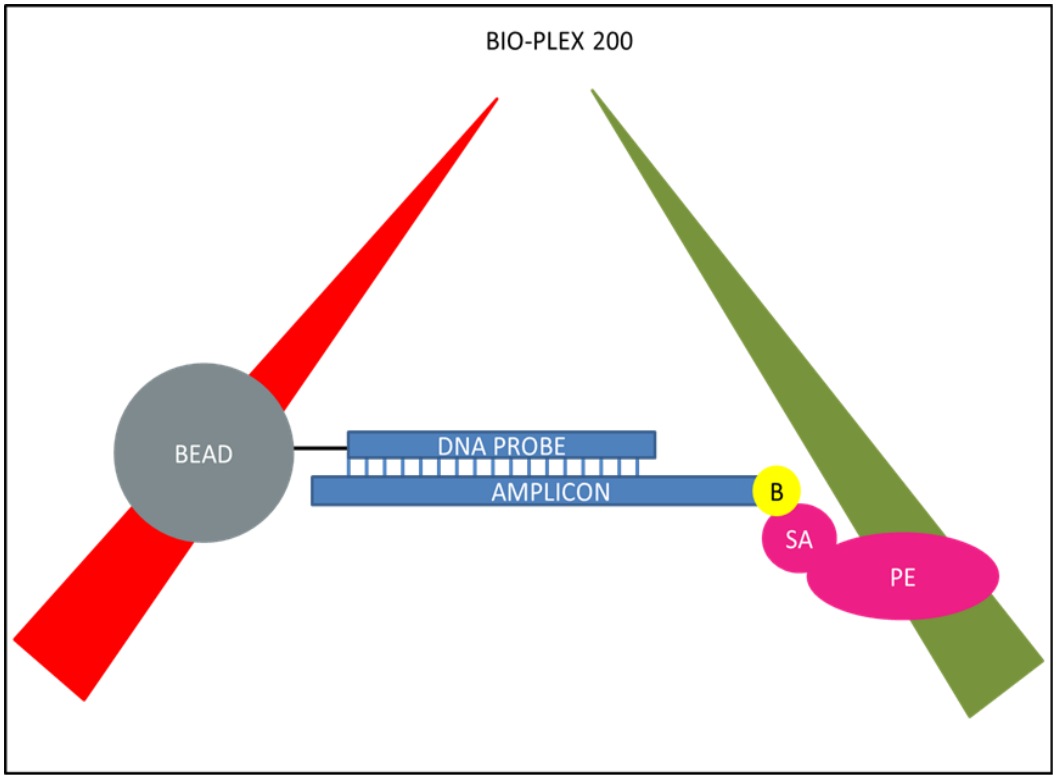
Figure I.1. O antigen gene cluster organization in *E. coli* serotypes.



O antigen gene clusters have different genes depending on the serotypes, but many of the genes are present in all the serotypes, but in different order.

Each box represents a gene in the O antigen gene cluster. Orange boxes represent housekeeping genes (*gal*F and *gnd*). Red boxes represent genes *wzx* and *wzy* (flippase and polymerase genes), target genes in the current study. An empty box represents an incomplete gene.

Figure I.2. – Schematic representation of the detection process using the Bio-Plex 200 ™ detector.



Red triangle represents red laser detecting the bead color/region. Green triangle represents green laser detecting the presence of biotin-SA-PE complex quantifying the reaction.

DNA probe is linked to the bead through a C12 attached to its 5'end.

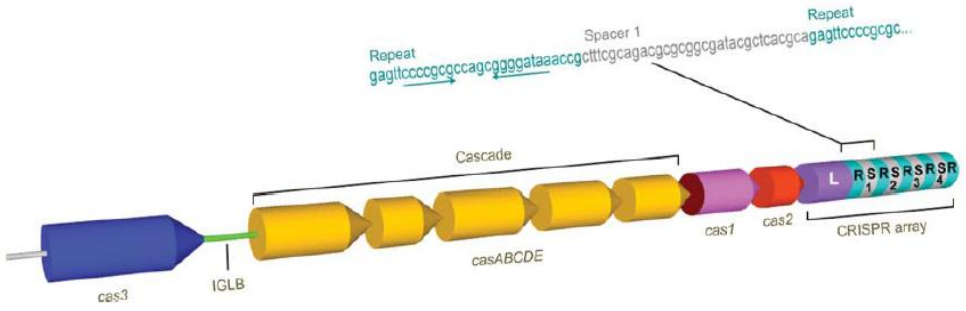
Reverse primer has a 5'modification (biotin) which allows the inclusion of the molecule in the amplicons.

Bead region and reaction verification are run at the same time.

SA: Streptoavidine

PE: Phicoeritrine

Figure I.3. Schematic diagram of the CRISPR I locus of *E. coli* K12.[1]



"The *cas* genes (*cas3*, *casABCDE*, *cas1* and *cas2*) are shown along with the coupled CRISPR array. Within the CRISPR array three distinctive elements are found: leader sequence (L), repeats (R) and spacers (S). One repeat and one spacer constitute one CRISPR unit (shown also as DNA sequence). The region colored light green wintergenic region between *casA* (*ygcL*) and *cas3* (*ygcB*) (IGLB)x is believed to contain promoters required for the expression of the *cas* genes."[1] (91).

[1] Figure and legend from Al Attar *et al.*, 2011 (91)

Chapter 2: Molecular serotyping of Shiga Toxin-producing *Escherichia coli* using suspension array (1)

2.1. Abstract

Shiga toxin-producing *E. coli* (STEC) cause serious foodborne diseases. Although *E. coli* O157:H7 has been the dominant STEC serotype, other serotypes have been involved in outbreaks and sporadic illnesses worldwide. In this work, a suspension array assay was developed to simultaneously serogroup the seven most prevalent STEC (O26, O45, O103, O111, O121, O145, and O157) in the US. A suspension array panel was designed targeting genes *wzx* or *wzy*; probes detecting Shiga toxin genes, stx_1 and stx_2, were included as well. For each serogroup in the assay, four to eleven isolates were used for testing and optimization. Fluorescence values of 59 STEC were 30 to >270 times greater than the signals of negative controls, demonstrating the method's effectiveness for the molecular serotyping of STEC.

2.2. Introduction

Shiga toxin-producing *Escherichia coli* (STEC) are a leading cause of foodborne diseases in the US (2-5). Although O157:H7 has been the major STEC serotype associated with outbreaks in the US, other important serogroups emerged in the past years (6, 7); serogroups O26, O45, O103, O111, O121, and O145 caused 70% of the cases of non-O157 STEC in the US between 2000 and 2006 (2, 8).

Additionally, certain non-O157 STEC have been identified as more common than O157 in other parts of the world (9). The US Department of Agriculture recently extended a zero-tolerance policy for *E. coli O157:H7* in raw beef products by declaring the six additional STEC serogroups as adulterants (10). Due to their potential virulence, timely detection and identification of major non-O157 STEC in clinical, food, and environmental samples is important to ensure public health.

Molecular methods have been developed for serotyping *E. coli* as an alternative for traditional serotyping (11-15). These methodologies target serogroup genes (*wzx* and/or *wzy*) that encode serogroup specific proteins forming the O antigen of Gram-negative bacteria (16).When specific sequences are detected, a serotype is attributed to the tested isolate. However, most assays are not suitable for the identification of multiple targets in a single reaction (17-19), or for the application to a large number of samples (13, 15, 20, 21). In this study, a bead-based suspension array (Bio-Plex ™) was explored as an alternative for discriminating among the seven STEC serogroups in a single reaction. In this assay, nucleic acid probes are linked to beads and hybridized to the target of interest (22), followed by a detection using a flow cytometry-like device that identifies the beads and quantifies the interaction with the target.

2.3. Objective

The present study aimed to develop a suspension array to identify seven STEC serogroups – O26, O45, O103, O111, O121, O145, and O157 – which most frequently cause human infections in the US.

2.4. Materials and Methods

2.4.1. DNA sequences

O antigen genes *wzx* (flipase) and *wzy* (polymerase) sequences were obtained for serotypes O26, O45, O103, O111, O121, O145, and O157 from GenBank data base. For a comprehensive detection of STEC, sequences for shiga toxin genes –*stx1* and *stx2*—were added to the design. Consensus sequences for each one of the targets were crafted from several sequences using BioEdit v.5.0.9.

2.4.2. Designing selective primers and probes.

Specific DNA primer pairs and probes were simultaneously designed using the software visual OMP (DNA software Inc., Ann Arbor, MI) for all the targets. Parameters included; oligonucleotide length of 25±3 bp, amplicon size of 350±150bp, melting temperature (Tm) of 60º C, no mismatches between oligonucleotide and target sequence, and no mishybridazation or crosshybridization to the target or between oligonucleotides. Probes were designed with an amino C12 modification at the 5'end for coupling to the micro beads (Bio-Rad, Hercules, CA) and biotin in the reverse primer to allow reaction detection by the lasers.

Multiplex-PCR reactions and hybridizations were simulated to search for any non-specific bindings between primers, probes, and amplified DNA using the visual OMP. Then, primers and probes were analyzed through BLAST (Basic Local Alignment Search Tool, http://blast.ncbi.nlm.nih.gov/) to guarantee specificity to its

target; any probe hybridizing any unwanted sequences was modified or replaced, ensuring it only matched the specific serotype.

A final hybridization simulation was run with the TMAC buffer option on (Tetra Methyl Ammonium Chloride, Sigma Aldrich, St. Luis, MO); TMAC buffer eliminates the difference in melting temperature due to different CG ratio, therefore only the probe's length determines its melting temperature. After simulations cleared any possible interaction, primers and probes selected were ordered from Biosearch technologies (Novato, CA).

2.4.3. Bacterial strains

A total of 103 bacterial strains were used to evaluate the panel specificity, including 59 STEC (4 to 11 isolates for each of the 7 STEC serogroups), 23 of other *E. coli* serogroups, and 21 of non-*E. coli* species including four *Shigella* spp., four *Salmonella* serotypes, *Proteus vulgaris, Pseudomonas aeruginosa, Enterobacter aerogenes, Citrobacter freundii, Hafnia alvei, Enterobacter cloacae, Acinetobacter baumannii,* three *Listeria* spp., *Streptococcus faecalis,* and *Bacillus subtilis* (Tables II.1 and II.2).

2.4.4. DNA isolation and amplification

Instagene matrix (Bio-Rad, Hercules, CA) was used for DNA extraction following manufacturer instructions (23). Briefly, colonies from an overnight pure culture were suspended in 1 mL of de-ionized water and then pelletized by

centrifugation (10,000 rpm for 2 minutes); supernatant was discarded, and 200uL of Instagene ™ matrix were added. The mixture was vortexed and incubated for 30 min at 56°C followed by an additional incubation step of 8 min at 100°C. Then, the sample was centrifuged (10.000 rpm for 2 minutes) and supernatant containing DNA was transferred to a new tube. Finally, DNA was standardized at ~100 ng/uL, aliquoted, and stored at -20°C until used.

HotStarTaq Plus Master mix (12.5 uL) (Qiagen, Valencia, CA) was used for PCR amplification; 40 nM of forward primers and 160 nM biotinalated reverse primers were added to the reaction. Previously extracted DNA was added in a volume of 1 uL of DNA, and nuclease free water was used to reach a final reaction volume of 25 uL. The PCR parameters were as follow: initial denaturation at 95 °C for 5 min; followed by 30 cycles of 94 °C for 30 s, 56 °C for 90 s, 72 °C for 90 s; then a final elongation at 72 °C for 10 min.

To verify amplification of the target, amplicons generated by PCR were detected by 2% gel electrophoresis. This step was run only during the implementation stage.

2.4.5. Probe to bead coupling

A previously described method was applied for probe to bead coupling (24). Briefly, 5 x 10^6 microbeads were resuspended in distilled water and transferred to a copolymer microfuge tube (USA scientific, Ocala, FL). Beads were pelleted by centrifugation and then re-suspended in 0.1 M MES buffer. Capture probes were suspended to 0.1 mM in distilled water, and 0.1 nmole of probes was added to the

beads. 2.5 µl of 10 mg/ml EDC (1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride) (Pierce biotechnology, Thermo scientific, Rockfort, IL) were added to each reaction to facilitate the coupling, and then the tube was incubated in the dark for 30 min. This step was repeated before adding one ml of 0.02 % Tween-20 to beads, followed by centrifugation and washing with 1 ml 0.1 % SDS. Beads were resuspended in 100 µl TE buffer, enumerated using BioRad bead counter, standardized for concentration, and stored in the dark at 4 °C.

2.4.6. Bead hybridization and detection protocol

For probe hybridization, coupled beads were diluted 1:100 in 1.5x TMAC buffer (1.5 M tetramethylammonium chloride, 75mM Tris, 6 mM EDTA, and 0.15% sarkosyl at pH 8.0) to create a bead pool, and 33 µl of the bead mixture were added to 5 µl of PCR product and 12 µl of TE buffer followed by 5 minute incubation at 94°C followed by 15 minute at 55°C. Streptavidin-Phycoerythrin (SA-PE, Invitrogen, Carsbad, CA) was diluted to 4 µg/ml in 1x TMAC and 75 µl were added to every well; mixture was incubated at hybridization temperature (55 °C) for 10 min.

Fluorescence detection was performed immediately using Bio-Plex 200 ™ (Bio-Rad) at 55° C. A red laser excites the dyes on the surface of the beads and allows the identification of the unique bead type; a green laser excites the fluorophore (SA-PE) linked to the biotin, allowing to quantify the binding of a specific probe.

Records for Median Fluorescence Intensity (MFI) from 100 beads per region per well, Positive MFI /Negative control MFI ratio (P/N) (25), and bead count were obtained from the Bio-plex ™ manager software (v 6.0.7). The bio-Plex 200 ™

Manager software provides an automatic calculation of the P/N ratios as follows: MFI from a probe in a sample is divided by the MFI of the same probe in the negative control. This methodology was previously described by other authors (25, 26). Data were analyzed using ANOVA with SAS program (v9.2).

2.5. Results

2.5.1. Primers and probes design

After multiple rounds of simulations and laboratory testing, nine probes and 18 primers were selected for the final suspension array panel (Table II.3). O antigen primers and probes were designed *de novo*; four of them targeted the *wzy* gene while three aimed for *wzx*. Previously described primers were used for *stx1* and *stx2* genes (27, 28), and new probes were designed based on predicted amplicons. Primers and probes ranged in size from 21 to 27 bp, and amplicon size ranged from 180 to 438 bp (Table II.3).

2.5.2. Multiplex PCR

From each STEC serogroup, four to eleven isolates were selected and amplified successfully with the designed primers. STEC stains showed one, two, or three bands depending on the presence of one or two shiga toxin genes. Amplicons generated were specific for each primer pair, and the DNA of each target was successfully detected by 2% gel electrophoresis (Fig II.1).

2.5.3. Suspension Array detection/Molecular serogrouping

The MFI produced by 100 microspheres of each one of the probes was homogeneous among positive samples to a same target, but there were large differences between targets. For instance, MFI average value for positive samples to probe O26 was 6479 (SD= 471), for probe O45 was12078 (SD=1010), and for probe O157 was 18722 (SD=890). Average values for negative samples were 140 (Fig. II. 2). MFI values of positive samples were significantly greater ($p<0.05$) than those of negative samples in each serogroup (Fig. II.2).

Similarly , average P/N ratios for positive samples/targets ranged depending on the target: 32.8±2.49 for O26 (n=11), 127.6±9.91 for O45 (n=5), 53.0±2.83 for O103 (n=5), 144.6±16.91 for O111 (n=5), 207.0±6.53 for O121 (n=4), 153.7±3.89 for O145 (n=6), 270.9±11.03 for O157 (n=11), 121.5±19.49 for *stx*$_1$ (n=48), and 30.7±1.83 for *stx*$_2$ (n=18). The average P/N ratio for non-specific interactions was 1.08. Since non-specific interactions never gave P/N values over 13, a cutoff ratio of 15 was set for serogroup/shigatoxin identification (Table II.4).

None of the 23 non-STEC *E. coli* nor 21 non-*E. coli* strains gave a positive reaction (P/N ratio ≥15) for any of the probes tested, except for *Shigella dysenteriae*, which was tested positive for *stx*$_1$ (Table II.5).

2.6. Discussion

In the present study, the method developed successfully identified the 7 most common STEC serogroups causing human disease in the US. Nine sets of primers and probes were combined in single well reactions and correctly identified all the

strains included in the panel. This is the first time a suspension array assay identifies not only strain serogroups but also shiga toxin genes, and it represents a reliable alternative for molecular serogrouping of *E. coli*.

DNA based methodologies have been previously used to serogroup *E. coli*. Multiplex PCR assays detect a limited number of targets due to low band resolution of gel electrophoresis (11, 13, 18), and real time PCR are also restricted due to the number of channels available for florescence reading. Some real time PCR assays run two to three reactions per sample to identify the seven most common STEC in the US (15, 21, 29). Conversely, this suspension array interrogates samples simultaneously for all nine targets at the same time. Microarrays also identify multiple *E. coli* serotypes at the same time (30, 31), but they are difficult to customize and require expensive equipment for reaction preparation and results reading. While suspension array also require a specialized detector to read results, any conventional thermocycler can perform the amplification step, and no other device is required for the analysis. Moreover, panel modifications are done by adding or subtracting primers and probes to the system. These advantages make suspension arrays a suitable option for STEC serogrouping.

Genes *wzx* and *wzy* were selected as targets. Both genes are located in the O antigen gene cluster which carries the information for the O antigen synthesis (16). Molecular analysis revealed these genes are conserved among strains of the same serogroup while differing between them, and they are broadly used for STEC serogrouping (18, 19, 31). In this study, these sequences demonstrated to be specific for their targets, with only minor cross reactions that did not interfered with final

results. Cross reactions in suspension arrays have been also described in other studies at similar levels (32).

Typical P/N ratios obtained in this study were higher than those reported elsewhere. Ratios in a *Salmonella* serogrouping panel ranged from 18 to 155 (26), and 8 to 23 in a *Campylobacter spp.* antimicrobial resistance assay (33). P/N values in this study ranged from 30 to 270, and consequently, the P/N ratio threshold was also set higher (P/N ≥15) than other studies to reduce false positives by discarding nonspecific interactions. Interestingly, P/N ratios varied depending on the probe/target combination (Fig.II.2), however, this did not interfere with correct serogroup identification. This phenomenon was also described by other authors, and it could not be attributed to any specific probe characteristic (26).

The panel was designed for the detection of almost every Shigatoxin gene subtype using two probes. Probe *stx1* presented some mismatches with subtype $stx1_d$ resulting in a reduced fluorescence for this interaction although higher than the threshold level. Probe *stx2* detects all subtypes but $stx2_f$ which sequence considerably differs from other stx_2 and has not been associated to human disease. Both issues can be addressed by the addition of additional primers and probes to the design, however, it was decided keeping the assay cost effective and not to add those targets for this version. Further studies will aim to design a complementary second panel that includes other top 20 O STEC serogroups.

Since the panel was design to identify O types from pure culture, food matrix interaction test were not applicable. Future plans include panel optimization for the detection of STEC from foods. Recently, a similar assay detected the top 10 STEC

serotypes, but excluded O157, *stx1* and *stx2* as targets (Lin *et al*., 2011). Since shiga toxin defines STECs, it was decided that it is fundamental to include them in a STEC serogrouping panel.

The use of high throughput technologies helps in a fast identification of targets of interest. In this design, samples are interrogated for nine targets at the same time, and it can be run in a 96 well format. Detection step takes less than one hour, so over 6 runs could be read per day. Adding new other probes the assay can be customized, improving the accuracy and sensitivity of the assay. Finally, once the panel is optimized for food samples it could help to an early identification of STEC and therefore, helping to prevent human disease.

2.7. Conclusion

In conclusion, this suspension array design provided a fast, reliable and improved alternative for the identification of important STEC serotypes, and can be useful to better understand the epidemiology of STEC infections and enhance outbreak investigations.

2.8. References

1. **Toro M, Najjar MB, Ju W, Brown E, Zhao S, Meng J.** 2013. Molecular serogrouping of Shiga Toxin-Producing *Escherichia coli* using suspension Array. Foodborne Pathog Dis. **10 (5)**:478-480.
2. **CDC.** 2010. Surveillance for foodborne disease outbreaks --- United States, 2007. MMWR Morb Mortal Wkly Rep **59:**973-979.
3. **Erickson MC, Doyle MP.** 2007. Food as a vehicle for transmission of Shiga toxin-producing *Escherichia coli*. J Food Prot **70:**2426-2449.
4. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2:**123-140.
5. **Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV.** 1999. Food-related illness and death in the United States. Emerg Infect Dis **5:**607-625.
6. **Bettelheim KA.** 2007. The non-O157 shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. Crit Rev Microbiol **33:**67-87.
7. **Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J.** 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J Clin Microbiol **42:**311-319.
8. **Uhlich GA, Sinclair JR, Warren NG, Chmielecki WA, Fratamico P.** 2008. Characterization of Shiga toxin-producing *Escherichia coli* isolates associated

with two multistate food-borne outbreaks that occurred in 2006. Appl Environ Microbiol **74:**1268-1272.

9. **Johnson KE, Thorpe CM, Sears CL.** 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin Infect Dis **43:**1587-1595.

10. **[USDA] United States Department of Agriculture FSIS.** 2012. Shiga toxin-producing *Escherichia coli* in certain raw beef products., Federal Register. Available at www.fsis.usda.gov/OPPDE/rdad/FRPubs/2010-0023FRN.pdf. Accessed May 2012.

11. **Fratamico PM, DebRoy C, Miyamoto T, Liu Y.** 2009. PCR detection of enterohemorrhagic *Escherichia coli* O145 in food by targeting genes in the *E. coli* O145 O-antigen gene cluster and the Shiga toxin 1 and Shiga toxin 2 genes. Foodborne Pathog Dis **6:**605-611.

12. **DebRoy C, Roberts E, Kundrat J, Davis MA, Briggs CE, Fratamico PM.** 2004. Detection of *Escherichia coli* serogroups O26 and O113 by PCR amplification of the *wzx* and *wzy* genes. Appl Environ Microbiol **70:**1830-1832.

13. **Debroy C, Roberts E, Valadez AM, Dudley EG, Cutter CN.** 2011. Detection of Shiga Toxin-Producing *Escherichia coli* O26, O45, O103, O111, O113, O121, O145, and O157 serogroups by multiplex polymerase chain reaction of the *wzx* gene of the O-Antigen gene cluster. Foodborne Pathog Dis **8:**651-652.

14. **Valadez AM, Debroy C, Dudley E, Cutter CN.** 2011. Multiplex PCR detection of Shiga toxin-producing *Escherichia coli* strains belonging to serogroups O157, O103, O91, O113, O145, O111, and O26 experimentally inoculated in beef carcass swabs, beef trim, and ground beef. J Food Prot **74:**228-239.

15. **Beutin L, Jahn S, Fach P.** 2009. Evaluation of the 'GeneDisc' real-time PCR system for detection of enterohaemorrhagic *Escherichia coli* (EHEC) O26, O103, O111, O145 and O157 strains according to their virulence markers and their O- and H-antigen-associated genes. J Appl Microbiol **106:**1122-1132.

16. **Samuel G, Reeves P.** 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res **338:**2503-2519.

17. **Wang L, Curd H, Qu W, Reeves PR.** 1998. Sequencing of *Escherichia coli* O111 O-antigen gene cluster and identification of O111-specific genes. J Clin Microbiol **36:**3182-3187.

18. **Fratamico PM, Briggs CE, Needle D, Chen CY, DebRoy C.** 2003. Sequence of the *Escherichia coli* O121 O-antigen gene cluster and detection of enterohemorrhagic *E. coli* O121 by PCR amplification of the *wzx* and *wzy* genes. J Clin Microbiol **41:**3379-3383.

19. **DebRoy C, Fratamico PM, Roberts E, Davis MA, Liu Y.** 2005. Development of PCR assays targeting genes in O-antigen gene clusters for detection and identification of *Escherichia coli* O45 and O55 serogroups. Appl Environ Microbiol **71:**4919-4924.

20. **Tsai TY, Lee WJ, Huang YJ, Chen KL, Pan TM.** 2006. Detection of viable enterohemorrhagic *Escherichia coli* O157 using the combination of immunomagnetic separation with the reverse transcription multiplex TaqMan PCR system in food and stool samples. J Food Prot **69:**2320-2328.

21. **Fratamico PM, Bagi LK, Cray WC, Narang N, Yan X, Medina M, Liu Y.** 2011. Detection by multiplex real-time polymerase chain reaction assays and isolation of Shiga Toxin-Producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 in ground beef. Foodborne Pathog Dis **8:**601-607.

22. **Dunbar SA.** 2006. Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. Clin Chim Acta **363:**71-82.

23. **Bio-Rad.** InstaGene Matrix: Instruction Manual. Bio-Rad, Hercules, CA.

24. **Fulton RJ, McDade RL, Smith PL, Kienker LJ, Kettman JR.** 1997. Advanced multiplexed analysis with the FlowMetrix system. Clin Chem **43:**1749-1756.

25. **Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P.** 2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol **45:**3323-3334.

26. **McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML, Fields PI.** 2011. Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. J Clin Microbiol **49:**565-573.

27. **Cebula TA, Payne WL, Feng P.** 1995. Simultaneous identification of strains of *Escherichia coli* serotype O157:H7 and their Shiga-like toxin type by mismatch amplification mutation assay-multiplex PCR. J Clin Microbiol **33:**248-250.

28. **Paton AW, Paton JC.** 1998. Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx1*, *stx2*, *eaeA*, enterohemorrhagic *E. coli hlyA*, *rfb*O111, and *rfb*O157. J Clin Microbiol **36:**598-602.

29. **Lin A, Sultan O, Lau HK, Wong E, Hartman G, Lauzon CR.** 2011. O serogroup specific real time PCR assays for the detection and identification of nine clinically relevant non-O157 STECs. Food Microbiol **28:**478-483.

30. **Liu Y, Fratamico P.** 2006. *Escherichia coli* O antigen typing using DNA microarrays. Mol Cell Probes **20:**239-244.

31. **Ballmer K, Korczak BM, Kuhnert P, Slickers P, Ehricht R, Hächler H.** 2007. Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. J Clin Microbiol **45:**370-379.

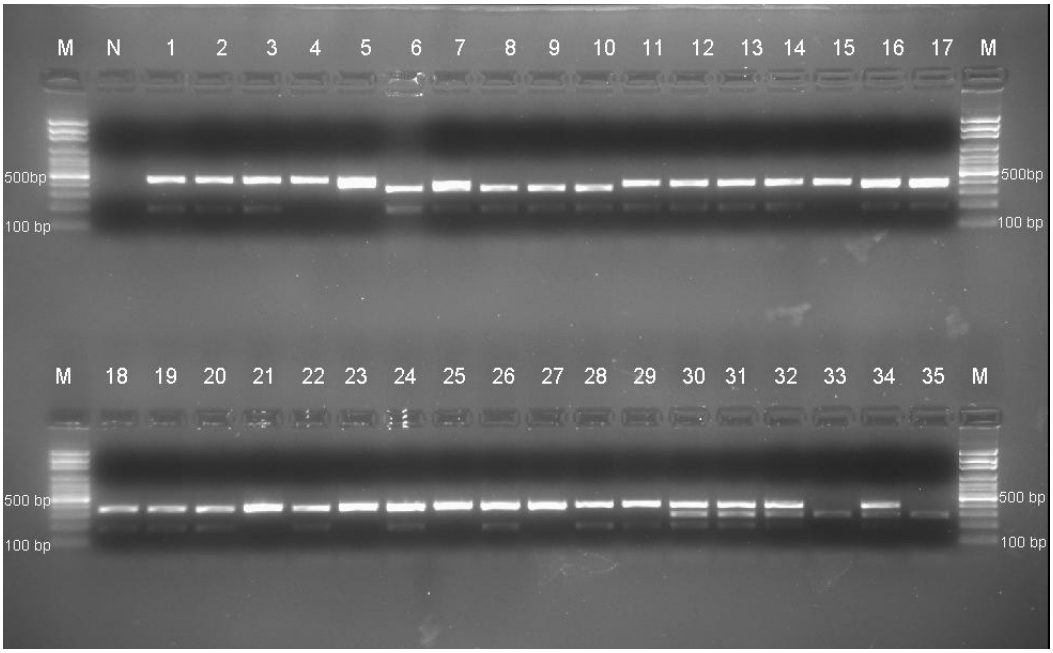
32. **Dunbar SA, Vander Zee CA, Oliver KG, Karem KL, Jacobson JW.** 2003. Quantitative, multiplexed detection of bacterial pathogens: DNA and protein applications of the Luminex LabMAP system. J Microbiol Methods **53:**245-252.

33. **Barco L, Lettini AA, Dalla Pozza MC, Ramon E, Fasolato M, Ricci A.** 2010. Fluoroquinolone resistance detection in *Campylobacter coli* and

Campylobacter jejuni by Luminex xMAP technology. Foodborne Pathog Dis **7:**1039-1045.

