

## ABSTRACT

Title of Document

GENOMIC ANALYSIS OF BACTERIOPHAGES  
FROM NON-O157 SHIGA TOXIN-PRODUCING  
*ESCHERICHIA COLI*

Shuai Tang, Master of Science, 2015

Directed By:

Professor, Jianghong Meng, NFSC

Shiga toxin-producing *Escherichia coli* (STEC) is the fiercest pathotype among all diarrheic *E. coli*. STEC O157:H7 has been a predominant serotype for STEC in the United States. However, an increasing number of cases of infections by STEC other than O157 have been reported in recent years. Shiga toxin (Stx) is the most important virulence factor of STEC and is encoded by *stx*, which is introduced into STEC genome by bacteriophages through gene transduction. A detailed understanding about Stx bacteriophages is necessary to reveal the emergence and pathogenicity of STEC. The very unstable genomes of Stx bacteriophages result in a dynamic phenotypic versatility including virulence, host cell repertoire and tolerance to adversities. Sequencing technology enables us to generate genomic sequence data of bacteriophages at an affordable cost. The project aimed at obtaining genomic DNA sequences of Stx bacteriophages of non-O157 STEC isolates of diverse serotypes. Thirteen bacteriophages were successfully induced from 83 STEC isolates of serotypes O74, O111, O121, O130, O163, O179 and O183. The bacteriophage DNA samples were collected and sequenced using MiSeq Desktop Sequencer (Illumina®, Inc). Automatically assembled sequences were manually compared to *E. coli* genome sequence available from NCBI (NC\_000913.3) to verify the reliability of the sequencing results. Nine verified bacteriophage sequences were aligned to two Stx bacteriophage genomes of NCBI (NC\_000924.1 and NC\_018846.1) and visible alignment results were obtained. A phylogenetic relationship of the nine phages and the two reference sequences was constructed and gene profiles of each sample sequences were identified. The comparative analysis indicated that recombination events occurred in probacteriophages showed traces. Similarity of bacteriophage genomes correlated to the serotypes of

host bacteria based on the comparison of phylogenetic tree and STEC serotypes. Gene identification results showed that nucleotide variance does not show region specificity, silent mutations are frequent in housekeeping genes and virulence genes are conservative among phage samples.

GENOMIC ANALYSIS OF BACTERIOPHAGES FROM NON-O157 SHIGA TOXIN-PRODUCING

*ESCHERICHIA COLI*

By

Shuai Tang

Thesis 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 Master of Science

2015

Advisory Committee:

Professor Jianghong Meng, Chair

Associate Professor Qin Wang

Assistant Professor Seong-Ho Lee

©Copyright by

Shuai Tang

2015

## Acknowledgements

First of all, I would like to express my sincere gratitude to my advisor Dr. Jianghong Meng for giving me an opportunity to study in the University of Maryland, College Park. Thanks for Dr. Meng's guidance in research and for his inspiration in life. Dr. Meng shows me the world of food safety and inspires me the prospects in this area. I cannot imagine how to complete my thesis without his guidance and help. I also want to express my deepest thanks to Dr. Seong-Ho Lee and Dr. Qin Wang for serving on my committee and supporting my research and study.

I also want to express my great appreciation to Dr. Lydia Rump of University of Maryland, College Park for her assistance through numerous discussions and suggestions about my research from experiment design to troubleshooting. I am also grateful to the kindness from Dr. Guojie Cao of University of Maryland, College Park, who guides me in sequencing technology and supports me in data analysis.

In addition, I acknowledge all the help and discussions from my previous and current labmates: Dr. Magaly Toro, Dr. Fei Wang, Dr. Baotao Liu, and Dr. Wentao Xu. It is a great honor for me to have worked in this wonderful team.

Most important, I want to thank my parents in China for their love and encouragements during my study. Thanks to my girlfriend Xiaochen Lin for her love, understanding and support during the two years of being separated in two countries.

# Table of Contents

Acknowledgements.....	ii
Table of Contents.....	iii
List of Tables.....	iv
List of Figures.....	v
CHAPTER 1 Introduction.....	- 1 -
1. Microbiology of Shiga toxin-producing <i>Escherichia coli</i> (STEC).....	- 1 -
2. Overview of Shiga toxin bacteriophage.....	- 6 -
3. DNA Sequencing Technology.....	- 13 -
Study Objectives.....	- 15 -
CHAPTER 2 Materials and Methods.....	- 16 -
1. Description of the non-O157 STEC isolates.....	- 16 -
2. Bacteriophage Genome Collection.....	- 16 -
3. Bacteriophage Genome Confirmation.....	- 19 -
4. Bacteriophage Genome Sequencing.....	- 20 -
5. Data Analysis.....	- 22 -
CHAPTER 3 Results.....	- 24 -
1. Induction of STEC Isolates.....	- 24 -
2. Bacteriophage Genome Extraction.....	- 26 -
3. DNA Cleaning Result.....	- 27 -
4. Global Alignment.....	- 28 -
5. Phylogenetic Relationship.....	- 38 -
6. Gene Annotation.....	- 38 -
CHAPTER 4 Discussion.....	- 40 -
1. Reasons of Low Induction Rate.....	- 40 -
2. Acquisition of Bacteriophage Genomes.....	- 42 -
3. Bacteriophage Genome Analysis.....	- 43 -
CHAPTER 5 Conclusion.....	- 46 -
Reference.....	- 47 -

## List of Tables

Table 2.1 Oligonucleotide Primers Used for PCR Tests .....	- 20 -
Table 3.1 Summary of Inducible non-O157 STEC Isolates .....	- 24 -
Table 3.2 Summary of non-O157 STEC Isolates with Reliable Bacteriophage Genome Sequences....	- 28 -

## List of Figures

Figure 1.1. Genomic Structure of Stx Bacteriophage .....	- 10 -
Figure 3.1. Induction rate of STEC isolates of different serotypes.....	- 25 -
Figure 3.2. Induction rate of STEC isolates of different origins.....	- 25 -
Figure 3.3. The electrophoresis result of PCR reaction targeting at <i>stx</i> .....	- 26 -
Figure 3.4. The electrophoresis result of PCR reaction targeting at <i>stx</i> -flanking fragment.....	27 -
Figure 3.5. Alignment between Bacteriophage Genome Came from STEC KSU 1234-1 to Bacteriophage 933W.....	- 29 -
Figure 3.6. Alignment between Bacteriophage Genome Came from STEC KSU 1234-1 to Bacteriophage p13374 .....	- 29 -
Figure 3.7. Alignment between Bacteriophage Genome Came from STEC KSU 7740-2 to Bacteriophage 933W.....	- 30 -
Figure 3.8. Alignment between Bacteriophage Genome Came from STEC KSU 7740-2 to Bacteriophage p13374 .....	- 30 -
Figure 3.9. Alignment between Bacteriophage Genome Came from STEC KSU 7756-1 to Bacteriophage 933W.....	- 31 -
Figure 3.10. Alignment between Bacteriophage Genome Came from STEC KSU 7756-1 to Bacteriophage p13374 .....	- 31 -
Figure 3.11. Alignment between Bacteriophage Genome Came from STEC KSU 7766-1 to Bacteriophage 933W.....	- 32 -
Figure 3.12. Alignment between Bacteriophage Genome Came from STEC KSU 7766-1 to Bacteriophage p13374 .....	- 32 -
Figure 3.13. Alignment between Bacteriophage Genome Came from STEC KSU 8266-1 to Bacteriophage 933W.....	- 33 -
Figure 3.14. Alignment between Bacteriophage Genome Came from STEC KSU 8266-1 to Bacteriophage p13374 .....	- 33 -
Figure 3.15. Alignment between Bacteriophage Genome Came from STEC KSU 9042-1 to Bacteriophage 933W.....	- 34 -

Figure 3.16. Alignment between Bacteriophage Genome Came from STEC KSU 9042-1 to Bacteriophage p13374 ..... - 34 -

Figure 3.17. Alignment between Bacteriophage Genome Came from STEC KSU 13802-1 to Bacteriophage 933W..... - 35 -

Figure 3.18. Alignment between Bacteriophage Genome Came from STEC KSU 13802-1 to Bacteriophage p13374 ..... - 35 -

Figure 3.19. Alignment between Bacteriophage Genome Came from STEC KDHE37 to Bacteriophage 933W..... - 36 -

Figure 3.20. Alignment between Bacteriophage Genome Came from STEC KDHE37 to Bacteriophage p13374 ..... - 36 -

Figure 3.21. Alignment between Bacteriophage Genome Came from STEC FE95381 to Bacteriophage 933W..... - 37 -

Figure 3.22. Alignment between Bacteriophage Genome Came from STEC FE95381 to Bacteriophage p13374 ..... - 37 -

Figure 3.23. Phylogenetic Tree of Sample Genomes and Two Reference Genomes ..... - 38 -

Figure 3.24. Existence rate of genes from reference genomes in the bacteriophage genome samples.. - 39 -

## List of Abbreviations

ATCC	American type culture collection
ATM	Amplicon tagment mix
BLAST	Basic local alignment search tool
Blastn	Basic local alignment search tool -- nucleotide
Blastp	Basic local alignment search tool -- protein
DC	Dendritic cell
ddNTP	Dideoxynucleotide
DLA	Double layer agar
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double strand DNA
EDTA	Ethylenediaminetetraacetic acid
EHEC	Enterohemorrhagic <i>E. coli</i>
EtBr	Ethidium bromide
Gb3	Globotriaosylceramide
HT1	Hybridization buffer
HS	High sensitivity

HUS	Humolytic uremic syndrome
ICTV	International committee on taxonomy of viruses
IL-1beta	Interleukin-1beta
KSU	Kansas State University
LB	Luria broth
LEE	Locus of enterocyte effacement
LNA1	Library normalization additives 1
LNB1	Library normalization beads 1
LNS1	Library normalization storage buffer 1
LNW1	Library normalization wash 1
LPS	Lipopolysaccharide
MEGA	Molecular evolutionary genetics analysis
NCBI	National center of biotechnology information
NGS	Next generation sequencing
NPM	Nextera PCR master mix
NT	Neutralize tagment buffer
OD	Optical density
OI	O-islands
OmpA	Outer membrane protein A
PAI	Pathogenicity island

PCR	Polymerase chain reaction
PEG8000	Polyethylene glycol 8000
psi	Pounds per square inch
RNA	Ribonucleic acid
RSB	Resuspension buffer
SBH	Sequencing by hybridization
SBL	Sequencing by ligation
SBS	Sequencing by synthesis
ssDNA	Single strand DNA
STEC	Shiga toxin-producing <i>E. coli</i>
Stx	Shiga toxin
T3SS	Type III secretion system
TAE	Tris-acetate-EDTA
THL	National Institute for Health and Welfare, Finland
Tir	Translocated intimin receptor
TMP	Trimethoprim
Tris-Cl	Hydroxymethylaminoethane chloride
UV	Ultraviolet

## CHAPTER 1 Introduction

### 1. Microbiology of Shiga toxin-producing *Escherichia coli* (STEC)

STEC were first isolated among pathogenic *E. coli* that showed toxicity on Vero cells in 1977 [1]. They were identified as foodborne pathogens when *E. coli* O157:H7 caused two outbreaks associated with ground beef in 1982. The main transmission route of STEC is through contaminated food and water, especially through not thoroughly cooked beef and unpasteurized milk. Ruminants are the natural carriers of STEC, among which cattle are the most important carrier. Although animal carriers do not show syndromes due to the lacking of Shiga toxin receptor, they are potential sources for human infections [2]. One study showed that arthropod could also serve as carriers in transmitting STEC [3]. In addition, STEC infections can occur through direct contact with animals and among people. Such transmissions have been reported at hospitals, nursing homes and daycare centers [4].

#### 1.1. The Epidemiologic Patterns of Different STEC Serotypes

The most important single serotype of STEC is O157:H7 [5]. *E. coli* O157:H7 is extremely virulent and widely spread [6]. Outbreaks caused by O157:H7 have occurred worldwide in the past 30 years. The control and treatment of *E. coli* O157 infections has been and will continue to be challenging in years to come [6].

Non-O157 STEC is capturing more and more attentions due to the rising foodborne poisoning cases. Currently, more than 200 serotypes of STEC other than O157 have been verified to cause human diseases [7]. The infection rate increased by 60.5% from 2002 to 2005 based on the investigating results over 35 countries [9]. The increase of O157 STEC infecting rate was 13% by comparison [8]. An O104:H4 STEC outbreak that spread over half Europe and North America and caused over 50 mortalities was the most destructive non-O157 outbreak ever [4].

The non-O157 STEC serotypes show geographical variance. In North America, the common non-O157 STEC serotypes contain O26, O45, O103, O111 and O121 [9]. While in Europe O45, O111 and

O121 STEC are rarely recorded; O103, O145 and O157: H<sup>-</sup> STEC are the other two common serotypes that are highly concerned in the Europe [10]. The dynamic patterns of the non-O157 STEC imply a future research direction in public health and food safety areas.

## 1.2. Virulence Factors and Pathogenicity of STEC

The pathogenicity of STEC is mainly expressed through relevant virulence factors. The principal factor is Shiga toxin (Stx) that is encoded by *stx* [11]. The pathogenicity of the STEC does not entirely rely on the activity of *stx*. Other virulence factors are also involved. Important relevant factors include intimin that is encoded by the *eae* located at chromosomal LEE pathogenicity island; fimbriae, membrane proteins, and lipopolysaccharides (LPS) which affect the adherence; and EHEC-hemolysin that is encoded by genes on pathogenic plasmid pO157 [5].

### 1.2.1. Shiga Toxin

Stx belongs to one of the most virulent toxins to humans [12]. By combining to endothelial cells at specific binding sites, Stx can present strong cytotoxicity, enterotoxicity, and neurotoxicity depending on the certain cells that being combined [13]. Consequently, multiple clinical conditions may occur and severe syndromes may happen. Current researches about the virulence factors mainly focus on the Stx. The Stx is a kind of compound toxin which is assembled by A and B subunits. The B subunit is a pentamer that serve as adhesive to bind to the galactose-(alpha 1-4)-galactose on the globotriaosylceramide (Gb3) receptors [14]. The binding initiates the endocytosis which allows the Stx to enter the host cell. The A subunit, which is the virulent unit, is then activated as RNA N-glycosidase to break the N-glycosidic bond of the 4324 adenine residue from the 5' ends on the 28S rRNA. The remove of the adenine nucleobase results in the disassembly of the 60S ribosome and the halt of protein synthesis [15].

The Stx family can be divided into 2 groups, Stx1 and Stx2. Immunologic cross reaction between the 2 groups is not detected and the STEC may contain either Stx1, Stx2 or both. Previous discoveries

have identified the possibility that one strain may carry 1 type of *stx1* and 2 types of *stx2* at the same time [16]. Although the 2 Stx groups do not share same antigenicity, the molecular structures, receptor structures and function mechanisms of them are highly similar. All *stx* locates in the probacteriophage genome and is able to spread among bacteria through transduction. Strains only carrying *stx2* have been proven to be more virulent than strains carrying both *stx1* and *stx2* or the *stx1*-carrying strains. According to the research focusing on the difference of cytotoxicity, the median lethal dose (LD50) of Stx2 on the microvascular endothelial cells of human glomerulus is 1/400 of the LD50 of Stx1, which indicates that the research on Stx2 is more significant for clinical practices [17]. Another research shown a significant association between *stx2* and the severity of diseases based on the univariate analysis and multivariate logistic regression analysis of 237 STEC isolates [18]. The statistical conclusion may be attributed to the higher transcription level of *stx2 in vivo* and the differences of binding sites on cell surface [18].

### 1.2.2. Adherence Factors

Adhering to gastrointestinal mucosa is the premise that STEC can show pathogenicity. The adherence is accomplished mainly through intimin that is encoded by *eae* with the assistance from other surface components and physical structures of bacteria surface, especially fimbriae and flagella.

#### 1.2.2.1. Intimin

Intimin is a 95kDa outer membrane protein encoded by the *eae* that is located in the Locus of Enterocyte Effacement (LEE). The binding of intimin to translocated intimin receptor (Tir) will initiate cellular signal transduction to reorganize the cytoskeleton and sequentially form specific lesions on cells. Consequently, bacteria are able to adhere to the gastrointestinal epithelium [19]. Both *eae* and *stx2* are widely detected in STEC strains; almost all STEC isolates from patients carry *eae*. The necessity of *eae* on pathogenicity is strengthened by the fact that the *eae*<sup>-</sup> O157:H7 mutant cannot adhere to epithelial cells unless the plasmids carrying normal *eae* are inducted [20]. Intimin is proved to be an essential factor which is involved in causing HUS, even the exact mechanism needs more detailed research [21].

#### 1.2.2.2. Fimbriae

Fimbriae are important to the adhesion of bacteria on the gastrointestinal epithelium. F18 fimbriae are identified as one of the major virulence factors that cause edema disease in weaned pigs. STECs that are responsible to edema disease adhere to gastrointestinal epithelium more intensely than general strains due to the contribution of F18 fimbriae variants [22]. A research revealed that the sensitivity of weaned pigs to STEC is significantly related to existence of F18 fimbriae receptor on gastrointestinal epithelium [23].

#### 1.2.2.3. Outer Membrane Protein

A research on outer membrane protein A (OmpA) of *E. coli* O157:H7 revealed that OmpA is responsible of the initiation of cytokine production by dendritic cells (DCs) as well as the migration of DCs toward infected tissues. Thus, OmpA is an important stimulator of immune response induced by STEC O157:H7 [24]. Outer membrane protein of STEC is responsible for the binding of bacterium to certain cells. Research revealed that the binding between STEC O157:H7 to HEp-2 cell is inhibited by specific outer membrane protein, while the adhesion of STEC O157:H7 to Henle407 cell is mediated by outer membrane protein [25]. The function of controlling adhesion makes outer membrane protein an important pathogenic factor.

#### 1.2.2.4. Lipopolysaccharide (LPS)

The research of the function of LPS on STEC adhesion is still unclear. Paton indicated that the O antigen on LPS increased adherence of STEC O113 and O157 to HEp-2 and Henle 407 cells in vitro [26]. In contrast, another research indicated that the adherence of LPS- mutants to HEp-2 cells were significantly enhanced, which implied that polysaccharide side chains are not required for attaching and effacing [28].