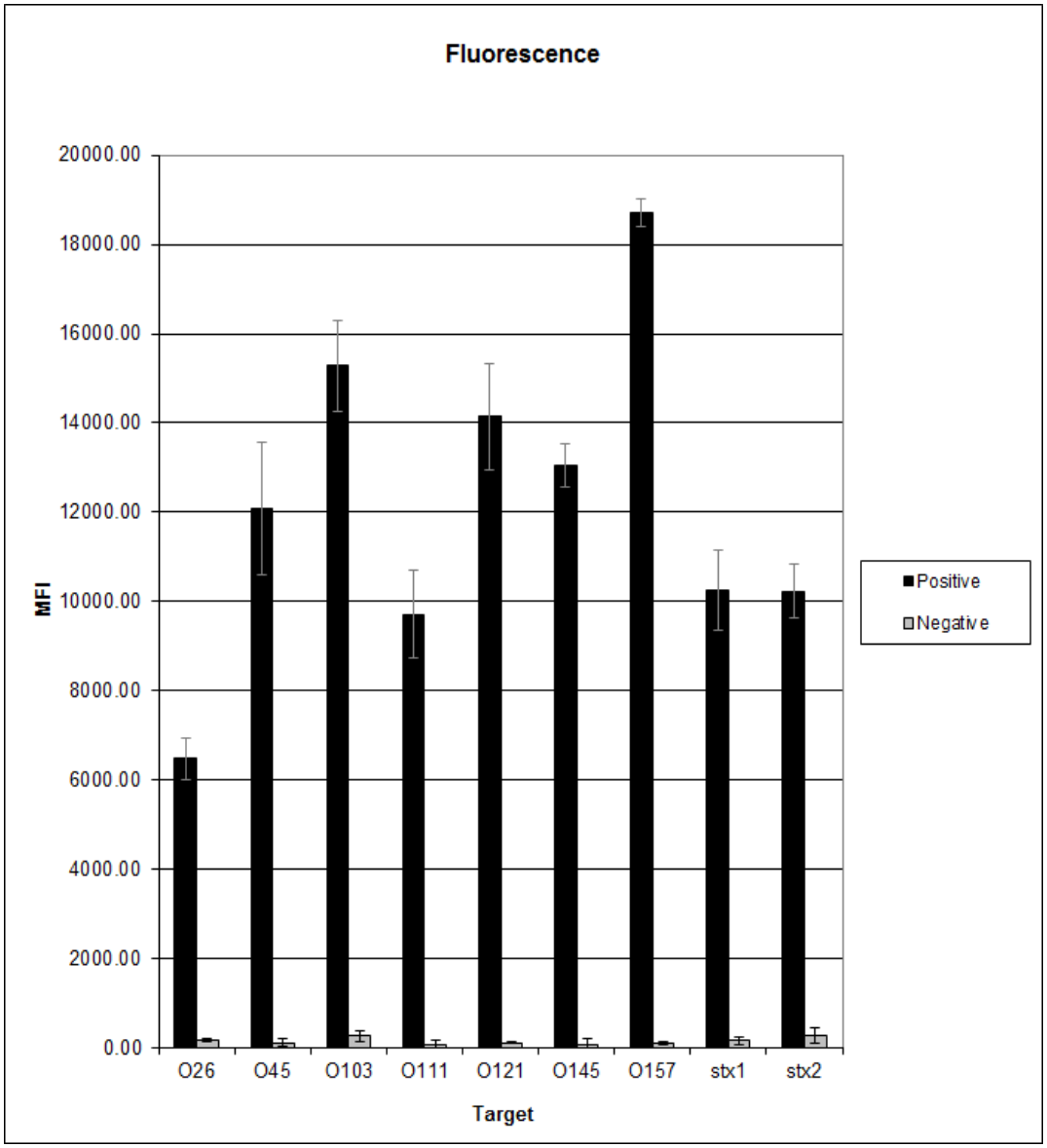
2.9. Tables and Figures

Figure II.1. Multiplex PCR results for some of the strains tested. Gel electrophoresis 2% agarose



M: 100 bp molecular weight marker. **N:** Negative control; **O26:** 1 to 5; **O45:** 6 to 10; **O103:** 11 to 15; **O111:** 16 to 20; **O121:** 21 to 24; **O145:** 25 to 29; **O157:** 30 to 35. **Stx1**: 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 29, 30, 31, 32, and 34. **Stx2**: 5, 7, 15, 16, 17, 21, 23, 24, 25, 26, 27, 30, 31, 32, and 34. Gel electrophoresis is not a step necessary for the assay, and it was run only during the design step in order to verify efficiency of the multiple PCR reaction.

Figure II.2. Positive to Negative (P/N) ratios averaged for studied isolates in one representative assay.



Bars represent average P/N ratio for the isolates representing a specific serotype or Shiga toxin. Lines represent standard deviation.

P/N ratio is calculated by dividing the fluorescence value of a probe in the sample divided by its fluorescence value in the negative control.

Table II.1. List of *E. coli* strains used for designing and optimizing suspension array assay.

Strain	Serotype	*stx1*	*stx2*	Strain	Serotype	*stx1*	*stx2*
CVM 9935	O26	+	-	**CVM 9602**	O111	+	-
CVM 9939	O26	+	-	**CVM 9610**	O111	+	-
CVM 9942	O26	+	-	**CVM 9611**	O111	+	-
CVM 9945	O26	+	-	**-1387**	O111:H8	+	-
CVM 9953	O26	+	-	**UMD 167**	O111:H11	+	-
CVM 9965	O26	+	-	**UMD 168**	O111:NM	+	+
TW 1200	O26:H2	-	-	**SJ13**	O111:NM	+	+
UMD 141	O26:H11	+	-	**UMD 164**	O111:NM	+	-
UMD 142	O26:H11	+	-	**TW 1676**	O121	+	-
TW 00971	O26:H11	+	-	**-8023**	O121:H19	-	+
SJ3	O26:H11	-	+	**TW 8868**	O121:H19	-	+
MI3	O45	+	-	**SJ19**	O121:H19	+	+
MI4	O45	+	-	**CVM 9785**	O145	+	-
UMD 144	O45:H2	+	-	**CVM 9790**	O145	-	+
SJ9	O45:H2	+	+	**CVM 9818**	O145	-	-
05-0645	O45:H2	+	-	**TW 7865**	O145:H28	-	+
CVM 9260	O103	+	-	**TW 08087**	O145: NM	+	-
CVM 9380	O103	+	-	**SJ23**	O145:NM	+	+
CVM 9385	O103	+	-	**CVM 3769**	O157	+	+
CVM 9397	O103	+	-	**CVM 3755**	O157	+	+
CVM 9398	O103	+	-	**BAA 460**	O157	+	+
CVM 9353	O103	+	-	**-43893**	O157	+	-
CVM 9354	O103	+	-	**TW 14359**	O157	-	+
TW 7920	O103:H2	+	-	**FDA 413**	O157	-	+
-4162	O103:H6	+	-	**CVM3764**	O157	+	+
SJ12	O103:H11	+	-	**EDL 933**	O157:H7	+	+
TW 7990	O103:NM	+	-	**-85170**	O157:H7	-	-
CVM 9591	O111	+	-	**UMD46**	O157:H7	+	+
CVM 9592	O111	+	-	**84-1097**	O157:H25	-	-
CVM 9596	O111	+	-				

*Strains of target serogroups: O26, O45, O103, O111, O121, O145, and O157.

Table II.2. List of strains used for exclusivity test for the suspension array.

E. coli	**Non *E. coli* strains**
E. coli K12	*Shigella flexnieri* ATCC12022
E. coli O1	*Shigella dysenteriae* CVM
E. coli O2	*Shigella bordii* CVM
E. coli O5	*Shigella sonnei* ATCC25931
E. coli O6	*Salmonella* Typhimurium LT-2
E. coli O8	*Salmonella* Typhimurium ATCC14028
E. coli O15	*Salmonella* Enterica ATCC13076
E. coli O18A	*Salmonella* Enteritidis
E. coli O22	*Proteus vulgaris* ATCC13315
E. coli O28	*Psudomonas aeruginosa* ATCC27853
E. coli O55	*Enterobacter aerogenes ATCC13048*
E. coli O88	*Citrobacter freundii* ATCC8090
E. coli O91	*Hafnia alvei* UMD375
E. coli O104	*Enterobacter cloacae* UMD190
E. coli O113	*Acinetobacter baumanii* ATCC19606
E. coli O118	*Listeria monocytogenes* ATCC13932
E. coli O128	*Listeria innocua* ATCC33090
E. coli O128ac	*Listeria grayii* ATCC19120
E. coli O142	*Staphylococcus aureus* ATCC29213
E. coli O146	*Streptococcus faecalis* ATCC19433
E. coli O153	*Bacillus subtilus* BIOREM 1-1
E. coli O174	
E. coli ATCC 25922	

Table II.3. Primer and probe sequences, gene target and amplicon size for different *E. coli* serogroups and for Shiga toxins genes (*stx1* and *stx2*) in the Suspension Array panel.

Serogroup	Oligo	Sequence	Gene	Product (bp)
O26	FWD	GTGTGTCTGGTTCGTATTTTTTATCTG	*wzx*	438
	REV	CCTTATATCCCAATATAGTACCCACCC		
	PROBE	AATAAAGCTAAAATTCAATGGGCGGAA		
O45	FWD	GGTCGATAACTGGTATGCAATATG	*wzx*	341
	REV	CTAGGCAGAAAGCTATCAACCAC		
	PROBE	CAACAGTTCTTGCAGACATGATC		
O103	FWD	TTATACAAATGGCGTGGATTGGAG	*wzy*	385
	REV	TGCAGACACATGAAAAGTTGATGC		
	PROBE	GAGCAGTTACGTCAATTACTGGCATG		
O111	FWD	CTTCGATGTTGCGAGGAATAATTC	*wzx*	362
	REV	GTGAGACGCCACCAGTTAATTGAAG		
	PROBE	CAAGAGTGCTCTGGGCTTCTATAGT		
O121	FWD	AGTGGGGAAGGGCGTTACTTATC	*wzy*	366
	REV	CAATGAGTGCAGGCAAAATGGAG		
	PROBE	CCGATATTCTAGTAGCCGTTATTTCAG		
O145	FWD	CCTGTCTGTTGCTTCAGCCCTTT	*wzy*	392
	REV	CTGTGCGCGAACCACTGCTAAT		
	PROBE	TTCATATTGGGCTGCCACTGATG		
O157	FWD	TCGTTCTGAATTGGTGTTGCTCA	*wzy*	278
	REV	CTGGTGTCGGAAAGAAATCGTTC		
	PROBE	GACAAGACGGAGAACAAAATGACTCAT		
stx1	FWD	ATAAATCGCCATTCGTTGACTAC[a]	*stx1*	180
	REV	AGAACGCCCACTGAGATCATC[a]		
	PROBE	GCATAGTGGAACCTCACTGAC		
stx2	FWD	GACCATCTTCGTCTGATTATTGAGC	*stx2*	385
	REV	GCGTCATCGTATACACAGGAGC[b]		
	PROBE	GGCGTTAATGGAGTTCAGTGG		

[a] *stx1* FWD and REV described by Paton and Paton (1998).

[b] *stx2* REV described by Cebula *et al.* (1995).

[c] Primers and probes were purchased from Biosearch Technologies.

FWD, forward primer; REV, reverse primer.

Table II.4. Serogroup, Shiga toxin profile, and positive/negative (P/N) ratios for *E. coli* strains used for developing STEC serogrouping panel.

				Strain Profile							***P/N Ratios***	
Strain	**Serotype/ Serogroup**	***stx1***	***stx2***	**O26**	**O45**	**O103**	**O111**	**O121**	**O145**	**O157**	***stx*$_1$**	***stx*$_2$**
CVM	O26	+	-	**32.62**	1.02	0.87	1.24	0.9	0.8	1.72	**132.22**	0.86
CVM	O26	+	-	**35.76**	0.95	0.84	0.97	0.93	0.71	0.54	**144.11**	0.61
CVM	O26	+	-	**31.75**	0.87	0.93	0.83	1.1	0.68	0.68	**137.04**	0.65
CVM	O26	+	-	**29.54**	1.16	1.06	0.94	1.05	0.75	1.14	**134.57**	0.8
CVM	O26	+	-	**35.6**	0.56	0.89	1.18	0.9	0.72	0.88	**151.86**	0.59
CVM	O26	+	-	**34.41**	0.88	0.85	0.81	0.73	0.71	0.9	**136.96**	0.61
TW	O26:H2	-	-	**36.8**	13	0.96	1.28	1.4	0.62	1.29	3.94	0.94
UMD	O26:H11	+	-	**29.95**	0.82	0.82	1.13	1.12	0.88	0.93	**117.57**	0.61
UMD	O26:H11	+	-	**30.31**	1.19	0.84	0.98	1	0.9	1.33	**121.92**	0.7
TW	O26:H11	+	-	**32.02**	0.81	0.85	0.99	0.94	0.84	1.08	**120.85**	0.6
SJ3	O26:H11	-	+	**32.56**	1.42	0.87	1.02	0.85	0.63	1.19	3.33	**28.63**
MI3	O45	+	-	0.76	**134.9**	0.91	1.16	0.97	1.05	1.64	**134.62**	0.65
MI4	O45	+	-	0.8	**129.8**	0.79	1.35	0.84	0.69	0.99	**127.75**	0.6
UMD	O45:H2	+	-	0.71	**136.5**	0.89	1.22	1.04	0.89	1.19	**138.35**	0.56
SJ9	O45:H2	+	+	0.71	**111.8**	2.3	0.78	0.84	0.68	1.1	**107.85**	**26.72**
05-0645	O45:H2	+	-	0.7	**124.9**	0.83	0.99	0.93	0.83	1.18	**125.14**	0.57
CVM	O103	+	-	0.73	0.72	**52.57**	0.93	0.58	0.58	0.87	**98.05**	0.69
CVM	O103	+	-	0.82	1	**46.45**	1.15	0.97	0.87	0.98	**94.63**	0.84
CVM	O103	+	-	0.77	0.68	**53.67**	0.87	0.86	0.77	0.91	**108.65**	0.73
CVM	O103	+	-	0.8	0.9	**57.92**	0.94	0.78	0.95	1.05	**117.52**	0.62
CVM	O103	+	-	0.84	0.77	**54.35**	0.8	0.87	0.9	1.17	**111.17**	0.7
CVM	O103	+	-	0.87	0.97	**54.7**	1.05	1.02	1	0.65	**94.08**	0.73
CVM	O103	+	-	0.87	0.92	**55.66**	1.09	0.98	0.75	0.96	**108.65**	0.7
TW	O103:H2	+	-	0.77	0.95	**49.38**	0.82	1.28	0.93	1.09	**102.45**	0.71
-4162	O103:H6	+	-	0.68	0.75	**49.78**	0.72	0.91	0.68	0.82	**92.38**	0.62

Cont.				***Strain Profile***							***P/N Ratios***	
Strain	**Serotype/ Serogroup**	***stx*$_1$**	***stx*$_2$**	**O26**	**O45**	**O103**	**O111**	**O121**	**O145**	**O157**	***stx*$_1$**	***stx*$_2$**
SJ12	O103:H11	+	-	0.75	0.98	**51.76**	1.08	0.95	0.95	0.91	**109.96**	0.61
TW	O103:NM	+	-	0.75	0.75	**49.81**	1.09	0.76	0.75	0.78	**100.21**	0.64
CVM	O111	+	-	0.51	1.13	0.49	**165.6**	0.92	0.53	1.09	**141.59**	0.62
CVM	O111	+	-	0.58	1.2	0.67	**151.2**	1.15	0.69	1.54	**136.5**	0.7
CVM	O111	+	-	0.66	1.05	0.5	**117.5**	1.11	0.83	0.9	**112.16**	0.69
CVM	O111	+	-	0.46	0.9	0.4	**157.6**	0.86	0.71	0.76	**141.15**	0.56
CVM	O111	+	-	0.53	1.08	0.44	**140.0**	1.49	0.68	0.84	**138.91**	0.87
CVM	O111	+	-	0.43	0.78	0.36	**166.2**	0.92	0.66	1.23	**143.53**	0.59
-1387	O111:H8	+	-	0.44	1.06	0.67	**154.1**	6.94	1.57	1.38	**130.65**	1
UMD	O111:H11	+	-	0.53	0.81	1.6	**139.9**	4.06	1.07	1.5	**127.74**	1.27
UMD	O111:NM	+	+	0.62	1.09	0.52	**125.8**	1.07	0.9	1.3	**124.29**	**29.23**
SJ13	O111:NM	+	+	0.4	0.84	0.4	**122.4**	1.14	0.62	0.7	**120.48**	**28.04**
UMD	O111:NM	+	-	0.52	0.91	0.6	**149.7**	2.33	0.78	0.99	**128.58**	0.62
TW	O121	+*	-	0.82	1.31	1.27	0.89	**216.0**	2.31	2.01	**36.42***	0.61
-8023	O121:H19	-	+	0.94	0.94	0.98	1.08	**206.5**	0.75	0.85	1.59	**29.17**
TW	O121:H19	-	+	1.04	0.91	0.94	0.91	**204.8**	0.96	1.58	1.06	**31.05**
SJ19	O121:H19	+	+	0.83	0.79	0.79	0.92	**200.5**	0.78	1.07	**121.56**	**30.94**
CVM	O145	-	-	0.89	0.91	0.84	0.97	1.42	**150.65**	0.95	0.88	0.9
CVM	O145	+	-	0.75	1.08	1.72	0.7	11.12	**156.38**	3.82	**118.84**	3.71
CVM	O145	-	+	0.93	0.89	0.97	0.91	2.67	**160.19**	1.41	1.3	**30.59**
TW	O145:H28	-	+	0.95	1.05	1.32	1.21	6.03	**149.85**	1.65	3.42	**34.79**
TW	O145: NM	+	-	0.73	0.94	2.02	0.66	5.21	**152.94**	8.41	**129.11**	1.53
SJ23	O145:NM	+	+	0.73	0.73	0.85	0.92	2.09	**152.42**	0.81	**119.81**	**30.38**
CVM	O157	+	+	0.77	0.63	0.85	1.09	0.73	0.64	**291.25**	**134.48**	**31.85**
CVM	O157	+	+	0.79	0.65	0.78	0.89	0.72	0.71	**279.95**	**129.04**	**31.74**
CVM37	O157	+	+	0.83	0.71	0.79	1.08	1.33	0.82	**270.72**	**117.61**	**31.12**
BAA	O157	+	+	0.84	0.69	0.89	0.96	1.29	0.55	**257.58**	**129.3**	**32.98**
-43893	O157	+	-	0.85	0.65	1.02	0.75	0.78	0.71	**261.95**	**143.95**	0.74

Cont.				***Strain Profile***						***P/N Ratios***		
Strain	**Serotype/ Serogroup**	***stx₁***	***stx₂***	**O26**	**O45**	**O103**	**O111**	**O121**	**O145**	**O157**	***stx*$_1$**	***stx*$_2$**
TW	O157	-	+	1.01	0.77	0.94	0.98	0.74	0.66	**260.87**	0.99	**31.43**
FDA	O157	-	+	0.92	0.69	1.05	0.9	0.93	0.53	**271.3**	1.13	**31.28**
EDL	O157:H7	+	+	0.85	0.54	1.01	0.88	0.94	0.89	**287.09**	**135.21**	**31.23**
-85170	O157:H7	-	-	1.02	0.75	0.96	0.94	1.5	0.85	**270.38**	1.34	1.19
UMD46	O157:H7	+	+	0.82	0.51	0.97	0.98	0.88	0.71	**261.9**	**104.43**	**31.7**
84-1097	O157:H25	-	-	0.95	2.63	1.34	0.99	2.69	1	**266.64**	2.13	1.21

*stx_{1d}

Table II.5. Inclusivity and exclusivity test results for the *Escherichia coli* O antigen Suspension Array panel.

Target	**O26**	**O45**	**O103**	**O111**	**O121**	**O145**	**O157**
E. coli Target serogroups*	11/11	5/5	11/11	11/11	4/4	6/6	11/11
Other *E. coli* serogroups**	0/23	0/23	0/23	0/23	0/23	0/23	0/23
Other bacterial species***	0/21	0/21	0/21	0/21	0/21	0/21	0/21

**E. coli* O26, O45, O103, O111, O121, O145, and O157.

**E. *coli* K12, O1, O2, O5, O6, O8, O15, O18, O22, O28, O55, O88, O91, O104, O113, O118, O128, O128ac, O142, O146, O153, and O174.

****Shigella flexnieri, S. dysenteriae, S. boydii, S. sonnei, Salmonella* Typhimurim *s*train LT2, *S.* Typhimurium ATCC14028, *S.* Enteritidis (2 strains), *Proteus vulgaris, Pseudomonas aeruginosa, Enterobacter aerogenes, Citrobacter freundii, Hafnia alvei, Enterobacter cloacae, Acinetobacter calcoaceticus, Listeria monocytogenes, L. innocua, L. grayii, Staphylococcus aureus, Streptococcus faecalis, and Bacillus subtilis*

Chapter 3: Characterization of CRISPR (Cluster Regularly Inter-spaced Short Palindromic Repeats) loci in Shiga toxin-producing *Escherichia coli*

3.1 Abstract

Shiga toxin-producing *Escherichia coli* (STEC) (n=194) representing 43 serotypes were examined for clustered regularly interspaced short palindromic repeats (CRISPR) arrays to study genetic relatedness among STEC serotypes. A subset of the strains (n=81) was further analyzed for I-E subtype *cas* and virulence genes to determine possible association of CRISPR elements with virulence. Four types of CRISPR arrays were identified. CRISPR1 and CRISPR2 were present in all strains tested. One strain also had both CRISPR3 and CRISPR4, whereas 193 strains displayed a short, combined array CRISPR3-4. A total of 3,353 spacers were identified, representing 528 distinct spacers. The average length of a spacer was 32 bp. Approximately half of the spacers (54%) were unique, and found in strains of less common serotypes. Overall, CRISPR spacer content correlated well with STEC serotypes, and identical arrays were shared between strains with different O types but the same H type (O26:H11, O103:H11, and O111:H11). There was no association between the presence of I-E subtype *cas* and virulence genes, but the total number of spacers correlated negatively with potential pathogenicity significantly ($p<0.05$). Fewer spacers were found in strains that had greater probability of causing outbreaks and disease than in those that less likely cause illness ($p<0.05$). These findings showed that CRISPR array content correlated well with STEC serotype, and that

CRISPR systems were inversely related to the presence of virulence determinants, although this relationship needs to be determined on a broader scale and a biologic link will need to be established.

3.2. Introduction

Shiga toxin-producing *Escherichia coli* (STEC) has been recognized as a human pathogen since the early 1980's, when two consecutive outbreaks of STEC serotype O157:H7 in contaminated beef patties sickened 47 people in the US (1). To date, over 400 additional serotypes have been associated to bacterial gastroenteritis worldwide (2) and there are estimated over 175,000 cases of STEC infections each year in the US alone (3). Depending on the ability to cause outbreaks and/or severe disease, Karmali *et al.* classified STEC serotypes into seropathotypes (SPT) A to E: SPT A causes outbreaks and disease at high rates and SPT E has not been linked to outbreaks or severe disease (4).

Clustered regularly interspaced short palindromic repeats (CRISPR) were first discovered in *E. coli* in 1987 (5) and have now been found in~45% of bacteria and ~ 90% of archaea (6-8). CRISPRs function as heritable and adaptive immune systems against mobile genetic elements (phages, plasmids, etc) (9-11), and are made of three components: a leader sequence that carries a promoter for transcription; CRISPR associated genes (*cas*) encoding proteins with multiple functions; and CRISPR arrays formed by repeats and spacers (12). While most repeats are typically indistinguishable in size and sequence within a defined locus, they are intercalated by non-repeated short sequences called spacers, which are of a constant number of nucleotides and unique and hypervariable within a locus (13). They originate from

mobile and invasive genetic elements incorporated into the array and subsequently serve as the sequence-specific recognition portion of the immune system (14-16).

Four CRISPR loci and two CRISPR types are described in *E. coli* (17-19): CRISPR1 is located between *iap* and *cysH*, and CRISPR2 is in the region between *ygcF* and *ygcE* (17, 19). They have identical consensus repeats (20). CRISPR1 associated proteins form the I-E type system. CRISPR3 and CRISPR4 also have identical consensus repeats, and are located between *clpA* and *infA.* CRISPR3 *cas* genes form CRISPR type I-F (17, 19, 20). Array size and content vary among CRISPR types and strains. It is not common that the four loci are present in a single *E. coli* isolate. CRISPR1 and CRISPR2 are most frequently found in *E. coli* (19, 21).

CRISPR arrays evolve by polarized acquisition of novel spacers and represent a chronological record of infectious assault on the bacterium from viral and other genetic elements. Distal spacers from the leader sequence are older and are common among strains, while newer spacers are closer to the leader and more strain specific. Occasionally, sporadic deletions of internal spacers do occur (22). Differences in spacer content would indicate variations in the host environment and geographical locations and may be useful in evolutionary and epidemiological studies (12). This variability makes CRISPR arrays suitable genetic markers for bacterial subtyping.

The primary biological role of CRISPR-*Cas* systems is to provide acquired immunity to protect the cell against mobile genetic elements such as viruses and plasmids (10, 11). Conversely, evolution of pathogenic strains is attributed to the acquisition of elements through lateral gene transfer, such as transposons, phages, genomic islands, and plasmids (23, 24). For example, genomic analysis of STEC

serogroups O26, O103, O111, and O157 revealed they have much larger genomes than non-pathogenic *E. coli*, mainly due to a large content of prophages and other integrative elements (25). It is expected that strains containing functional CRISPR systems restrict the acquisition of mobile genetic elements, and that strains with the most complex and active CRISPR systems would have a lower susceptibility to infections by mobile genetic elements (19). However, studies on the relationship of CRISPR systems and the acquisition of genetic mobile elements resulted in different findings. While an inverse relationship was reported between the presence of *cas* and virulence factors in *Enterococcus* spp., no correlation was found between CRISPR and the presence of plasmids containing antimicrobial resistance genes in *E. coli* (26, 27). The hypothesis in this study was that CRISPR arrays are a suitable marker for STEC serotyping, and that there is an inverse relationship between the presence of CRISPR elements and virulence determinants in STEC.

3.3. Objective

Studies of the CRISPR loci in a variety of STEC are limited. The aim of this study was to describe CRISPR arrays of 194 STEC strains of 43 serotypes, to investigate the relationship between arrays in important serotypes and to explore potential relationship between CRISPR elements and virulence genes.

3.4. Materials and Methods

3.4.1. Strain collection

A set of 190 STEC from the food safety laboratory collection at University of Maryland strains collection were analyzed, including 30 O26, 30 O103, 41 O111, 6 O45, 4 O121, 6 O145, 12 O157, and a variable number of strains of other serogroups (Table III.1). The strains were isolated from a variety of geographical location and sources, including humans, cattle and beef products, sheep, goat, deer, okapi, goat, and produce. Collection dates range from 1976 up through 2010.

3.4.2. DNA isolation

Genomic DNA was extracted from a pure culture after streaking on LB agar and incubating at 35º C for 24 h, using Instagene matrix (Bio-Rad, Hercules, CA). Briefly, 1-2 colonies were suspended in 1 mL of ultrapure water and centrifuged. Supernatant was discarded and 200 uL of Instagene matrix was added followed by incubation at 56°C for 15 min and at 94°C for 8 min. After centrifugation, the supernatant containing DNA was stored at -20°C until use.

3.4.3. PCR and DNA sequencing

CRISPR array sequences were obtained through PCR and Sanger sequencing using previously described primers (21). PCR reactions consisted of 1 uL of bacterial DNA mixed with HotStarTaq Plus Master mix (12.5 µl) (Qiagen, Valencia, CA), 10 pM of forward and reverse primers and water to reach a final reaction volume of 25

µl. PCR parameters included: initial denaturation at 95 °C for 5 min, 10 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 90 s for 10 cycles followed by 25 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 90 s plus a 10-s cycle elongation for each successive cycle (21). PCR products were sequenced by MCLAB (South San Francisco, CA) from both ends using Applied Biosystems ® fluorescent dye terminator technology in an ABI 3730xl sequencer with the same PCR primers.

3.4.4. CRISPR array sequence analysis

Sequences were assembled with the Geneious software v. 6.0.5 (New Zealand). Arrays were extracted using the "clean sequence tool" enclosed in a macro script/database provided by DuPont, as previously described (28). The tool detected repeats listed in a repeat database and automatically separated repeats and the intercalated short sequences –spacers— into different columns Data were subsequently formatted to a graphic representation of each spacer and repeat based on their sequence (28). To corroborate array sequence, each sequence was tested using the CRISPRfinder (http://crispr.u-psud.fr/Server/) program online (29). In addition, CRISPR sequences of four major STEC serogroups (O26, O103, O111, O157) and *E. coli* K12 were obtained from NCBI and included in the analysis (Table III.1).

To analyze arrays, strains were arranged based on the presence of common consecutive spacers from the distal end to the leader sequence. Strains with the longest series of spacers on their array were designated as "anchors", which were used as a guide for organizing strains in clusters.

3.4.5. Protospacers analysis

Spacer identity was determined using a standalone blast program (blast+ 2.2.27) against the NCBI non redundant (nr) nucleotide collection. Protospacers were defined as homologous sequences with an e value <1.10e-5 and less than 10% difference in sequence length (21). Self matches to *E. coli* CRISPR loci sequences were omitted.

3.4.6. Nucleotide sequence accession number

Sequences identified were submitted to GenBank with accession numbers from KF522692 to KF523262.

3.4.7. I-E subtype *cas* screening

A seropathotype (4) balanced subset of 81 strains were selected based on a previous study (30) to screen for the presence of *cas1* and *cas2*, which are markers of the I-E system (Table III.1). Primers *cas1*FW (5' –CGCCTGCATTATGCTCGAAC-3'), *cas1*REV (5'-CATTTTGCGCACCACCTTCA- 3'), *cas2*FW (5'-ATGAGCATGGTCGTGGTTGT- 3'), and *cas2*REV (5' –CCCATCCAAATCCACCGGAA- 3') were designed based on whole genome sequencing (WGS) of 24 strains using Geneoius v. 6.0.5. In separate reactions for I-E subtype *cas1* and *cas2*, 12.5 µl of HotStarTaq Plus Master mix (Qiagen) were mixed with 10 pM of forward and reverse primers (Invitrogen, Carlsbad, CA), 1 µl of bacterial DNA and water for a final reaction volume of 25 µl. PCR parameters were:

initial denaturation at 95 °C for 5 min, 30 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 90 s and a final extension at 72 °C for 10 min.

3.4.8. I-E subtype *cas* analysis.

A maximum likelihood phylogenetic tree was constructed based on the concatenated sequence of system I-E subtype *cas* genes (*cas1, cas2, cas3', cse1, cse2, cas6e*, *cas7*, *cas5*) (20) of 16 STEC strains previously sequenced (31) and eight publically available *E. coli* sequences (GenBank)(Table III.1). The tree was constructed using Mega 5.1(32) with 1,000 bootstrap iterations, and *E. coli* K12 was used as outgroup. Pairwise distance matrix was calculated based on total 1,014 SNPs to display the evolutionary divergence between different groups on the phylogenetic tree (Mega 5.1 with 1,000 bootstrap replications).

3.4.9. Virulence genes screening

The presence of selected virulence genes, *stx1*, *stx2*, *eae*, *hly*A, *pagC, sen, nleB, efa-1, efa-2, terC, ureC, iha, aidA-1, nle2-3, nleG6-2, nleG5-2* ,*irp2* and *fyuA*, was obtained for a subset (Table III.1) of strains from a previous study (30).

3.4.10. Statistical Analysis

Data were analyzed with SSPS v20. ANOVA or Kruskal-Wallis test was performed, when suitable. P values <0.05 were considered statistically significant.

3.5. Results

In the current work, CRISPR arrays of 194 STEC strains of 43 representative serotypes were screened and characterized, and also potential association between CRISPRs and virulence genes were evaluated.

3.5.1. CRISPR array

Four types of CRISPR arrays were identified among the 194 STEC strains and *E. coli* K-12. CRISPR1 and CRISPR2 were present in all 195 strains tested. One strain (95-3322) also had CRISPR3 and CRISPR4, whereas 193 strains displayed a short, combined array CRISPR3-4 (Table III.2). The length of CRISPR1 and CRISPR2 arrays varied from 1 to 20 spacers, with most having 5 or 7 spacers. The unique CRISPR3 and CRISPR4 arrays of strain 95-3322 were of 11 and 6 spacers in length, respectively, whereas the combined array CRISPR3-4 typically had only one spacer (Table III.2). Nearly 90% STEC (173/195) carried an additional array in the I-E system located at 0.5 kb from CRISPR2 (19). This array (CRISPR2b) had one spacer and its sequence was conserved among strains (Data not shown).

CRISPR1 was less polymorphic than CRISPR2. Most CRISPR1 arrays (94%; 184/195) shared an ancestral (first) spacer, and many strains (64%; 125/195) also shared the second oldest spacer (Fig. III.1), indicating a common origin. However, CRISPR2 did not share the first spacer and many shared only the second spacer. Both arrays had numerous deletions of spacers, mostly at 2 or 3 spacers. Interestingly, despite the observation that the older spacer of CRISPR1 was shared by 184 strains, the first repeat was only shared by 151 strains (Data not shown). For most of the

combined arrays, CRISPR3-4 had only one spacer (95%; 180/195), and this same spacer was present in 145 strains across different serotypes, reflecting common origin.

3.5.2. Spacer diversity

A total of 3,353 spacers were identified, from which 528 were distinct. The average length of a spacer was 32 bp, ranging from 30 to 35 bp. Approximately half of the 528 spacers (54%) were unique (Table III.2) and found in strains of less common serotypes (Fig. III.1). Many strains shared spacers in the same CRISPR loci, but no spacers were shared between CRISPR loci (Fig. III.1).

Ten of the 528 spacers had identity with plasmids from *Salmonella* Enteritidis or *E. coli* (i.e., protospacers). These spacers occurred in 13 strains. Additionally, one spacer showed identity to Enterobacteria phage P7 and was present in 12 of the 13 strains (Table III.3). Most spacers (8/10) with known protospacers formed part of CRISPR1, and some strains (7/13) had more than one of these spacers in their array (Table III.4). For example, strains XDN 4854 and XDN 5545 contained five and four of these spacers in CRISPR1, respectively. CRISPR3 and CRISPR4, found only in strain 95-3322, carried one spacer with known protospacers, and it also carried one of the spacers in CRISPR1. Locations of these spacers in the array were random, from position 1 to position 19. Most strains harboring these spacers were of uncommon serotypes, and five of them were not even typeable serologically (Table III.4). These

homologies to phage and plasmids sequences are consistent with a role that CRISPR play in resisting mobile genetic elements as previously described in literature (9, 10).

3.5.3. Array organization by serotype

CRISPR arrays were organized based on the spacer content of anchor strains which are defined as those that contain all spacers for a group/cluster in correct order representing ancestral strains. Although a universal anchor was not identified, four clusters were established, each one with one anchor (Fig. III.1). The first cluster was formed by O145 strains, and anchored by a human isolate, 07865, (O145:H28). The second group was anchored by CVM 9591 (O111:H11) isolated from cow in 1995. The cluster included two subgroups: O111: H8 and O111: NM in a block, and O26:H11, O103:H11, and O111:H11, among others, in a second group. The third cluster was more diverse, formed by several serotypes including O45:H2, O103:H2, O103:H25, O91:H21, and O91:H14. This group was anchored by CVM 9340 (O103:H25) from human. Last group was also very diverse, anchored by 08023 (O121:H19). Strains of less common serotypes did not forme clusters. Since CRISPR1 and CRISPR2 co-clustered, the same arrangement was achieved using either one as a guide (Fig. III.1). This was consistent with a parallel evolution of the two CRISPR loci over time.

Strain clustering based on CRISPR spacer content correlated well with STEC serotype status. For instance, serotype O111:H8 formed a large cluster of 29 strains that had almost identical spacer contents with only a few minor deletions of 1 or 2 spacers in CRISPR1 and CRISPR2. Similar findings were observed among serotypes

O26:H11, O103:H2 and O157:H7. Unique, long CRISPR arrays were present in less common STEC serotypes (Fig. III.1). These data underscore the notion that CRISPR elements may serve as reliable discriminatory signatures at least down to the level of serotype for STEC strain lineages.

It was notable that spacer content seemed to correlate well with strains retaining the same H antigen type, but not necessarily with strains having the same O group. For example, O103:H2 did not share any spacer in CRISPR2 with O103:H11, although they did have common ancestral spacers in CRISPR1 (3/12). However, O103:H2 clustered together with O45:H2, and contained identical spaces in CRISPR1 up to the fourth spacer where O103:H2 had additional eight spacers. Similarly, O45:H2 and O103:H2 differed only by one spacer deletion in CRISPR2 (Fig. III.2). On the other hand, only nine of 17 spacers were shared by strains between O111:H8 and O111:H11, whereas strains of O26:H11 and O103:H11had practically identical arrays, forming a sub-cluster based on antigen H11 (Fig III.1). Taken together, these data may point to H antigen loci as more phylogenetic stable while O antigen alleles appear to be shuffled in the evolution of STEC clades (31).

3.5.4. Correlation between CRISPR content and occurrence of virulence genes

Previous reports indicate an inverse correlation between the presence of virulence genes and the distribution of *cas* genes in *Enterococcus fecalis* (*E. fecalis*) (27). Therefore, a subset of strains (n=81) of different STEC seropathotypes (Table III.1) was analyzed for virulence genes (30) and the presence of I-E subtype *cas* genes. While most strains (91%) had *cas1*, all STEC strains carried *cas2*. Because of

such high positive rates, there was no significant difference in the presence of I-E subtype *cas* among different seropathotypes. Similarly, no association between the presence of I-E subtype *cas* and virulence genes was observed.

A significant difference was observed in the total content of spacers between strains of different seropathotypes ($p<0.05$) (Fig III.2a), and fewer spacers were found in strains that had greater probability of causing outbreaks (SPT A and B) compared to those with lower probability ($p<0.05$) (Fig. III.2b). Similarly, strains with higher potential of causing severe disease (SPT A, B and C) had fewer spacers than those with lower potential (SPT D and E) ($p<0.05$) (Fig. III.2c). An association between the number of spacers and the presence of certain virulence genes was also observed. For example, *eae*-positive strains had significantly fewer spacers than *eae*-negative strains ($p<0.05$). Other virulence genes including *pagC, sen, terC, ureC, nleB, nle2-3*, *nleG6-2,* and *nleG5-2* also showed the same significant relationship with the number of spaces. However, the opposite relationship was seen with genes *fyuA* and *irp*2, and no association was detected between the number of spacers and the presence of *hly*A, *aidA-1*, *iha*, *efa1*, *efa*2, *stx1* and *stx2* (Data not shown). Interestingly, strains containing both *stx* genes showed significantly fewer spacers than those with only one of them ($p<0.05$) (Fig. III.2d).

3.5.5.I-E subtype *cas* phylogeny

To investigate the relationship between CRISPR and the evolutionary history of strains, a maximum likelihood phylogenetic tree was reconstructed based on the concatenated sequence of the I-E *cas* system genes extracted from 24 *E. coli* (Fig.

III.3). Strains were grouped into four major clades except *E. coli* K12 which was used as outgroup. All O157:H7 strains formed a single clade, whereas O103:H2 strains belonged to another cluster. However, an O103:H25 strain (CVM9340) appeared in a separated clade. Interestingly, the remaining strains of serotypes O111:H11, O111:H8 and O26:H11 all clustered together, indicating a closer phylogenetic relationship and more conserved I-E subtype *cas* alleles among them.

Additionally, pairwise distance matrix of SNP differences (data not shown) supported phylogeny results of maximum likelihood analysis. For example, SNPs differences between group H8 and H11, group O103:H25, O103:H2, and group O157:H7 were 14, 74, and 100 SNPs, respectively (Fig III.3).

3.6. Discussion

In this study, the occurrence and content of CRISPR loci in STEC was determined, and conservation among strains of the same serotype (O and H antigen type combination) but not between serogroups (*i.e.*, only O antigen type) was observed. However, in certain cases, strains of different serotypes but with the same H type shared identical CRISPR sequences, suggesting such serotypes might have common ancestors based on H type but not in O groups (Fig. III.1). This may provide a genetic basis for the specific detection and tracking of particular *E coli* strains in the environment or in the food supply. In addition, a significant negative association was observed between the number of spacers (an indicator of CRISPR systems activity) and the pathogenic potential of STEC strains as indicated by their seropathotype (4) – a find heretofore undescribed among STEC strains.

Other studies also demonstrated the relationship between CRISPR array content and serotypes. Delannoy *et al.* reported the presence of specific CRISPR polymorphisms related to O:H serotypes of STEC O26:H11; O45:H2; O103:H2; O104:H4, O111:H8; O145:H28, and O157:H7 which were useful to differentiate these serotypes (33, 34). However, they reported numerous cross reactions: primers for O145:H28 reacted with strains O28:H28, and primers detecting O103:H2 and O45:H2 altogether also cross-reacted with strains O128:H2 and O145:H2, among others (33). These data showed similar CRISPR array characteristics; identical arrays were shared by strains of different serotypes having the same H antigen (O26:H11, O103:H11, and O111:H11; O45:H2 and O103:H2), and arrays of strains of the same serogroups with different H types seemed unrelated (O103:H2 and O103:H11) further underscoring the linkage between CRISPR array and H antigen alleles (Fig. III.1). Similarly, Yin *et al.* confirmed a relationship between CRISPR polymorphism and serotypes, but also described a strong conservation of CRISPR arrays within isolates of the same H type including H7, H2 and H11 (18). A previous study on the evolutionary history of non-O157 STEC by WGS showed O26:H11 and O111:H11 grouped together, also suggesting strains with the same H antigens may have common ancestors (31).

Furthermore, the concatenated sequence of their I-E subtype *cas* genes did not discriminated between strains of the serotypes O26:H11, O111:H8 and O111:H11, reflecting close relatedness of those serotypes. Ju *et al.* also demonstrated strains with H8/H11antigens formed a major clade on the whole genome wide phylogenetic tree but displayed closer relatedness with O103:H2 strains (31). In contrast, group H8/

H11was closer to strains of serotype O103:H25 than to strains of O103:H2 based on I-E subtype *cas* sequences. Thus, concatenated I-E subtype *cas* genes could not be used to determine the same phylogenies as found in genomic comparisons among serotypes (35).

The role of CRISPR as an immune system against mobile genetic elements has been previously reported (9, 36), and since many virulence determinants are acquired through mobile genetic elements (25) it is expected that strains with more active CRISPR systems carry less virulence genes and other mobile genetic elements. Some studies focused on the role of CRISPR systems in acquisition of virulence determinants with contradictory results, notably in *E. coli*. Specifically, one study showed the acquisition of plasmids carrying antimicrobial genes was not related to the presence of the CRISPR system (26). In contrast, another study found that CRISPR-Cas systems were inversely correlated with the presence of acquired antibiotic resistance in *E. fecalis* strains (37). In *Enterococcus spp.*, an inverse correlation between the presence of two virulence genes and the distribution of *cas* genes was reported, and less virulence genes were detected when *cas* genes were present (27). In the present study, it was found that *cas* genes were not related to the presence of virulence markers in STEC (35); however, statistically significant differences in the total number of spacers between strains of seropathotype A and other seropathotypes indicated the most pathogenic serotypes had less spacers than those of other serotypes (Fig. III.2), suggesting a negative correlation between CRISPR activity and propensity for pathogenic trait acquisition. This was consistent with the documented role of CRISPR-Cas immune systems in limiting the uptake of

genetic material derived from mobile and invasive elements such as phages and plasmids. Similarly, the presence of some virulence genes was related to lower spacer content. Notably, when two *stx* genes were present, strains had significantly less spacers than those having one ($p<0.05$). This would challenge recent studies in *E. coli* suggesting that CRISPR systems would not work as a cell defense system against alien genetic elements (21, 38). Conversely, High Pathogenicity Island (HPI) genes *irp2* and *fyuA* presented an inverse relationship; however, its contribution to virulence seems irrelevant (30). Therefore, further studies are necessary across a broad range of genetic strains to assess the relationship between CRISPR-Cas systems and virulence in *E. coli.*

Four CRISPR arrays have been identified in *E. coli,* but all four are rarely found in a single isolate (17, 19). Similarly to what was found by Yin *et al*. (18), data in this study showed that the type I-E CRISPR-*Cas* system (CRISPR1 and CRISPR2) was most widely distributed in STEC. One strain (95-3322), however, carried the four arrays, and the remaining strains carried a shorter, combined CRISPR3-4 array as previously described (19) which is associated with the fusion of the remaining sections of loci 3 and 4 when I-F genes, originally located between the two loci, are deleted (19). To confirm the absence of I-F *cas* genes, the region between primers C3Fw (*clpA* target) and C4 Rev (*infA* target) was sequenced. In most cases, the fragment produced was of ~800 bp instead of the expected ~3000bp when I-F subtype *cas* genes are present (179/190) (19). The absence of *cas* genes and repeats among these motifs would suggest a relatively minor role for CRISPR system I-F in STEC "immune" function.

Protospacers for 10 CRISPR spacers from STEC were identified. Most protospacers (9/10) were located in *Salmonella* and *E. coli* plasmids, including multiple sequences from the same plasmid (Table III.2), for both CRISPR1 and CRISPR2. Additionally, a spacer showing identity with Enterobacteria phage P7 was found in 12 out of the 13 strains for which matching protospacers were identified (Table III.2). Yin *et al.* also observed multiple spacer sequences originated from the same origin (18), and previously Datsenko *et al.* demonstrated that mutated motif stimulates the acquisition of more spacers from the same target to strengthen immunity against the element (39).

3.7. Conclusion

The current study provides novel insights into the occurrence and role of CRISPR-Cas systems in STEC serogroups (O26, O103 and O111) as well as several additional uncommon serotypes. CRISPR array sequence analysis suggests H antigen might have been acquired more ancestrally than O antigen since arrays are shared by strains with the same H antigen but not among strains with the same O antigen. Alternatively, stability among H antigens in STECs may also point to a more vertical inheritance pattern and less promiscuity than O antigen evolution, known to be dappled by numerous horizontal gene transfer events throughout its radiation in *E. coli* (40). Also, the relationship between CRISPR elements and pathogenicity traits in STEC needs to be studied to determine whether they have a causal relationship or whether a formal balancing selection drives acquisition of the two. Further studies using additional and genetically diverse strains would provide a better understanding of the CRISPR-Cas system in STEC, and *E. coli* as a whole. CRISPR arrays and other

genetic markers could be used to differentiate high risk STEC from low risk strains, thereby provide useful tools for the control of STEC infections, and insights into their genetic content and phenotypic traits.

3.8. References

1. **Wells JG, Davis BR, Wachsmuth IK, Riley LW, Remis RS, Sokolow R, Morris GK.** 1983. Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype. J. Clin. Microbiol. **18 (3):**512-520.
2. **Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J.** 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J. Clin. Microbiol. **42 (1):**311-319.
3. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerging Infect. Dis. **17 (1):**7-15.
4. **Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Reid-Smith R, Isaac-Renton J, Clark C, Rahn K, Kaper JB.** 2003. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J. Clin. Microbiol. **41(11):**4930-4940.

5. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A.** 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J. Bacteriol. **169 (12):**5429-5433.

6. **Barrangou R.** 2013. CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip Rev RNA **4 (3):**267-278.

7. **Barrangou R, Horvath P.** 2012. CRISPR: New Horizons in Phage Resistance and Strain Identification. Annu Rev Food Sci Technol **3:**143-162.

8. **Bhaya D, Davison M, Barrangou R.** 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu. Rev. Genet. **45:**273-297.

9. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315 (5819):**1709-1712.

10. **Garneau JE, Dupuis M, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S.** 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468 (7320):**67-71.

11. **Marraffini LA, Sontheimer EJ.** 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322 (5909):**1843-1845.

12. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327 (5962):**167-170.

13. **Al-Attar S, Westra ER, van der Oost J, Brouns SJJ.** 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biological Chemistry **392 (4):**277-289.

14. **Bolotin A, Quinquis B, Sorokin A, Ehrlich SD.** 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151(Pt 8):**2551-2561.

15. **Mojica FJ, Díez-Villaseñor C, García-Martínez J, Soria E.** 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. **60 (2):**174-182.

16. **Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321 (5891):**960-964.

17. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5 (6):**e11126.

18. **Yin S, Jensen MA, Bai J, Debroy C, Barrangou R, Dudley EG.** 2013. The evolutionary divergence of Shiga Toxin-Producing *Escherichia coli* is reflected in Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) spacer composition. Appl. Environ. Microbiol. **79 (18):**5710-5720.

19. **Díez-Villaseñor C, Almendros C, García-Martínez J, Mojica FJ.** 2010. Diversity of CRISPR loci in *Escherichia coli*. Microbiology **156 (Pt 5):**1351-1361.

20. **Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. **9(6):**467-477.

21. **Touchon M, Charpentier S, Clermont O, Rocha EP, Denamur E, Branger C.** 2011. CRISPR Distribution within the *Escherichia coli* Species is not suggestive of immunity-associated diversifying selection. J. Bacteriol. **193 (10):**2460-2467.

22. **Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S.** 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. **190 (4):**1390-1400.

23. **Lawrence JG.** 2005. Common themes in the genome strategies of pathogens. Curr. Opin. Genet. Dev. **15 (6):**584-588.

24. **Coombes BK, Gilmour MW, Goodman CD.** 2011. The evolution of virulence in non-O157 shiga toxin-producing *Escherichia coli*. Front Microbiol **2:**90.

25. **Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **106 (42):**17939-17944.

26. **Touchon M, Charpentier S, Pognard D, Picard B, Arlet G, Rocha EP, Denamur E, Branger C.** 2012. Antibiotic resistance plasmids spread among natural isolates of *Escherichia coli* in spite of CRISPR elements. Microbiology **158 (Pt 12):**2997-3004.

27. **Lindenstrauß AG, Pavlovic M, Bringmann A, Behr J, Ehrmann MA, Vogel RF.** 2011. Comparison of genotypic and phenotypic cluster analyses of virulence determinants and possible role of CRISPR elements towards their incidence in *Enterococcus faecalis* and *Enterococcus faecium*. Syst. Appl. Microbiol. **34 (8):**553-560.

28. **Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R.** 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. **190 (4):**1401-1412.

29. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35:**W52-57.

30. **Ju W, Shen J, Toro M, Zhao S, Meng J.** 2013. Distribution of Pathogenicity Islands OI-122, OI-43/48, OI-57 and High Pathogenicity Island (HPI) in Shiga Toxin-producing *Escherichia coli*. Appl. Environ. Microbiol. **79 (11):**3406-3412.

31. **Ju W, Cao G, Rump L, Strain E, Luo Y, Timme R, Allard M, Zhao S, Brown E, Meng J.** 2012. Phylogenetic analysis of non-O157 Shiga toxin-

producing *Escherichia coli* by whole genome sequencing. J. Clin. Microbiol. **50 (12):**4123-4127.

32. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28 (10):**2731-2739.

33. **Delannoy S, Beutin L, Fach P.** 2012. Use of CRISPR sequence polymorphisms for specific detection of enterohemorrhagic *E. coli* (EHEC) strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28 and O157:H7 by real time PCR. J Clin Microbiol. **50 (12):** 4035-4040.

34. **Delannoy S, Beutin L, Burgos Y, Fach P.** 2012. Specific Detection of Enteroaggregative Hemorrhagic *Escherichia coli* O104:H4 Strains by Use of the CRISPR Locus as a Target for a Diagnostic Real-Time PCR. J. Clin. Microbiol. **50 (11):**3485-3492.

35. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8 (2):**e55687.

36. **Edgar R, Qimron U.** 2010. The *Escherichia coli* CRISPR system protects from λ lysogenization, lysogens, and prophage induction. J. Bacteriol. **192 (23):**6291-6294.

38. **Palmer KL, Gilmore MS.** 2010. Multidrug-resistant Enterococci lack CRISPR-*cas*. MBio **1 (4):**e00227-10.

38. **Mojica FJ, Díez-Villaseñor C.** 2010. The on-off switch of CRISPR immunity against phages in *Escherichia coli*. Mol Microbiol **77 (6):**1341-1345.

39. **Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E.** 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun **3:**945.

40. **Tarr PI, Schoening LM, Yea YL, Ward TR, Jelacic S, Whittam TS.** 2000. Acquisition of the *rfb-gnd* cluster in evolution of *Escherichia coli* O55 and O157. J Bacteriol **182 (21):**6183-6191.

3.9. Figures and Tables

Figure III.1. Arrays CRISPR1 and CRISPR2 for STEC strains in the study.

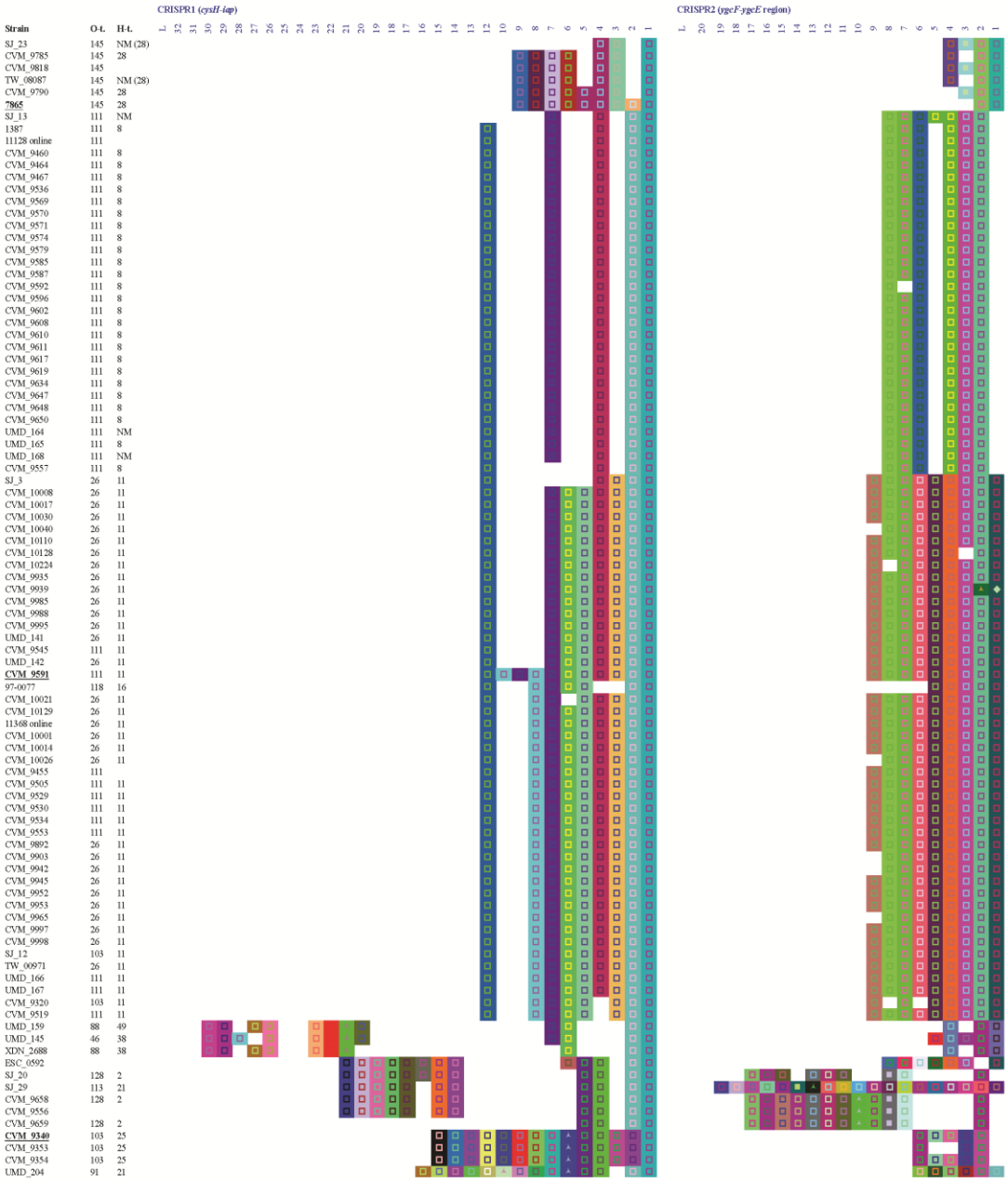
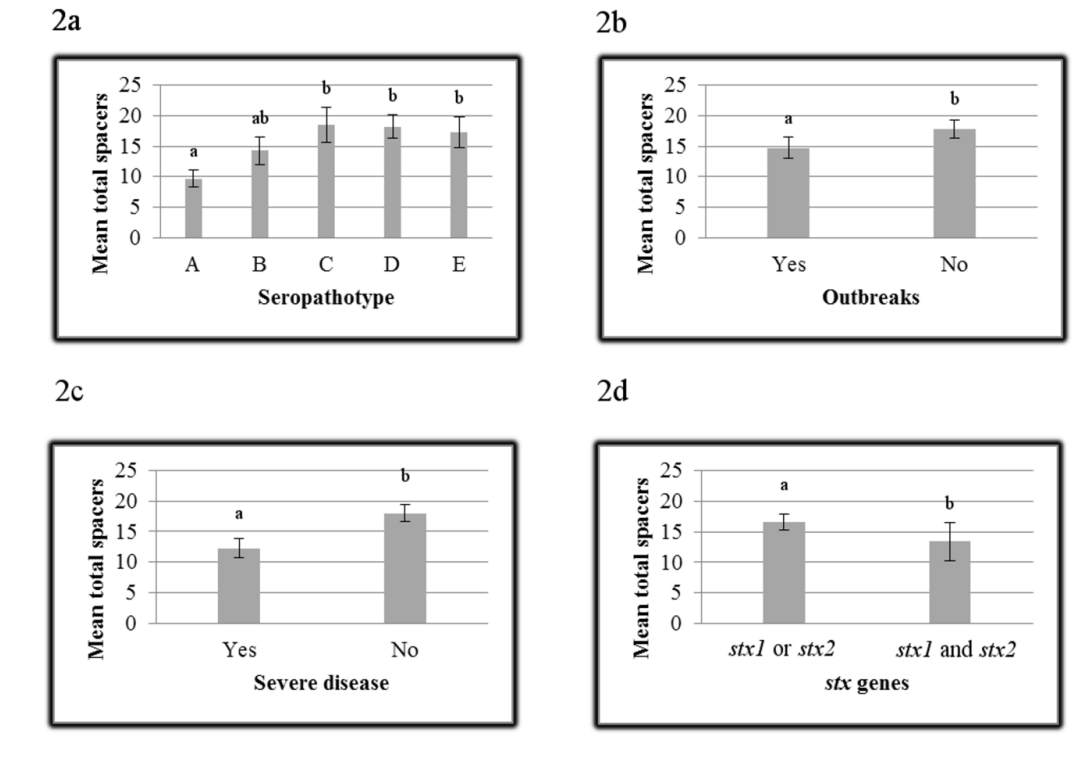


Figure III.1. Arrays CRISPR1 and CRISPR2 for STEC strains in the study (cont).