#### 1.2.2.5. Enterohemorrhagic *E. coli* (EHEC) Hemolysin

EHEC hemolysin is the most important virulence factor encoded by pO157 plasmid. The vulnerable targets of hemolysin includes lymphocytes, granulocytes and erythrocytes. Hemolysin forms pores on the membrane of target cells leading to cell death. EHEC hemolysin is the specific toxin produced by STEC [30]. Detailed function of EHEC hemolysin is the induction of interleukin-1beta (IL-1beta) to mediate the expression of Gb3 receptor, which results in elevating erythrocyte lysis [30].

### 1.3. Differences between STEC Serotypes

The first identified STEC outbreak caused by *E. coli* O157:H7 in 1982, USA. The result of sample investigation on isolates recorded since 1980s shown that this pathogen was an unusual newcomer during that time. In the following decade, *E. coli* O157:H7 caused outbreaks in UK, Europe continent, Africa and New Zealand sequentially and became a common pathogen detected from meat, produce and water resource [31]. The O157:H7 STEC got reputation soon because of the ability to cause outbreaks and Hemolytic Uremic Syndrome (HUS). The violent virulence of O157:H7 STEC arises concern that non-O157 serotypes have the potential to infect human beings either. Until now, about 150 non-O157 serotypes have been confirmed to be associated with sporadic and epidemic infections that caused watery diarrhea, bloody diarrhea and even HUS and death [32].

The term STEC indicates the *E. coli* which contains at least 1 Shiga toxin gene that encodes products responsible to HUS in the bacterial genome. Although some non-O157 STEC cause symptoms similar to O157:H7 STEC, other non-O157 STEC have been confirmed unable to cause HUS, outbreaks even clinical manifestations [33]. The pathogenic diversity obscures the identification of high-risk STECs from patients, animal samples, food products and environment. The clinical spectrum of non-O157 STEC infection is still an unclear problem. The current molecular genetics methods cannot reliably identify the risk level of non-O157 STEC. Although a trustworthy identification method will be significant for public health, the research gets little meaningful improvement after the world health organization assigned the repaid identification of virulent non-O157 STEC as a public health priority 16 years ago [34].

#### 1.4. Virulence of non-O157 STEC

The sequencing result of 3 non-O157 STEC serotypes, the O26, O111 and O103 serotypes, prompted the revival of the notion of parallel evolution. The 3 above serotypes are emerging agents of outbreak and HUS in many countries. These 3 serotypes individually acquired mobile genetic elements of mobile genetic elements in a similar way discovered in *E. coli* O157:H7 that underwent several genetic steps during evolutionary pathway. Although the exact track of getting exogenous genes is unknown, the lambdoid phages and related integrative elements involved in the evolution were identified in the research of *E. coli* O157:H7 regardless of the fact that the 3 non-O157 serotypes separately established functional T3SS [35]. In the process of emerging Stx virulence, about 30 T3SS effectors were acquired. The most interesting part is the horizontally acquired non-LEE-encoded *nle* genes since the maintenance of this gene in the genome is related to the selective forces that the bacterium suffers [35].

Researches revealed that unrelated effectors may work cooperatively to result in a desired effect on the host cell or may oppose effects induced by other effectors [37]. In addition, the interaction of different effectors follows a hierarchical order. The end of the delivery pathway of effectors is the intimin receptor, Tir, which is necessary for successful secretion of T3SS products [38]. Moreover, since the encoded virulent material would be meaningless when secretion system is malfunctioned, an assumption is that the selective force will drive the acquisition of effector genes of similar functions to replace the useless secretion machinery. Until now, no decisive results can prove a hierarchy of achieving effectors and the maintenance of T3SS effectors under neutral pressure. The questions about the origin of emerging non-O157 STEC, the effect of evolutionary direction and whether the evolutionary direction relies on specific environmental and animal reservoirs require more genetic information [37].

#### 2. Overview of Shiga toxin bacteriophage

Bacteriophages are virus that infect and replicate within a bacterium. Traditionally, a main attention has been paid to virus infect either animal or plant. However, bacteriophages are showing an

emerging significance in public health, veterinary medicine and agronomy with the improvement of molecular biology and virology. Stx bacteriophages, the stx carrier of Shiga toxin-producing *Escherichia coli* (STEC), is one of the most famous and learned bacteriophage [40]. The high pathogenicity of STEC motivates the investigation of Stx bacteriophages since STEC outbreaks appeared. The first isolated Stx bacteriophages include Stx1 bacteriophage H19B from STEC O26:H11 H19 and Stx2 bacteriophage 933W from STEC O157:H7 933W [41]. All isolated Stx bacteriophages contain double-strand DNA genome and are similar to lambdoid bacteriophage morphologically. Currently, a complete whole genome sequence has been completed on Stx bacteriophage 933W, VT2-sakai, H19B and VT1-sakai [44].

## 2.1. Life Cycle of Bacteriophages

Bacteriophages may enter either a lytic cycle or a lysogenic cycle after contacting to a bacterium. Through exploiting bacterial nutrition and energy, a lytic cycle results in the death of the infected bacteria and the release of numerous offspring bacteriophages. In contrast, a lysogenic bacteriophage will either integrates the bacteriophage gene into the bacterial genome or maintain bacteriophage gene inside the bacterium as a plasmid. The lysogenic bacteriophage gene will synchronize with the life cycle of the host bacterium. The actions of the bacteriophage gene are simultaneous to the bacterial genetic materials without causing disruption of the host bacterium. A bacteriophage gene that being integrated into bacterial genome is a probacteriophage [48]. Therefore, lysogenic bacteriophages introduce exogenous DNA fragments into bacteria and results in a change in the pathogenicity and phenotype if the integrated genes were related to virulence factors. The genetic diversity and pathogenic variation caused by the interaction between Stx bacteriophage and bacteria are agreed to be the ultimate source of STEC variation [49].

## 2.2. Bacteriophage Taxonomy

The classification of virus is based on the nucleic acid types, host reservoir, morphological and physiological patterns. Currently, seven orders, 96 families, 22 subfamilies, 420 genera and 2618 species

of viruses have been categorized by the International Committee on Taxonomy of Viruses (ICTV) [50]. Bacteriophages belong to either Caudovirales or Nidovirales in the order level. All discovered Stx bacteriophages belong to Caudovirales.

### 2.3. Stx Bacteriophage Morphology

Caudovirales can be divided into three genus: Myoviridae, Podoviridae and Siphoviridae [50]. Caudovirales species may infect either bacteria or archaea. Caudovirales species contain linear dsDNA range from 18kb to 500kb. A virus particle is assembled by a protein envelope head containing linear dsDNA and tail fibers which recognize specific binding receptors on the surface of sensitive bacteria. Once tail fibers attach to receptors of bacteria surface, the bottom of bacteriophage head will bind reversibly to bacteria, and bacteriophage genetic material will be injected through bacterial membrane. The diameter of a typical Caudovirale head ranges from 45 to 170nm, with a molecular weight ranges from  $20 \times 10^6$  to  $600 \times 10^6$ . The types of structural protein components of different Caudovirales species range from 7 to 49. The survival pH for most Caudovirales species ranges from 5 to 9. Caudovirales species have diverse heat sensitivity, most species can be inactivated by heating from 55 °C to 75 °C within 30 minutes. Many Caudovirales species are resistant to ultraviolet light and chloroform [51].

All genus under Caudovirales contain Stx bacteriophage species. Stx bacteriophages show diverse shapes and forms. Stx1 bacteriophage  $\Phi$ H19 was described being assembled by a head of 117nm length and 67nm width and tails of 117nm length and 12nm width. Stx1 bacteriophage  $\Phi$ 30480 has a head of 74nm length and width plus short tails of 26nm in length. A bacteriophage induced from STEC O26 shown irregular icosahedral head and long tails, while bacteriophages from either STEC O157 or STEC O29 have regular icosahedral head and short tails [52]. The most learned Stx bacteriophage 933W contains a regular icosahedral head of about 70nm length, short contractile tails of 27nm in length and 13nm in width [53].

Stx bacteriophages show diverse forms. Either regular or irregular icosahedronic head of the bacteriophage have been observed under electron microscope. The tails of Stx bacteriophages also show diversity. Identified Stx bacteriophages may be categorized into Siphoviridae (regular length, non-contractile tail), Myoviridae (regular length, contractile tail) or Podoviridae (short, non-contractile tail) [50]. Bacteriophages with dissimilar morphologic patterns also differ in genotype and genome sizes [54]. The diversity of Stx bacteriophage impedes the understanding of STEC evolution.

## 2.4. Genome structure and regulation of gene expression

### 2.4.1. Stx Bacteriophage Genome

The first Stx bacteriophage genome sequence was on Stx2 bacteriophage 933W in 1999 [44]. After that, Stx bacteriophages isolated from STEC O26, O29 and O157 were sequenced. The genome size of Stx bacteriophages ranges from 47kb to 70kb [53]. As a member of lambdoid bacteriophage family, Stx bacteriophage show a close genomic structure to other lambdoid bacteriophages. Stx bacteriophages and lambdoid bacteriophages share homologous *cl*, *cro*, *cII*, *O*, *P* and *ren*. The variances mainly happen in late genes and telomerase gene [56]. A probacteriophage genetic map of Stx2 bacteriophage Φ361 and lambdoid bacteriophage shows the remarkable similarity between the genome structures. *stx* locates downstream of the Q and upstream of the lysis genes S and R. The counterparts of *stx* are late genes in lambdoid genome [50].

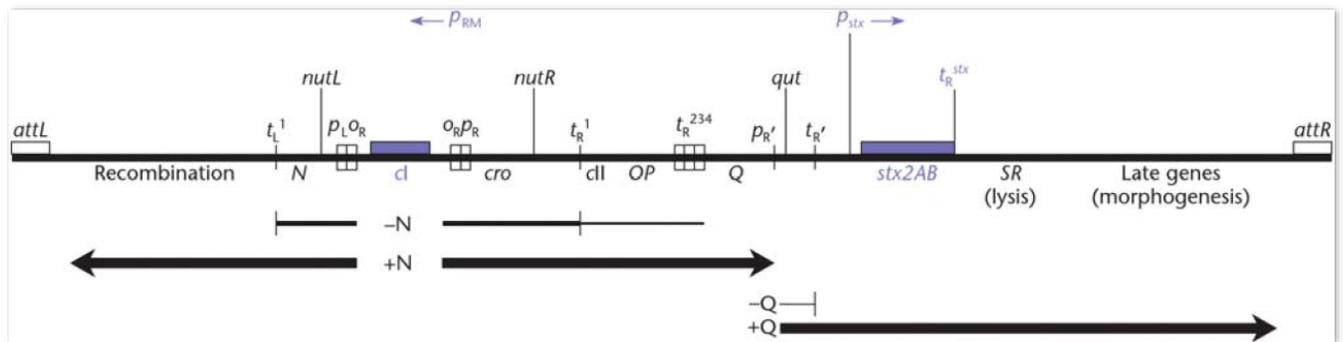


Figure 1.1. Genomic Structure of Stx Bacteriophage

#### 2.4.2. Regulation of Gene Expression in Stx Bacteriophage

Transcription of Stx bacteriophages can be divided into three periods: early, medium and late period. After bacteriophage genome is injected into bacterial host, early transcription is immediately initiated by promoter PL and PR and direct to left and right, respectively. One transcript of left side transcription is an antirepressor N protein. N protein represses terminators of early transcription and enables the initiation of medium transcription to synthesize cII and cIII products. The productivity of CII protein directly decide which cycle the bacteriophage will enter. High concentration of CII motivates bacteriophage DNA to integrate into bacterial genome and represses lysis as well. The activation of stx depends on the encoding of antirepressor Q protein, which activates the transcription of late genes by helping RNA polymerase to go through late transcription terminator. The expression of late genes initiates a lytic cycle and results in the proliferation of offspring lytic bacteriophages and the destruction of bacterial membrane [49]. Shiga toxin encoded in this period will enter environment after cell lysis without a specific secretion mechanism [29]. However, the evolutionary force drives virulence gene to associate with *Q* gene is not clarified.

#### 2.5. Stx Bacteriophage Evolution

*stx* of STEC isolates mainly derive from Stx probacteriophages of Stx probacteriophage-like elements. Although the time and reason that Stx probacteriophage-like elements lost complete genetic structure are unclear, these fragments fail to enter lytic cycle [37]. On the other hand, complete Stx probacteriophages may switch to lytic bacteriophages under certain conditions and release individual viral particles [38]. Multiple genetic recombination events may lead to the coexistence of multiple Stx probacteriophages and Stx probacteriophage-like elements in one bacterial genome. Recombination events destruct the completion of probacteriophage genes or mediate the emergence of chimera probacteriophage from multiple genetic sources [57]. The complete genome sequence on isolated Stx bacteriophages revealed the occurrence of assembly between different Stx bacteriophages. Three genes of the Stx2 bacteriophage  $\Phi$  Min27 isolated from STEC O157:H7 was identified as homologous to SFV

bacteriophage from *Shigella dysenteriae* Sf8401 [58]. Although most Stx probacteriophage will immunize the host bacterium from being infected by other homologous bacteriophages, some Stx bacteriophages lack this characteristic [59]. The possible explanations include the variance of immune region on Stx probacteriophage genome and the existence of antirepressor similar to bacteriophage P22 [60]. Even an antirepressor *ant* homologous to bacteriophage P22 has been discovered in Stx bacteriophage 933W, VT2-a probacteriophage of Sakai O157:H7, NP\_049506 and NP\_050545, the expression and protein activity of the *ant* homologue are not clearly studied yet [61].

Heterology of Stx bacteriophage genomes has been discovered and some hypothetical evolution mechanisms have been presented. When Herold treated STEC O157:H7 EDL933 with norfloxacin to activate the expression of late genes of Stx probacteriophage BP933, another Stx probacteriophage CP933V-like element and a non-Stx probacteriophage were activated simultaneously [62]. The concurrent induction of multiple probacteriophages or probacteriophage-like elements is a prerequisite of the assembly of chimera lytic bacteriophage. The newly assembled Stx bacteriophage may be covered in heterologous protein envelope with heterologous tail fibers [44]. An enlarged host repertoire will be available since the attachment between lytic bacteriophages and host bacteria due to the recognition of bacteriophage tail fiber to specific receptor on bacterial membrane. A successful infection into a novel bacterium allows a Stx bacteriophage to face diverse probacteriophages regularly existing in the host genome. A following recombination will result in highly dynamic variation. The compatibility between heterologous genetic materials makes a rapid variation possible in Stx bacteriophage evolution [37].

## 2.6. *stx*-involved Interactions

A possible source of the virulence of pathogens is external mobile genetic elements such as probacteriophage, transposons, plasmids and genomic islands. The encoding products of pathogenicity island (PAI) is a class of genomic islands that are related to host colonization and diseases. A flexible gene pool in the PAI is involved in evolution and virulence potential as well as the genetic signature of new and emerging pathogens [63]. In STEC, locus of enterocyte effacement (LEE) is a kind of

chromosomal PAI that encodes type III secretion system (T3SS), which is then involved in the attaching and effacing lesion. T3SS is especially a key genetic determinant for successful colonization and persistence of the microorganism inside animal reservoirs as well as a critical virulence factor after human beings are infected through zoonotic transmission [65]. Since the natural hosts of STEC are ruminant animals, a plausible process is that the origin of STEC type III secretion was not highly pathogenic, the current virulent T3SS is the upshot according to variant ecological backgrounds. The evolutionary pathway was described as “coincidental evolution” that indicates the presence of virulent factors came from purposes other than virulence [37]. The result of genetic screen indicates that the substrate of T3SS – the effectors – are responsible to the persistence of both O157 and non-O157 STEC in cattle. In addition, the significance of some effector molecules in colonization, transmission and virulence have been characterized, which reveals the relationship between survival in animals and human virulence of T3SS substrates [66]. Furthermore, the research suggests that the type III secrete-competent strains in animal population may potentially be high-virulent strains for human beings following zoonotic, foodborne and environmental transmission and therefore of great public health significance [37].

The mobile genetic elements in the genome sequence of O157:H7 STEC are not limited in LEE PAI. Many of O-islands (OI) which were considered as lineage specific segments that exist in O157 but not in K-12. A main source of OI is bacteriophage that integrated into bacterial chromosome after virus transduction. These elements immediate colonization and infection after invading human beings. A part of individual genes in OI have been confirmed to be related to epidemic potential and disease severity caused by non-O157 STEC [67]. But until now, the exact function of mobile genetic elements to STEC is still not clear and the thorough identification on non-O157 virulence determinants have not been done yet [37].

## 2.7. Pathogenic functions of Stx bacteriophage

Stx bacteriophages exist under many conditions such as river, sewage and soil [68]. The source of the environmental Stx bacteriophages may be feces containing STEC or Stx bacteriophages. Stx

bacteriophages are more stable than bacteria of the same sample under normal conditions and are more resistant to sterilizer and high pressure than bacteria [68]. The infecting ability of Stx bacteriophages remains stable too. The stability of Stx bacteriophages partly causes stx-containing bacteria ubiquitous. Some gentle hygienic protocols may be invalid to inactivate Stx bacteriophages [69]. The emergence of Shiga toxin-producing bacteria may also be available if Stx bacteriophage is consumed and infected by gastrointestinal bacteria [70].

### 2.8. Induction of Stx Bacteriophage

Low level of certain antibiotics may induce stx-containing bacteria to release Stx bacteriophages and secrete Shiga toxin [71]. Effective inducers include trimethoprim (TMP) and ciprofloxacin [72]. Mitomycin C is widely used in Stx bacteriophage researches and other antibiotics, such as olaquinox and carbadox have been reported to trigger the release of Stx bacteriophage and Shiga toxin secretion [72]. In addition, Stx bacteriophage can also be induced by norepinephrine, hydrogen peroxide and neutrophil, etc [74]. The promotion of bacteriophage induction motivates a wavier antibiotic treatment for both patients and domestic animals. An elevated HUS incidence has been reported in STEC infected children who underwent antibiotics treatment [77]. However, a consensus about the application of antibiotics has not been set yet.

### 3. DNA Sequencing Technology

Sequencing technology enables researchers to obtain genetic materials of species rapidly and exactly, which are critical in revealing physiological mechanisms, molecular interactions and phylogenetic evolution.