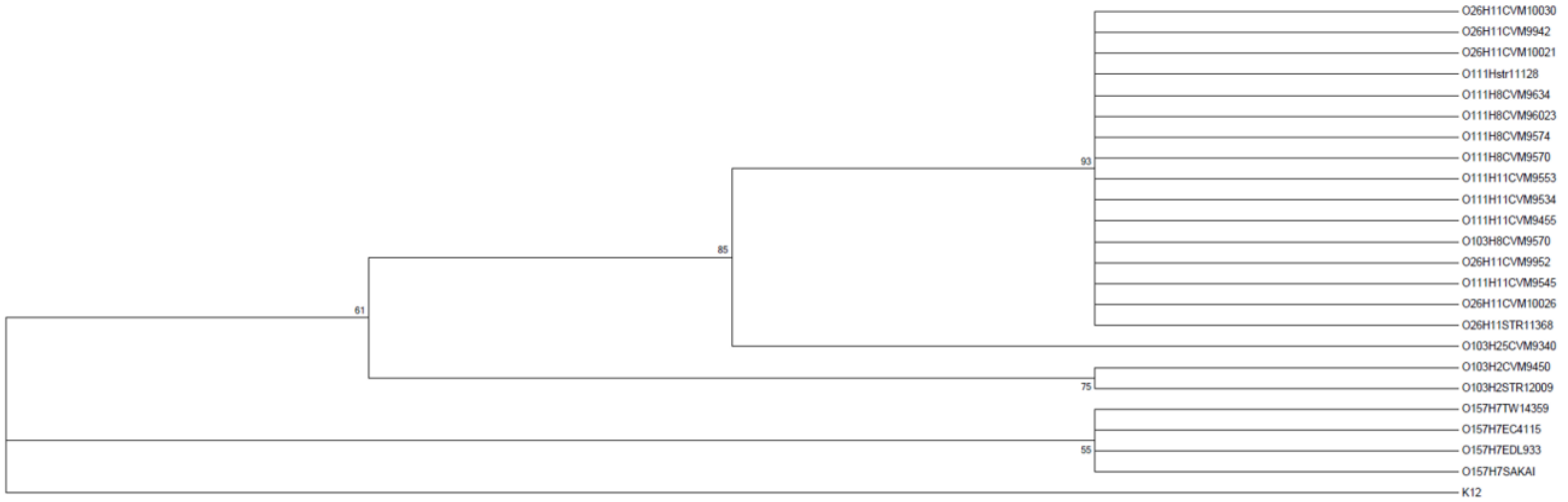
Left block represent CRISPR1 and right block represents CRISPR2 for the same strains in the same order. Only spacers are shown and are represented by colored squares. Same color/figure combination represents same nucleotide sequence. Spacers located to the right are older spacers, and spacers located to the left are newer spacers. Column labeled “L” indicates leader sequence position. Red strain names indicate anchor strain. Sequences were extracted using a proprietary macro designed by DuPont. Same software was used for the representation of spacers and repeats (28). Except by *E. coli* K12, all other 194 strains were STEC (*stx1*, *stx2* or *stx1* and *stx2* positive).

Figure III.2. Total spacer content depending on strain seropathotype, ability to cause outbreak, severe disease, and *stx* genes content.



Bars represent total spacer count (CRISPR1, CRISPR2a and CRISPR2b, CRISPR3-4, and CRISPR 3 and CRISPR4) by a) each of the five seropathotypes, b) potential ability to cause outbreak and c) potential ability to cause severe disease based on Karmali classification (4) and d) *stx* genes content. Lines represent ±2 standard error. Statistic tests revealed significant differences ($p<0.05$) between seropathotypes. Significant differences ($p<0.05$) were also detected between ability to cause outbreaks, severe disease and *stx* genes content.

Figure III.3. Maximum likelihood phylogenetic tree based on concatenated sequences of type I-E *cas* genes for 24 STEC.



Concatenated sequences of type I-E *cas* genes (*cas1, cas2, cas3', cse1, cse2, cas6e, cas7, cas5*) were obtained from a previous sequencing project (31) (16 strains) and publically available sequences for 8 *E. coli.* Maximum likelihood phylogenetic tree based crafted with Mega 5.1(32) with 1,000 bootstrap replications. *E. coli* K12 was selected as outgroup strain.

Table III.1: O group and H type, year of isolation and Origin of strains used in CRISPR study

Strain	O Group	H type	Year Isolation	Origin
1880	55	7	-	-
CVM_10001	26	11	1993	Human
CVM_10008	26	11	1993	Human
CVM_10014	26	11	1994	-
CVM_10017	26	11	1995	Cow
CVM_10040	26	11	1995	Human
CVM_10110	26	11	1997	Cow
CVM_10128	26	11	1997	Human
CVM_10129	26	11	1997	Human
CVM_9260	103	2	2000	Deer
CVM_9301	103	2	1987	Goat
CVM_9318	103	2	1990	Cow
CVM_9322	103	2	1990	Cow
CVM_9328	103	2	1992	Human
CVM_9380	103	2	1996	Human
CVM_9385	103	2	1996	Human
CVM_9397	103	2	1997	Cow
CVM_9398	103	2	1997	Cow
CVM_9400	103	2	1997	Cow
CVM_9410	103	2	1997	Cow
CVM_9415	103	2	1997	Cow
CVM_9422	103	2	1997	Cow
CVM_9434	103	-	1997	Cow
CVM_9439	103	2	1997	Human
CVM_9449	103	2	1999	Human
CVM_9453	103	2	1999	Human
CVM_9460	111	8	2000	Cow
CVM_9464	111	8	2000	Cow
CVM_9467	111	8	1976	Cow
CVM_9505	111	11	1987	Cow
CVM_9519	111	11	1987	Cow
CVM_9529	111	11	1987	Cow
CVM_9536	111	8	1988	-
CVM_9556	111	-	1990	Cow
CVM_9557	111	8	1990	Cow
CVM_9569	111	8	1993	Cow

Strain	O Group	H type	Year Isolation	Origin
CVM_9571	111	8	1993	Cow
CVM_9579	111	8	1993	Human
CVM_9585	111	8	1995	Cow
CVM_9587	111	8	1995	Cow
CVM_9591	111	11	1995	Cow
CVM_9592	111	8	1995	Human
CVM_9596	111	8	1996	Cow
CVM_9608	111	8	1997	Cow
CVM_9610	111	8	1997	Cow
CVM_9611	111	8	1997	Cow
CVM_9617	111	8	1997	Cow
CVM_9619	111	8	1997	Human
CVM_9647	111	8	1999	Cow
CVM_9650	111	8	1999	Human
CVM_9658	128	2	2000	Sheep
CVM_9659	128	2	2000	Sheep
CVM_9790	145	28	1995	Human
CVM_9818	111	-	1997	Cow
CVM_9892	26	11	2000	Cow
CVM_9935	26	11	1981	Antelope
CVM_9939	26	11	1982	Cow
CVM_9945	26	11	1984	Cow
CVM_9953	26	11	1985	Pig
CVM_9965	26	11	1988	Cow
CVM_9985	26	11	1991	-
CVM_9988	26	11	1992	Cow
CVM_9995	26	11	1993	Human
CVM_9997	26	11	1993	Human
CVM_9998	26	11	1993	Human
ESC0589	NT	-	2007	Produce
ESC0592	NT	-	2008	Produce
ESC0593	45	2	2008	Produce
ESC0599	NT	52	2004	Produce
ESC0601	121	19	2004	Produce
ESC0620	NT	-	-	Produce
K46A236a	8	30	-	-
K47A282a	8	2	-	-

Strain	O Group	H type	Year Isolation	Origin
MI4	45	2	-	-
SJ13	111	NM	-	-
SJ4	2	25	-	-
TW_7920	103	2	-	-
UMD131	OR	9	-	Human
UMD137	9	-	-	-
UMD141	26	11	-	Calf
UMD144	45	2	-	-
UMD146	55	7	-	-
UMD161	103	2	-	Human
UMD162	103	2	-	Steer
UMD163	103	2	-	Calf
UMD164	111	NM	-	Human
UMD168	111	NM	-	Calf
UMD204	91	21	-	Sheep
UMD206	5	NM	-	Sheep
WT119-2	91	14	2009	Pork
WT126-1	91	7	2009	Beef
WT380-3	NT	-	2009	Pork
WT396-5	91	-	2009	Pork
XDN4854	NT	10	2005	Beef
XDN5545	NT	7	2005	Beef
XDN5578	NT	46	2005	Beef
90.0327[a]	22	8	-	Cow
97.0077[a]	118	-	-	-
971[a]	26	11	-	-
1387[a]	111	8	-	Calf
1659[a]	157	7	-	Human
4162[a]	103	6	-	-
5645[a]	45	2	-	-
5906[a]	55	7	-	-
7865[a]	145	28	-	Human
8023[a]	121	19	-	-
493/89[a]	157	NM	1989	Human
5412/89[a]	157	NM	1989	Human
90-3158[a]	146	21	-	-
94-3024[a]	104	21	-	-

Strain	O Group	H type	Year Isolation	Origin
95-3322[a]	22	5	-	-
96-3305[a]	128	45	-	-
AA1[a]	174	8	-	-
CVM_9320[a]	103	11	1990	Cow
CVM_9338[a]	103	2	1993	Human
CVM_9353[a]	103	25	1995	Cow
CVM_9354[a]	103	25	1995	Cow
CVM_9530[a]	111	11	1987	Pig
CVM_9648[a]	111	8	1999	Cow
CVM_9652[a]	128	16	2000	Okapi
CVM_9653[a]	128	16	2000	Okapi
CVM_9763[a]	128	16	1999	Okapi
CVM_9785[a]	145	28	1995	Cow
CVM_9903[a]	26	11	1977	Cow
EC96012[a]	157	7	1996	Human
EC96038[a]	157	7	1996	Human
EC97144[a]	157	7	1997	Human
ESC0603[a]	36	14	2006	Produce
ESC0608[a]	73	18	2009	Produce
ESC0609[a]	116	21	2009	Produce
ESC0610[a]	113	36	2009	Produce
ESC0613[a]	168	8	2009	Produce
ESC0615[a]	113	21	2009	Produce
FDA413[a]	157	7	-	-
H1085C[a]	157	NM	2003	-
H2687[a]	157	NM	2003	-
H56909[a]	157	NM	1999	-
MI3[a]	45	2	-	-
SJ12[a]	103	11	-	-
SJ18[a]	121	17	-	-
SJ20[a]	128	2	-	-
SJ23[a]	145	NM	-	-
SJ29[a]	113	21	-	Human
SJ3[a]	26	11	-	-
SJ31[a]	113	21	-	-
SJ9[a]	45	2	-	-
TW_08087[a]	145	28	-	Human

Strain	O Group	H type	Year Isolation	Origin
TW_08868[a]	121	19	1999	Human
TW_7990[a]	103	2	-	-
UMD135[a]	5	NM	-	-
UMD142[a]	26	11	-	-
UMD145[a]	46	38	-	Beef
UMD159[a]	88	49	-	Beef
UMD160[a]	91	21	-	Cow
UMD165[a]	111	8	-	Calf
UMD166[a]	111	11		Calf
UMD167[a]	111	11	-	Calf
UMD170[a]	113	21	-	Human
UMD173[a]	125	NM	-	-
UMD217[a]	50	7	-	-
WT219-5[a]	91	14	2009	Pork
WT348-1[a]	91	14	2009	Pork
WT419-1[a]	91	14	2010	Pork
XDN11682[a]	83	8	2006	Beef
XDN15018[a]	15	27	2007	Beef
XDN15432[a]	83	8	2007	Beef
XDN20177[a]	8	16	2002	Beef
XDN23765[a]	NT	2	2003	Beef
XDN2688[a]	88	38	2004	Beef
XDN2746[a]	83	8	2004	Beef
XDN5789[a]	15	16	2005	Beef
CVM_9340[ab]	103	25	1993	Human
CVM_9534[ab]	111	11	1988	Cow
CVM_9574[ab]	111	8	1993	Human
EC4115[ab]	157	7	2006	Human
Sakai[acb]	157	7	1996	Human
TW14359[ab]	157	7	2006	Human
CVM_10021[b]	26	11	1995	Cow
CVM_10026[b]	26	11	1995	Cow
CVM_10030[b]	26	11	1995	Cow
CVM_10224[b]	26	11	1997	Human
CVM_9450[b]	103	-	1999	Human
CVM_9455[b]	111	-	2000	-
CVM_9545[b]	111	11	1988	Cow

Strain	O Group	H type	Year Isolation	Origin
CVM_9553[b]	111	11	1990	Cow
CVM_9570[b]	111	8	1993	Cow
CVM_9602[b]	111	8	1996	Human
CVM_9634[b]	111	8	1998	Cow
CVM_9942[b]	26	11	1983	Cow
CVM_9952[b]	26	11	1985	Cow
EDL933[b]	157	7	1982	Beef
11128[bc]	111	-	2001	Human
11368[bc]	26	11	2001	Human
12009[bc]	103	2	2001	Human
K12[bc]	OR	48	-	-

-: Data not available.
[a]: used for CRISPR and pathogenicity study
[b]: Used for *cas* genes sequencing study
[c]: Sequences obtained from NCBI

Table III.2. General characteristics of CRISPR arrays from *E. coli* (n=195)

Characteristic	CRISPR1	CRISPR2a	CRISPR2b	CRISPR3	CRISPR4	CRISPR3-4
Number of isolates with array	195	195	186	1	1	193
Number of Unique arrays	78	79	6	1	1	6
Spacers in array						
Range	1-20	1-20	0-1	11	6	1-13
Average	9	7	1	11	6	1
Mode	5	7	1	11	6	1
Total spacers	1612	1349	157	11	6	218
Different spacers	258	230	1	11	6	22
Unique spacers	128	123	0	11	6	15
Spacers Length (bp)						
Average	32	32	32	32	32	32
Min	31	30	32	32	32	28
Max	34	35	32	32	33	34
Protospacers detected	4	4	0	1	1	0

Table III.3**:** Spacers sequences with protospacers, location, hits and E-value.

Sequence	**Location (array)**	**Hits**	**Description**	**E-value**
CCAAATAGTCCCCAACATCCTGCCGTAAATAT	C1	CP003279.1	S. Typhi str. P-stx-12 plasmid	3.00E-08
		AM412236.1	S. Paratyphi A IncH1 multiple drug resistance	3.00E-08
		AF250878.1	S. Typhi R27 plasmid	3.00E-08
CCAGCCGTTCAGTATTGCCGGTGTCAGCAAAA	C1	JX182975.1	*Citrobacter freundii* plasmid pNDM-CIT.	3.00E-08
		CP003279.1	S. Typhi str. P-stx-12 plasmid	3.00E-08
GGTAAAAACACGGTCTGAACCGACATTCATGT	C1	AP010962.1	*E. coli* O111:H- str. 11128 plasmid pO111_2	3.00E-08
		AF503408.1	Enterobacteria phage P7	3.00E-08
GGTATGAATCCTTTCCCTGGTTTATTAAACGT	C1	CP003279.1	S. Typhi str. P-stx-12 plasmid	3.00E-08
		AP010961.1	*E. coli* O111:H- str. 11128 plasmid pO111_1	3.00E-08
		AM412236.1	S. Paratyphi A IncH1 multiple drug resistance, serovar Paratyphi A	3.00E-08
		DQ449578.1	*K. pneumoniae* strain NK245 plasmid pK245	3.00E-08
		AF239689.1	S. Typhimurium plasmid pSA1 RepE gene.	3.00E-08

Sequence	Location (array)	Hits	Description	E-value
		AF250878.1	S. Typhi R27 plasmid	3.00E-08
		AL513383.1	S. Typhi str. CT18 plasmid pHCM1	3.00E-08
		Y00547.1	Plasmid R27	3.00E-08
		CP000966.1	*K. pneumoniae* 342 plasmid pKP91	7.00E-06
CAGGTAAAACTGCCTGATTTCATCGTTCCAAG	C2	CP003279.1	S. Typhi str. P-stx-12 plasmid	3.00E-08
		AM412236.1	S. Paratyphi A IncH1 multiple drug resistance, serovar Paratyphi A	3.00E-08
		AF250878.1	S. Typhi R27 plasmid	3.00E-08
GATTGATACCGTCGGTCGGTTAGTGGATTACT	C2	CP003279.1	S. Typhi str. P-stx-12 plasmid	3.00E-08
		AP010961.1	*E. coli* O111:H- str. 11128 plasmid pO111_1 DNA	3.00E-08
		AB366440.1	S. Choleraesuis plasmid pMAK1 strain: L-2454	3.00E-08
		AM412236.1	S. Paratyphi A IncH1 multiple drug resistance	3.00E-08
		AF250878.1	S. Typhi R27 plasmid	3.00E-08
		AL513383.1	S. Typhi str. CT18 plasmid pHCM1	3.00E-08
GCTTAATAAACGAGGAGAATACCGATGTTTTT	C2	CP003279.1	S. Typhi str. P-stx-12 plasmid	3.00E-08

Sequence	Location (array)	Hits	Description	E-value
		AP010961.1	*E. coli* O111:H- str. 11128 plasmid pO111_1	3.00E-08
		AB366440.1	S. Choleraesuis plasmid pMAK1 strain: L-2454	3.00E-08
		AM412236.1	S. Paratyphi A IncH1 multiple drug resistance	3.00E-08
		AF250878.1	Salmonella Typhi R27 plasmid	3.00E-08
		AL513383.1	S. Typhi str. CT18 plasmid pHCM1	3.00E-08
TACCTGCTTAATACATTGCGGGTTAACGTTCT	C2	CP003279.1	S. Typhi str. P-stx-12 plasmid	3.00E-08
		AP010961.1	E. coli O111:H- str. 11128 plasmid pO111_1	3.00E-08
		FJ183736.1	S. Typhi plasmid IncHI1 HCM1.116 locus allele 2 genomic sequence	3.00E-08
		FJ183735.1	S. Typhi plasmid IncHI1 HCM1.116 locus allele 1 genomic sequence	3.00E-08
		AB366440.1	S. Choleraesuis plasmid pMAK1 DNA, complete genome, strain: L-2454	3.00E-08
		AM412236.1	S. Paratyphi A IncH1 multiple drug resistance	3.00E-08
		AF250878.1	S. Typhi R27 plasmid	3.00E-08
		AL513383.1	S. Typhi str. CT18 plasmid pHCM1	3.00E-08

Sequence	Location (array)	Hits	Description	E-value
CTGAACGTTGAAGAGTGCGACCGTCTCTCCTT	C3	JX997935.1	*E. coli* strain 3A11 plasmid pHN3A11	3.00E-08
		JX627737.1	*E. coli* plasmid pFOS-HK151325	3.00E-08
		JX486126.1	Uncultured bacterium plasmid pEFC36a	3.00E-08
		JX445149.1	S. Heidelberg plasmid pSH146_87	3.00E-08
		CP003290.1	*E. coli* O104:H4 str. 2011C-3493 plasmid pESBL-EA11	3.00E-08
		JF927996.1	*E. coli* plasmid pXZ	3.00E-08
		JQ432559.1	*E. coli* plasmid pHK23a	3.00E-08
		FR851303.1	*E coli* APEC strain 7122 (O78:K80:H9) plasmid pChi7122-2	3.00E-08
		JF274993.1	S. Typhimurium plasmid pSal8934a	3.00E-08
		JN232517.1	*E. coli* strain 7A8 plasmid pHN7A8	3.00E-08
		CP002732.1	*E. coli* UMNK88 plasmid pUMNK88	
		CP002731.1	*E. coli* UMNK88 plasmid pUMNK88_91	

Sequence	Location (array)	Hits	Description	E-value
TTAAAATAAATGCAACGGACAAAGAAGCCATT	C4	GU371926.1	*E. coli* plasmid pEC_B24, complete sequence	3.00E-08
		FJ664747.1	*E. coli* strain TA445 plasmid colBM colicin M gene cluster; cma gene, partial sequence; colicin B gene cluster, complete sequence; and colicin B activity	3.00E-08
		M16816.1	Plasmid pF166 (from E.coli) cba gene encoding colicin B activity	3.00E-08

Table III.4. Location of spacers with protospacers in STEC strains.

Strain	Serotype	CRISPR	Length of the array	Number Spacers with Protospacers	Location in the array (from leader sequence)			
90_0327	O22:H8	1	11	2	2	7		
95_3322	O22:H5	1	9	1	2			
		3	11	1	1			
		4	6	1	6			
ESC_0589	NT	1	7	2	2	3		
ESC_0608	O73:H18	1	12	1	2			
ESC_0613	O168:H8	1	7	1	2			
UMD_131	OR:H9	2	12	3	10	11	12	
XDN_11682	O83:H8	1	3	1	2			
XDN_15432	O83:H8	1	10	2	7	2		
XDN_23765	ONT:H2	1	6	1	2			
XDN_2746	O83:H8	1	3	1	2			
XDN_4854	ONT:H10	1	19	5	2	8	17	19
		2	8	1	8			
XDN_5545	ONT:H7	1	19	4	2	7	17	19
XDN_5578	ONT:H46	1	9	2	2	7		

Chapter 4: CRISPR ARRAY IN *Salmonella enterica* subspecies *enterica* (*S. enterica*).

4.1 Abstract

CRISPR arrays of 221 strains of *Salmonella enterica* subspecies *enterica* (*S. enterica*) of 53 serotypes were sequenced and analyzed for spacer content to define their relationship with serotype. In deep study of the CRISPR-cas system in 50 *S. enterica* serotype Bareilly (*S.* Bareilly) strains was performed to understand intra serotype variations. It was found that both CRISPR arrays (CRISPR1 and CRISPR2) correlate well with serotypes in *S. enterica*, although some serotypes displayed more than one type of array. Two types of CRISPR2 arrays in *S.* Bareilly were found, demonstrating intra serotype variation when increasing the sample size. Additionally, it was found that CRISRP-cas system reflects *S.* Bareilly phylogeny similar to MLST, notably *cas* genes, but a link between array information to food of isolation or geographical origin of the isolate was not found. In conclusion, CRISPR array would be useful for designing of molecular serotyping assays, but a wider range of strains should be added to the analysis to include the whole variation in *S. enterica*.

4.2 Introduction

Salmonella enterica subspecies *enterica* (*S. enterica*) causes over one million diseases every year in the U.S, and is one of the most prevalent foodborne pathogens in the world (1). Frequently associated to foodborne outbreaks (1), it has been isolated from diverse types of food and environments (2); serotypes Enteritidis, Typhimurium, Newport, and Javiana are the most frequently associated to illnesses in the US (3), but over 1500 *S. enterica* serotypes are described to date (4).

Serotyping is one of the most used tools in outbreaks investigations (5), however, traditional serotyping may take over five days (6), it is common to find untypable serotypes (7), and it requires the production of hundreds of antisera in laboratory animals (6). As an alternative, molecular typing methods deliver similar results in reduced time (6, 8, 9); therefore, the identification of new molecular markers is fundamental to develop new and improved typing methods which will help to improve public health (10).

Recently, a new potential marker has been described: CRISPR-cas systems. They are putative bacterial immune systems (11, 12) that consist of three parts; an array formed by a variable number of *repeats* (29 nt repeated sequences) intersperse with *spacers* (non-repeated short sequences), a group of genes codifying for CRISPR associated proteins (*cas*), and a *leader* sequence with the promoter for transcription (13). *Spacers* would be acquired from foreign genetic elements –such as plasmids and phages— and incorporated into the array after a failed infection (14, 15). They would work as probes that detect and prevent an infection with a previously known mobile genetic element. When a match is found, *cas* proteins will digest the invading

element, inactivating it (13, 15, 16). Since invasions and immunization events depend on the bacterium environment, spacer content should be variable depending on the ecological environment where a lineage evolved (17, 18). This characteristic could be used to differentiate isolates from different origin, or to describe different cell lineages (12, 19).

Some authors described *E. coli* and *Salmonella* CRISPR arrays as slow evolving systems (20), but other studies concluded CRISPR arrays could be potentially used for serotyping of Shigatoxin producing *Escherichia coli* (STEC) and *Salmonella* (21-23); however, the variation of CRISPR arrays across *Salmonella* serotypes is not fully understood. Also, CRISPR-cas system differences among strains of the same serotypes are not fully explored, especially for some emerging serotypes such as *Salmonell*a *enterica* subspecies *enterica* serotype Bareilly (*S.* Bareilly) which recently reached the CDC top 15 *Salmonella* serotypes that most frequently cause human disease (3).

The hypothesis in this study were that CRISPR arrays differ between serotypes, and that variations might also be found among strains of the same serotype. Also, that different components of the CRISPR-cas system might reflect the phylogeny of a strain in a serotype. To answer these questions, CRISPR arrays for 53 different *Salmonella* serotypes were sequenced and characterized, and the CRISPR-cas system of 50 *S.* Bareilly strains were studied.

4.3 Objective

To characterize *S. enterica* CRISPR's arrays in selected serotypes and the CRISPR system of *S.* Bareilly.

4.4 Materials and Methods

4.4.1. Strain collection

A total of 221 *S. enterica* were selected from FDA *Salmonella* culture collection. Strains from 53 different serotypes were used to study CRISPR array inter serotypes differences (Table IV.1).

To study CRISPR-cas systems intra serotype variation, 50 *S.* Bareilly strains with different PFGE pattern were used. Strains were isolated in different dates, sources and geographic locations (Table IV.2).

4.4.2. DNA isolation

DNA was isolated using InstaGene matrix (Bio-Rad, Hercules, CA) following manufacturer instructions. Briefly, 1 to 2 colonies were suspended in 1 ml of ultrapure water. Suspension was centrifuged and supernatant discarded. Then, 200ul of Instagene matrix were added, and the suspension was incubated first at 56°C for 30 minutes, and then at 100°C for 8 minutes. The suspension was centrifuged, and the supernatant, containing DNA, was extracted and frozen until use.

4.4.3. PCR and DNA sequencing *Salmonella* enterica serotypes

CRISPR1 and CRISPR2 were amplified and sequenced using previously described primers (24) (Table IV.3). PCR conditions included an initial denaturation at 95°C for 10 minutes and 28 cycles of denaturation at 94°C, annealing at 55°C, and 1:30 sec of elongation at 72°C, plus a final elongation at 72°C for 15 minutes, and infinite hold at 4°C.

PCR products were sequenced by MCLAB (San Francisco, California) with the same group of primers (Table IV.3). Additionally, internal primers were designed to sequence longer DNA arrays for the following serotypes: Bareilly, Braenderup, Heildelberg, Montevideo, Muenster, Newport, Seftenberg, Tennessee, and Typhimurium (Table IV.3).

4.4.4. CRISPR array detection, extraction and analysis

CRISPR array sequences were assembled from at least one forward and one reverse read using Geneious v.6.15 created by Biomatters. Additional internal readings were used for some isolates.

CRISPR sequences were individually extracted using a Macro tool designed by DuPont. The program identifies repeats and spacers, and also creates a graphic representation of spacers and repeats as previously described (25).

CRISPR arrays were manually organized in clusters based on the presence of common consecutive repeats and spacers from the distal end to the leader sequence. Anchor strains –strains with the longest series of spacers and arrays in a group of strains— were used as guide for array organization.

4.4.5. Protospacer analysis

Each one of the spacers detected was compared against NCBI phages and plasmids database (January 2013) with the BLAST plug in utility from Geneious v6.15 (26). Protospacers were homologous sequences with an e value <1.10e-5 and less than 10% difference in sequence length (27). Matches to CRISPR sequences were omitted.

4.4.6. I-E *cas* system analysis of *Salmonella* Bareilly

I-E *cas* genes from *S.* Bareilly were extracted from whole genome sequences data provided by FDA. Genes were extracted using the BLAST add on installed in Geneious 6.5.5 (26). *cas* genes were concatenated in the following order: *cas1, cas2, cse3, cse5e, cse4, cse2, cse1* and *cas3*; low quality and partial sequences were excluded from the analysis. Then, sequences were aligned with Clustal W (28) plug in for Geneious with default options, and a consensus phylogenetic tree was crafted using Jukes Kantor model with the Neighbor-joining method, with 1,000 replications. *S.* Typhi CT18 was selected as outgroup. Additionally, leader sequences for both CRISPR arrays were extracted and analyzed. Alignment and phylogenic tree were constructed using the same parameters described above.

4.4.7. *Salmonella* Bareilly phylogenetic tree with Multilocus Sequence Typing (MLST) genes.

To understand the relationship between the CRISPR-cas system and phylogenetic evolution of the strains, housekeeping genes for *Salmonella enterica* (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*), which reflect phylogeny (29), were extracted from the whole genome sequence and concatenated using Geneious. Sequences were aligned with Clustal W (28) plug in for Geneious with default options, and a consensus tree was constructed with 1,000 bootstrap iterations, and *S.* Typhi CT18 was used as outgroup.