The classical Sanger sequencing was developed in 1977. Sanger sequencing is based on chain-terminating dideoxynucleotides (ddNTP) to synthesize DNA fragments of random lengths during in vitro DNA replication. When ddNTP is involved in the replication of DNA strand, a lack of 2'-OH and 3'-OH fails the formation of phosphodiester bond and therefore to terminate DNA elongation. Radionuclide-

labeled ddNTPs which are detectable using autoradiograph can reflect nucleobase type of certain positions according to electrophoresis map [78].

An improvement based on Sanger sequencing in the 1980s was to replace radionuclide labels with fluorescent labels and a following automatic sequencer developed [79]. Other sequencing theory developed in 1990s include pyrosequencing, sequencing by ligation (SBL), and sequencing by hybridization (SBH), etc [80]. These technologies became the basis of next generation sequencing (NGS) technology.

Although classical sequencing technology enabled the sequencing from virus genes to human genome map draft, the complication and cost blocked its wide utilization. In the first decade of 21<sup>st</sup> century, the development of 454 sequencing from Roche, Solexa sequencing from Illumina and SOLiD sequencing from ABI marked the emergence of NGS technology [83]. Comparing to Sanger sequencing which cost 3 billion dollars and three years to complete human genome project, SOLiD may complete sequencing human individual genome within one week [83]. Since NGS technology saves time through shortening target fragments being sequenced, it is proper to re-sequence known genomes. Classical sequencing technology is still needed when sequencing completely new genomes [84].

This project used MiSeq sequencer (Illumina) based on Solexa sequencing technology, which is the most commonly used technology in prokaryotic sequencing.

- 1) DNA library establishment

DNA sample is fragmented into short fragments range from 200bp -500bp. Overhangs are added to the ends of each fragment to establish ssDNA library.

- 2) Attachment

A flow cell which contains 8 channels on the surface. Each channel contains fixed overhangs that may bind to overhangs added on each DNA fragments. Thus DNA fragments will attach to the flow cell in a bridge-like form.

### 3) Bridge Amplification

PCR reaction system will be worked on the flow cell. PCR polymerase recognizes overhangs to initiate DNA replication and amplify the ssDNA fragments to dsDNA.

### 4) Denature and Amplification Cycles

The amplified dsDNAs are denatured to ssDNAs and re-amplified repeatedly. Same ssDNAs will be amplified and generate clusters which contain 500 to 1000 clones.

### 5) Sequencing

Solexa sequencing applies Sequencing by Synthesis (SBS) technology. The fluorescent labeled dNTPs used in reaction system are 3'-OH modified to terminate following elongation. After synthesis is over, the labeled dNTP will be recorded. Then the fluorescent label and the 3'-OH modification will be cleaved chemically to enable further sequencing reaction.

An improved paired-end sequencing method was developed in 2009, where two ends of each DNA fragment in the DNA library are modified with overhangs [85]. The template of first sequencing cycle will be washed off after first sequencing. Paired-end module will lead the complement chain to fix at the identical position to generate template cluster for a second SBS cycle [85].

The reading length of paired-end sequencing method arrives at  $2 \times 75$ bp. Each cycle of Solexa sequencing will result in 20.5 Gb to 25 Gb information within 10 days. Currently, the error rate of Solexa technology is about 1% to 1.5% due to nucleotide substitution [86]. But Solexa can measure the length of homopolymer perfectly since only one dNTP is added in each synthesis cycle.

## Study Objectives

1. To obtain non-O157 Stx bacteriophage genome sequences.
2. To explore the genomic variance of Stx bacteriophage samples.
3. To study the biological functions of Stx bacteriophage genes.

## CHAPTER 2 Materials and Methods

### 1. Description of the non-O157 STEC isolates

83 non-O157 STEC isolates were granted from Dr. T.G. Nagaraja of Kansas State University (KSU), plus 3 non-O157 STEC isolates were granted from National Institute for Health and Welfare, Finland (THL). The 86 non-O157 STEC isolates can be categorized into 24 groups according to O-serotypes (O6, O8, O15, O22, O38, O39, O74, O84, O88, O91, O96, O111, O113, O116, O117, O121, O130, O141, O145, O146, O153, O163, O171, O172, O179 and O183). 81 isolates were collected in 1999 from Kansas (all came from KSU), in which contains 2 human clinical strains and 79 strains from either cattle feces or water sample. The other 2 isolates of KSU were collected in 2013 from cattle samples in Nebraska. The 3 THL isolates were isolated in 2008 from cattle samples in Burkina Faso. *stx2* were detected from all 86 isolates; 64 isolates contain *stx1* while 22 isolates are not.

### 2. Bacteriophage Genome Collection

#### 2.1. Bacteriophage Induction

Non-O157 STEC isolates were picked up by sterilized cotton tips and were streaked on LB agar (Fisher, Pittsburg, PA) plate. The plates were incubated overnight at 35°C in the incubator. Single colony was selected and inoculated in 30mL LB broth (Fisher, Pittsburg, PA). Measure the optical density (OD600) of the bacterial solution hourly after incubating the solution at 37°C with shaking at 150 rpm. Adding mitomycin C solution to a final concentration of 0.5 µg/mL when OD600 reaches 0.3. The solution is then incubated under the same condition overnight. After centrifuge the solution at 10,000 × g for 30 minutes at 4°C, supernatant is then passed through sterilized 0.22 µm membrane filter to receive the mix of bacteriophages.

#### 2.2. Bacteriophage Stock Preparation

Prepare LBM broth by adding 20% maltose solution (w/v) into autoclaved LB broth to a final concentration of 0.5% (v/v). The maltose solution was filtered through 0.22  $\mu\text{m}$  membrane filter before using. Commercial *E. coli* K-12 strain (Strain PTA-7555, ATCC, Manassas, VA) is recovered from frozen tube by incubating in 30mL LB broth overnight at 37°C with shaking at 150 rpm. 1mL K-12 solution is centrifuged at 5000  $\times g$  for 5 minutes, discard supernatant and resuspend the pellet with 100  $\mu\text{L}$  LBM broth. 100  $\mu\text{L}$  bacteriophage mixture is mixed with the K-12 resuspension then incubated for 20 minutes at 37°C with shaking at 150 rpm. LB top agar was prepared beforehand by mixing granulated agar to LB broth to a final concentration of 0.7% (w/v). Preheat the LB top agar and 0.1M calcium chloride solution to 47°C in the waterbath during incubating the bacteriophage and K-12 mixture, add 1.5mL molten LB top agar and 1.5 mL calcium chloride solution together into the mixture then immediately pour on the labeled LB agar plate. Gently swirl the plate to distribute the top agar and then put the plate in room temperature for 30 minutes for top agar solidification. Invert the double-layer agar (DLA) plate and incubate the plate at 35°C for at least 16 hours. Check for clear circular bacteriophage plaques after 16 hour incubation, some bacteriophage samples need extended incubation period to allow plaque formation.

Prepare SM buffer with gelatin for bacteriophage resuspension. To prepare 1 liter of SM buffer with gelatin, dissolve the 5.8g NaCl and 2g  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  in 800 ml of  $\text{H}_2\text{O}$ ; add the 50mL Tris-Cl and 5mL 2% (w/v) gelatin, then adjust the volume to 1 liter with  $\text{H}_2\text{O}$ . Sterilize the buffer by autoclaving for 20 minutes at 15 psi (1.05  $\text{kg}/\text{cm}^2$ ) on liquid cycle. Allow the plaques to grow until the diameter reaches about 3mm. Add 5mL SM buffer in the DLA plate. Shake the plates at 80 rpm at room temperature. Transfer the SM buffer containing bacteriophage particles into a 15mL centrifuge tube, add 5  $\mu\text{L}$  chloroform then store the bacteriophage lysate in 4°C refrigerator.

### 2.3. Bacteriophage Genome Extraction

Genome extraction was done with Bacteriophage DNA Isolation Kit (Norgen Biotek, Thorold, Ontario, Canada). Transfer 1mL of bacteriophage lysate into a 1.5mL microtube. Bacterial host DNA was digested by adding 10  $\mu$ L RNA-free DNase I (Norgen Biotek, Thorold, Ontario, Canada) to the microtube and incubate for 15 minutes at room temperature. DNase I was inactivated by heating up to 75°C for 5 minutes. 500  $\mu$ L Lysis Buffer B of the DNA Isolation Kit was added in the reaction system with the following 10 seconds vortex. The system was then incubated at 65°C for 15 minutes. 320  $\mu$ L isopropanol (molecular biology grade, Fisher, Pittsburg, PA) was added after the incubation. Assemble a spin column (Norgen Biotek, Thorold, Ontario, Canada) to a 1.5mL microtube, transfer 650  $\mu$ L suspension from the reaction system then centrifuge at 6,000  $\times$  g for 1 minute. Flowthrough was discarded and repeat the centrifuge until the all solution of the reaction system went through the column. Wash the column by adding 400  $\mu$ L Washing Solution A (Norgen Biotek, Thorold, Ontario, Canada) in the column to centrifuge at 6,000  $\times$  g for 1 minute, discard the flowthrough then repeat washing for another 2 times. Replace the collection tube with another 1.5mL microtube for DNA collection. 75  $\mu$ L Elution Buffer B was added in the column then centrifuge at 6,000  $\times$  g for 1 minute. Flowthrough was collected as bacteriophage DNA sample. DNA concentration was measured immediately after DNA samples were collected.

Qubit<sup>®</sup> 2.0 Fluorometer System was applied to measure DNA concentration. Qubit<sup>®</sup> working solution was prepared by diluting Qubit<sup>®</sup> dsDNA HS Reagent 1:200 in Qubit<sup>®</sup> dsDNA HS Buffer. Mixing 190  $\mu$ L Qubit<sup>®</sup> working solution with 10  $\mu$ L DNA sample thoroughly by vortexing for 2-3 seconds. The DNA sample tubes were inserted into Qubit<sup>®</sup> 2.0 Fluorometer and the concentration was measured with the dsDNA High Sensitivity option. DNA samples with a concentration higher than 5ng/ $\mu$ L were stored in -20°C freezer for PCR confirmation and DNA sequencing.

For DNA samples with low concentration, bacteriophage lysate was amplified in LB broth and collected through polyethylene glycol precipitation. Infect bacteriophage to K-12 suspension following the procedures described above. The mixture of bacteriophage and bacteria was transferred into 35mL

sterilized LB broth to incubate overnight at 37°C with shaking at 150 rpm. Terminate the incubation if the solution was clear and cell precipitate visible at the bottom of the container, or incubation was extended if opaque solution was received. Solution went through sterilized 0.22 µm membrane filter to remove bacterial residues. 700 µL chloroform was added in the solution then incubate for 30 minutes at room temperature with shaking at 150 rpm. Polyethylene glycol 8000 (PEG8000) was added in the solution to reach a final concentration of 10% (w/v). Shake the solution at room temperature at 250 rpm until PEG8000 was completely dissolved. NaCl (Analytical grade, Fisher, Pittsburg, PA) was added in the solution to reach a final concentration of 1M, then swirl the container until salt dissolves. Store the solution on ice for at least 1 hour. Centrifuge the solution at 10,000 ×g for 10 minutes at 4°C. Discard the supernatant, resuspend the pellet with 1.5mL SM buffer and transfer the resuspension to a 2mL microtube. DNase I (Invitrogen Cat. No. 18068015) was added in the resuspension to reach a final concentration of 1 µg/mL then incubate at 35°C for 30 minutes. DNase I was inactivated by adding 25mM EDTA solution to a final concentration of 1 µL/mL then heated to 65°C for 10 minutes. The amplified bacteriophage stock was extracted by Bacteriophage DNA Isolation Kit again to receive an acceptable DNA concentration.

### 3. Bacteriophage Genome Confirmation

All genomic DNA samples were confirmed through PCR test. Two pairs of primers were applied to detect the existence of *stx2* and *stx2*-flanking fragment in the harvested DNA samples. The PCR reaction system were prepared with GoTaq® DNA Polymerase Kit (Promega, Madison, WI). The reaction system for each pair of primers was identical, containing 2.5 µL 5× Green GoTaq® Flexi Buffer, 7 µL MgCl<sub>2</sub> Solution (25mM), 2 µL PCR Nucleotide Mix (10mM), 1.5 µL primer solution (upstream primer 1 µM, downstream primer 1 µM), 0.25 µL GoTaq® Hot Start Polymerase, 2.5 µL template DNA and 9.25 µL Nuclease-Free Water. The thermocycle program included an initial denaturation at 94°C for 10 minutes, followed by 30 cycles of denaturation (94°C for 30 seconds), annealing (55°C for 60 seconds) and extension (72°C for 90 seconds), ending with a final extension at 72°C for 10 minutes.

Primer Name	Primer Sequence	Specificity	Product Size	Reference
stx2-F	5'-CCATGACAACGGACAGCAGTT-3'	<i>stx2</i>	779	V. Gannon <i>et al.</i> 1992
stx2-R	5'-CCTGTCAACTGAGCAGCACTTTG-3'			
422	5'-CATGAAGAAGATGTTTATGGCG-3'	<i>stx</i> -flanking region	311	A. Unkmeir <i>et al.</i> 2000
457	5'-CACCCCGTTCTCATCCGTCATG-3'			

Table 2.1 Oligonucleotide Primers used for PCR Tests

The existence of target genes was tested through agarose gel electrophoresis. Prepare 1×TAE buffer by diluting commercial 10×TAE buffer by ultrapure water. Agarose gel was prepared by mixing 1% of agarose into 1×TAE buffer (w/v). The agarose suspension was heated in microwave oven using max level for 2 minutes until agarose was dissolved. The suspension was left at room temperature for 10 minutes until the suspension temperature dropped to about 50°C (the flask touchable by hands). Adding 2µL ethidium bromide (EtBr) into 100mL suspension. The suspension was poured in the gel casting tray, place the comb into the tray then wait until the gel solidified. The comb was removed and the gel was transferred into electrophoresis chamber then pour 1×TAE buffer into the chamber to submerge the gel. 20µL of each DNA sample was transferred to separate wells in the agarose gel, plus 10µL DNA ladder standard in one well of each row. The agarose gel with DNA samples loaded was placed in electric field of 100 volts for 30 minutes. The gel was then moved to UV camera to verify the existence of the target genes through looking for DNA bands of certain size.

#### 4. Bacteriophage Genome Sequencing

The verified bacteriophage genomes were normalized to 0.2ng/µL by diluting the genome with ultrapure water. Nextera XT Sample Prep Kit (Illumina, San Diego, CA) was used to make bacteriophage genome library. Each sample was mixed with 10µL TD buffer and 5µL ATM in separate wells of a 96-well plate. The samples were heated at 55°C for 5 minutes, then held at 10°C for DNA tagmentation

reaction. 5  $\mu$ L NT buffer was added to each well as the samples dropped to 10°C then the plate was placed at room temperature for tagmentation halting.

Fifteen microliters NPM was added to each well, as well as 5  $\mu$ L i5 index and 5  $\mu$ L i7 index. The tagmented DNA samples were amplified by PCR reaction. Thermocycling protocol included an initial denaturation at 72°C for 3 minutes followed by 95°C for 30 seconds, 12 cycles of denaturation (95°C for 10 seconds), annealing (55°C for 30 seconds) and elongation (72°C for 30 seconds), the final elongation step was 72°C for 5 minutes.

AMPure XP beads were applied to extract PCR products from the PCR reaction system. 25  $\mu$ L beads were added in each well, then the 96-well plate was shook at 1800 rpm for 2 minutes. The plate was left at room temperature for 5 minutes to enable beads attaching to DNA products. The 96-well plate was placed on a magnetic stand for 2 minutes until beads were completely attracted by magnet. Supernatant was discarded, then 200  $\mu$ L 80% ethanol was added to wash the beads. After shaking at 1800 rpm for 2 minutes, the plate was placed on the magnetic stand and supernatant was discarded. The ethanol wash needs to be done for 2 times in total. After the beads were washed, the pellet was air dried on the magnetic stand for 15 minutes to guarantee a complete evaporation of ethanol. The plate was removed from the magnetic stand and 52.5  $\mu$ L RSB was added to each well. The plate was shook at 1800 rpm for 2 minutes then left at room temperature for 2 minutes to enable RSB attaching to DNA products. The plate was placed on the magnetic stand for 2 minutes, 45  $\mu$ L of supernatant from each well was transferred to another 96-well plate to complete PCR product clean up.

Twenty microliters of cleaned PCR product was mixed with 45.83  $\mu$ L LNA1 and 8.33  $\mu$ L LNB1. 45  $\mu$ L of the mixture was transferred to another 96-well plate. The plate was shook at 1800 rpm for 30 minutes. The plate was then placed on a magnetic stand and the supernatant was discarded after beads were completely attracted by the magnet. The pellet was washed by LNW1 for 2 times: 45  $\mu$ L of LNW1 was added in each well and the plate was shook at 1800 rpm for 5 minutes. The plate was placed on the

magnetic stand and the supernatant was discarded after beads were completely attracted by the magnet. After the second washing was over, 30  $\mu$ L 0.1N NaOH solution was added in each well, the beads were resuspended by shaking at 1800 rpm for 5 minutes. After placing the plate on the magnetic stand until the supernatant became clear, 30  $\mu$ L supernatant was transferred to another plate and 30  $\mu$ L LNS1 was mixed with supernatant to complete PCR product normalization.

Thirty microliters normalized PCR product of each sample was mixed together in a reagent reservoir. 24  $\mu$ L of the pooled sample was added to a 1.5mL microtube containing 576  $\mu$ L HT1 buffer. The solution was mixed by vortexing for 2 to 3 seconds. The tube was heated at 96°C for exactly 2 minutes, then the tube was immediately transferred to ice-water bath for 5 minutes to complete the library preparation. The library was loaded in a MiSeq 500 V2 reagent cartridge. The cartridge and a flow cell were loaded in the MiSeq sequencer. Nextera XT option was selected in the MiSeq sequencing.

## 5. Data Analysis

The automatic results of MiSeq sequencer were FASTQ format. The data was evaluated, assembled and converted to FASTA format by CLC Genomics Workbench. After original data was imported in the software, the data was trimmed by using quality score threshold of 0.001 to exclude poor-quality sequences. Trimmed data was assembled by de novo sequencing with contig length of 300. The program generated FASTA files and annotation report as well.

Original FASTA sequences contained bacteriophage contigs and *E. coli* K-12 contigs as well. Bacterial contigs were recognized by aligning FASTA files to *E. coli* K-12 chromosomal genome (*Escherichia coli* strain K-12 substrain MG1655, Reference sequence: NC\_000913.3) available from National Center for Biotechnology Information (NCBI) using local BLAST 2.2.24. Blast-nucleotide (Blastn) was chosen to align each sequence with NC\_000913.3. Sequences were set as queries and NC\_000913.3 as target. Non-default parameters included modified E-value threshold of 0.001 and percent identity of 80%. Matched contigs were deleted from the sequence. A reverse alignment was

performed by using NC\_000913.3 as query and sequences as targets, matched contigs of the targets were deleted. The remaining contigs were globally aligned to NC\_000913.3 by Mauve 2.4.0 to receive a visible alignment result. The highly matched contigs were also deleted. The remaining contigs were aligned to 2 Stx bacteriophage genomes from NCBI (Enterobacteria bacteriophage 933W, Reference sequence: NC\_000924.1; Escherichia bacteriophage P13374, Reference sequence: NC\_018846.1). Blastn was chose to align by using sequences as queries and Stx bacteriophage genomes as targets using the parameter described above, highly-matched contigs were reserved. Poorly-matched contigs were aligned to Escherichia nucleotide collection using online BLAST with an E-value threshold of 0.01. The contigs without a clear source were saved into the sequences to complete bacteriophage genome clean up. Cleaned bacteriophage genomes were globally aligned to the 2 reference Stx bacteriophage genomes by Mauve 2.4.0 to receive a visible alignment result.

Multiple alignment of sample bacteriophage genomes and 2 reference genomes (NC\_000924.1 and NC\_018846.1) was performed by ClustalW. After all columns containing gaps were deleted, the remaining alignment result was saved to construct phylogenetic tree using maximum likelihood algorithm by Molecular Evolutionary Genetics Analysis 6.0 (MEGA6.0).

Local Blast was applied to confirm the existence of genes in sample bacteriophage genomes. Nucleotide sequence of genes from NC\_000924.1 and NC\_018846.1 were set as queries to align to sample bacteriophage genomes by Blastn with E-value threshold of 0.001 and percent identity of 90%. Sample bacteriophage genomes were translated to protein sequences by MEGA6.0. The protein sequences of genes from NC\_000924.1 and NC\_018846.1 were set as queries to align to sample bacteriophage genomes by Blast-protein (Blastp) with E-value threshold of 0.001 and percent identity of 90%.

## CHAPTER 3 Results

### 1. Induction of STEC Isolates

13 STEC isolates were tested to be inducible. The total induction rate was 15.66%. All 3 THL isolates were inducible. The induction rate of KSU isolates was 12.05%. Inducible isolates included 7 O-serotypes: O74, O111, O121, O130, O163, O179 and O183. 6 inducible isolates are O111 serotypes and 2 are O163 serotypes. Statistical analysis included the induction rate according to serotypes and origins.

Isolate ID	Serotype	Origin	Location	Isolating Time	stx1	stx2
<b>KSU1234-1</b>	O111	Cattle	Kansas, USA	1999	+	+
<b>KSU7740-2</b>	O163	Water-creek	Kansas, USA	1999	+	+
<b>KSU7756-1</b>	O111	Cattle	Kansas, USA	1999	+	+
<b>KSU7766-1</b>	O111	Water-pond	Kansas, USA	1999	+	+
<b>KSU8266-1</b>	O111	Cattle	Kansas, USA	1999	+	+
<b>KSU9042-1</b>	O111	Water-pond	Kansas, USA	1999	-	+
<b>KSU13802-1</b>	O163	Cattle	Kansas, USA	1999	+	+
<b>KDHE8</b>	O111	Human	Kansas, USA	1999	+	+
<b>KDHE31</b>	O121	Human	Kansas, USA	1999	+	+
<b>KDHE37</b>	O130	Human	Kansas, USA	1999	-	+
<b>FE95381</b>	O74	Bovine intestine	Burkina Faso	11/11/2008	+	+
<b>FE95383</b>	O179	Bovine intestine	Burkina Faso	10/30/2008	+	+
<b>FE95610</b>	O183	Beef meat	Burkina Faso	12/16/2008	+	+

Table 3.1. Summary of Inducible non-O157 STEC Isolates

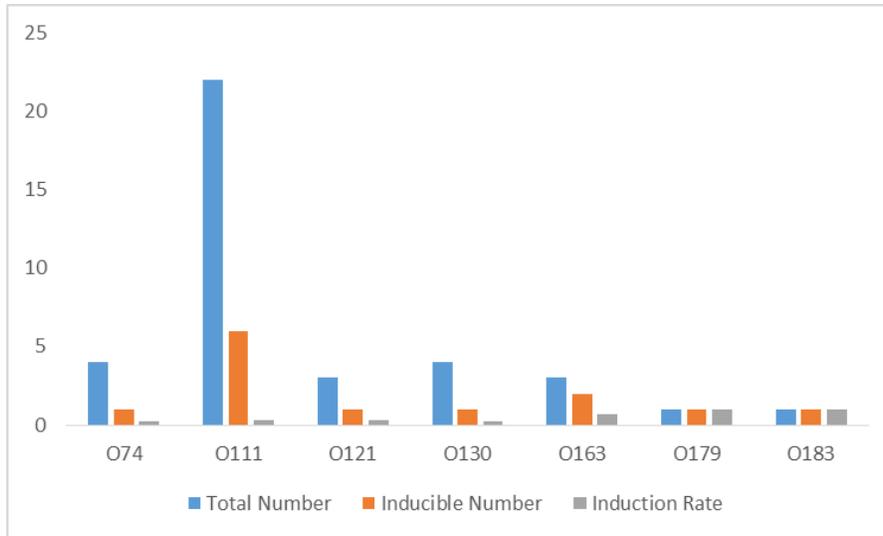


Figure 3.1. Induction rate of STEC isolates of different serotypes. Blue bars: Number of STEC isolates of certain serotypes. Orange bars: Number of inducible isolates of certain serotypes. Grey bars: Induction rate. Rates of each serotype from left to right are 25% (1 out of 4), 27.3 (6/22), 33.3% (1/3), 25% (1/4), 66.7% (2/3) 100% (1/1), and 100% (1/1), respectively. Serotypes without any inducible strains are not shown in the figure.

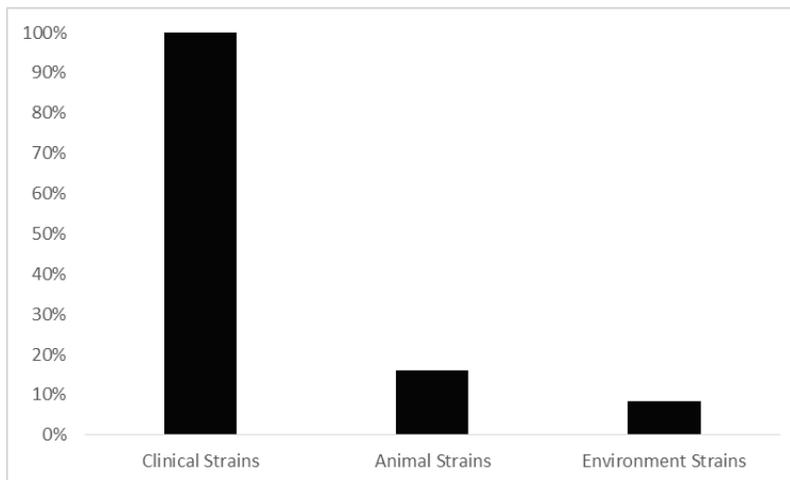


Figure 3.2. Induction rate of STEC isolates of different origins. All clinical strains were inducible, the induction rates of animal strains and environmental strains were 16.0% and 8.3%, respectively.

## 2. Bacteriophage Genome Extraction

Bacteriophage samples were successfully collected from all inducible STEC isolates. Genomes were tested by PCR reaction after being extracted from bacteriophage samples. Existence of *stx* gene and was confirmed in all 13 bacteriophage sample genomes. *stx*-flanking fragment was not detected from the bacteriophage sample came from STEC isolate KDHE37, and the other 12 sample genomes all contain the fragment.

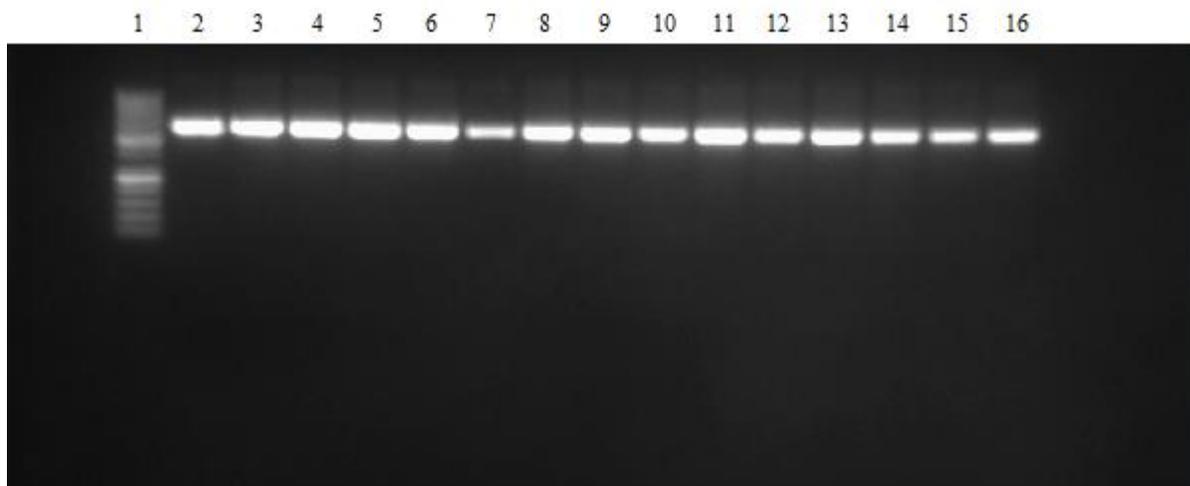


Figure 3.3. The electrophoresis result of PCR reaction targeting at *stx*. Line 1: 100bp DNA ladder standard; Line 2 and 3: STEC chromosome as positive controls (STEC O157:H7 strain and STEC O26 strain); Line 4: Bacteriophage of KSU 1234-1; Line 5: Bacteriophage of KSU 7740-2; Line 6: Bacteriophage of KSU 7756-1; Line 7: Bacteriophage of KSU 7766-1; Line 8: Bacteriophage of KSU 8266-1; Line 9: Bacteriophage of KSU 9042-1; Line 10: Bacteriophage of KSU 13802-1; Line 11: Bacteriophage of KDHE 8; Line 12: Bacteriophage of KDHE 31; Line 13: Bacteriophage of KDHE 37; Line 14: Bacteriophage of FE95381; Line 15: Bacteriophage of FE95383; Line 16: Bacteriophage of FE 95610.

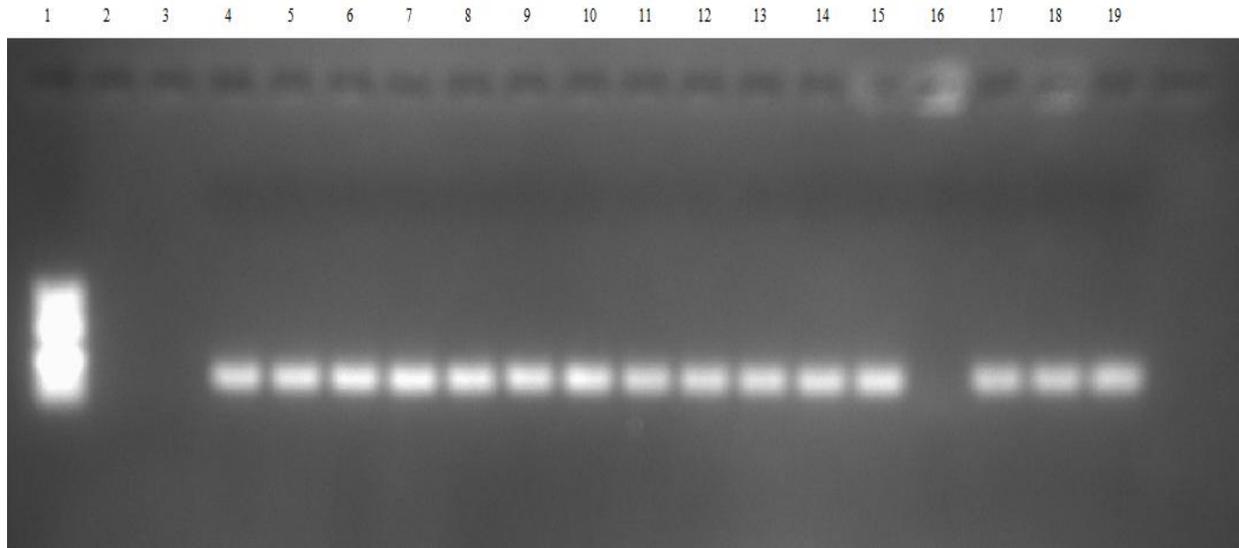


Figure 3.4. The electrophoresis result of PCR reaction targeting at *stx*-flanking fragment. From line 1 to line 19 are: 100bp DNA ladder standard, negative control (PCR reaction system without DNA template), negative control (*E. coli* K-12 chromosome), positive control (STEC O157:H7 chromosome), positive control (STEC O26 chromosome), positive control (STEC O45 chromosome), bacteriophage of KSU 1234-1, bacteriophage of KSU 7740-2, bacteriophage of KSU 7756-1, bacteriophage of KSU 7766-1, bacteriophage of KSU 8266-1, bacteriophage of KSU 9042-1, bacteriophage of KSU 13802-1, bacteriophage of KDHE 8, bacteriophage of KDHE 31, bacteriophage of KDHE 37, bacteriophage of FE95381, bacteriophage of FE95383 and bacteriophage of FE 95610, respectively.

### 3. DNA Cleaning Result

Nine bacteriophage sample genomes were acquired after deleting contigs recognized as K-12 chromosomal fragments. The other 4 genomes only contained bacterial sequences and entire genomes were deleted after cleaning protocol.

<b>Isolate ID</b>	<b>Serotype</b>	<b>Origin</b>	<b>Location</b>	<b>Isolating Time</b>	<b>stx1</b>	<b>stx2</b>
<b>KSU1234-1</b>	O111	Cattle	Kansas, USA	1999	+	+
<b>KSU7740-2</b>	O163	Water-creek	Kansas, USA	1999	+	+
<b>KSU7756-1</b>	O111	Water-pond	Kansas, USA	1999	+	+
<b>KSU7766-1</b>	O111	Water-pond	Kansas, USA	1999	+	+
<b>KSU8266-1</b>	O111	Cattle	Kansas, USA	1999	+	+
<b>KSU9042-1</b>	O111	Water-pond	Kansas, USA	1999	-	+
<b>KSU13802-1</b>	O163	Cattle	Kansas, USA	1999	+	+
<b>KDHE37</b>	O130	Human	Kansas, USA	1999	-	+
<b>FE95381</b>	O74	Bovine intestine	Burkina Faso	11/11/2008	+	+

Table3.2. Summary of non-O157 STEC Isolates with Reliable Bacteriophage Genome Sequences

#### 4. Global Alignment

Cleaned bacteriophage sample genomes were aligned to 2 reference genomes, NC\_000924.1 and NC\_018846.1, by Mauve 2.4.0. NC\_000924.1 is the genome sequence of bacteriophage 933W, which was induced from STEC O157:H7. The sequence was acquired in 1999, which was the first sequenced Stx bacteriophage genome. NC\_018846.1 is the genome sequence of bacteriophage p13374, which was induced from STEC O104:H4 that caused the 2011 Germany *E. coli* O104:H4 outbreak. The visible alignment results shown the difference of similarity when a sample genome aligned to traditional bacteriophage genome against to a novel Stx bacteriophages.

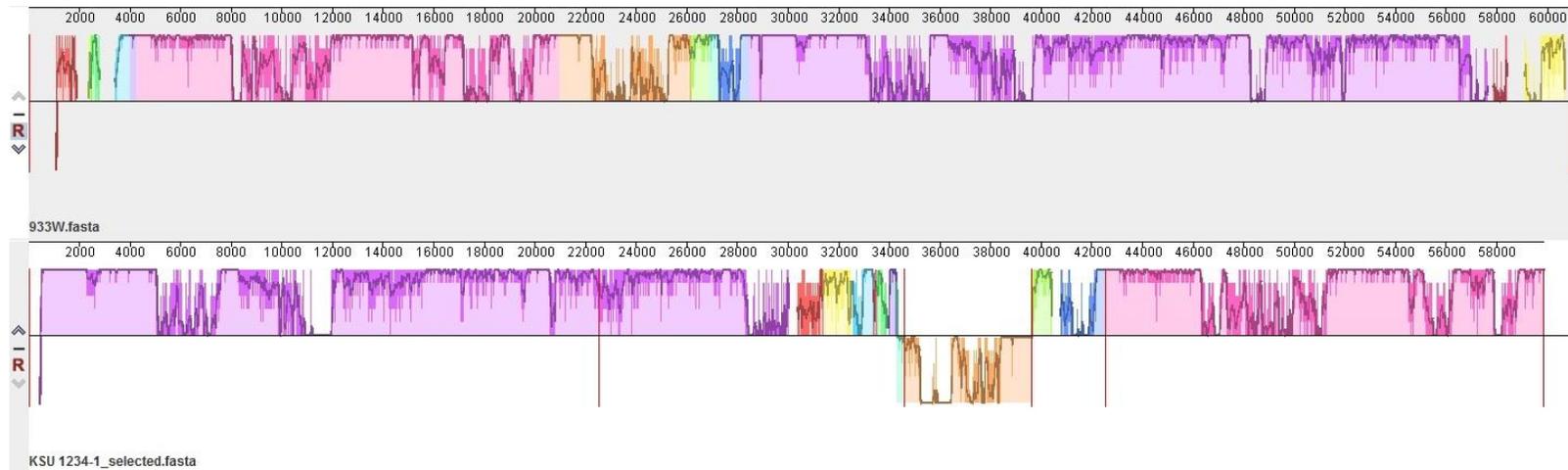


Figure 3.5. Global Alignment between Bacteriophage Genome Came from STEC KSU 1234-1 to Bacteriophage 933W.