4.4.8. *S.* Bareilly antimicrobial resistance test

Antimicrobial minimum inhibitory concentrations (MIC) of *Salmonella* Bareilly isolates were determined via Sensitre automated microbial susceptibility system (Trek Diagnostic system, Westlake, OH) and interpreted according to the CLSI MIC Interpretative standards. *Escherichia coli* ATCC 35218, *Enterococcus faecalis* ATCC 29212, *Staphyloccoccus aureus* ATCC 29213 and *Pseudomonas aeruginosa* ATCC 27853 were used as controls in antimicrobial MIC determinations. The following antimicrobials were tested: Amikacin (Ami), Amoxicillin/clavulanic acid (Aug), Ampicillin (Amp), Cefoxitin (Fox), Ceftiofur (Tio), Ceftriaxone (Axo), Chloramphenicol (Chl), Ciprofloxacin (Cip), Gentamicin (Gen), Kanamycin (Kan), Nalidixic acid (Nal), Streptomycin (Str), Sulfamethoxazole (Smx), Tetracycline (Tet), and Trimethoprim/sulfamethoxazole (Cot).

4.5 Results

In this study, CRISPR arrays of 221 *Salmonella enterica* strains of 53 serotypes were described, and CRISPR-cas system of 50 strains of *S.* Bareilly were characterized.

4.5.1 CRISPR array in *Salmonella sp*.

Arrays CRISPR1 and CRISPR2 were present in all 241 screened strains, as previously described (20, 30). Arrays were manually organized based on the presence of common, consecutive spacers. As a consequence, organization based on CRISPR1 also organized arrays in CRISPR2 (Figure IV.1).

CRISPR1 displayed longer arrays than CRISPR2 (Fig. IV.1). *S. enterica* serotype Muenster CRISPR1 array reached 62 spacers, while the longest CRISPR2 array only had 37 spacers (*S. enterica* serotype Kentucky) (Fig. IV.1). Moreover, the average length of CRISPR1 was longer than CRISPR2, although array length and content varied depending on serotypes (Fig. IV.2, Fig. IV 3, and Table IV.4).

A universal anchor strain in CRISPR1 or CRISPR2 was not identified, but some arrays could be used as anchors for some clusters. In general, same serotype strains clustered together, and their arrays differed only based on spacer/repeats deletions but not in their sequences; For example, every *S. enterica* serotype Agona strain had exactly the same sequence in CRISPR1, and a half of strains lacked only one spacer in CRISPR2 (4/7). Most *S. enterica* serotype Montevideo were identical in CRISPR1 (18/22) and 21/22 were identical in CRISPR2 as well. On the other hand, some serotypes presented more than one type of array that did not arrange together: *S.*

enterica serotype Newport presented 2 types of arrays that did not share any spacer in CRISPR1 or CRISPR2 (Fig IV.1), as well as *S. enterica* serotype Muenchen (Fig. IV. 1).

Strains of different serotypes generally displayed different arrays, often forming individual clusters; however, some serotypes also shared some spacers among them; in CRISPR2 *S. enterica* serotypes Agona, Abony and Tallahassee shared the two older spacers, and their first spacer was also present in *S. enterica* serotype Tennessee. *S. enterica* serotypes Braenderup, Virchow, and Anatum shared two ancestral spacers, and the former also had five other spacers in common. (Fig IV.1). A cluster of strains formed by *S. enterica* serotypes Newport, Typhimurium, and Heildelberg shared the three older spacers, and additionally, they also had other common spacers in different positions, indicating deletions of spacers.

Although CRISPR1 co-clustered with CRISPR2, organization deferred some times. For example, *S. enterica* serotype Tallahassee did not share any spacer with *S. enterica* serotype Agona and Abony in CRISPR1, and *S. enterica* serotype Newport displayed totally different arrays from *S. enterica* serotypes Typhimurium and Heidelberg in CRISPR1 (Fig IV.1).

4.5.2 CRISPR array in *Salmonella* Bareilly.

S. Bareilly CRISPR arrays were organized in two different groups. CRISPR1 showed almost the same spacers in the same order for all the strains, but some differences were observed due to spacer deletion. Oppositely, CRISPR2 arrays organized in two well distinct clusters that did not share any spacer between them.

Strain SAL2887 was identified as an anchor strain for group A, but an anchor strain for group B could not be identified since 24/26 strains displayed identical arrays (Fig. IV.4). When organizing strains based on CRISPR2 arrays, almost every group B strain lacked the last, newer 4 spacers in CRISPR1 (23/24), and those strains also presented a difference sequence in the first, more ancestral repeat (Fig IV. 4). Interestingly, strain SAL2890 presented pattern more similar to group A in CRISPR2 while grouped with cluster B based on CRISPR 1 organization (Fig IV.4).

A relationship between the type of array and other strain's characteristics such as type of food the strain was isolated from or geographical origin was not identified. Moreover, antimicrobial resistance test indicated all *S.* Bareilly strains were susceptible to every antimicrobial tested; therefore, a potential relationship between CRISPR and this characteristic was not suitable for study.

4.5.3. Spacers diversity

A total of 6744 spacers in *Salmonella enterica* CRISPR arrays were found; 19% of spacers were different, and 8.9% were present only once (Table IV.3).

Spacer's average length was 32 bp in both arrays, but CRISPR 1 spacers had more range of variation (30 to 44 bp) than CRISPR2 (Table IV. 3). Interestingly, six spacers were present in both loci. Protospacers for 20 spacers were found; all of them were originated from phage.

Strains of *Salmonella* Bareilly displayed over 1800 spacers, but only 55 were distinct and none of them was unique. The average length of spacers was 32 bp with

only 1 bp difference between the longest and the shortest spacer, and we detected only one protospacer, which had identity with a phage (Data not shown).

Some strains also presented some insertion sequences that were not identified as spacers in their arrays; strains *S. enterica* serotypes Typhimurium and Heildelberg presented a 13 nt sequence in the middle of CRISPR1 arrays (position 10, aprox), and some of the *S. enterica* serotype Paratyphi B strains displayed the same insertion. *S. enterica* serotype Kentucky displayed a 50 nt insertion in position 7 out of 20 spacers also in CRISPR1. No insertions were detected in CRISPR2, however, a G residue intercalated 2 repeats in some *S. enterica* serotype Newport strains (Figure IV.1).

4.5.4. *S.* Bareilly phylogeny based on Multi locus sequence type genes and CRISPR system.

Concatenated sequences of *S.* Bareilly housekeeping genes (*aroC, dnaN, hemD, hisD, purE, sucA*, and *thrA*) arranged in two different clusters (Fig. IV.5). Seventy 70 SNPs were identified which were present across all genes sequences with frequencies of 6 (*aroC*), 13 (*dnaN*), 7 (*hemD*), 2 (*hisD*), 28 (*purE*), 8 (*sucA*), and 6 (*thrA*) (data not shown).

S. Bareilly concatenated *cas* genes (*cas1, cas2, cse3, cse5e, cse4, cse2, cse1* and *cas3)* clustered in a two-branch phylogenetic tree (Fig. IV.6) as well as the leader sequence for array CRISPR2 (Fig IV.7). *cas* genes for 39 out of the 41 analyzed strains presented only 6 SNPs among them. However, two strains (2917 and 2923) displayed low conservation areas in *cas2*. Phylogenetic tree crafted with data from

CRISPR1 leader sequence (Fig IV.8) showed polytomy due to a 100% of sequence conservation among *S.* Bareilly strains, and no SNPs were detected.

Agreement between housekeeping genes and CRISPR systems clusters was correct in 46 out of the 48 cases analyzed; strains clustering in MLST group A also grouped together in *cas* genes, leader sequence2 and CRISPR2 groups A. *Salmonella* Bareilly strains 2890 and 2892 showed disagreement in their grouping: Strain 2890 classified as group A for housekeeping genes, but it did not group with other strains in by *cas* genes tree, and displayed the same distance from both groups However, this strain CRISPR array clustered in group A and leader sequence group A. Interestingly, strain 2892 grouped in cluster A for housekeeping genes, but its complete CRISPR-cas system grouped into cluster B.

4.6 Discussion

In the present work, it was found that CRISPR arrays of *S. enterica* strains correlate well with their serotype, and that some serotypes present more than one type of array. As a consequence, it was confirmed that these sequences are useful to design molecular serotyping techniques. Additionally, it was concluded that *S.* Bareilly phylogeny is reflected by their CRISPR-cas system. Notably, *cas* genes cluster in a similar fashion that housekeeping genes do. This information could help to understand the evolution of the CRISPR-cas system in *S.* Bareilly and be used as an example for the study of the system in other *Salmonella* serotypes. The composition of the CRISPR array was not linked to geographical regions or type of food the *S.* Bareilly strain was originated. These, plus the low rate of variation in arrays, made us

to conclude that CRISPR arrays do not provide enough information for foodborne outbreaks tracking. *S.* Bareilly did not display antimicrobial resistance. This is the first report on *S.* Bareilly strains diversity.

Other studies have described that CRISRP arrays in *Salmonella enterica* are highly conserved at serotype level (23), and therefore they can be used to detect, identify and distinguish *Salmonella* serotypes; however, information in the loci is not enough to differentiate at the strain level (23, 31). Previously, Lui *et al.* (24) analyzed 171 strains from the nine most common *S. enterica* serotypes and concluded that it is necessary to include additional genetic markers to differentiate *S. enterica* to the strain level. This study included a more complete range of serotypes (over 200 strains from 53 *S. enterica* serotypes), and as a result, it was found a more diverse range of arrays. Liu *et al*. found arrays ranging from 2 to 25 spacers in CRISPR1 and CRISPR2. Array length reached of up to 88 spacer's (Figure IV.1). Also, they found 166 and 188 unique spacers in their collection for CRISPR1 and CRISPR2 respectively, and over 300 unique spacers were found in both loci. It is expected that increasing the range of strains included in the analysis will raise a more diverse arrays and spacer's variety. Fricke *et al.* (30) also concluded that array content correlated well with serotypes, and that phylogenetic distances in *Salmonella* are partially reflected in the CRISPR array: they described that *S. enterica* serotype Newport and *S. enterica* serotype Saintpaul strains presented more than one type of arrays. *S. enterica* serotype Newport and *S. enterica* serotype Muenchen presented two types of array; different lineages have been previously reported based on MLST and whole genome sequencing in *S. enterica* serotype Newport (32, 33); however, *S. enterica*

serotype Muenchen phylogenetic diversity have not been previously reported. As a conclusion, CRISPR arrays could be used to differentiate strain serotype, however, a larger database is necessary to include a higher range of intra serotype variation.

The diversity of *S.* Bareilly has not been addressed in literature yet. This work was focused on their CRISPR-cas system to determine whether it helps to discriminate between strains with different PFGE profiles (data not shown). CRISPR1 presented only one sequence patterns, while CRISPR2 showed two different profiles; therefore, CRISPR arrays were not as discriminating as PFGE. Moreover, it was discovered CRISPR array profiles could not be associated to the isolate geographical origin or food of recovery. The relationship between other CRISPR-cas system components and strain phylogeny as defined by housekeeping genes was also studied (29). *cas* genes and CRISPR2's leader sequence presented a phylogenetic organization similar to that presented by housekeeping genes. Since CRISPR1's leader sequence and array were non variable, they were not suitable to define phylogenies.

Di Marzio *et al.* (34) studied the relationship between antibiotic resistance in *Salmonella* Typhimurium and CRISPR sequence. They found an association between different types of CRISPR-MLVA alleles for some antimicrobial resistance phenotypes, but the evolution rate of CRISPR did not evolve as fast as antimicrobial resistance. They conclude that antimicrobial resistance is not directly mediated by CRISPR. In this study, all the strains were susceptible to all the antimicrobials tested, therefore, the relationship could not be analyzed. Studies in *S.* Bareilly antimicrobial resistance show inconsistent results. *S.* Bareilly isolated from imported foods showed

resistance to antimicrobials (Ampicillin, Chloramphenicol, Sulfamethoxazole, and Tetracyclin) (35), and it showed high levels of resistance against Tetracycline (83.3%), Neomycin (50%) and Streptomicyn (16.7%) when isolated from pasture raised chicken in the US (36). In India, seafood isolates from 2003 to 2007 of *S.* Bareilly were susceptible to all antimicrobials tested (Ampicillin, Nalidixic acid, Ciprofloxacin, Chloramphenicol, Gentamicyn, Sulfamethizol, Cephalexin, Streptomycin, Kanamycin, Oxytetraciclyn and Carbenicillin) (37), and plasmid related to antimicrobial resistance were not detected in these strains. Similarly, studied strains did not display antimicrobial resistance to any antimicrobial tested. Consistently with these results, most of the samples came from south Asian countries (43/50), explaining at least partially these results. It would be interesting to study the genetic make-up of these strains, especially for plasmids contain to understand why these strains do not display antimicrobial resistance.

The presence of serotypes with more of one CRISPR array pattern may raise questions about serotyping misclassification. In a simultaneous study, many studied strains were subject to whole genome sequencing, and *S. enterica* serotype Muenchen strains demonstrated that they were correctly classified at a whole genome level (data not shown). On the other hand, serotypes with only one type of array might be the result of a highly skewed sample: most serotypes were underrepresented with only one strain in the analysis. Other authors have also found one general CRISPR array pattern for each serotype, but it cannot be discarded that those results are due to an incomplete sampling. Based on the findings in *S.* Bareilly CRISPR arrays, future studies should focus on adding to the CRISPR database a variety of strains from

different serotypes but also from different geographic areas, foods origin, and hosts to include the whole range of array variation to create am more complete database. This is of special importance in the current, globalized commerce.

The use of CRISPR array content as the only information piece for outbreak investigation does not seem suitable for *S. enterica*. Liu *et al.* (31) developed a system adding two virulence genes and achieve enough differentiation. Similarly, it would be interesting to determine whether some housekeeping genes could provide such as discrimination: Gene *purE* showed 26 SNPs in a 399 bp. Fragment, and sequence *dnaN* presented 10 SNPs in 500 bp sequence. These two genes are potential candidates to evaluate a system that, using CRISPR array information, will accurate subtype *S enterica* serotypes to the strain level.

4.7 Conclusion

The conclusion of this study indicates that CRISPR arrays reflect serotype classification of *S. enterica*, and therefore would be useful for designing of molecular serotyping assays. However, a wider range of strains should be added to the analysis to include the whole variation in *S. enterica*. Also, it was determined that CRISRP-cas system reflects *S.* Bareilly phylogeny at a level similar to MLST. However, a better understanding of the correlation between CRISPR array and whole genome sequence is needed to decide if CRISPR is useful to study *S. enterica* phylogeny.

4.8 References

1. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.
2. **Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ.** 2013. Outbreak-associated *Salmonella enterica* serotypes and food Commodities, United States, 1998-2008. Emerg Infect Dis **19:**1239-1244.
3. **[CDC] Centers for Disease Control and Prevention.** 2012. Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet Surveillance Report for 2011 (Final Report). Available in: http://www.cdc.gov/foodnet/PDFs/2011_annual_report_508c.pdf, accessed on January 2014.
4. **Grimont PA, Weil F-X.** 2007. Antigenic formulae of the *Salmonella* serovars, 9th Edition ed, Paris, France: WHO Collaborative Centre for Reference and Research on *Salmonella*.
5. **Boyle EC, Bishop JL, Grassl GA, Finlay BB.** 2007. *Salmonella*: from pathogenesis to therapeutics. J Bacteriol **189:**1489-1495.
6. **McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML, Fields PI.** 2011. Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. J Clin Microbiol **49:**565-573.
7. **[CDC] Centers for Disease Control and Prevention.** 2011. National Enteric Disease Surveillance: *Salmonella* Surveillance Overview. Available in:

http://www.cdc.gov/nationalsurveillance/PDFs/NationalSalmSurveillOverview_508.pdf, accessed January 2014.

8. **Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P.** 2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol **45:**3323-3334.

9. **Shi C, Singh P, Ranieri ML, Wiedmann M, Moreno Switt AI.** 2013. Molecular methods for serovar determination of *Salmonella*. Crit Rev Microbiol. [Epub ahead of print, Posted online on November 14, 2013]. Available in: http://informahealthcare.com/doi/abs/10.3109/1040841X.2013.837862, accessed January 2014.

10. **Wattiau P, Boland C, Bertrand S.** 2011. Methodologies for *Salmonella enterica* ssp *enterica* subtyping: gold standards and alternatives. Appl Environ Microbiol **77** (22):7877-7885.

11. **Barrangou R.** 2013. CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip Rev RNA **4:**267-278.

12. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167-170.

13. **Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321:**960-964.

14. **Bolotin A, Quinquis B, Sorokin A, Ehrlich SD.** 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151:**2551-2561.

15. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

16. **Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ.** 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol **18:**529-536.

17. **Tyson GW, Banfield JF.** 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol **10:**200-207.

18. **Karginov FV, Hannon GJ.** 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell **37:**7-19.

19. **Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV, Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R.** 2008. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. PLoS One **3:**e2652.

20. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.

21. **Delannoy S, Beutin L, Fach P.** 2012. Use of CRISPR sequence polymorphisms for specific detection of enterohemorrhagic *E. coli* (EHEC) strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28 and O157:H7 by real time PCR. J Clin Microbiol **50** (12): 4035-4040.

22. **Yin S, Jensen MA, Bai J, Debroy C, Barrangou R, Dudley EG.** 2013. The evolutionary divergence of Shiga Toxin-Producing *Escherichia coli* is reflected in Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Spacer Composition. Appl Environ Microbiol **79:**5710-5720.

23. **Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, de Romans S, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth-Jeanjean S, Achtman M, Brisse S, Sola C, Weill FX.** 2012. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. PLoS One **7:**e36995.

24. **Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG.** 2011. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella* enterica subsp. enterica. Appl Environ Microbiol **77:**1946-1956.

25. **Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R.** 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol **190:**1401-1412.

26. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

27. **Touchon M, Charpentier S, Clermont O, Rocha EP, Denamur E, Branger C.** 2011. CRISPR Distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. J Bacteriol **193:**2460-2467.

28. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-2948.

29. **Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A, Dougan G, Harrison LH, Brisse S, Group SEMS.** 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog **8:**e1002776.

30. **Fricke WF, Mammel MK, McDermott PF, Tartera C, White DG, Leclerc JE, Ravel J, Cebula TA.** 2011. Comparative Genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage vvolution. J Bacteriol **193:**3556-3568.

31. **Liu F, Kariyawasam S, Jayarao BM, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG.** 2011. Subtyping *Salmonella enterica* serovar Enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). Appl Environ Microbiol **77:**4520-4526.

32. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8:**e55687.

33. **Harbottle H, White DG, McDermott PF, Walker RD, Zhao S.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization *of Salmonella enterica* serotype Newport isolates. Journal of Clinical Microbiology **44:**9.

34. **Dimarzio M, Shariat N, Kariyawasam S, Barrangou R, Dudley EG.** 2013. Antibiotic resistance in *Salmonella* Typhimurium associates with CRISPR sequence type. Antimicrob Agents Chemother. **57** (9): 4282-4289.

35. **Zhao S, McDermott PF, Friedman S, Qaiyumi S, Abbott J, Kiessling C, Ayers S, Singh R, Hubert S, Sofos J, White DG.** 2006. Characterization of antimicrobial-resistant *Salmonella* isolated from imported foods. J Food Prot **69:**500-507.

36. **Melendez SN, Hanning I, Han J, Nayak R, Clement AR, Wooming A, Hererra P, Jones FT, Foley SL, Ricke SC.** 2010. *Salmonella enterica* isolates from pasture-raised poultry exhibit antimicrobial resistance and class I integrons. J Appl Microbiol **109:**1957-1966.

37. **Kumar R, Surendran PK, Thampuran N.** 2009. Analysis of antimicrobial resistance and plasmid profiles in *Salmonella* serotypes associated with tropical seafood of India. Foodborne Pathog Dis **6:**621-625.

4.9 Figures and Tables

Figure IV.1a. CRISPR arrays of *Salmonella enterica* serotypes.

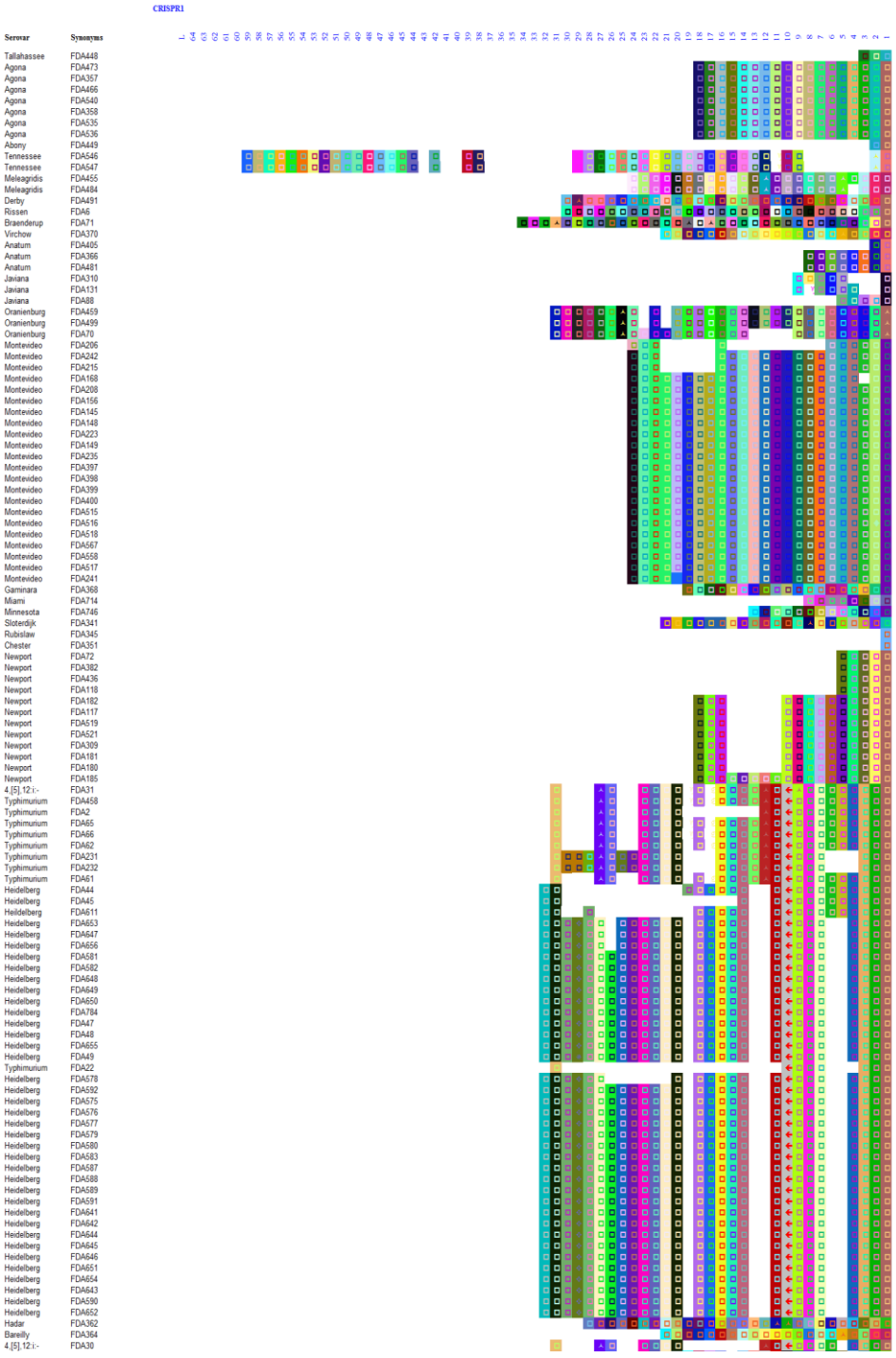


Figure IV.1b. CRISPR arrays of *Salmonella enterica* serotypes (cont)

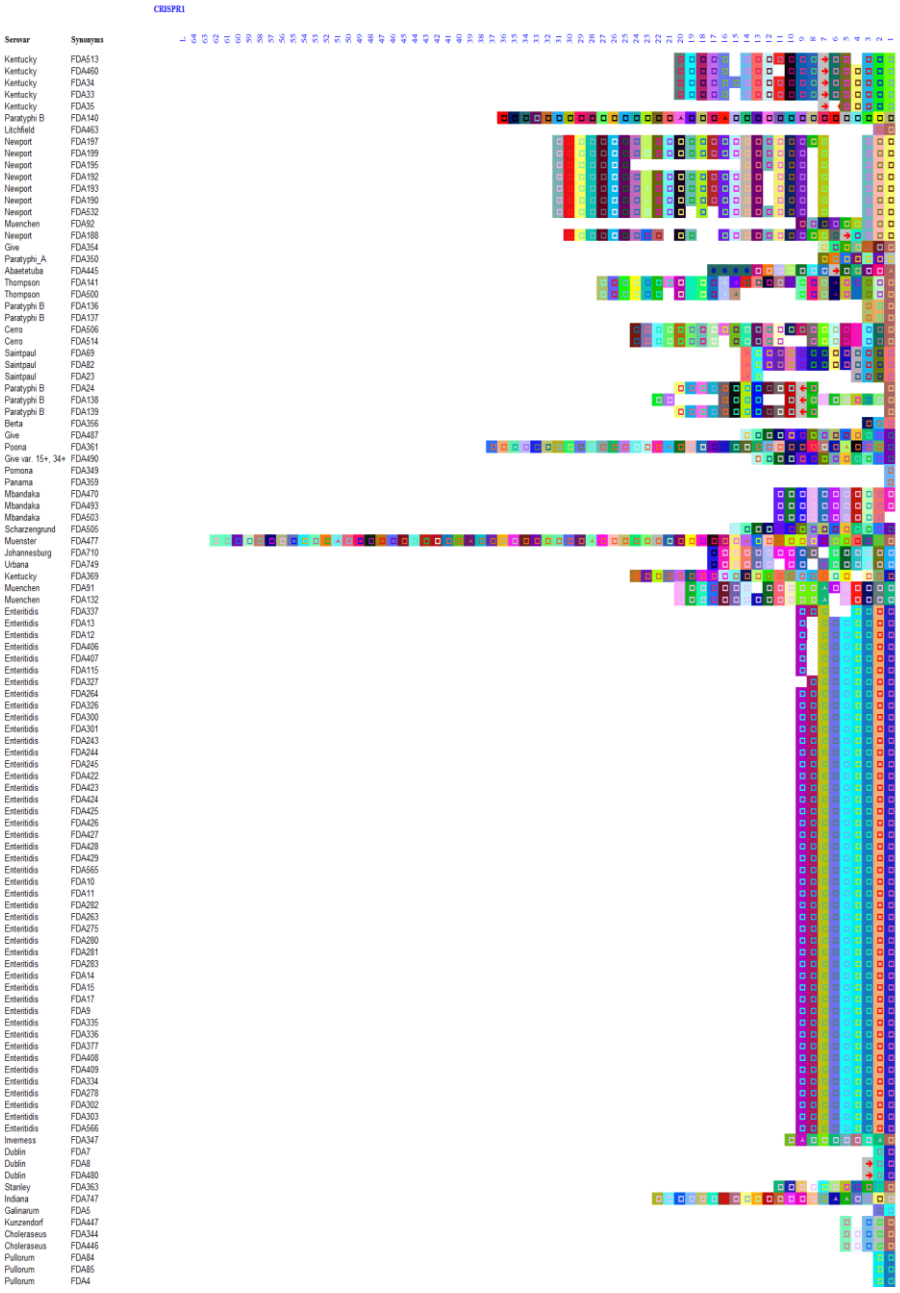


Figure IV.1c. CRISPR arrays of *Salmonella enterica* serotypes (cont)

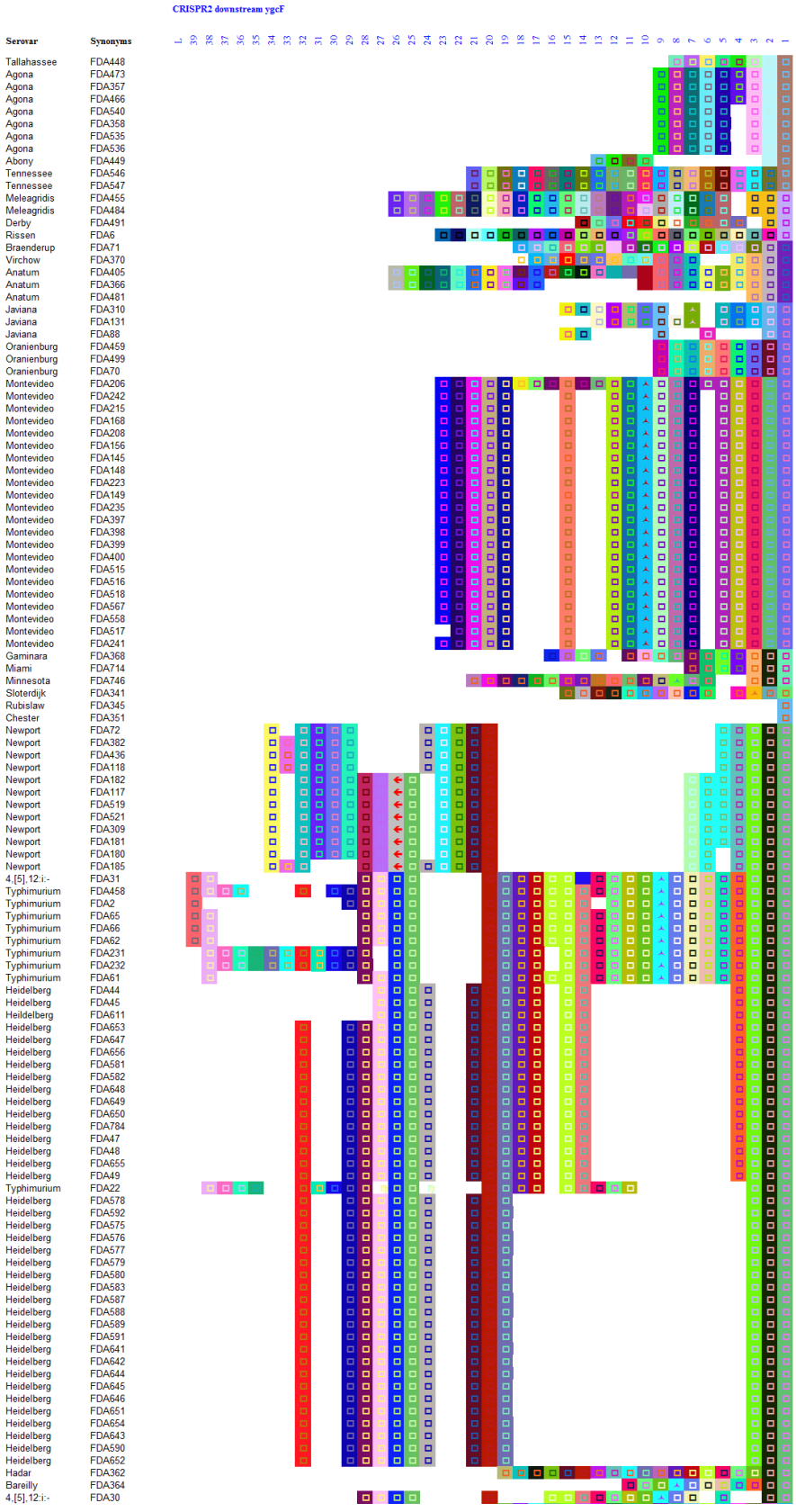


Figure IV.1d. CRISPR arrays of *Salmonella enterica* serotypes (cont).

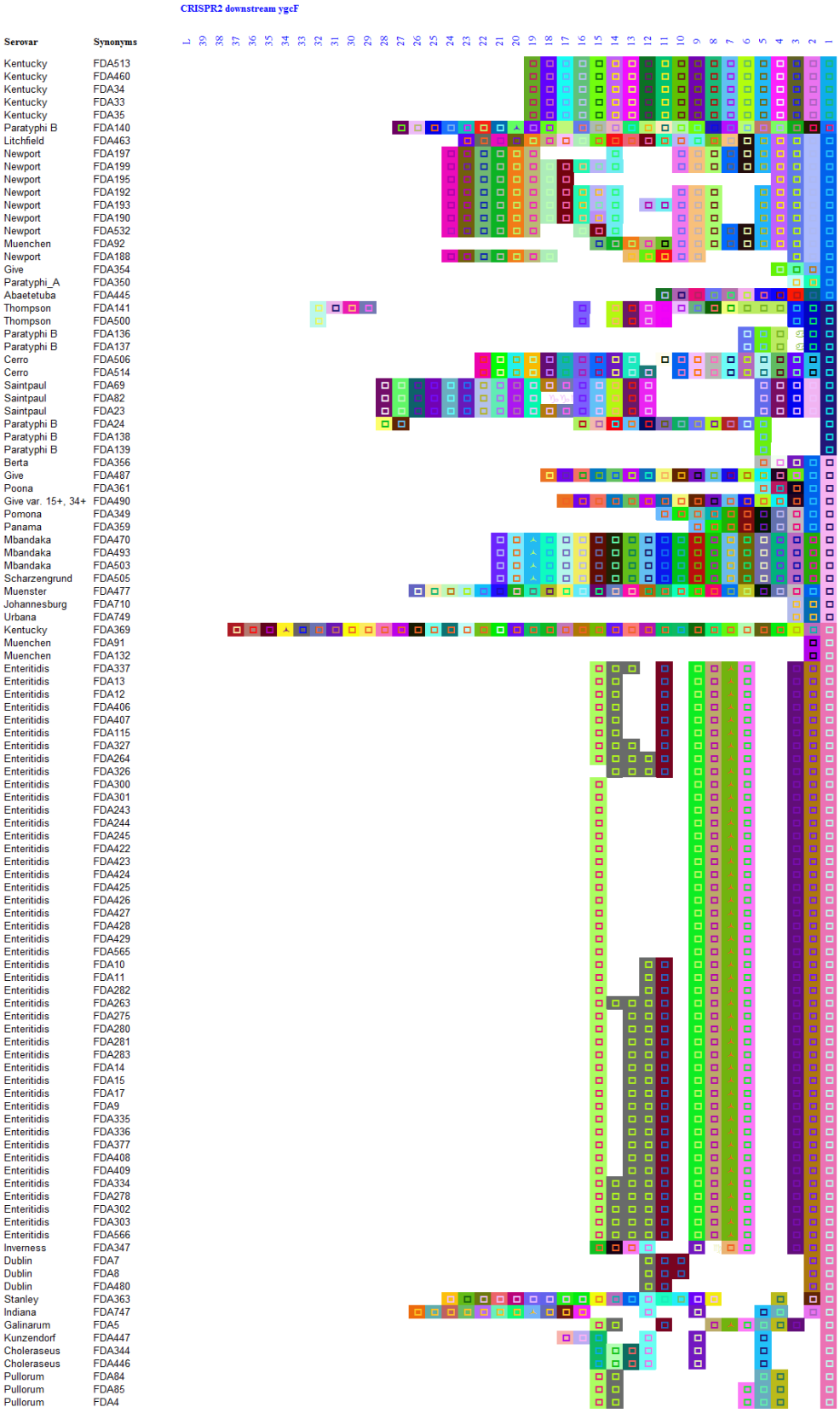
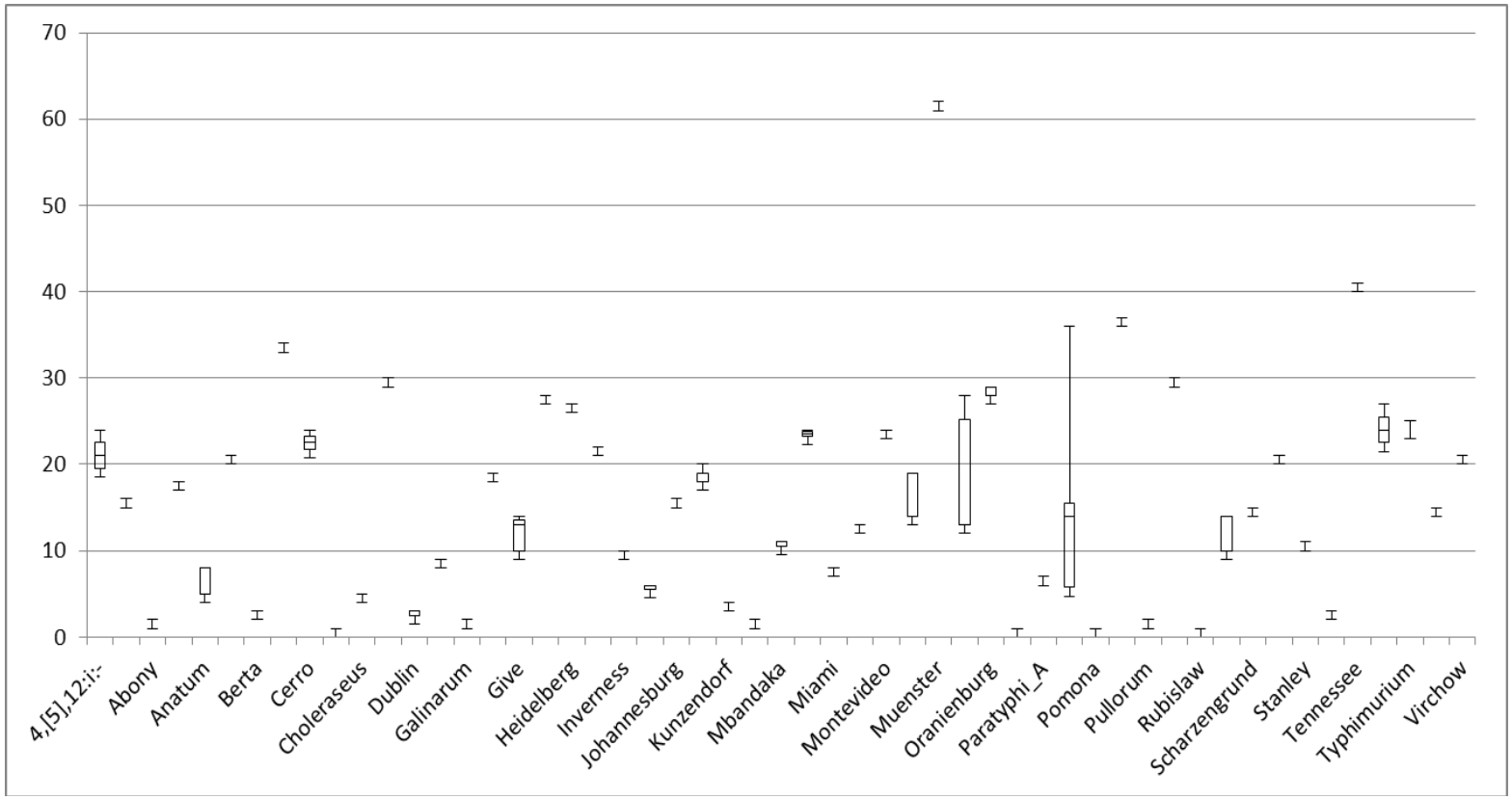
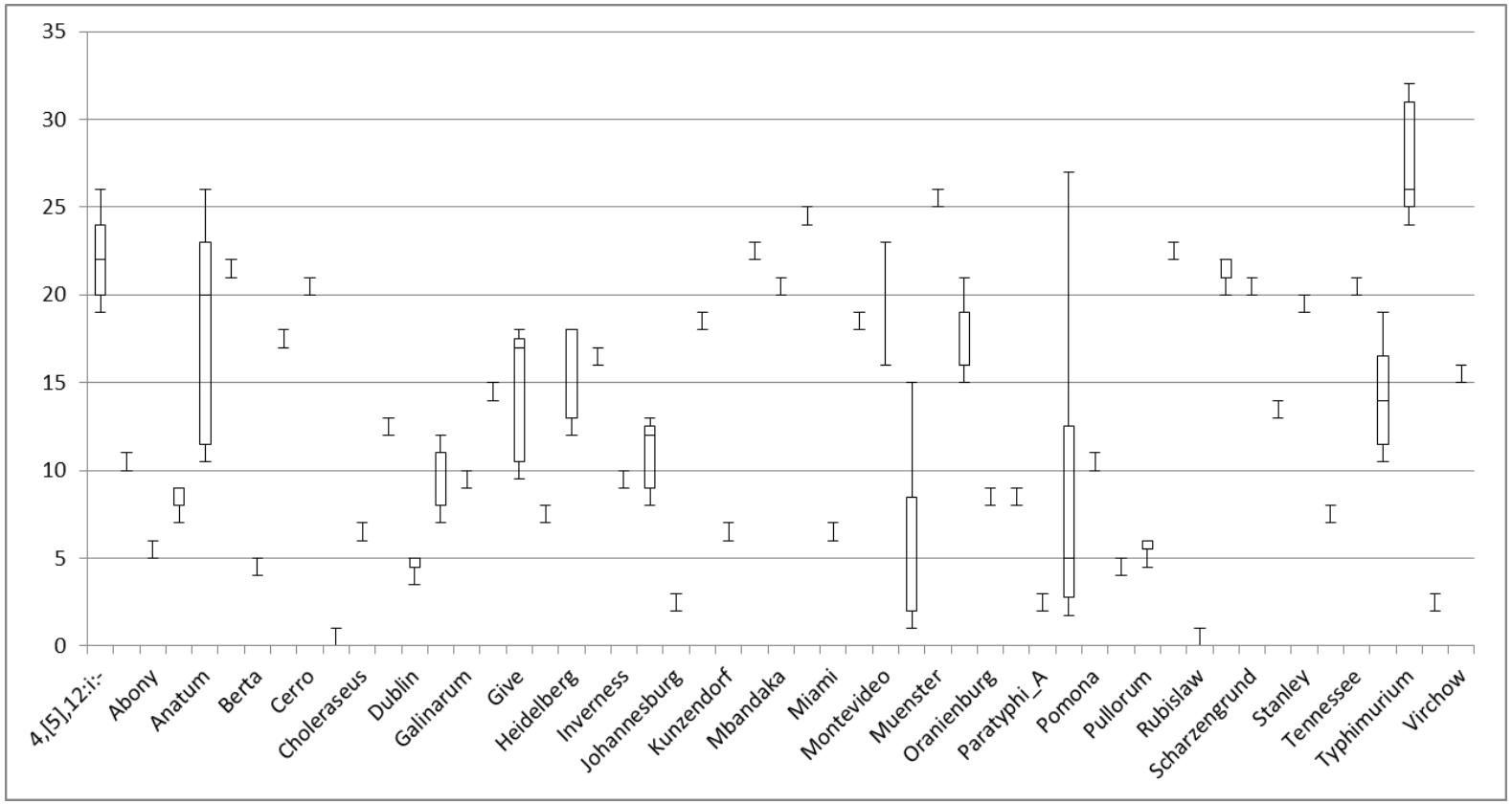


Figure IV.2. CRISPR 1 array length (spacers) by serotype in *Salmonella enterica.**



*Based on Pettengill *et al.*, not yet published.

Figure IV.3. CRISPR 2 array length (spacers) by serotype in *Salmonella enterica**.



Based on Pettengill *et al.*, not yet published.

Figure IV.4. Arrays CRISPR1 and CRISPR2 of *Salmonella* Bareilly

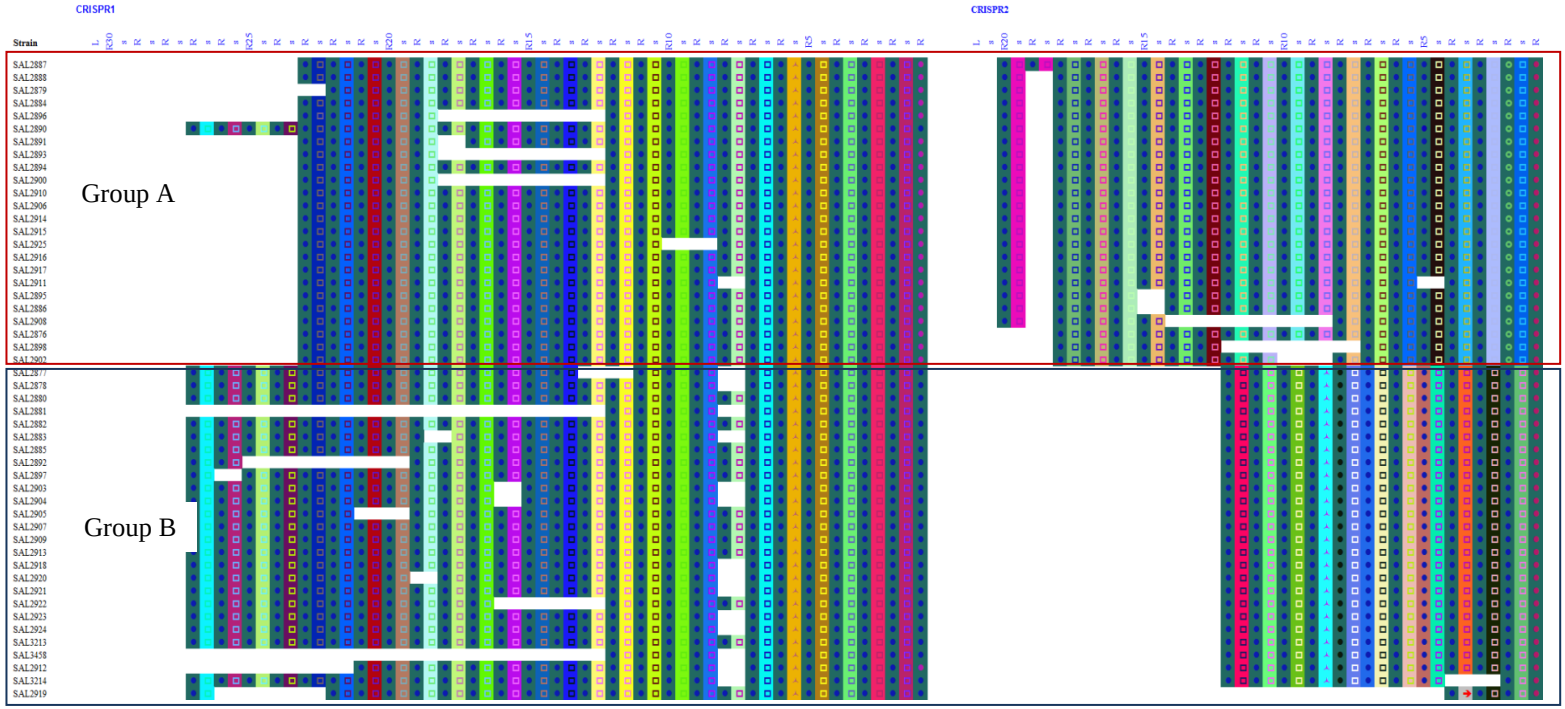
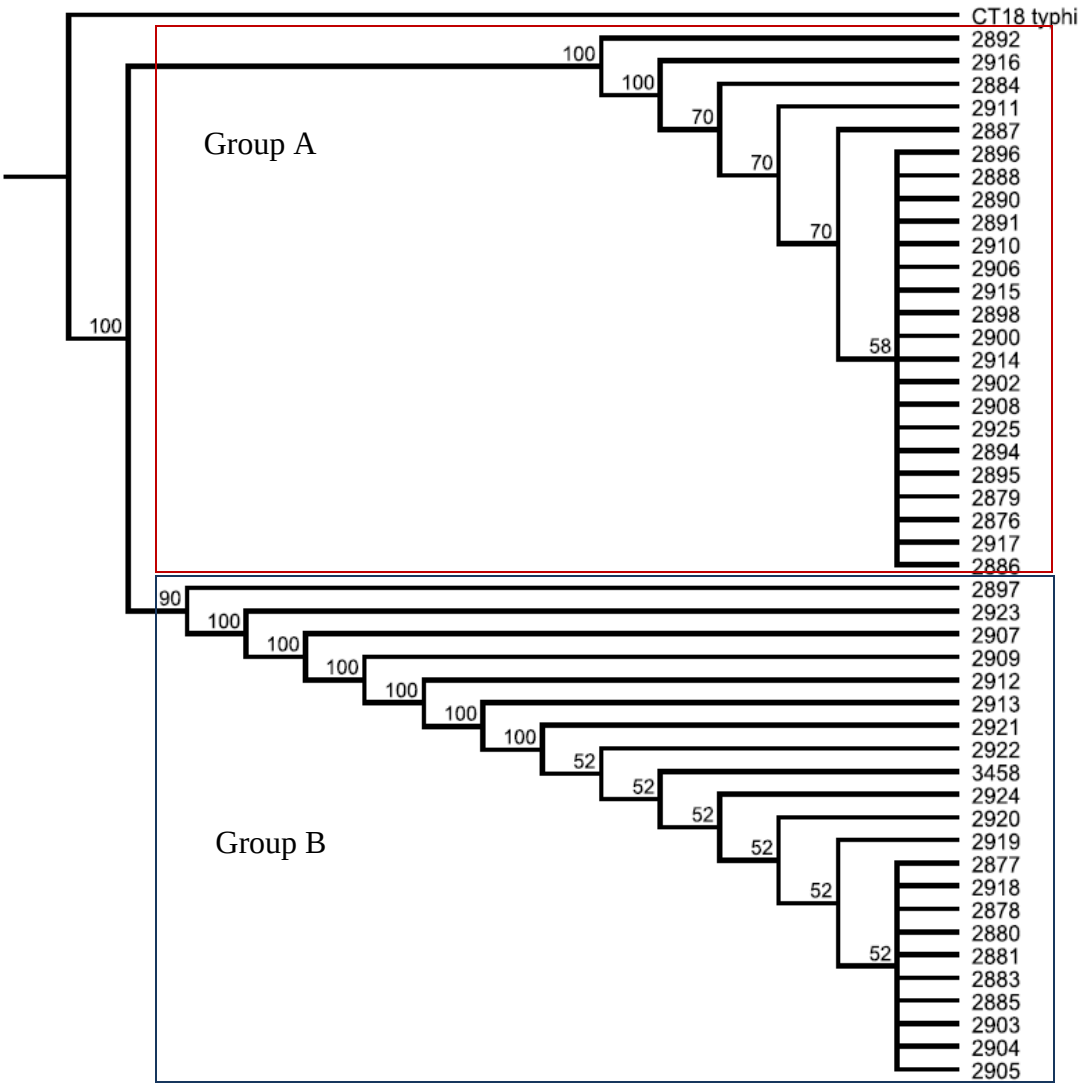
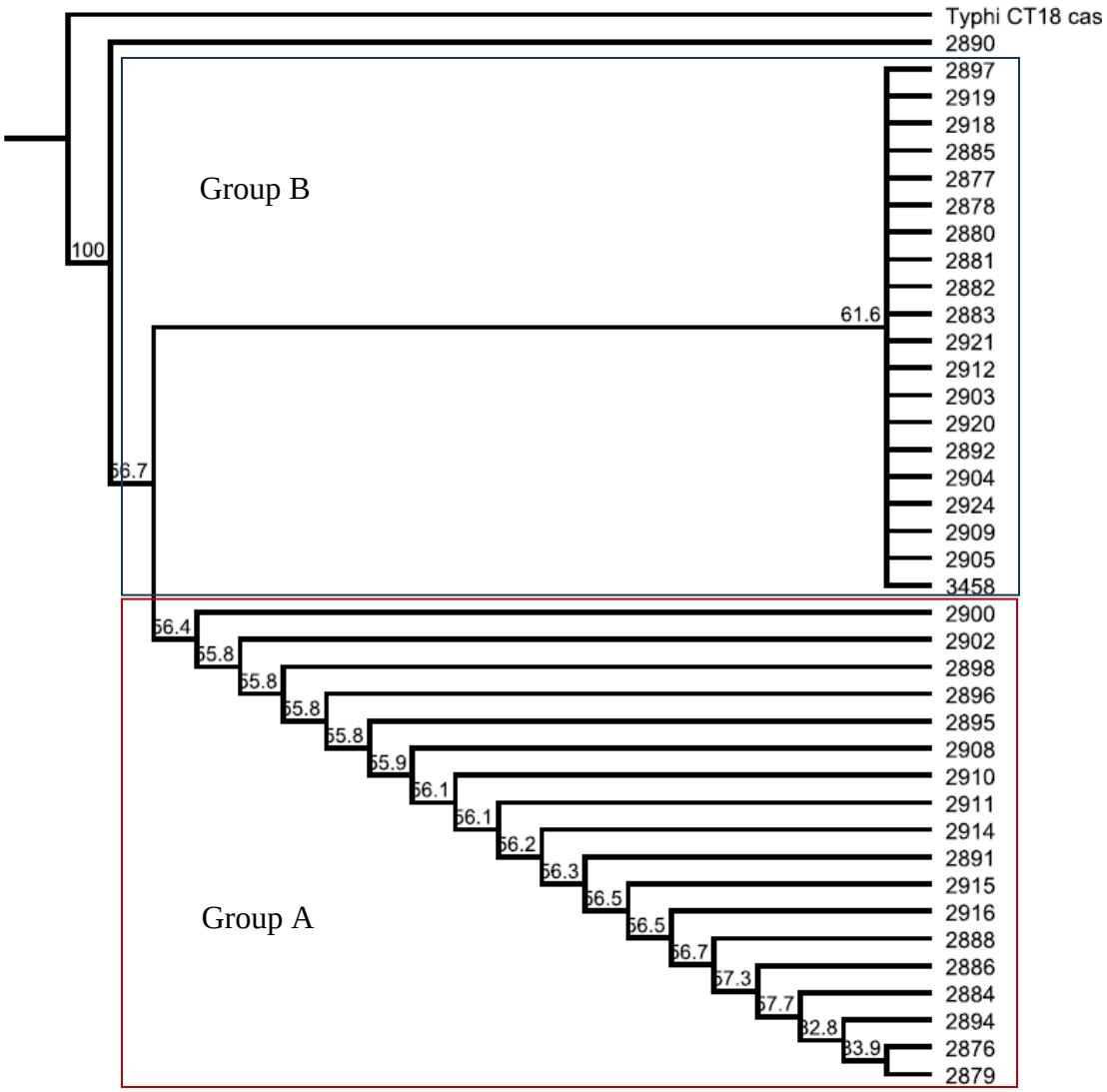


Figure IV.5. Maximum likelihood phylogenetic tree based on concatenated sequences of housekeeping genes of *Salmonella* Bareilly strains.



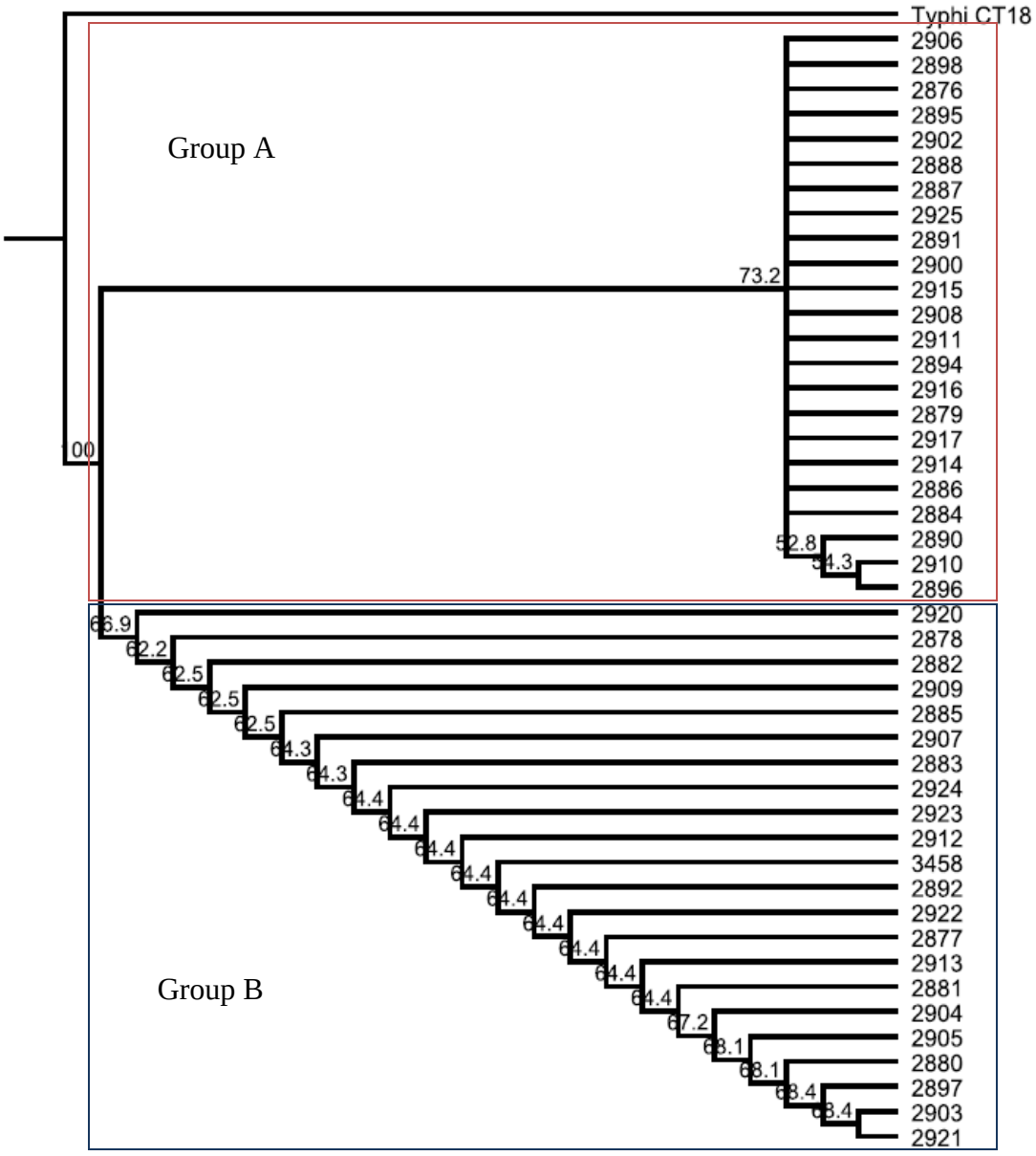
Concatenated sequences of *S.* Bareilly housekeeping genes (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*) were obtained from whole genome sequence. Maximum likelihood phylogenetic tree based crafted with Mega 5.1 with 1,000 bootstrap replications. *Salmonella* Typhi CT18 was selected as outgroup strain.

Figure IV.6. Maximum likelihood phylogenetic tree based on concatenated sequences of type I-E *cas* genes for *Salmonella* Bareilly strains.