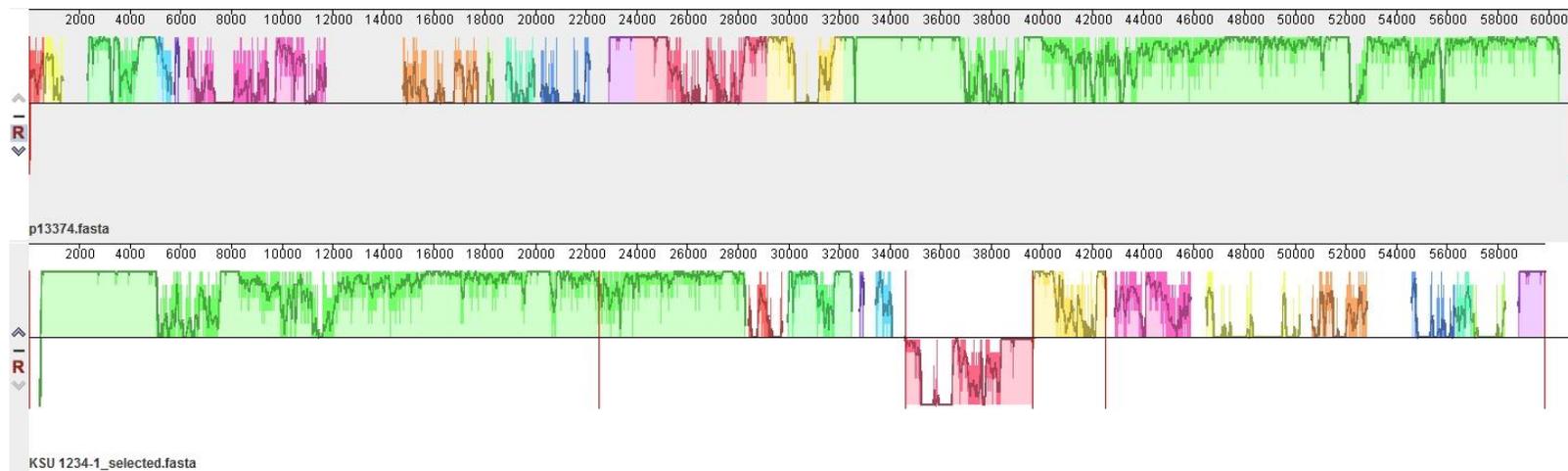


Figure 3.6. Global Alignment between Bacteriophage Genome Came from STEC KSU 1234-1 to Bacteriophage p13374.

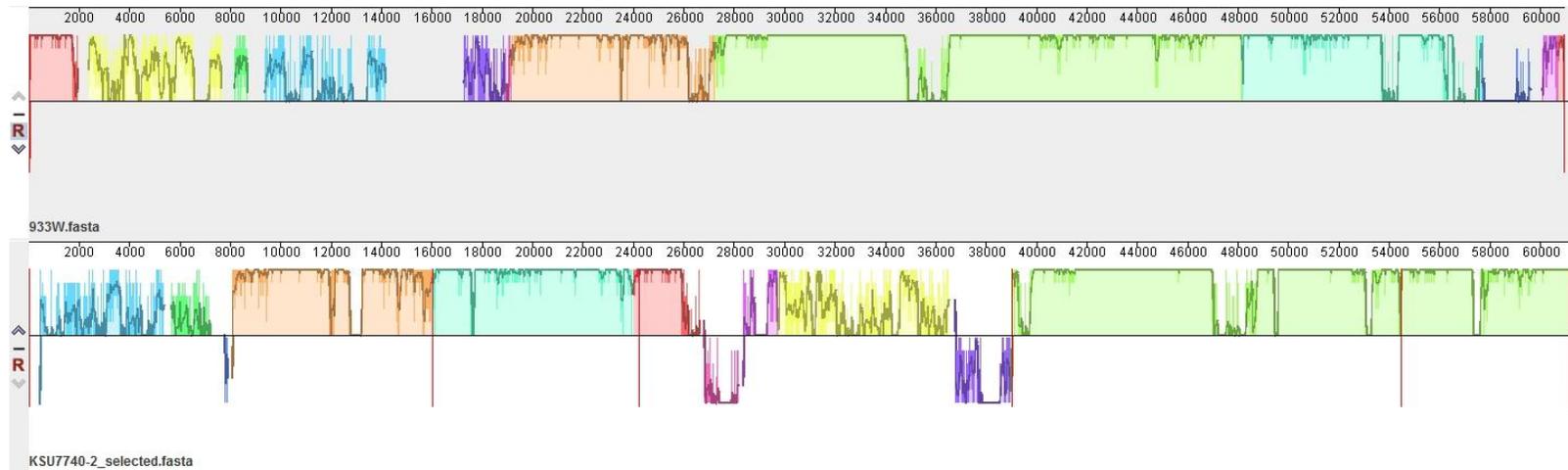


Figure 3.7. Global Alignment between Bacteriophage Genome Came from STEC KSU 7740-2 to Bacteriophage 933W.

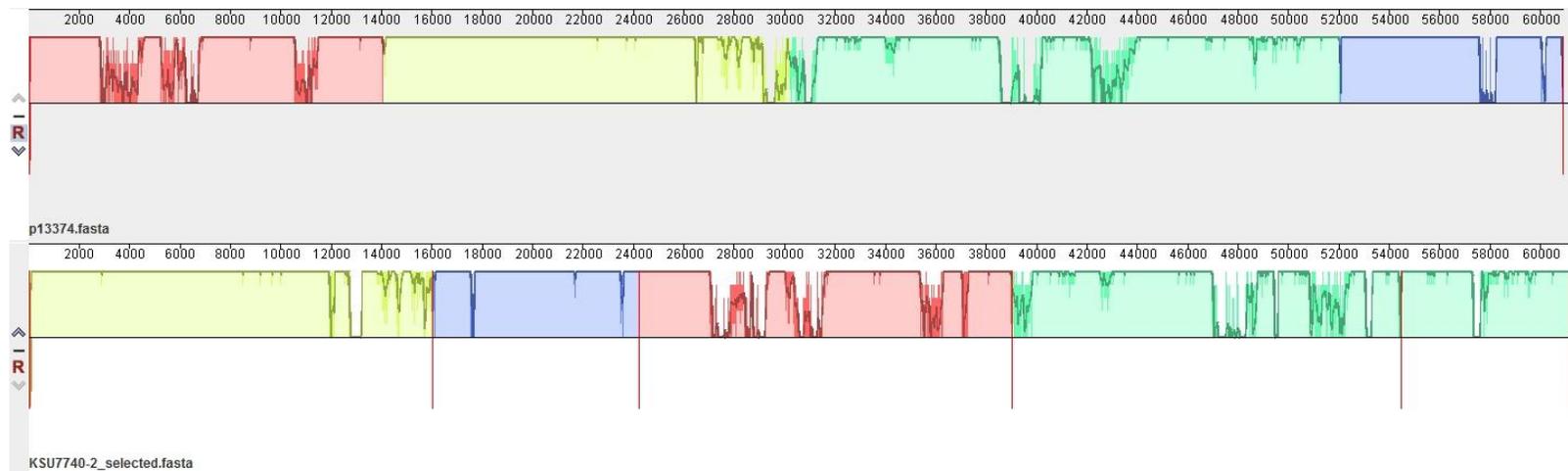


Figure 3.8. Global Alignment between Bacteriophage Genome Came from STEC KSU 7740-2 to Bacteriophage p13374.

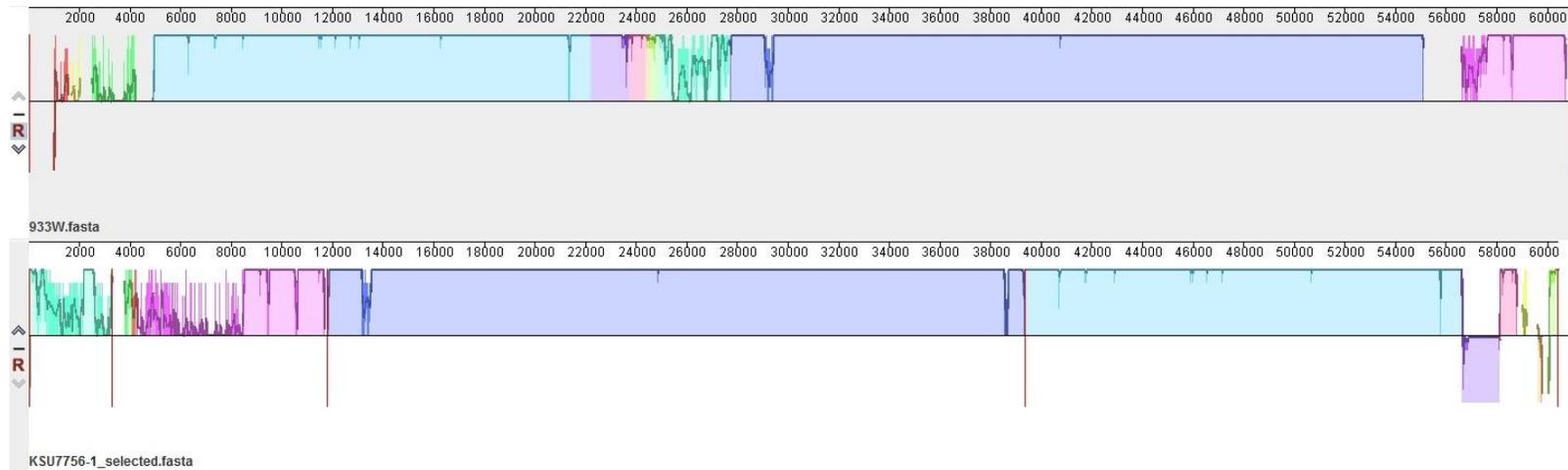


Figure 3.9. Global Alignment between Bacteriophage Genome Came from STEC KSU 7756-1 to Bacteriophage 933W.

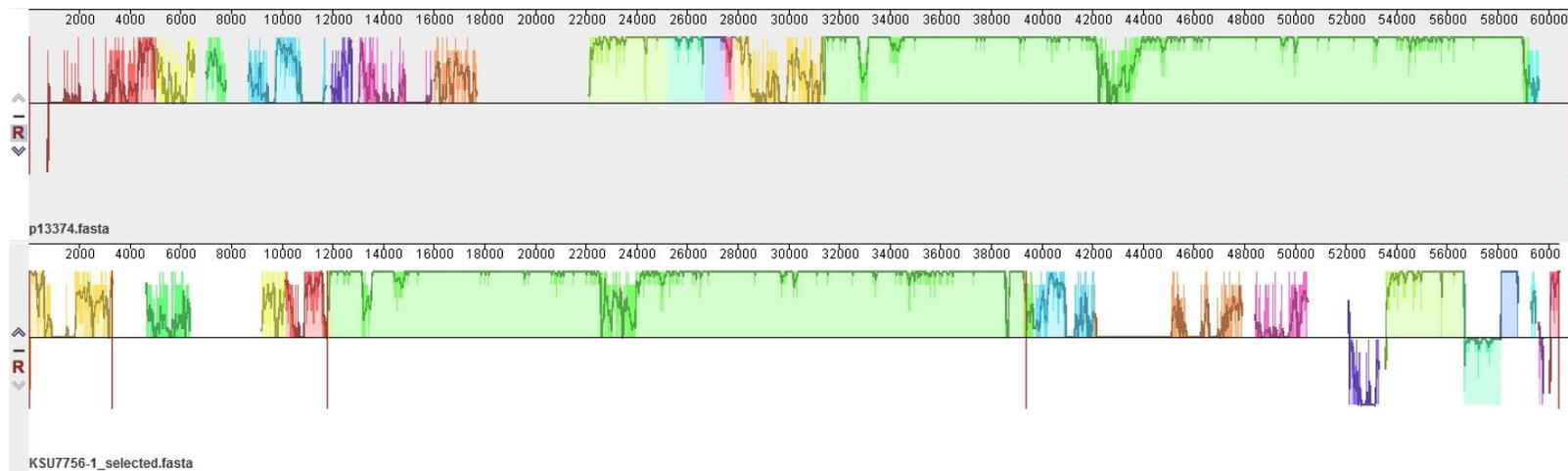


Figure 3.10. Global Alignment between Bacteriophage Genome Came from STEC KSU 7756-1 to Bacteriophage p13374.

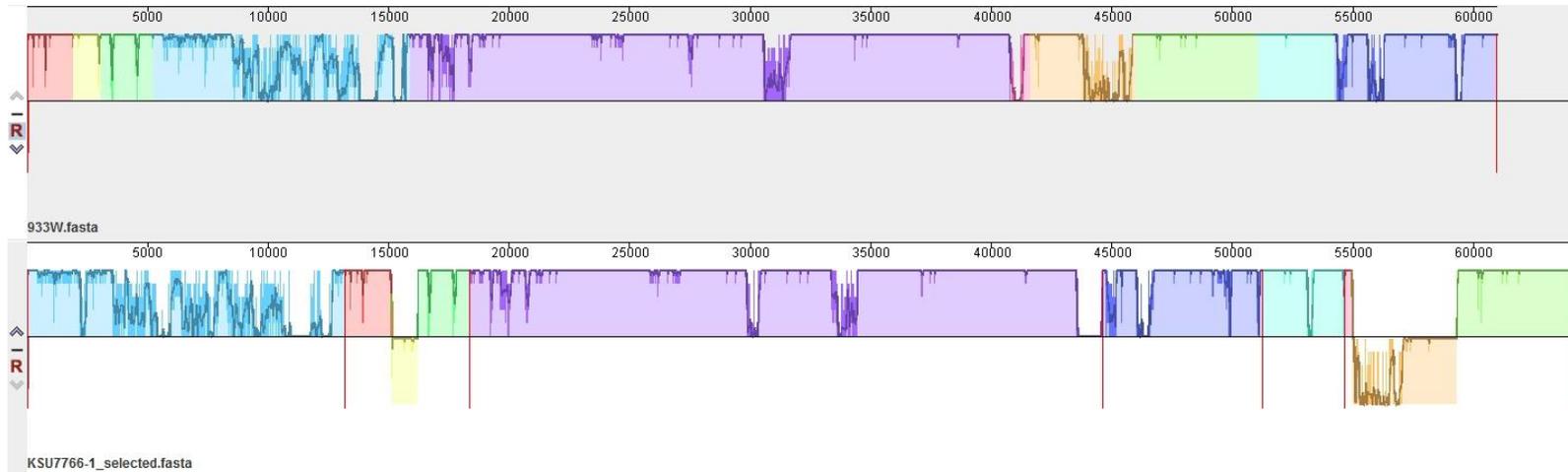


Figure 3.11. Global Alignment between Bacteriophage Genome Came from STEC KSU 7766-1 to Bacteriophage 933W.

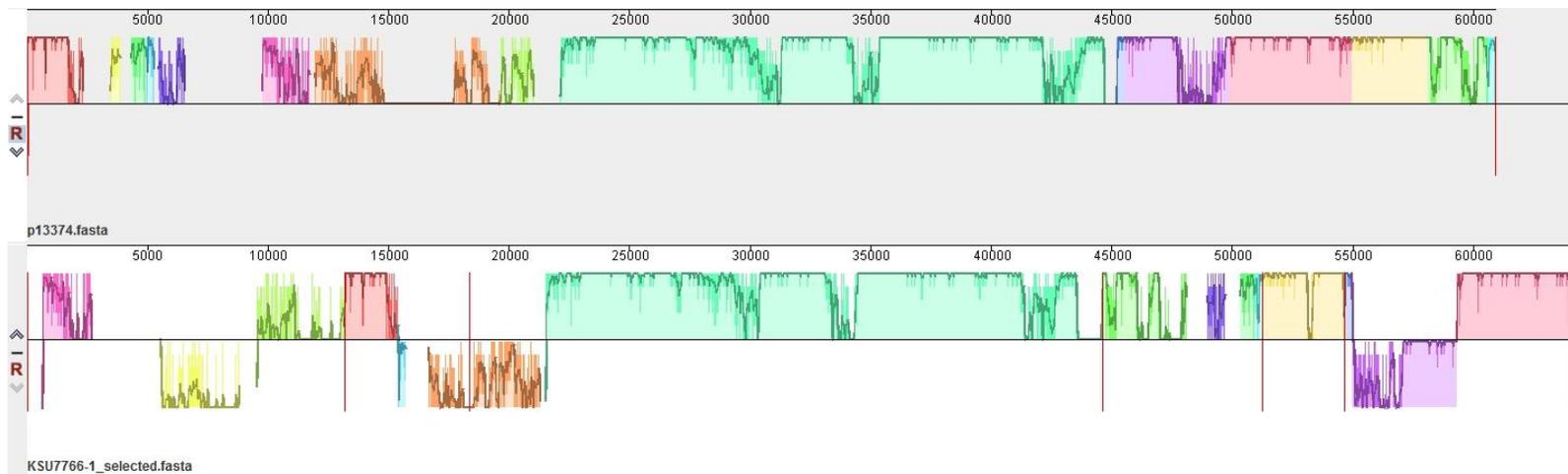


Figure 3.12. Global Alignment between Bacteriophage Genome Came from STEC KSU 7766-1 to Bacteriophage p13374.

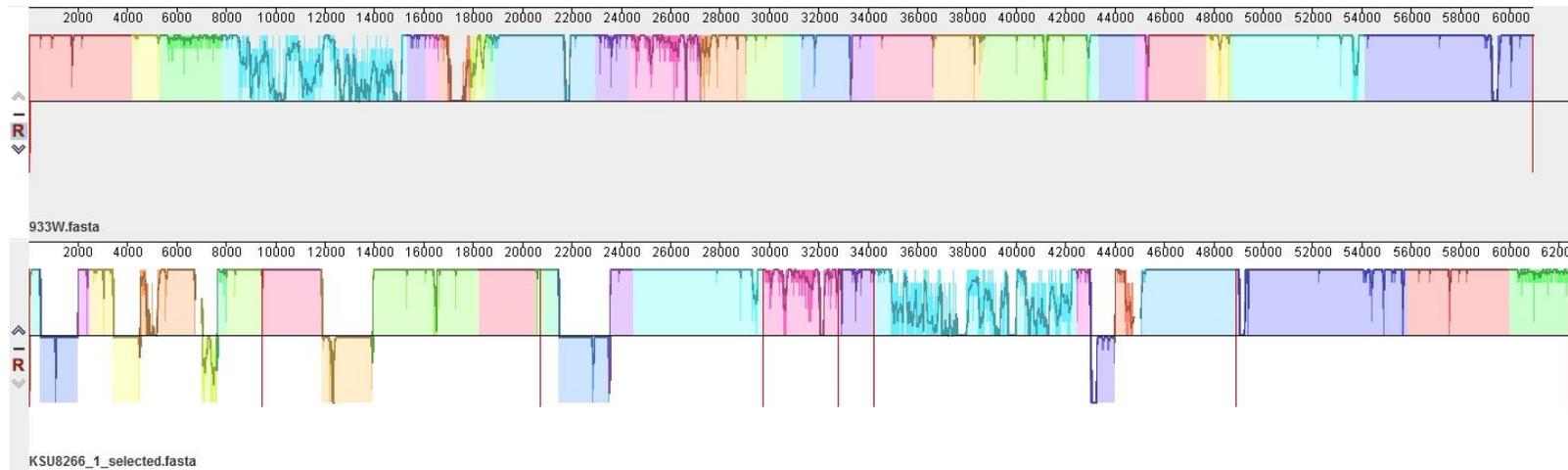


Figure 3.13. Global Alignment between Bacteriophage Genome Came from STEC KSU 8266-1 to Bacteriophage 933W.

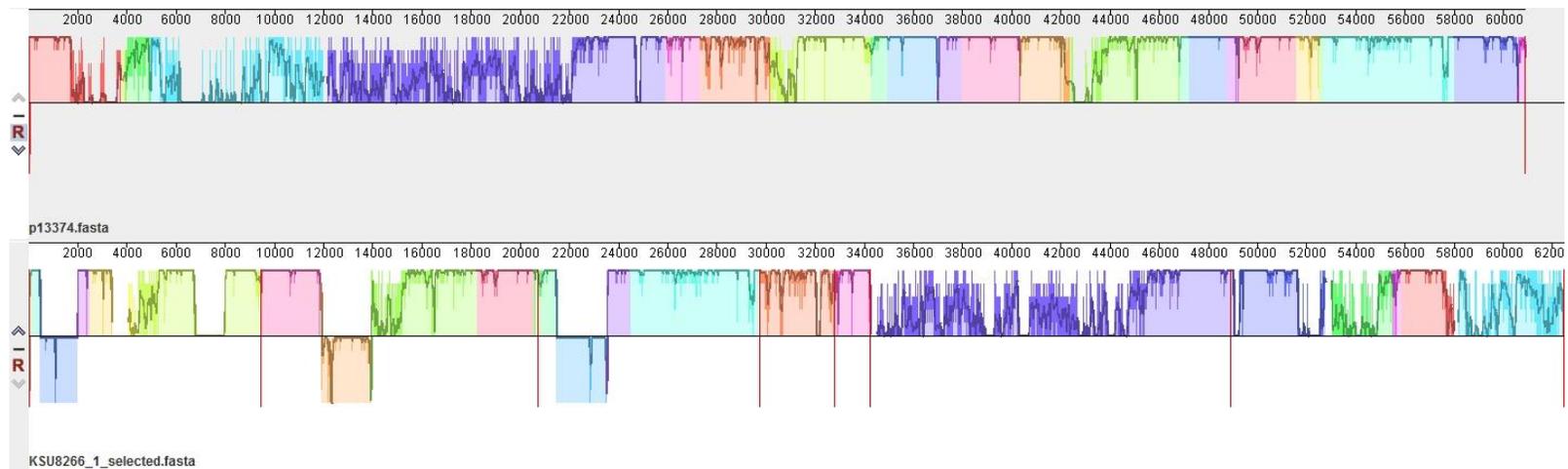


Figure 3.14. Global Alignment between Bacteriophage Genome Came from STEC KSU 8266-1 to Bacteriophage p13374.

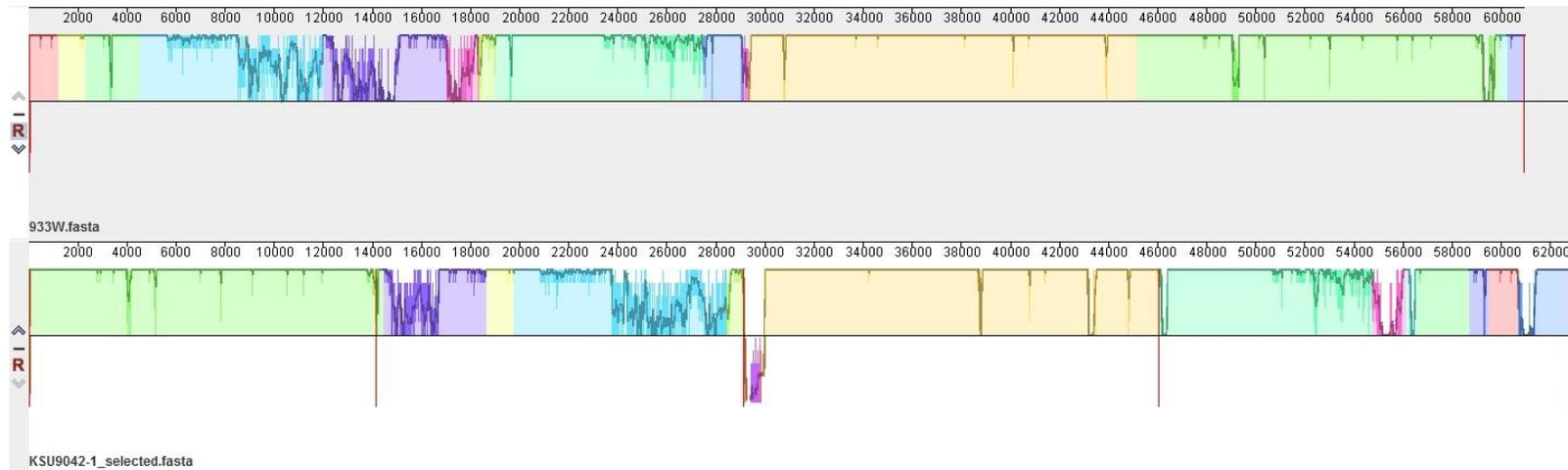


Figure 3.15. Global Alignment between Bacteriophage Genome Came from STEC KSU 9042-1 to Bacteriophage 933W.

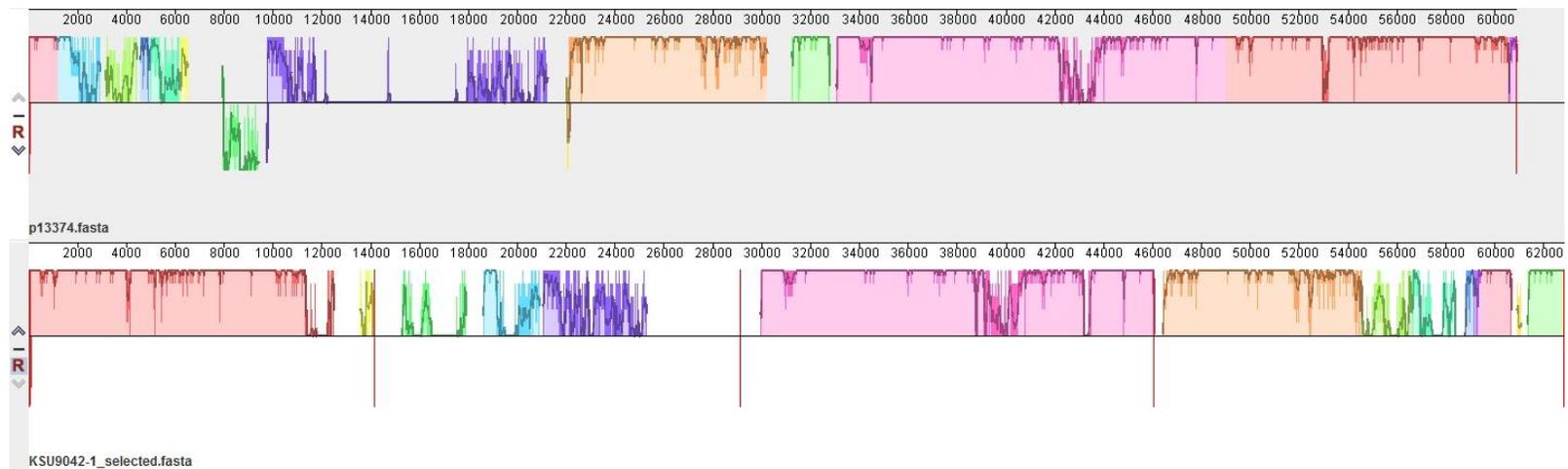


Figure 3.16. Global Alignment between Bacteriophage Genome Came from STEC KSU 9042-1 to Bacteriophage p13374.

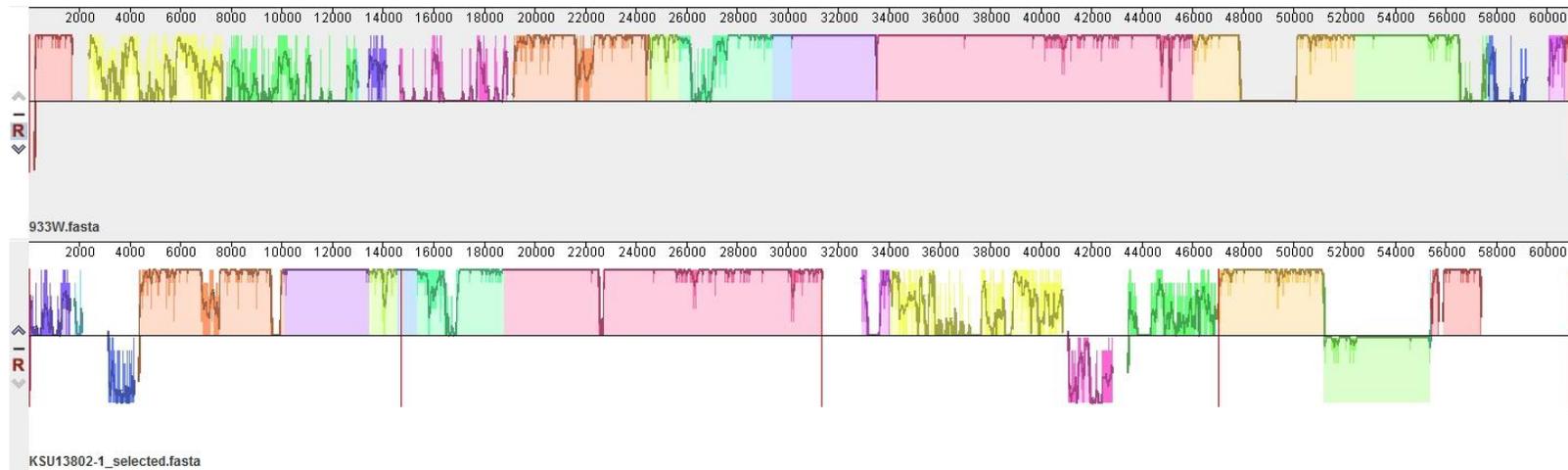


Figure 3.17. Global Alignment between Bacteriophage Genome Came from STEC KSU 13802-1 to Bacteriophage 933W.

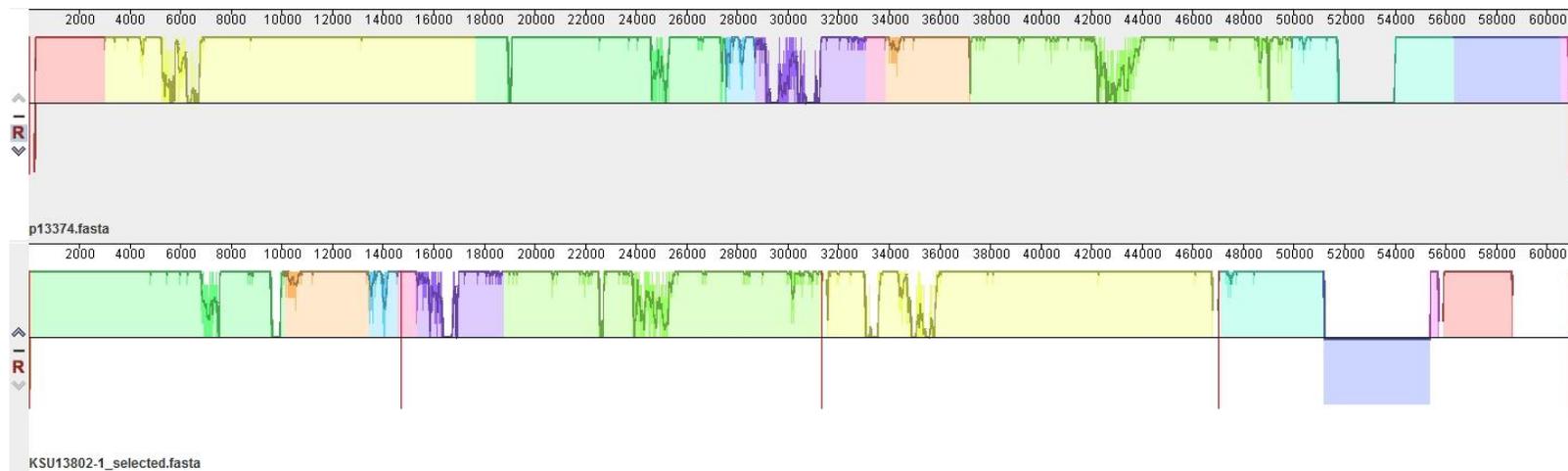


Figure 3.18. Global Alignment between Bacteriophage Genome Came from STEC KSU 13802-1 to Bacteriophage p13374.

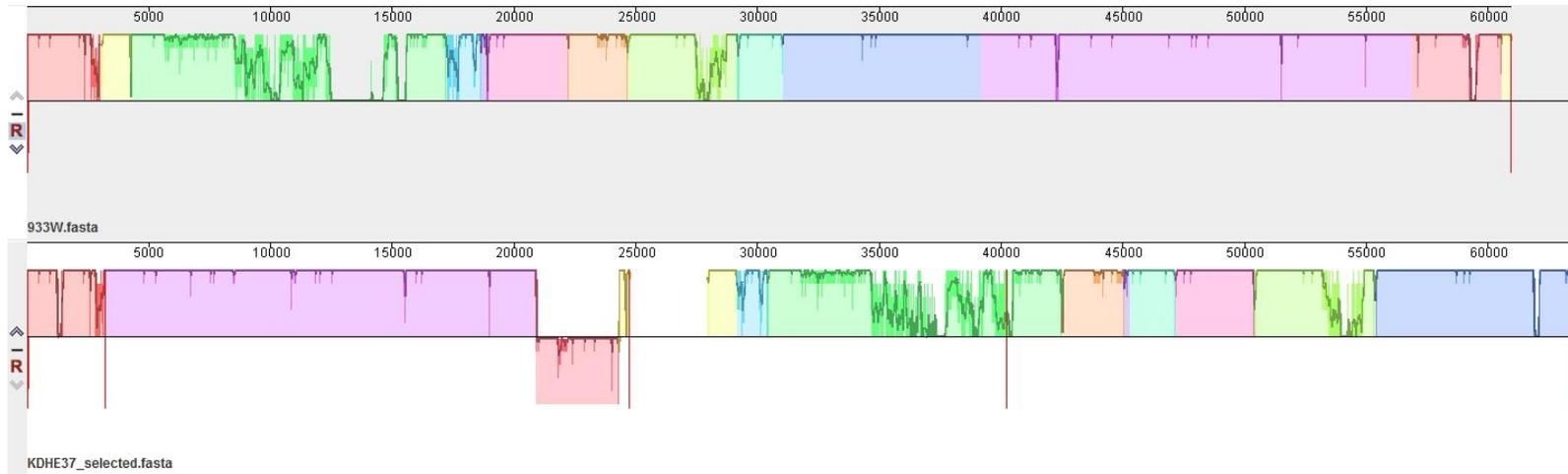


Figure 3.19. Global Alignment between Bacteriophage Genome Came from STEC KDHE 37 to Bacteriophage 933W.

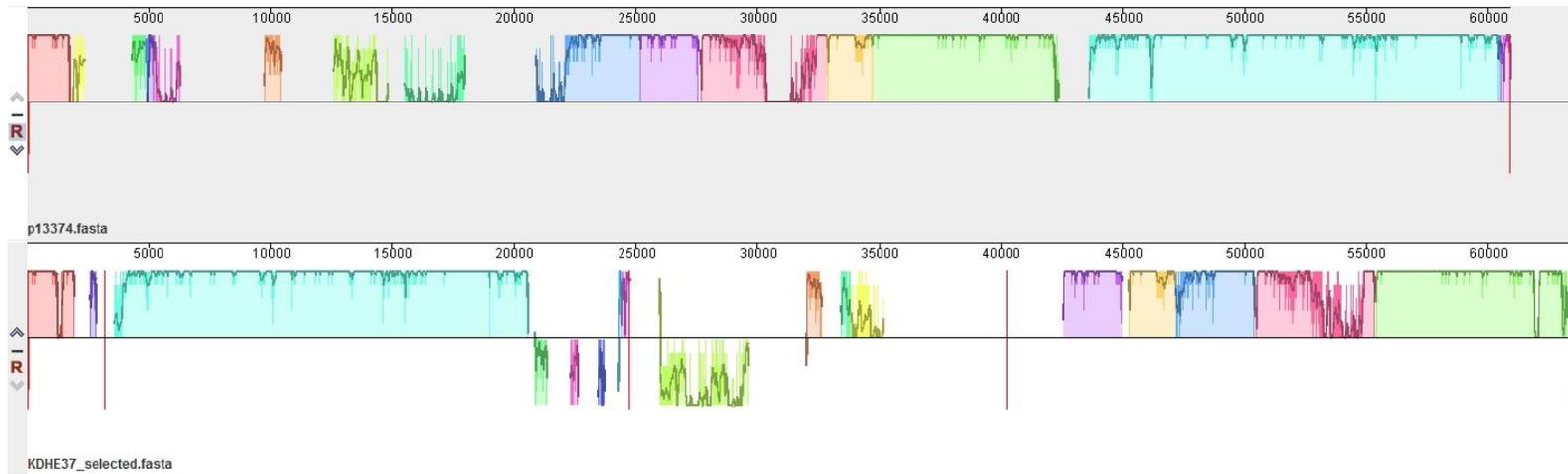


Figure 3.20. Global Alignment between Bacteriophage Genome Came from STEC KDHE 37 to Bacteriophage p13374.