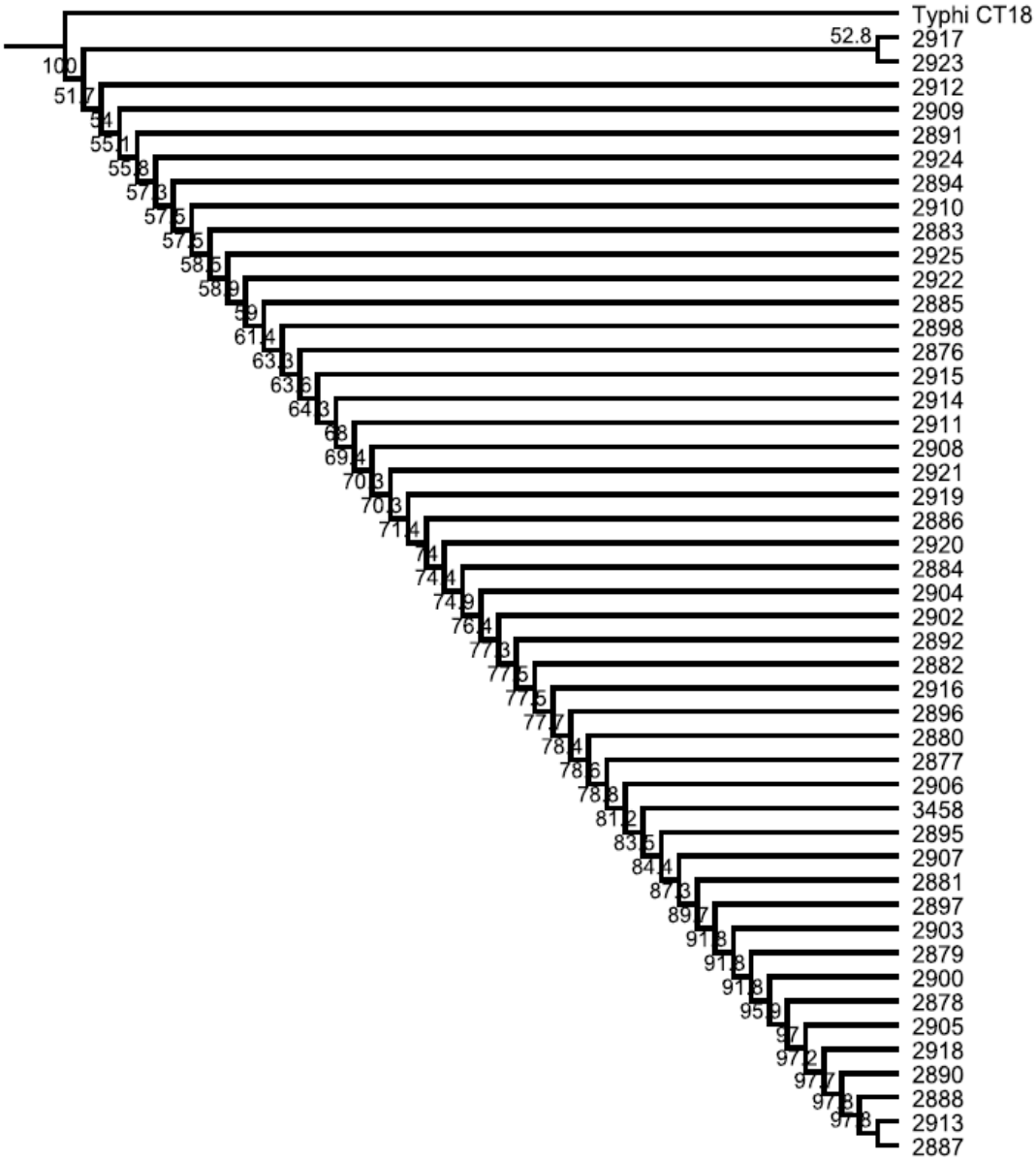
Concatenated sequences of *S.* Bareilly type I-E *cas* genes (*cas1, cas2, cse3, cse5e, cse4, cse2, cse1* and *cas3*) were obtained from whole genome sequence. Maximum likelihood phylogenetic tree based crafted with Mega 5.1 with 1,000 bootstrap replications. *Salmonella* Typhi CT18 was selected as outgroup strain.

Figure IV.7. Maximum likelihood phylogenetic tree based on CRISPR2 Leader sequences for *Salmonella* Bareilly strains.



Leader sequences for CRISPR2 were obtained from whole genome sequence. Maximum likelihood phylogenetic tree based crafted with Mega 5.1 with 1,000 bootstrap replications. *Salmonella* Typhi CT18 was selected as outgroup strain.

Figure IV.8. Maximum likelihood phylogenetic tree based on CRISPR1 Leader sequences for *Salmonella* Bareilly strains.



Leader sequences for CRISPR1 were obtained from whole genome sequence. Maximum likelihood phylogenetic tree based crafted with Mega 5.1 with 1,000 bootstrap replications. *Salmonella* Typhi CT18 was selected as outgroup strain.

Table IV.1. *Salmonella enterica* serotypes and number of strains in CRISPR sequencing study.

	Serotype	**# strains**
1	4,[5],12:i:-	2
2	Abaetetuba	1
3	Abony	1
4	Agona	7
5	Anatum	3
6	Bareilly	1
7	Berta	1
8	Braenderup	1
9	Cerro	2
10	Chester	1
11	Choleraseus	2
12	Derby	1
13	Dublin	3
14	Enteritidis	45
15	Galinarum	1
16	Gaminara	1
17	Give	3
18	Hadar	1
19	Heidelberg	38
20	Indiana	1
21	Inverness	1
22	Javiana	3
23	Johannesburg	1
24	Kentucky	5
25	Kunzendorf	1
26	Litchfield	1
27	Mbandaka	3

28	Meleagridis	2
29	Miami	1
30	Minnesota	1
31	Montevideo	22
32	Muenchen	3
33	Muenster	1
34	Newport	20
35	Oranienburg	3
36	Panama	1
37	Paratyphi B	6
38	Paratyphi_A	1
39	Pomona	1
40	Poona	1
41	Pullorum	3
42	Rissen	1
43	Rubislaw	1
44	Saintpaul	3
45	Scharzengrund	1
46	Sloterdijk	1
47	Stanley	1
48	Tallahassee	1
49	Tennessee	2
50	Thompson	2
51	Typhimurium	9
52	Urbana	1
53	Virchow	1
	Total	**221**

Table IV.2. *Salmonella* Bareilly isolates for CRISPR study

ID	Date of Isolation	Source	Country
SAL2876	7/29/1997	Whisker fish	Vietnam
SAL2877	7/11/1998	Frozen undeveined shrimp	India
SAL2879	12/22/1998	Frozen raw whole fish	Vietnam
SAL2880	12/26/1998	Frozen shrimp	India
SAL2881	5/12/1999	Frozen raw shrimp	India
SAL2882	2/25/2000	Frozen Raw Peeled Shrimp	India
SAL2883	8/7/2001	Frozen Whole Tilapia	Thailand
SAL2884	2/28/2002	Frozen Crab with Claws	Sri Lanka
SAL2885	10/18/2003	Coriander powder	India
SAL2886	4/16/2004	Fennel Seeds	United Arab Emirates
SAL2887	9/17/2004	Sand Goby Fish	Vietnam
SAL2888	12/28/2000	Frozen Shrimp	India
SAL2890	2/4/2004	Kheer mix	Pakistan
SAL2891	3/16/2004	Coriander powder	India
SAL2892	5/13/2005	Irrigation water	USA
SAL2893	1/13/2006	Frozen Fish	India
SAL2894	3/11/1997	Raw shrimp	Vietnam
SAL2895	2/16/2007	RED CHILI POWDER	Pakistan
SAL2896	12/5/2000	Crushed Chilies	India
SAL2897	4/21/2001	Sesame Seed	India
SAL2898	9/7/2004	Chili Powder	Thailand
SAL2900	3/9/2006	Coriander	Mexico
SAL2902	8/17/2007	Coconut	India
SAL2903	11/15/2001	Hilsa Fish	Thailand
SAL2904	12/4/2001	Frozen Rock Lobster Tails	United Arab Emirates
SAL2905	7/31/2002	Turmeric Powder	India
SAL2906	5/8/2003	Ginger powder	India

ID	Date of Isolation	Source	Country
SAL2907	2000-10-017	Shrimp	India
SAL2908	2/21/2006	Ground red pepper	USA
SAL2909	10/16/2007	Punjabi chole (king spice)	India
SAL2910	11/16/2006	Shrimp shell on	Sri Lanka
SAL2911	6/26/2007	Organic black pepper	India
SAL2912	11/17/2001	Cayenne Pepper	India
SAL2913	6/25/1996	scallops	Indonesia
SAL2914	7/8/1997	pabda fish	Bangladesh
SAL2915	10/11/1997	Frozen Rohu fish	India
SAL2916	3/26/1998	shrimp	India
SAL2917	11/7/1998	Cumin powder	India
SAL2918	3/13/2001	Coriander	Bangladesh
SAL2919	9/16/1999	Coriander powder	India
SAL2920	8/16/2002	Lobster Tails	Taiwan
SAL2921	2/7/2003	Frozen Baila	Bangladesh
SAL2922	5/31/2005	Chili Powder	India
SAL2923	11/12/2005	Fresh Water Fish (Bacha)	Bangladesh
SAL2924	4/5/2006	Fish stomach	Vietnam
SAL2925	9/8/2006	Chili powder	India
SAL3213	2/16/2008	Octopus	India
SAL3214	3/12/2008	Frozen shrimp	Thailand
SAL3458	5/1/2008	Frozen Yellowfin Tuna steaks	Indonesia

Table IV. 3. Primers used for *Salmonella enterica* serotypes amplification and sequencing.

Name	Sequence	Use
CRISPR1 FW	GATGTAGTGCGGATAATGCT	Forward primer for both amplification and sequencing
CRISPR1 REV1	GGTTTCTTTTCTTCCTGTTG	Reverse primer for both amplification and sequencing*
CRISPR1 REV2	GGTTTCTTTTCTTCCTGTTG	Reverse primer for both amplification and sequencing*
CRISPR2 FW	ACCAGCCATTACTGGTACAC	Forward primer for both amplification and sequencing
CRISPR2 REV	ATTGTTGCGATTATGTTGGT	Reverse primer for both amplification and sequencing
Bareilly C1 FWB	AGTTCAACAAACACCACGACG	Internal sequencing
Bareilly C1 REVB	CAGCTGAACAAACGTCAGGC	Internal sequencing
Braenderup C1 R1	ATATCGGAATTCAGCGCGGT	Internal sequencing
Montevideo A C1R1	CGTCCTGTGGAACCGGTTTA	Internal sequencing
Montevideo B C1F1	GGGATAAACCGGCTTCCCAA	Internal sequencing
Newport B C1R1	GGAACACGATGAGCAACACG	Internal sequencing
Typhimurium /Heildelberg C1R1	CACAATCACTGCGGGGGTAT	Internal sequencing
Seftenberg C1F1	GGGATAAACCGGGCTGACAA	Internal sequencing
Seftenberg C1R1	AAAACC CCGACCAGTTTTGC	Internal sequencing
Tennessee C1F1	CAACCTTTCGCGCTAATGGTG	Internal sequencing
Tennessee C1R1	CAAGTGGCAGCAGAACACAC	Internal sequencing
Seftenberg C2F1	GTTTAAACGCCTTGCCGTGT	Internal sequencing
Seftenberg C2R1	GGGAACACGGTTTTTCGTCG	Internal sequencing
Mntevideo C2F1	TTGAGCGTTTCTCGGCTGAT	Internal sequencing
Newport B C2R1	TAGAGCACCGCGGTTTATCC	Internal sequencing
Typhimurium C2F1	TTCATAGTGCCCGTGTTCCC	Internal sequencing
Typhimurium C2F4	GATCCTCAACGGTCAGGCTG	Internal sequencing

Name	Sequence	Use
Tennessee C2R2	TTAGCCGCTCGGTTTATCCC	Internal sequencing
NewportA C2R2	GGCGCGGGGAACACTATAAT	Internal sequencing
Muenster C2R1	GCGGGGAACACATACAGGAA	Internal sequencing
Newport A C2R1	TAACCTGCGGTTTATCCCCG	Internal sequencing

Table IV.4. General characteristics of CRISPR arrays from *Salmonella enterica* serotypes (n=224).

Characteristic	CRISPR1	CRISPR2	Combined
Number arrays	221	221	22
Number of non identical arrays	97	92	111
Spacers in array			
Range	1-62	1-32	2-88
Average	17	14	31
Mode	9	13	40
Total spacers	3658	3086	6744
Different spacers	744	565	1302*
Unique spacers	345	248	590
Spacers Length (bp)			
Average	32	32	32
Min	30	32	30
Max	44	33	44
Protospacers detected	13	7	20

*Six spacers were present in both loci

Chapter 5: Summary of findings, implications and future Studies

Foodborne pathogens are an important source of infection for thousands of people in the word. WHO estimates one third of the world population suffers from foodborne diseases each year, and in the US it is estimated that about 50 million of cases occur annually. Early detection of contaminated food and water can help to prevent cases; therefore, the detection and evaluation of new molecular markers is fundamental to develop new and improved detection and identification methods. The objective of this research was to develop and analyze alternative techniques and markers for the detection of *Salmonella enterica* and Shigatoxin producing *E. coli* (STEC).

In chapter II, a new suspension array assay was developed. The assay identifies and discriminates between the seven most important STEC from pure culture in a single reaction. Probes targeting genes *wzx* or *wzy*, serogroup specific genes, identified whether an isolate belonged to serotypes O26, O45, O103, O111, O121, O145, or O157. The assay also included probes targeting genes *stx1* and *stx2* (Table II.3). Positive samples were clearly distinguished from negative samples (Table II.4), and a cut off ratio of 15 was set to differentiate between a positive and a negative isolate.

The detection of STEC isolates from serogroups O26, O45, O103, O111, O121, O145, and O157 is important because they are considered adulterants when found in meat and meats products. The use of fast molecular techniques that detect contamination with this pathogen contribute to the reduction of the number of cases

of human disease due to seize of contaminated products; the use of suspension arrays seems a promising alternative for developing fast detection and identification techniques. This is the first time a suspension array assay identifies not only serogroups but also Shiga toxin genes, which allows to classify isolates as STEC. The method developed represents a reliable alternative for molecular serotyping of *E. coli*, and can be useful to better understand the epidemiology of STEC infections and enhance outbreak investigations. Further studies will aim to design a complementary second panel that includes the next top 20 O STEC serogroups, as well as optimizing the panel for the detection of STECs from foods.

In chapter III, CRISPR arrays of 194 STEC strains were described and characterized, and a potential relationship between array length and virulence markers was explored. It was found that strains of a same serotyped clustered together, but strains of the same O serogroups did not necessarily formed blocks (Fig 1). Instead, some strains of different serogroup but the same H type clustered together and share most of their spacers. These findings suggest that H antigen genes would be more phylogenetic stable than O antigen, or that H antigen was acquired more ancestrally than O antigen in STEC. This has been recently suggested by other researchers and suggests a shift in the paradigm of STEC evolution.

CRISPR function is to prevent the invasion of horizontal gene transfer elements, and many of virulence determinants were acquired in that way in *E. coli.* The hypothesis that there is an inverse correlation between the presence of CRISPR elements and virulence markers was tested, but the presence of I-E subtype cas genes

was not linked to virulence elements. However, array length was related to several virulence characteristics; strains of seropathotypes more likely to cause outbreak and severe disease had longer CRISPR that those of other seropathotypes. Strains with one type of Shigatoxin genes had longer arrays than those with two, and shorter arrays were a marker for the presence of some virulence genes. This relationship may suggest that CRISPR systems had interfered with the acquisition of virulence traits in STEC. Further studies, using additional and genetically diverse strains, would provide a better understanding of the CRISPR-*Cas* system in STEC, and *E. coli* as a group. If the relationship between CRISPR array length and presence of virulence markers is true, array length could be used to predict high risk STEC strains and could be a useful tool for the control of STEC infections.

In chapter IV, CRISPR arrays of 221 *Salmonella enterica* subspecies *enterica* (*S. enterica*) were described, and the CRISPR-cas system of the emerging serotype *S. enterica* serotype Bareilly (*S.* Bareilly) was analyzed. Similarly to STEC strains, CRISPR arrays of *Salmonella enterica* correlate well with serotypes, and could be used to develop molecular serotyping assays. However, some serotypes displayed more than one type of array. It has been suggested that CRISPR arrays reflect phylogeny of *Salmonella enterica*. One of the serotypes presented different patters was Salmonella Newport, and it has been documented that convergent evolution created this serotype from different *Salmonella* lineages.

CRISPR-cas analysis of *S.* Bareilly revealed that there are two patterns in array 2; those arrays did not share any spacer, but all CRISPR1 arrays had the same

spacer pattern. The variation in CRISPR arrays was also reflected in leader sequences and *cas* genes. Interestingly, variations corresponded with those showed by housekeeping genes, which reflect phylogenetic relationships (29). This suggests the presence of two lineages of *Salmonella* Bareilly, a finding that has not been previously described in literature.

CRISPR arrays seem suitable for developing molecular serotyping methods for *S. enterica*, but it is necessary to include a higher number of strains from different geographical locations, food vehicle and hosts, thus creating a more comprehensive database of CRISPR arrays and spacers. Also, the low rate of variation of CRISPR arrays in different *Salmonella* serotypes, especially in a large sample of *Salmonella* Bareilly with different PFGE pattern, geographical origin and food vehicle, confirms that CRISPR arrays are not suitable for *Salmonella* subtyping unless other genetic elements are included in the comparison.

Master References

References in Chapter I

1. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.
2. [**CDC] Centers for Disease Control and Prevention.** 2011. Surveillance for foodborne disease outbreaks --- United States, 2008. MMWR Morb Mortal Wkly Rep **60:**1197-1202.
3. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2:**123-140.
4. **Erickson MC, Doyle MP.** 2007. Food as a vehicle for transmission of Shiga toxin-producing *Escherichia coli*. J Food Prot **70:**2426-2449.
5. **Eurosurveillance.** 2011. EFSA publishes scientific report on the public health risk of Shiga-toxin producing *Escherichia coli* (STEC) in fresh vegetables. Euro Surveill **16** (23). Available at: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19888. Accessed January 7, 2014.
6. **Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N Engl J Med **308:**681-685.

7. **Johnson KE, Thorpe CM, Sears CL.** 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin Infect Dis **43:**1587-1595.

8. **Bettelheim KA.** 2007. The non-O157 shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. Crit Rev Microbiol **33:**67-87.

9. **Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J.** 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J Clin Microbiol **42:**311-319.

10. **Griffin P.** 2007. CDC perspective on non-O157 Shiga toxin-producing *E. coli* (STEC) in the United States. Available at: http://www.fsis.usda.gov/wps/wcm/connect/6200a224-5afa-48c8-a8c5-e87cb9c391f1/Non-0157_STEC_Griffin.pdf?MOD=AJPERES, accessed on January 2014.

11. **Gould LH, Bopp C, Strockbine N, Atkinson R, Baselski V, Body B, Carey R, Crandall C, Hurd S, Kaplan R, Neill M, Shea S, Somsel P, Tobin-D'Angelo M, Griffin PM, Gerner-Smidt P.** 2009. Recommendations for diagnosis of shiga toxin--producing *Escherichia coli* infections by clinical laboratories. MMWR Recomm Rep **58:**1-14.

12. **Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, Strockbine NA.** 2005. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983-2002. J Infect Dis **192:**1422-1429.

13. **[USDA] United States Department of Agriculture, Food Safety and Inspection Service.** 2012. Shiga Toxin Producing *Escherichia coli* in certain raw beef products. Available at: http://www.fsis.usda.gov/OPPDE/rdad/FRPubs/2010-0023FRN.pdf, accessed May2012.

14. **Nataro JP, Kaper JB.** 1998. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev **11:**142-201.

15. **Johannes L, Römer W.** 2010. Shiga toxins--from cell biology to biomedical applications. Nat Rev Microbiol **8:**105-116.

16. **Karmali MA, Petric M, Lim C, Fleming PC, Steele BT.** 1983. *Escherichia coli* cytotoxin, haemolytic-uraemic syndrome, and haemorrhagic colitis. Lancet **2:**1299-1300.

17. **Hussein HS, Bollinger LM.** 2008. Influence of selective media on successful detection of Shiga toxin-producing *Escherichia coli* in food, fecal, and environmental samples. Foodborne Pathog Dis **5:**227-244.

18. **Gyles CL.** 2007. Shiga toxin-producing *Escherichia coli*: an overview. J Anim Sci **85:**E45-62.

19. **Karmali MA, Steele BT, Petric M, Lim C.** 1983. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. Lancet **1:**619-620.

20. **Call DR, Brockman FJ, Chandler DP.** 2001. Detecting and genotyping *Escherichia coli* O157:H7 using multiplexed PCR and nucleic acid microarrays. Int J Food Microbiol **67:**71-80.

21. **Auvray F, Lecureuil C, Taché J, Leclerc V, Deperrois V, Lombard B.** 2007. Detection, isolation and characterization of Shiga toxin-producing *Escherichia coli* in retail-minced beef using PCR-based techniques, immunoassays and colony hybridization. Lett Appl Microbiol **45:**646-651.

22. **DebRoy C, Fratamico PM, Roberts E, Davis MA, Liu Y.** 2005. Development of PCR assays targeting genes in O-antigen gene clusters for detection and identification of *Escherichia coli* O45 and O55 serogroups. Appl Environ Microbiol **71:**4919-4924.

23. **Samuel G, Reeves P.** 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res **338:**2503-2519.

24. **Stenutz R, Weintraub A, Widmalm G.** 2006. The structures of *Escherichia coli* O-polysaccharide antigens. FEMS Microbiol Rev **30:**382-403.

25. **Fratamico PM, Briggs CE, Needle D, Chen CY, DebRoy C.** 2003. Sequence of the *Escherichia coli* O121 O-antigen gene cluster and detection of enterohemorrhagic *E. coli* O121 by PCR amplification of the wzx and wzy genes. J Clin Microbiol **41:**3379-3383.

26. **Wang L, Curd H, Qu W, Reeves PR.** 1998. Sequencing of *Escherichia coli* O111 O-antigen gene cluster and identification of O111-specific genes. J Clin Microbiol **36:**3182-3187.

27. **Fratamico PM, DebRoy C, Miyamoto T, Liu Y.** 2009. PCR detection of enterohemorrhagic *Escherichia coli* O145 in food by targeting genes in the *E.*

coli O145 O-antigen gene cluster and the shiga toxin 1 and shiga toxin 2 genes. Foodborne Pathog Dis **6:**605-611.

28. **DebRoy C, Roberts E, Kundrat J, Davis MA, Briggs CE, Fratamico PM.** 2004. Detection of *Escherichia coli* serogroups O26 and O113 by PCR amplification of the *wzx* and *wzy* genes. Appl Environ Microbiol **70:**1830-1832.

29. **Debroy C, Roberts E, Valadez AM, Dudley EG, Cutter CN.** 2011. Detection of Shiga Toxin-Producing *Escherichia coli* O26, O45, O103, O111, O113, O121, O145, and O157 serogroups by multiplex polymerase chain reaction of the *wzx* gene of the O-Antigen gene cluster. Foodborne Pathog Dis **8:**651-652.

30. **Ren Y, Liu B, Cheng J, Liu F, Feng L, Wang L.** 2008. Characterization of *Escherichia coli* O3 and O21 O antigen gene clusters and development of serogroup-specific PCR assays. J Microbiol Methods **75:**329-334.

31. **Valadez AM, Debroy C, Dudley E, Cutter CN.** 2011. Multiplex PCR detection of Shiga toxin-producing *Escherichia coli* strains belonging to serogroups O157, O103, O91, O113, O145, O111, and O26 experimentally inoculated in beef carcass swabs, beef trim, and ground beef. J Food Prot **74:**228-239.

32. **Hsu CF, Tsai TY, Pan TM.** 2005. Use of the duplex TaqMan PCR system for detection of Shiga-like toxin-producing *Escherichia coli* O157. J Clin Microbiol **43:**2668-2673.

33. **Tsai TY, Lee WJ, Huang YJ, Chen KL, Pan TM.** 2006. Detection of viable enterohemorrhagic *Escherichia coli* O157 using the combination of immunomagnetic separation with the reverse transcription multiplex TaqMan PCR system in food and stool samples. J Food Prot **69:**2320-2328.

34. **Beutin L, Jahn S, Fach P.** 2009. Evaluation of the 'GeneDisc' real-time PCR system for detection of enterohaemorrhagic *Escherichia coli* (EHEC) O26, O103, O111, O145 and O157 strains according to their virulence markers and their O- and H-antigen-associated genes. J Appl Microbiol **106:**1122-1132.

35. **Fratamico PM, Bagi LK, Cray WC, Narang N, Yan X, Medina M, Liu Y.** 2011. Detection by multiplex real-time polymerase chain reaction assays and isolation of shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 in ground beef. Foodborne Pathog Dis **8:**601-607.

36. **Lin A, Sultan O, Lau HK, Wong E, Hartman G, Lauzon CR.** 2011. O serogroup specific real time PCR assays for the detection and identification of nine clinically relevant non-O157 STECs. Food Microbiol **28:**478-483.

37. **Chen S, Zhao S, McDermott PF, Schroeder CM, White DG, Meng J.** 2005. A DNA microarray for identification of virulence and antimicrobial resistance genes in *Salmonella* serovars and *Escherichia coli*. Mol Cell Probes **19:**195-201.

38. **Liu Y, Fratamico P.** 2006. *Escherichia coli* O antigen typing using DNA microarrays. Mol Cell Probes **20:**239-244.

39. **Ballmer K, Korczak BM, Kuhnert P, Slickers P, Ehricht R, Hächler H.** 2007. Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. J Clin Microbiol **45:**370-379.

40. **Bugarel M, Beutin L, Martin A, Gill A, Fach P.** 2010. Micro-array for the identification of Shiga toxin-producing *Escherichia coli* (STEC) seropathotypes associated with hemorrhagic colitis and hemolytic uremic syndrome in humans. Int J Food Microbiol **142:**318-329.

41. **Nolan JP, Sklar LA.** 2002. Suspension array technology: evolution of the flat-array paradigm. Trends Biotechnol **20:**9-12.

42. **Moalic V.** 2004. Technologie Luminex (TM): principe, applications, et perspectives. *Immuno-analyse & Biologie Spécialisée* **19:**7.

43. **Dunbar SA.** 2006. Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. Clin Chim Acta **363:**71-82.

44. **Baums IB, Goodwin KD, Kiesling T, Wanless D, Diaz MR, Fell JW.** 2007. Luminex detection of fecal indicators in river samples, marine recreational water, and beach sand. Mar Pollut Bull **54:**521-536.

45. **Dunbar SA, Vander Zee CA, Oliver KG, Karem KL, Jacobson JW.** 2003. Quantitative, multiplexed detection of bacterial pathogens: DNA and protein applications of the Luminex LabMAP system. J Microbiol Methods **53:**245-252.

46. **Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P.** 2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol **45:**3323-3334.

47. **Griffith JF, Cao Y, McGee CD, Weisberg SB.** 2009. Evaluation of rapid methods and novel indicators for assessing microbiological beach water quality. Water Res **43:**4900-4907.

48. **Barco L, Lettini AA, Pozza MCD, Ramon E, Fasolato M, Ricci A.** 2010. Fluoroquinolone Resistance Detection in *Campylobacter coli* and *Campylobacter jejuni* by Luminex® xMAP™ Technology. Foodborne Pathog Dis 7 (9): 1039-1045.

49. **McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML, Fields PI.** 2011. Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. J Clin Microbiol **49:**565-573.

50. **Su LH, Chiu CH.** 2007. *Salmonella*: clinical importance and evolution of nomenclature. Chang Gung Med J **30:**210-219.

51. **Boyle EC, Bishop JL, Grassl GA, Finlay BB.** 2007. *Salmonella*: from pathogenesis to therapeutics. J Bacteriol **189:**1489-1495.

52. **Casadesús J.** 2011. *Salmonella*: from basic science to clinical issues. Future Microbiol **6:**133-135.

53. **White DG, Zhao S, Sudler R, Ayers S, Friedman S, Chen S, McDermott PF, McDermott S, Wagner DD, Meng J.** 2001. The isolation of antibiotic-resistant *Salmonella* from retail ground meats. N Engl J Med **345:**1147-1154.

54. **Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ.** 2013. Outbreak-associated *Salmonella enterica* serotypes and food Commodities, United States, 1998-2008. Emerg Infect Dis **19:**1239-1244.

55. **Hanning IB, Nutt JD, Ricke SC.** 2009. Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention measures. Foodborne Pathog Dis **6:**635-648.

56. [**CDC] Centers for Disease Control and Precention.** 2013. Outbreaks Involving *Salmonella*. Available at: http://www.cdc.gov/salmonella/outbreaks.html, accessed on January 2014.

57. [**CDC] Centers for Disease Control and Prevention.** 2013. National Enteric Disease Surveillance: *Salmonella* annual report, 2011. Available at: http://www.cdc.gov/ncezid/dfwed/pdfs/salmonella-annual-report-2011-508c.pdf, accessed on January 2014.

58. **McQuiston JR, Fields PI, Tauxe RV, Logsdon JM.** 2008. Do *Salmonella* carry spare tyres? Trends Microbiol **16:**142-148.

59. **Hong Y, Liu T, Lee MD, Hofacre CL, Maier M, White DG, Ayers S, Wang L, Berghaus R, Maurer JJ.** 2008. Rapid screening of *Salmonella enterica* serovars Enteritidis, Hadar, Heidelberg and Typhimurium using a serologically-correlative allelotyping PCR targeting the O and H antigen alleles. BMC Microbiol **8:**178.

60. [**CDC] Centers for Disease Control and Prevention.** 2011. PulseNet. Available at: http://www.cdc.gov/pulsenet/, accessed on January 2014.