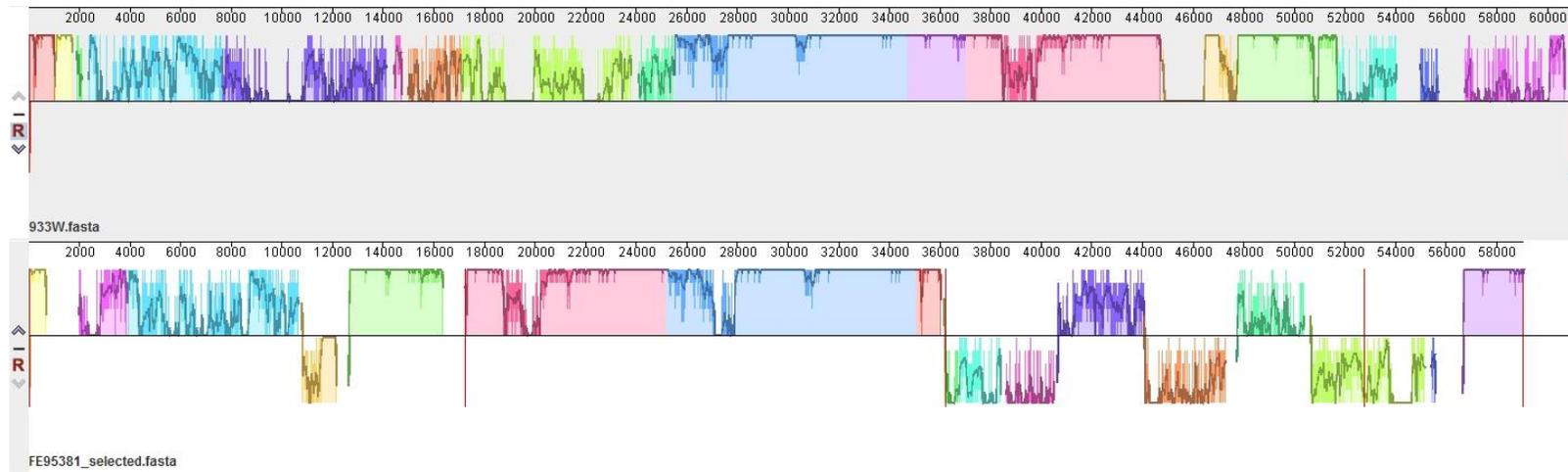


Figure 3.21. Global Alignment between Bacteriophage Genome Came from STEC FE 95381 to Bacteriophage 933W.

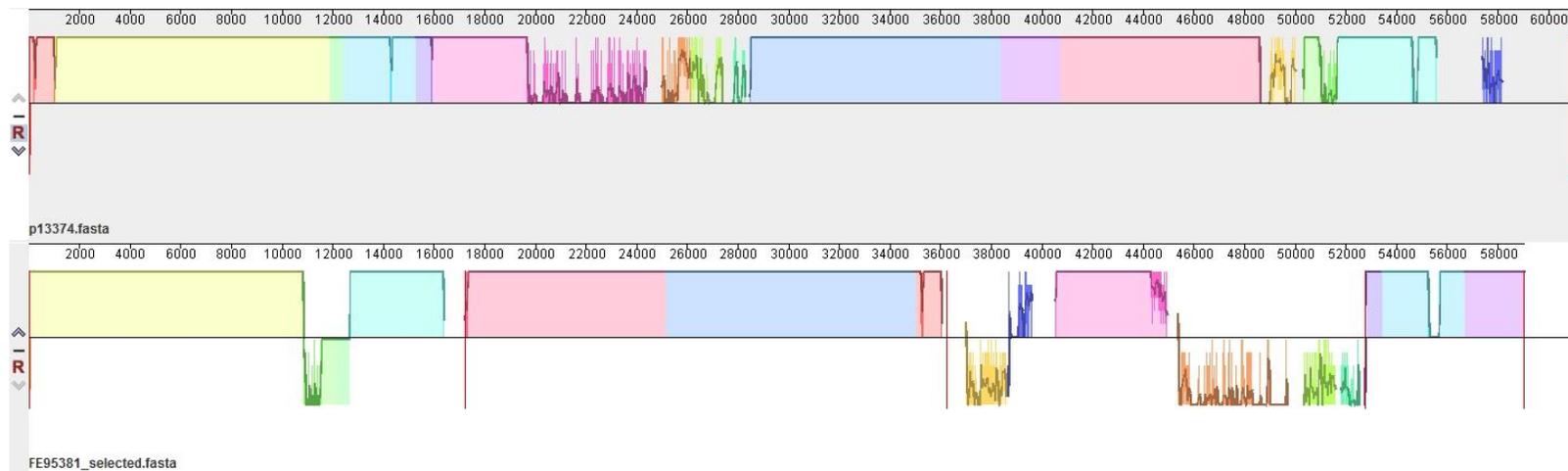


Figure 3.22. Global Alignment between Bacteriophage Genome Came from STEC FE 95381 to Bacteriophage p13374.

## 5. Phylogenetic Relationship

Phylogenetic tree was constructed using maximum likelihood algorithm. 9 sample genomes and the 2 reference genomes were grouped into 2 branches. The traditional bacteriophage reference was clustered with 6 sample genomes, which including 5 sample genomes from the O111 STEC isolates and the other one from O130; the novel bacteriophage reference was grouped with 3 strains, including all sample genomes from the O130 STEC isolates and another from O74.

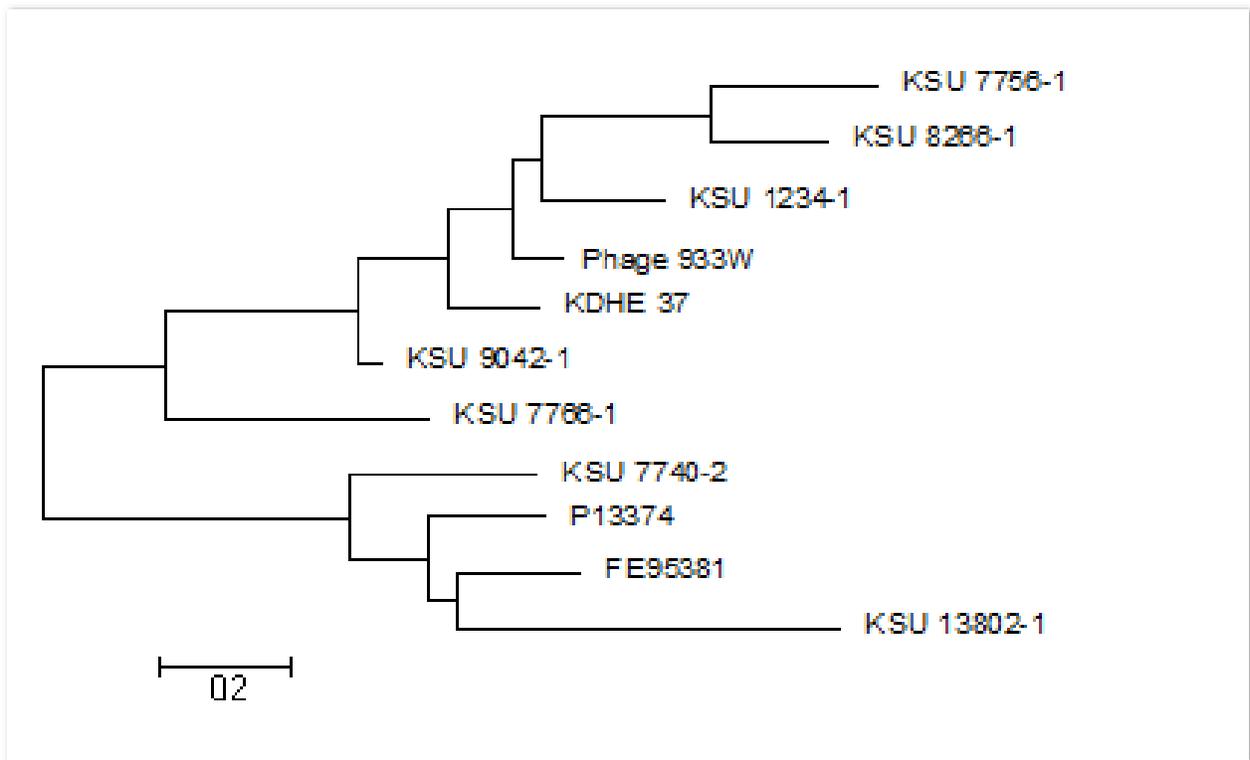


Figure 3.23. Phylogenetic Tree of Sample Genomes and 2 Reference Genomes.

## 6. Gene Annotation

Genes of sample genomes were identified by aligning to the genes of reference genomes. The existence of genes of each sample genome was summarized. Both nucleotide sequences and translated-protein sequences were compared.

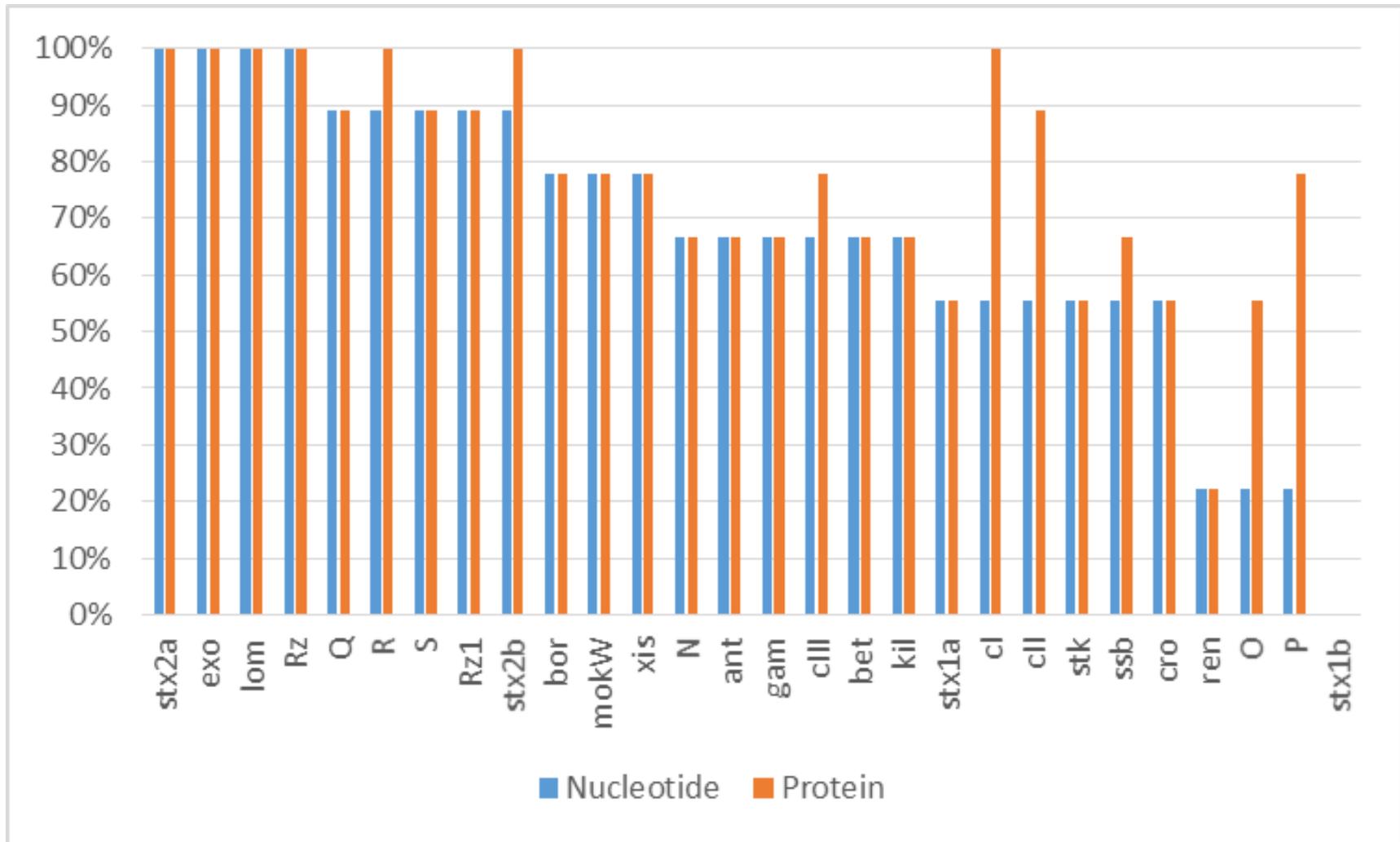


Figure 3.24. Existence rate of genes from reference genomes in the bacteriophage genome samples. Blue bars: results based on nucleotide sequence comparison. Orange bars: results based on protein sequence comparison.

## CHAPTER 4 Discussion

### 1. Reasons of Low Induction Rate

Bacteriophage induction is the primary part of the project. Although 86 STEC strains were collected, only 13 bacteriophage samples were successfully induced. Various inducing methods have been tried including an increased mitomycin C concentration, elongated incubating time and ultraviolet irradiation was used as an alternative method to induce bacteriophages. All these methods turned out failed to induce bacteriophages out of the STEC strains. By analyzing the background of STEC strains, strains isolated of different origins shown difference in induction rate.

According to Figure 2, all these clinical strains were successfully induced. In contrast, the induction rate of STEC from other sources were significantly lower than clinical strains. Currently, the most efficient method to identify STEC strains is to detect the existence of *stx* in the bacterial genome. However, PCR test is unable to confirm if the *stx* is a latent gene or a piece of DNA residue. Ideally, the *stx* will be activated when the host bacterium lives in an adverse condition. The expression of *stx* begins after the linear probacteriophage exits from bacterial chromosome and form a circular DNA. Failure to reform the circular DNA will not activate the expression of *stx*. Hence, the produce of Shiga toxin is the sequential events happened after the probacteriophage is induced to enter lytic cycle. However, a probacteriophage can hardly maintain genome originality because of the attack from bacterial DNA repair system and recombination events with other similar bacteriophage DNA. Any mutation occurred in genes responsible for the separation from bacterial genome or the formation from linear to circular DNA will inactivate all activities of the bacteriophage, which can be considered as caused the bacteriophage to die. Since the mutation happens randomly, a positive *stx* PCR test is not enough to confirm the activity of the Stx bacteriophage. The housekeeping genes of the bacteriophage may have been mutated and the *stx* thus became a useless DNA residue too.

The mutation of Stx bacteriophage may attribute to the low induction rate of the project. Although all strains carry *stx*, only clinical strains had proved the activity of the Stx bacteriophage it carried through the virulence it shown in patients. The mitomycin C treatment used in the project helped to select STEC containing functional Stx probacteriophage. For other STEC strains, the existence of *stx* indicates that they were infected by Stx bacteriophage and those no longer functional bacteriophages will not be a latent threat to them anymore. Although those Stx bacteriophage are unable to be induced, sequencing on those malfunctioned probacteriophages will still be valuable. These sequences will show the information about housekeeping genes of Stx bacteriophages, which are even more critical to bacteriophages than virulence genes. In this case, the knowledge on these housekeeping genes maybe more useful in the regulation and control of Stx bacteriophages than focusing on virulence genes.

The induction rate of recently isolated STEC strains is higher than the old strains. Commonly, STEC strains are stored at -80°C, which can be considered as an extreme condition for both bacteria and the co-existing probacteriophages. Although low temperature significantly inhibits the metabolisms, activities related to stress reactions are able to be activated when the storage time is long enough. Most strains in the study were isolated more than 15 years ago (Isolated in 1998). The long-term low temperature stress may have resulted in abnormal genetic activities that result in noninducible probacteriophages.

The relationship between serotypes and induction rate is weak according to the results. One deficiency in the study is that the number of certain serotypes is too low to show statistical significance. Although the study included 24 serotypes, all serotypes had less than 4 strains except STEC O111 which had 22 strains. On the other hand, genes involved in probacteriophage recovery are located in early gene zone, which are conservative due to their importance in controlling essential activities. Serotypes of STEC cells indicate divert protein profiles on cell membrane, which may related to the attachment and infection of bacteriophage particles but not a strong factor that affect the maintenance and recovery of probacteriophages inside the bacterial chromosome.

## 2. Acquisition of Bacteriophage Genomes

Although PCR results indicated a reliable genome extraction of inducible bacteriophage samples, the contamination of *E. coli* K-12 chromosome inside the bacteriophage genomes greatly impeded the quality of the sequencing results. Original data ranges from 2.7 Mbp to 4.8Mbp, which is much larger than Stx bacteriophage genome size (60 – 80 Kbp). Although the quality of sequencing was acceptable, the coverage of sequencing reads was not high enough to separate the bacteriophage from the bacterial genomes. A mixed profile of contigs can only be separated by comparing each contigs to the whole bacterial genome rather than assembly which is basically an inter-contig comparison process.

By comparing raw data to *E. coli* K-12 sequence, the average of genome size dropped from 3.6 Mbp to 750 Kbp. However, 4 bacteriophage sample genomes were found that all contigs belong to K-12 sequences. The result inferred that these genome samples only contain a very low proportion of bacteriophage genome among the bacterial genomes. In addition, these 4 bacteriophage samples were collected through broth incubation, in which bacterial DNA was supposed to be digested by adding DNase I. Since a large amount of bacteria was lysed during bacteriophage incubation, the bacterial genome concentration in the final product maybe too high to be completely digested by DNase. The other samples that were extracted from plaques contained a higher proportion of bacteriophage particles, thus the bacteriophage genomes were read during sequencing. Even though, the bacterial genomes were still involved in the sequencing results.

The contamination of host cells may be a strong obstacle for bacteriophage sequencing. Pure bacteriophage stocks can hardly be acquired because bacteria cannot be treated as vulnerable victims to bacteriophages in community level. Due to short life cycle, bacteria are able to make prompt responses to generate bacteriophage-resistant individuals. Therefore the resistant mutants are not discriminable from bacteriophage particles when collecting bacteriophages. In this study, extended incubation of clear broth containing bacteriophages and bacteria mixture sometimes resulted in an opaque solution when

amplifying bacteriophages using broth. This is a visible result that bacteriophages – as an evolutionary force – directed the prevalence of bacteriophage resistant phenotype within several life cycles.