61. **Liu F, Kariyawasam S, Jayarao BM, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG.** 2011. Subtyping *Salmonella enterica* serovar enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). Appl Environ Microbiol **77:**4520-4526.

62. **Wattiau P, Boland C, Bertrand S.** 2011. Methodologies for *Salmonella enterica* ssp *enterica* subtyping: gold standards and alternatives. Appl Environ Microbiol. 77 (22): 7877-7885.

63. **Aggarwal P, Sarkar R, Singh M, Grover BD, Anand BR, Chowdhuri AN.** 1983. *Salmonella* Bareilly infection in a paediatric hospital of New Delhi. Indian J Med Res **78:**22-25.

64. **Gupta P, Talwar V, Revathi G, Kumar A.** 1997. Nosocomial *Salmonella* Bareilly septicemia: a nursery outbreak. Indian Pediatr **34:**144-146.

65. **Mccullough NB, Eisele CW.** 1951. Experimental human salmonellosis. III. Pathogenicity of strains of *Salmonella* Newport, *Salmonella* Derby, and *Salmonella* Bareilly obtained from spray-dried whole egg. J Infect Dis **89:**209-213.

66. **Singh G, Sharma NC, Jayasheela M, Saxena SN.** 1988. A study of the epidemiology of *Salmonella* Bareilly in India using a new phage-typing system. Epidemiol Infect **100:**221-225.

67. **Prapasarakul N, Pulsrikarn C, Vasaruchapong T, Lekcharoen P, Chanchaithong P, Lugsomya K, Keschumras N, Thanomsuksinchai N, Tanchiangsai K, Tummaruk P.** 2012. *Salmonella* serovar distribution in

cobras (Naja kaouthia), snake-food species, and farm workers at Queen Saovabha Snake Park, Thailand. J Vet Diagn Invest **24:**288-294.

68. **Melendez SN, Hanning I, Han J, Nayak R, Clement AR, Wooming A, Hererra P, Jones FT, Foley SL, Ricke SC.** 2010. *Salmonella enterica* isolates from pasture-raised poultry exhibit antimicrobial resistance and class I integrons. J Appl Microbiol **109:**1957-1966.

69. **Rhoades HE.** 1946. *Salmonella* bareilly in quail. North Am Vet **27:**419.

70. **Kudaka J, Itokazu K, Taira K, Iwai A, Kondo M, Susa T, Iwanaga M.** 2006. Characterization of *Salmonella* isolated in Okinawa, Japan. Jpn J Infect Dis **59:**15-19.

71. **Wilson IG.** 2002. *Salmonella* and *Campylobacter* contamination of raw retail chickens from different producers: a six year survey. Epidemiol Infect **129:**635-645.

72. **[CDC] Centers for Disease Prevention and Control.** 2012. Multispate outbreak of *Salmonella* Bareilly and *Salmonella* Nchanga infections associated with a raw scrapped ground tuna product (final update). Available at: http://www.cdc.gov/salmonella/bareilly-04-12/index.html?utm_source=Website+Signup&utm_campaign=f4e381310d-Monday_Morning_Light_April_16_2012_13_2012&utm_medium=email, accessed on January 2014.

73. **de Jong B, Oberg J, Svenungsson B.** 2007. Outbreak of salmonellosis in a restaurant in Stockholm, Sweden, September - October 2006. Euro Surveill **12:**E13-14.

74. **Cleary P, Browning L, Coia J, Cowden J, Fox A, Kearney J, Lane C, Mather H, Quigley C, Syed Q, Tubin-Delic D, team oc.** 2010. A foodborne outbreak of *Salmonella* Bareilly in the United Kingdom, 2010. Euro Surveill **15** (48) pii 19732.

75. **Sharma NC, Bhatia R, Singh S, John PC, Kumar S, Singh H.** 1994. Bacteriophage typing in *Salmonella* Bareilly. Epidemiol Infect **112:**45-49.

76. **Saxena MK, Singh VP, Lakhcharua BD, Taj G, Sharma B.** 2002. Strain differentiation of Indian isolates of *Salmonella* by ERIC-PCR. Res Vet Sci **73:**313-314.

77. **Zhao S, McDermott PF, Friedman S, Qaiyumi S, Abbott J, Kiessling C, Ayers S, Singh R, Hubert S, Sofos J, White DG.** 2006. Characterization of antimicrobial-resistant *Salmonella* isolated from imported foods. J Food Prot **69:**500-507.

78. **Kumar R, Surendran PK, Thampuran N.** 2009. Analysis of antimicrobial resistance and plasmid profiles in *Salmonella* serovars associated with tropical seafood of India. Foodborne Pathog Dis **6:**621-625.

79. **Kunin V, Sorek R, Hugenholtz P.** 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol **8:**R61.

80. **Sorek R, Kunin V, Hugenholtz P.** 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol **6:**181-186.

81. **Karginov FV, Hannon GJ.** 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell **37:**7-19.

82. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

83. **Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S.** 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol **190:**1390-1400.

84. **van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ.** 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci **34:**401-407.

85. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167-170.

86. **Perez-Rodriguez R, Haitjema C, Huang Q, Nam KH, Bernardis S, Ke A, Delisa MP.** 2011. Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. Mol Microbiol **79:**584-599.

87. **Touchon M, Charpentier S, Clermont O, Rocha EP, Denamur E, Branger C.** 2011. CRISPR Distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. J Bacteriol **193:**2460-2467.

88. **Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV.** 2002. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res **30:**482-496.

89. **Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, Gagarinova A, Pogoutse O, Brown G, Binkowski A, Phanse S,**

Joachimiak A, Koonin EV, Savchenko A, Emili A, Greenblatt J, Edwards AM, Yakunin AF. 2011. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol Microbiol **79:**484-502.

90. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A.** 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol **169:**5429-5433.

91. **Al-Attar S, Westra ER, van der Oost J, Brouns SJJ.** 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biological Chemistry **392:**277-289.

92. **Haft DH, Selengut J, Mongodin EF, Nelson KE.** 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol **1:**e60.

93. **Mojica FJ, Díez-Villaseñor C, Soria E, Juez G.** 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol **36:**244-246.

94. **Díez-Villaseñor C, Almendros C, García-Martínez J, Mojica FJ.** 2010. Diversity of CRISPR loci in *Escherichia coli*. Microbiology **156:**1351-1361.

95. **Deveau H, Garneau JE, Moineau S.** 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol **64:**475-493.

96. **Shah SA, Garrett RA.** 2011. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol **162:**27-38.

97. **Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol **9:**467-477.

98. **Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E.** 2011. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature **477:**486-489.

99. **Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ.** 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol **18:**529-536.

100. **Marraffini LA, Sontheimer EJ.** 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322:**1843-1845.

101. **Edgar R, Qimron U.** 2010. The *Escherichia coli* CRISPR system protects from λ lysogenization, lysogens, and prophage induction. J Bacteriol **192:**6291-6294.

102. **Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A.** 2010. Transcription profile of Thermus thermophilus CRISPR systems after phage infection. J Mol Biol **395:**270-281.

103. **Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321:**960-964.

104. **Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K.** 2010. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. Molecular Microbiology **77:**1367-1379.

105. **Jore MM, Brouns SJ, van der Oost J.** 2011. RNA in defense: crisprs protect prokaryotes against mobile genetic elements. Cold Spring Harb Perspect Biol. 1; 4 (6).

106. **Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K.** 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci U S A **108:**10098-10103.

107. **Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV, Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R.** 2008. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. PLoS One **3:**e2652.

108. **Tyson GW, Banfield JF.** 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol **10:**200-207.

109. **Vale PF, Little TJ.** 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. Proc Biol Sci **277:**2097-2103.

110. **Pourcel C, Salvignol G, Vergnaud G.** 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology **151:**653-663.

111. **Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J.** 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol **35:**907-914.

112. **Gori A, Bandera A, Marchetti G, Degli Esposti A, Catozzi L, Nardi GP, Gazzola L, Ferrario G, van Embden JD, van Soolingen D, Moroni M, Franzetti F.** 2005. Spoligotyping and *Mycobacterium tuberculosis*. Emerg Infect Dis **11:**1242-1248.

113. **Pride DT, Sun CL, Salzman J, Rao N, Loomer P, Armitage GC, Banfield JF, Relman DA.** 2011. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. Genome Res **21:**126-136.

114. **Price EP, Smith H, Huygens F, Giffard PM.** 2007. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. Appl Environ Microbiol **73:**3431-3436.

115. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.

116. **Delannoy S, Beutin L, Burgos Y, Fach P.** 2012. Specific detection of enteroaggregative hemorrhagic *Escherichia coli* O104:H4 strains by use of the crispr locus as a target for a diagnostic real-time PCR. J Clin Microbiol **50:**3485-3492.

117. **Delannoy S, Beutin L, Fach P.** 2012. Use of CRISPR sequence polymorphisms for specific detection of enterohemorrhagic *E. coli* (EHEC)

strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28 and O157:H7 by real time PCR. J Clin Microbiol.

118. **Yin S, Jensen MA, Bai J, Debroy C, Barrangou R, Dudley EG.** 2013. The evolutionary divergence of shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. Appl Environ Microbiol **79:**5710-5720.

119. **Fricke WF, Mammel MK, McDermott PF, Tartera C, White DG, Leclerc JE, Ravel J, Cebula TA.** 2011. Comparative Genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. J Bacteriol **193:**3556-3568.

120. **Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc Natl Acad Sci U S A **106:**17939-17944.

121. **Palmer KL, Gilmore MS.** 2010. Multidrug-resistant enterococci lack CRISPR-cas. MBio **12; 1** (4).

122. **Garneau JE, Dupuis M, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S.** 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468:**67-71.

123. **Lindenstrauß AG, Pavlovic M, Bringmann A, Behr J, Ehrmann MA, Vogel RF.** 2011. Comparison of genotypic and phenotypic cluster analyses of

virulence determinants and possible role of CRISPR elements towards their incidence in *Enterococcus faecalis* and *Enterococcus faecium*. Syst Appl Microbiol. 34 (8): 553-560.

124. **Touchon M, Charpentier S, Pognard D, Picard B, Arlet G, Rocha EP, Denamur E, Branger C.** 2012. Antibiotic resistance plasmids spread among natural isolates of *Escherichia coli* in spite of CRISPR elements. Microbiology **158:**2997-3004.

125. **Dang TN, Zhang L, Zöllner S, Srinivasan U, Abbas K, Marrs CF, Foxman B.** 2013. Uropathogenic *Escherichia coli* are less likely than paired fecal *E. coli* to have CRISPR loci. Infect Genet Evol **19C:**212-218.

126. **Swarts DC, Mosterd C, van Passel MW, Brouns SJ.** 2012. CRISPR interference directs strand specific spacer acquisition. PLoS One **7:**e35888.

127. **Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS.** 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature **497:**254-257.

128. **Gunderson FF, Cianciotto NP.** 2013. The CRISPR-associated gene *cas2* of *Legionella pneumophila* is required for intracellular infection of amoebae. MBio **4:**e00074-00013.

129. **Louwen R, Horst-Kreft D, de Boer AG, van der Graaf L, de Knegt G, Hamersma M, Heikema AP, Timms AR, Jacobs BC, Wagenaar JA, Endtz HP, van der Oost J, Wells JM, Nieuwenhuis EE, van Vliet AH, Willemsen PT, van Baarlen P, van Belkum A.** 2013. A novel link between

Campylobacter jejuni bacteriophage defence, virulence and Guillain-Barré syndrome. Eur J Clin Microbiol Infect Dis **32:**207-226.

References in Chapter II

1. **Toro M, Najjar MB, Ju W, Brown E, Zhao S, Meng J.** 2013. Molecular serogrouping of Shiga Toxin-Producing *Escherichia coli* using suspension Array. Foodborne Pathog Dis. **10 (5)**:478-480.
2. **CDC.** 2010. Surveillance for foodborne disease outbreaks --- United States, 2007. MMWR Morb Mortal Wkly Rep **59:**973-979.
3. **Erickson MC, Doyle MP.** 2007. Food as a vehicle for transmission of Shiga toxin-producing *Escherichia coli*. J Food Prot **70:**2426-2449.
4. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2:**123-140.
5. **Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV.** 1999. Food-related illness and death in the United States. Emerg Infect Dis **5:**607-625.
6. **Bettelheim KA.** 2007. The non-O157 shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. Crit Rev Microbiol **33:**67-87.
7. **Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J.** 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J Clin Microbiol **42:**311-319.

8. **Uhlich GA, Sinclair JR, Warren NG, Chmielecki WA, Fratamico P.** 2008. Characterization of Shiga toxin-producing *Escherichia coli* isolates associated with two multistate food-borne outbreaks that occurred in 2006. Appl Environ Microbiol **74:**1268-1272.

9. **Johnson KE, Thorpe CM, Sears CL.** 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin Infect Dis **43:**1587-1595.

10. **[USDA] United States Department of Agriculture FSIS.** 2012. Shiga toxin-producing *Escherichia coli* in certain raw beef products., Federal Register. Available at www.fsis.usda.gov/OPPDE/rdad/FRPubs/2010-0023FRN.pdf. Accessed May 2012.

11. **Fratamico PM, DebRoy C, Miyamoto T, Liu Y.** 2009. PCR detection of enterohemorrhagic *Escherichia coli* O145 in food by targeting genes in the *E. coli* O145 O-antigen gene cluster and the Shiga toxin 1 and Shiga toxin 2 genes. Foodborne Pathog Dis **6:**605-611.

12. **DebRoy C, Roberts E, Kundrat J, Davis MA, Briggs CE, Fratamico PM.** 2004. Detection of *Escherichia coli* serogroups O26 and O113 by PCR amplification of the *wzx* and *wzy* genes. Appl Environ Microbiol **70:**1830-1832.

13. **Debroy C, Roberts E, Valadez AM, Dudley EG, Cutter CN.** 2011. Detection of Shiga Toxin-Producing *Escherichia coli* O26, O45, O103, O111, O113, O121, O145, and O157 serogroups by multiplex polymerase chain

reaction of the *wzx* gene of the O-Antigen gene cluster. Foodborne Pathog Dis **8:**651-652.

14. **Valadez AM, Debroy C, Dudley E, Cutter CN.** 2011. Multiplex PCR detection of Shiga toxin-producing *Escherichia coli* strains belonging to serogroups O157, O103, O91, O113, O145, O111, and O26 experimentally inoculated in beef carcass swabs, beef trim, and ground beef. J Food Prot **74:**228-239.

15. **Beutin L, Jahn S, Fach P.** 2009. Evaluation of the 'GeneDisc' real-time PCR system for detection of enterohaemorrhagic *Escherichia coli* (EHEC) O26, O103, O111, O145 and O157 strains according to their virulence markers and their O- and H-antigen-associated genes. J Appl Microbiol **106:**1122-1132.

16. **Samuel G, Reeves P.** 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res **338:**2503-2519.

17. **Wang L, Curd H, Qu W, Reeves PR.** 1998. Sequencing of *Escherichia coli* O111 O-antigen gene cluster and identification of O111-specific genes. J Clin Microbiol **36:**3182-3187.

18. **Fratamico PM, Briggs CE, Needle D, Chen CY, DebRoy C.** 2003. Sequence of the *Escherichia coli* O121 O-antigen gene cluster and detection of enterohemorrhagic *E. coli* O121 by PCR amplification of the *wzx* and *wzy* genes. J Clin Microbiol **41:**3379-3383.

19. **DebRoy C, Fratamico PM, Roberts E, Davis MA, Liu Y.** 2005. Development of PCR assays targeting genes in O-antigen gene clusters for

detection and identification of *Escherichia coli* O45 and O55 serogroups. Appl Environ Microbiol **71:**4919-4924.

20. **Tsai TY, Lee WJ, Huang YJ, Chen KL, Pan TM.** 2006. Detection of viable enterohemorrhagic *Escherichia coli* O157 using the combination of immunomagnetic separation with the reverse transcription multiplex TaqMan PCR system in food and stool samples. J Food Prot **69:**2320-2328.

21. **Fratamico PM, Bagi LK, Cray WC, Narang N, Yan X, Medina M, Liu Y.** 2011. Detection by multiplex real-time polymerase chain reaction assays and isolation of Shiga Toxin-Producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 in ground beef. Foodborne Pathog Dis **8:**601-607.

22. **Dunbar SA.** 2006. Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. Clin Chim Acta **363:**71-82.

23. **Bio-Rad.** InstaGene Matrix: Instruction Manual. Bio-Rad, Hercules, CA.

24. **Fulton RJ, McDade RL, Smith PL, Kienker LJ, Kettman JR.** 1997. Advanced multiplexed analysis with the FlowMetrix system. Clin Chem **43:**1749-1756.

25. **Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P.** 2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol **45:**3323-3334.

26. **McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML, Fields PI.** 2011. Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. J Clin Microbiol **49:**565-573.

27. **Cebula TA, Payne WL, Feng P.** 1995. Simultaneous identification of strains of *Escherichia coli* serotype O157:H7 and their Shiga-like toxin type by mismatch amplification mutation assay-multiplex PCR. J Clin Microbiol **33:**248-250.

28. **Paton AW, Paton JC.** 1998. Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx1*, *stx2*, *eaeA*, enterohemorrhagic *E. coli hlyA*, *rfb*O111, and *rfb*O157. J Clin Microbiol **36:**598-602.

29. **Lin A, Sultan O, Lau HK, Wong E, Hartman G, Lauzon CR.** 2011. O serogroup specific real time PCR assays for the detection and identification of nine clinically relevant non-O157 STECs. Food Microbiol **28:**478-483.

30. **Liu Y, Fratamico P.** 2006. *Escherichia coli* O antigen typing using DNA microarrays. Mol Cell Probes **20:**239-244.

31. **Ballmer K, Korczak BM, Kuhnert P, Slickers P, Ehricht R, Hächler H.** 2007. Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. J Clin Microbiol **45:**370-379.

32. **Dunbar SA, Vander Zee CA, Oliver KG, Karem KL, Jacobson JW.** 2003. Quantitative, multiplexed detection of bacterial pathogens: DNA and protein applications of the Luminex LabMAP system. J Microbiol Methods **53:**245-252.

33. **Barco L, Lettini AA, Dalla Pozza MC, Ramon E, Fasolato M, Ricci A.** 2010. Fluoroquinolone resistance detection in *Campylobacter coli* and *Campylobacter jejuni* by Luminex xMAP technology. Foodborne Pathog Dis **7:**1039-1045.

References in Chapter III

1. **Wells JG, Davis BR, Wachsmuth IK, Riley LW, Remis RS, Sokolow R, Morris GK.** 1983. Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype. J. Clin. Microbiol. **18 (3):**512-520.
2. **Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J.** 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J. Clin. Microbiol. **42 (1):**311-319.
3. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerging Infect. Dis. **17 (1):**7-15.
4. **Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Reid-Smith R, Isaac-Renton J, Clark C, Rahn K, Kaper JB.** 2003. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J. Clin. Microbiol. **41(11):**4930-4940.

5. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A.** 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J. Bacteriol. **169 (12):**5429-5433.

6. **Barrangou R.** 2013. CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip Rev RNA **4 (3):**267-278.

7. **Barrangou R, Horvath P.** 2012. CRISPR: New Horizons in Phage Resistance and Strain Identification. Annu Rev Food Sci Technol **3:**143-162.

8. **Bhaya D, Davison M, Barrangou R.** 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu. Rev. Genet. **45:**273-297.

9. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315 (5819):**1709-1712.

10. **Garneau JE, Dupuis M, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S.** 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468 (7320):**67-71.

11. **Marraffini LA, Sontheimer EJ.** 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322 (5909):**1843-1845.

12. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327 (5962):**167-170.

13. **Al-Attar S, Westra ER, van der Oost J, Brouns SJJ.** 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biological Chemistry **392 (4):**277-289.

14. **Bolotin A, Quinquis B, Sorokin A, Ehrlich SD.** 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151(Pt 8):**2551-2561.

15. **Mojica FJ, Díez-Villaseñor C, García-Martínez J, Soria E.** 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. **60 (2):**174-182.

16. **Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321 (5891):**960-964.

17. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5 (6):**e11126.

18. **Yin S, Jensen MA, Bai J, Debroy C, Barrangou R, Dudley EG.** 2013. The evolutionary divergence of Shiga Toxin-Producing *Escherichia coli* is reflected in Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) spacer composition. Appl. Environ. Microbiol. **79 (18):**5710-5720.

19. **Díez-Villaseñor C, Almendros C, García-Martínez J, Mojica FJ.** 2010. Diversity of CRISPR loci in *Escherichia coli*. Microbiology **156 (Pt 5):**1351-1361.

20. **Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. **9(6):**467-477.

21. **Touchon M, Charpentier S, Clermont O, Rocha EP, Denamur E, Branger C.** 2011. CRISPR Distribution within the *Escherichia coli* Species is not suggestive of immunity-associated diversifying selection. J. Bacteriol. **193 (10):**2460-2467.

22. **Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S.** 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. **190 (4):**1390-1400.

23. **Lawrence JG.** 2005. Common themes in the genome strategies of pathogens. Curr. Opin. Genet. Dev. **15 (6):**584-588.

24. **Coombes BK, Gilmour MW, Goodman CD.** 2011. The evolution of virulence in non-O157 shiga toxin-producing *Escherichia coli*. Front Microbiol **2:**90.

25. **Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T.** 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **106 (42):**17939-17944.

26. **Touchon M, Charpentier S, Pognard D, Picard B, Arlet G, Rocha EP, Denamur E, Branger C.** 2012. Antibiotic resistance plasmids spread among natural isolates of *Escherichia coli* in spite of CRISPR elements. Microbiology **158 (Pt 12):**2997-3004.

27. **Lindenstrauß AG, Pavlovic M, Bringmann A, Behr J, Ehrmann MA, Vogel RF.** 2011. Comparison of genotypic and phenotypic cluster analyses of virulence determinants and possible role of CRISPR elements towards their incidence in *Enterococcus faecalis* and *Enterococcus faecium*. Syst. Appl. Microbiol. **34 (8):**553-560.

28. **Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R.** 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. **190 (4):**1401-1412.

29. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35:**W52-57.

30. **Ju W, Shen J, Toro M, Zhao S, Meng J.** 2013. Distribution of Pathogenicity Islands OI-122, OI-43/48, OI-57 and High Pathogenicity Island (HPI) in Shiga Toxin-producing *Escherichia coli*. Appl. Environ. Microbiol. **79 (11):**3406-3412.

31. **Ju W, Cao G, Rump L, Strain E, Luo Y, Timme R, Allard M, Zhao S, Brown E, Meng J.** 2012. Phylogenetic analysis of non-O157 Shiga toxin-

producing *Escherichia coli* by whole genome sequencing. J. Clin. Microbiol. **50 (12):**4123-4127.

32. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28 (10):**2731-2739.

33. **Delannoy S, Beutin L, Fach P.** 2012. Use of CRISPR sequence polymorphisms for specific detection of enterohemorrhagic *E. coli* (EHEC) strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28 and O157:H7 by real time PCR. J Clin Microbiol. **50 (12):** 4035-4040.

34. **Delannoy S, Beutin L, Burgos Y, Fach P.** 2012. Specific Detection of Enteroaggregative Hemorrhagic *Escherichia coli* O104:H4 Strains by Use of the CRISPR Locus as a Target for a Diagnostic Real-Time PCR. J. Clin. Microbiol. **50 (11):**3485-3492.

35. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8 (2):**e55687.

36. **Edgar R, Qimron U.** 2010. The *Escherichia coli* CRISPR system protects from λ lysogenization, lysogens, and prophage induction. J. Bacteriol. **192 (23):**6291-6294.

38. **Palmer KL, Gilmore MS.** 2010. Multidrug-resistant Enterococci lack CRISPR-*cas*. MBio **1 (4):**e00227-10.

38. **Mojica FJ, Díez-Villaseñor C.** 2010. The on-off switch of CRISPR immunity against phages in *Escherichia coli*. Mol Microbiol **77 (6):**1341-1345.

39. **Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E.** 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun **3:**945.

40. **Tarr PI, Schoening LM, Yea YL, Ward TR, Jelacic S, Whittam TS.** 2000. Acquisition of the *rfb-gnd* cluster in evolution of *Escherichia coli* O55 and O157. J Bacteriol **182 (21):**6183-6191.

References in Chapter IV

1. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.

2. **Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ.** 2013. Outbreak-associated *Salmonella enterica* serotypes and food Commodities, United States, 1998-2008. Emerg Infect Dis **19:**1239-1244.

3. **[CDC] Centers for Disease Control and Prevention.** 2012. Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet Surveillance Report for 2011 (Final Report). Available in: http://www.cdc.gov/foodnet/PDFs/2011_annual_report_508c.pdf, accessed on January 2014.

4. **Grimont PA, Weil F-X.** 2007. Antigenic formulae of the *Salmonella* serovars, 9th Edition ed, Paris, France: WHO Collaborative Centre for Reference and Research on *Salmonella*.

5. **Boyle EC, Bishop JL, Grassl GA, Finlay BB.** 2007. *Salmonella*: from pathogenesis to therapeutics. J Bacteriol **189:**1489-1495.

6. **McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML, Fields PI.** 2011. Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. J Clin Microbiol **49:**565-573.

7. **[CDC] Centers for Disease Control and Prevention.** 2011. National Enteric Disease Surveillance: *Salmonella* Surveillance Overview. Available in: http://www.cdc.gov/nationalsurveillance/PDFs/NationalSalmSurveillOverview_508.pdf, accessed January 2014.

8. **Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P.** 2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol **45:**3323-3334.

9. **Shi C, Singh P, Ranieri ML, Wiedmann M, Moreno Switt AI.** 2013. Molecular methods for serovar determination of *Salmonella*. Crit Rev Microbiol. [Epub ahead of print, Posted online on November 14, 2013].

10. **Wattiau P, Boland C, Bertrand S.** 2011. Methodologies for *Salmonella enterica* ssp *enterica* subtyping: gold standards and alternatives. Appl Environ Microbiol **77** (22):7877-7885.

11. **Barrangou R.** 2013. CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip Rev RNA **4:**267-278.

12. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167-170.

13. **Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321:**960-964.

14. **Bolotin A, Quinquis B, Sorokin A, Ehrlich SD.** 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151:**2551-2561.

15. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

16. **Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ.** 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol **18:**529-536.

17. **Tyson GW, Banfield JF.** 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol **10:**200-207.

18. **Karginov FV, Hannon GJ.** 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell **37:**7-19.

19. **Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV, Balakhonov SV, Wang X, Song Y, Anisimov AP,**

Vergnaud G, Yang R. 2008. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. PLoS One **3:**e2652.

20. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.

21. **Delannoy S, Beutin L, Fach P.** 2012. Use of CRISPR sequence polymorphisms for specific detection of enterohemorrhagic *E. coli* (EHEC) strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28 and O157:H7 by real time PCR. J Clin Microbiol **50** (12): 4035-4040.

22. **Yin S, Jensen MA, Bai J, Debroy C, Barrangou R, Dudley EG.** 2013. The evolutionary divergence of Shiga Toxin-Producing *Escherichia coli* is reflected in Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Spacer Composition. Appl Environ Microbiol **79:**5710-5720.

23. **Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, de Romans S, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth-Jeanjean S, Achtman M, Brisse S, Sola C, Weill FX.** 2012. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. PLoS One **7:**e36995.

24. **Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG.** 2011. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella* enterica subsp. enterica. Appl Environ Microbiol **77:**1946-1956.

25. **Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R.** 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol **190:**1401-1412.

26. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

27. **Touchon M, Charpentier S, Clermont O, Rocha EP, Denamur E, Branger C.** 2011. CRISPR Distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. J Bacteriol **193:**2460-2467.

28. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-2948.

29. **Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A, Dougan G, Harrison LH, Brisse S, Group SEMS.** 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog **8:**e1002776.

30. **Fricke WF, Mammel MK, McDermott PF, Tartera C, White DG, Leclerc JE, Ravel J, Cebula TA.** 2011. Comparative Genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage vvolution. J Bacteriol **193:**3556-3568.

31. **Liu F, Kariyawasam S, Jayarao BM, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG.** 2011. Subtyping *Salmonella enterica* serovar Enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). Appl Environ Microbiol **77:**4520-4526.

32. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8:**e55687.

33. **Harbottle H, White DG, McDermott PF, Walker RD, Zhao S.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization *of Salmonella enterica* serotype Newport isolates. Journal of Clinical Microbiology **44:**9.

34. **Dimarzio M, Shariat N, Kariyawasam S, Barrangou R, Dudley EG.** 2013. Antibiotic resistance in *Salmonella* Typhimurium associates with CRISPR sequence type. Antimicrob Agents Chemother. **57** (9): 4282-4289.

35. **Zhao S, McDermott PF, Friedman S, Qaiyumi S, Abbott J, Kiessling C, Ayers S, Singh R, Hubert S, Sofos J, White DG.** 2006. Characterization of antimicrobial-resistant *Salmonella* isolated from imported foods. J Food Prot **69:**500-507.

36. **Melendez SN, Hanning I, Han J, Nayak R, Clement AR, Wooming A, Hererra P, Jones FT, Foley SL, Ricke SC.** 2010. *Salmonella enterica* isolates from pasture-raised poultry exhibit antimicrobial resistance and class I integrons. J Appl Microbiol **109:**1957-1966.

37. **Kumar R, Surendran PK, Thampuran N.** 2009. Analysis of antimicrobial resistance and plasmid profiles in *Salmonella* serovars associated with tropical seafood of India. Foodborne Pathog Dis **6:**621-625.