Since a significant large amount of contigs contained both bacterial and bacteriophage sequence fragments, only very strict comparing threshold can be applied when comparing to L-12 chromosome to avoid the potential deletion of bacteriophage contigs. The side effect of this strategy was the remaining of some non-bacteriophage origin contigs in the genome, which requires the comparison of remaining contigs to available Stx bacteriophage genomes. Two main factors are responsible for the combination of bacteriophage and bacteria fragments into one contig. First, contig assembly based on reference genomes strongly relies on the quality of previous sequenced data. Failed to distinguish extrinsic DNA fragments from a genome is not uncommon in gene annotation. These errors possibly lead to incorrect assembly of newly sequenced data. Another possible factor is the high frequency of gene recombination events between probacteriophages and their host chromosome. Bacteriophage genome carrying bacteria DNA fragments tends to be assembled together and clustered to other bacterial fragments due to the high reliability of bacteria sequences.

Cleaned genomes contain 4 – 6 contigs, genome size ranges from 60 Kbp to 70 Kbp. Some deleted contigs also contain fragments that are highly matched to Stx bacteriophages. However, the contigs are not reserved if the length of match is shorter than 500 bp since the sequencing kit used in the study generates reads of 500 bp. Shorter matches indicate a linkage to fragments of non-bacteriophage origins. The deletion of these domains may affect the quality of genomic analysis in a limited range. However, if these contigs were reserved, the prevalence of non-bacteriophage fragments would be even more turbulent due to the small genome size of bacteriophages.

### 3. Bacteriophage Genome Analysis

The global alignment results indicate that 7 bacteriophage sample genomes contain at least one complementary region inside the genome. These regions are evidences of gene recombination events due

to high similarity. 8 of the 9 bacteriophage genomes were from STEC strains isolated in 1998, which was very close the time that bacteriophage 933W was sequenced. But the alignment results still show divergence especially in the region of 9000 bp to 12000 bp, which is the region of immediate early genes. The region also includes *cI*, which is the only activating gene after the bacteriophage inserts into bacterial chromosome. The failure of *cI* expression will disturb the maintenance of probacteriophage and result in recovery of probacteriophage. In this study, *cI* may be affected by mitomycin C, which disturbs nucleotide activities and hence induces the variance of the gene to activate the conversion of life cycle.

Another divert region according to the alignment results is the late gene region. Gene variation, gene recombination and gene loss events can be recognized in this region. Late gene region encodes secondary metabolites including toxins, proteins that compose outer envelope and tails. A probacteriophage is tolerant to variations occurred in this region since these genes are temporarily silent. Even bacteriophage in lytic cycle can tolerate mutations and combinations of late gene regions since essential regulations will not be affected.

In this study, genomes are divided into two groups. All 5 bacteriophage sample genomes came from STEC O111 serotypes are clustered together with bacteriophage 933W sequence, while the 2 bacteriophage sample genomes came from STEC O163 serotypes are grouped to bacteriophage p13374. Bacteriophage infection starts with the recognition between bacteriophage tail proteins to specific receptors on cell surface. Serotypes that link to cell surface antigens indicate a variety of surface proteins can only be attached by tails of certain bacteriophage variants. Thus a link between the serotypes of STEC hosts and the phylogenetic relationship is constructed.

The phylogenetic tree also reflects that the 2 references used in this study are divert. Although bacteriophage p13374 was isolated from a very violent STEC strain, the pathogenicity of STEC is not solely rely on the expression of Shiga toxin. The efficiency of bacteriophage p13374 in producing Shiga toxin possibly differs from bacteriophage 933W since the upstream fragments of *stx* are not identical.

However, over-produced Shiga toxin also reduces microbial load, the final pathogenic efficiency is unknown by just focusing on Shiga toxin.

Significant differences are shown in gene identification results using nucleotide and protein sequences. Protein sequences are more conservative than nucleotide sequences, which indicates a large amount of silent mutation events and replacement of similar segment events. This condition is especially significant for essential regulation genes including delayed early genes *cI* and *cII* that determine the life cycle of bacteriophage particles, as well as *O* and *P* which activate DNA replication after probacteriophage recovers from lysogeny.

Although the original information of STEC isolates shows that some isolates contain *stxI*, the sequences indicate that none of them contain a complete *stxI*. According to previous studies, the coexistence of *stx1* and *stx2* was detected, while was not a prevalent condition. Because of the similarity of these 2 genes, they need to compete with each other for the specific locus as well as related regulation elements. Based on this study, *stx1* tends to loss from the genome due to the fact that all 5 *stxI*-containing bacteriophage samples only reserve *stx1a*. Another possible condition is that the pore formed by Shiga toxin 2 subunit B complex is compatible to the entrance of both Shiga toxin 1 subunit A and Shiga toxin 2 subunit A. Thus the loss of *stx1b* benefits the bacteriophage due to the relieved gene load.

## CHAPTER 5 Conclusion

Although the definition of Shiga toxin-producing *Escherichia coli* is the existence of *stx*, the identification of *stx* from microorganisms only indicates the potential of producing Shiga toxin. A small portion of STECs from animal and environment origins remain the ability to recover Stx probacteriophage to encode *stx*. The long term latency of Stx probacteriophages inside a host chromosome gives sufficient chances for genetic variation events to occur. Many mutations result in the inactivation of bacteriophages, some variations are directed by evolutionary stresses such as the interaction with host. Mutation frequency of housekeeping genes does not differ to other genes significantly, the rate of silent mutation of housekeeping genes is higher than late genes and protein-encoding genes. *stx1* and *stx2* locate in the same locus inside Stx bacteriophage genome, and the coexistence of the 2 virulent genes is not a stable condition and *stx1* tends to loss.

Although sequencing technology is increasingly powerful, bacteriophage genome sequencing has its own obstacles including the low purity of bacteriophage particle sources, the limited database of reference bacteriophage genomes and the negligence of correct annotating genes in previous microbial genomes. Stx bacteriophage genomes are not well studied and the sequences acquired in this study can be analyzed in further research when a larger sample size becomes available.

## Reference

- [1] Konowalchuk, J., J. I. Speirs, and S. Stavric. "Vero response to a cytotoxin of *Escherichia coli*." *Infection and immunity* 18.3 (1977): 775-779.
- [2] Beutin, L., et al. "Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals." *Journal of Clinical Microbiology* 31.9 (1993): 2483-2488.
- [3] Degan, Patrick H., and Nancy A. Moran. "Diverse bacteriophage-encoded toxins in a protective insect endosymbiont." *Applied and environmental microbiology* 74.21 (2008): 6782-6791.
- [4] Frank, Christina, et al. "Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104: H4 outbreak in Germany." *New England Journal of Medicine* 365.19 (2011): 1771-1780.
- [5] Adamu, M. T., et al. "A Review on *Escherichia coli* O157: H7-The Super Pathogen." *Health* 5.2 (2014): 118-134.
- [6] Riley, Lee W., et al. "Hemorrhagic colitis associated with a rare *Escherichia coli* serotype." *New England Journal of Medicine* 308.12 (1983): 681-685.
- [7] Bosilevac, Joseph M., and Mohammad Koohmaraie. "Prevalence and characterization of non-O157 Shiga toxin-producing *Escherichia coli* isolates from commercial ground beef in the United States." *Applied and environmental microbiology* 77.6 (2011): 2103-2112.
- [8] Gould, L. Hannah, et al. "Increased recognition of non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States during 2000–2010: Epidemiologic features and comparison with *E. coli* O157 infections." *Foodborne pathogens and disease* 10.5 (2013): 453-460.
- [9] van Elsas, Jan Dirk, et al. "Survival of *Escherichia coli* in the environment: fundamental and public health aspects." *The ISME journal* 5.2 (2011): 173-183.
- [10] Wang, Fei, et al. "Current Trends in Detecting Non-O157 Shiga Toxin-Producing *Escherichia coli* in Food." *Foodborne pathogens and disease* 10.8 (2013): 665-677.

- [11] Krüger, Alejandra, and Paula MA Lucchesi. "Shiga toxins and *stx* bacteriophages: highly diverse entities." *Microbiology* 161.Pt 3 (2015): 451-462.
- [12] Melton-Celsa, Angela R. "Shiga toxin (Stx) classification, structure, and function." *Microbiology spectrum* 2.2 (2014).
- [13] Sandvig, Kirsten, et al. "Endocytosis and retrograde transport of Shiga toxin." *Toxicon* 56.7 (2010): 1181-1185.
- [14] Tironi-Farinati, Carla, et al. "Intracerebroventricular Shiga toxin 2 increases the expression of its receptor globotriaosylceramide and causes dendritic abnormalities." *Journal of neuroimmunology* 222.1 (2010): 48-61.
- [15] Obrig, Tom G. "*Escherichia coli* Shiga toxin mechanisms of action in renal disease." *Toxins* 2.12 (2010): 2769-2794.
- [16] Fuller, Cynthia A., et al. "Shiga toxin subtypes display dramatic differences in potency." *Infection and immunity* 79.3 (2011): 1329-1337.
- [17] Pacheco, Alline R., and Vanessa Sperandio. "Shiga toxin in enterohemorrhagic *E. coli*: regulation and novel anti-virulence strategies." *Frontiers in cellular and infection microbiology* 2 (2012).
- [18] Kawano, Kimiko, et al. "Relationship between *stx* Genotype and Stx2 Expression Level in Shiga Toxin-Producing *Escherichia coli* O157 Strains." *Japanese journal of infectious diseases* 65.4 (2012): 322-325.
- [19] Stevens, Joanne M., Edouard E. Galyov, and Mark P. Stevens. "Actin-dependent movement of bacterial pathogens." *Nature Reviews Microbiology* 4.2 (2006): 91-101.
- [20] Tzipori, Saul, et al. "The role of the *eaeA* gene in diarrhea and neurological complications in a gnotobiotic piglet model of enterohemorrhagic *Escherichia coli* infection." *Infection and Immunity* 63.9 (1995): 3621-3627.

- [21] Mora, Azucena, et al. "Seropathotypes, phylogroups, Stx subtypes, and intimin types of wildlife-carried, Shiga toxin-producing *Escherichia coli* strains with the same characteristics as human-pathogenic isolates." *Applied and environmental microbiology* 78.8 (2012): 2578-2585.
- [22] Moonens, Kristof, et al. "Nanobody mediated inhibition of attachment of F18 fimbriae expressing *Escherichia coli*." *PloS one* 9.12 (2014): e114691.
- [23] Vögeli, P., et al. "Genes specifying receptors for F18 fimbriated *Escherichia coli*, causing oedema disease and postweaning diarrhoea in pigs, map to chromosome 6." *Animal genetics* 27.5 (1996): 321-328.
- [24] Torres, Alfredo G., et al. "Outer membrane protein A of *Escherichia coli* O157: H7 stimulates dendritic cell activation." *Infection and immunity* 74.5 (2006): 2676-2685.
- [25] Foster, Debora Barnett, et al. "Enterohemorrhagic *Escherichia coli* Induces Apoptosis Which Augments Bacterial Binding and Phosphatidylethanolamine Exposure on the Plasma Membrane Outer Leaflet." *Infection and immunity* 68.6 (2000): 3108-3115.
- [26] Herold, Sylvia, James C. Paton, and Adrienne W. Paton. "Sab, a novel autotransporter of locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* O113: H21, contributes to adherence and biofilm formation." *Infection and immunity* 77.8 (2009): 3234-3243.
- [27] Ogierman, Monica A., Adrienne W. Paton, and James C. Paton. "Up-Regulation of Both Intimin and *eaeA*-Independent Adherence of Shiga Toxigenic *Escherichia coli* O157 by *ler* and Phenotypic Impact of a Naturally Occurring *ler* Mutation." *Infection and immunity* 68.9 (2000): 5344-5353.
- [28] Cockerill, F. I. I. I., et al. "Polysaccharide side chains are not required for attaching and effacing adhesion of *Escherichia coli* O157: H7." *Infection and immunity* 64.8 (1996): 3196-3200.
- [29] Gyles, C. L. "Shiga toxin-producing *Escherichia coli*: An overview." *Journal of animal science* 85.13\_suppl (2007): E45-E62.
- [30] Zhang, Xiaoi, et al. "Enterohemorrhagic *Escherichia coli* specific enterohemolysin induced IL-1 $\beta$  in human macrobacteriophages and EHEC-induced IL-1 $\beta$  required activation of NLRP3 inflammasome." (2012): e50288.

- [31] Tarr, Phillip I., Carrie A. Gordon, and Wayne L. Chandler. "Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome." *The Lancet* 365.9464 (2005): 1073-1086.
- [32] Monaghan, Áine, et al. "Serotypes and virulotypes of non-O157 shiga-toxin producing *Escherichia coli* (STEC) on bovine hides and carcasses." *Food microbiology* 32.2 (2012): 223-229.
- [33] Martínez-Castillo, Alexandre, and Maite Muniesa. "Implications of free Shiga toxin-converting bacteriophages occurring outside bacteria for the evolution and the detection of Shiga toxin-producing *Escherichia coli*." *Frontiers in cellular and infection microbiology* 4 (2014).
- [34] Mingle, Lisa A., et al. "Enhanced identification and characterization of non-O157 Shiga toxin-producing *Escherichia coli*: a six-year study." *Foodborne pathogens and disease* 9.11 (2012): 1028-1036.
- [35] Ogura, Yoshitoshi, et al. "Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*." *Proceedings of the National Academy of Sciences* 106.42 (2009): 17939-17944.
- [36] Bugarel, Marie, Lothar Beutin, and Patrick Fach. "Low-density microarray targeting non-locus of enterocyte effacement effectors (*nle* genes) and major virulence factors of Shiga toxin-producing *Escherichia coli* (STEC): a new approach for molecular risk assessment of STEC isolates." *Applied and environmental microbiology* 76.1 (2010): 203-211.
- [37] Coombes, Brian K., Matthew W. Gilmour, and Chelsey D. Goodman. "The evolution of virulence in non-O157 Shiga toxin-producing *Escherichia coli*." *Frontiers in microbiology* 2 (2011).
- [38] Xu, Xuefang, et al. "Lysogeny with Shiga toxin 2-encoding bacteriophages represses type III secretion in enterohemorrhagic *Escherichia coli*." *PLoS Pathog* 8.5 (2012): e1002672.
- [39] Lewis, Steven B., et al. "Enterohemorrhagic *Escherichia coli* colonization of human colonic epithelium *in vitro* and *ex vivo*." *Infection and immunity* 83.3 (2015): 942-949.

- [40] Herold, Sylvia, Helge Karch, and Herbert Schmidt. "Shiga toxin-encoding bacteriophages—genomes in motion." *International journal of medical microbiology* 294.2 (2004): 115-121.
- [41] Acheson, David WK, et al. "In vivo transduction with Shiga toxin 1-encoding bacteriophage." *Infection and immunity* 66.9 (1998): 4496-4498.
- [42] Bielaszewska, Martina, et al. "Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic *Escherichia coli* O26 infection in humans." *Applied and environmental microbiology* 73.10 (2007): 3144-3150.
- [43] Strockbine, Nancy A., et al. "Two toxin-converting bacteriophages from *Escherichia coli* O157: H7 strain 933 encode antigenically distinct toxins with similar biologic activities." *Infection and immunity* 53.1 (1986): 135-140.
- [44] Plunkett, Guy, et al. "Sequence of Shiga Toxin 2 Bacteriophage 933W from *Escherichia coli* O157: H7: Shiga Toxin as a Bacteriophage Late-Gene Product." *Journal of Bacteriology* 181.6 (1999): 1767-1778.
- [45] Makino, Kozo, et al. "Complete nucleotide sequence of the probacteriophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157: H7 derived from the Sakai outbreak." *Genes & genetic systems* 74.5 (1999): 227-239.
- [46] Yokoyama, Katsushi, et al. "Complete nucleotide sequence of the probacteriophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157: H7 strain derived from the Sakai outbreak." *Gene* 258.1 (2000): 127-139.
- [47] Perna, Nicole T., et al. "Genome sequence of enterohaemorrhagic *Escherichia coli* O157: H7." *Nature* 409.6819 (2001): 529-533.
- [48] Krebs, Jocelyn E., Elliott S. Goldstein, and Stephen T. Kilpatrick. *Lewin's genes XI*. Jones & Bartlett Learning, 2014.
- [49] Snyder, Larry, et al. *Molecular genetics of bacteria*. American Society of Microbiology, 2013.
- [50] Francki, Richard Ignacy Bartolomiej, et al., eds. *Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses*. Virology Division of the

*International Union of Microbiological Societies*. Vol. 2. Springer Science & Business Media, 2012.

- [51] Ackermann, Hans-W. "Tailed bacteriophages: the order Caudovirales." *Advances in virus research* 51 (1999): P135-P202.
- [52] Willshaw, G. A., et al. "Heterogeneity of *Escherichia coli* bacteriophages encoding Vero cytotoxins: comparison of cloned sequences determining VT1 and VT2 and development of specific gene probes." *Journal of general microbiology* 133.5 (1987): 1309-1317.
- [53] Müller, Ines, et al. "Spread of a Distinct Stx2-Encoding Bacteriophage." *J. Virol* 86.19 (2012): 10444.
- [54] Steyert, Susan R., et al. "Comparative genomics and stx bacteriophage characterization of LEE-negative Shiga toxin-producing *Escherichia coli*." *Frontiers in cellular and infection microbiology* 2 (2012).
- [55] Beutin, Lothar, et al. "Spread of a distinct Stx2-encoding bacteriophage prototype among *Escherichia coli* O104: H4 strains from outbreaks in Germany, Norway, and Georgia." *Journal of virology* 86.19 (2012): 10444-10455.
- [56] Datz, Martina, et al. "Analysis of the enterohemorrhagic *Escherichia coli* O157 DNA region containing lambdoid bacteriophage gene p and Shiga-like toxin structural genes." *Applied and environmental microbiology* 62.3 (1996): 791-797.
- [57] Eppinger, Mark, et al. "Genomic anatomy of *Escherichia coli* O157: H7 outbreaks." *Proceedings of the National Academy of Sciences* 108.50 (2011): 20142-20147.
- [58] Su, L. K., et al. "Lysogenic infection of a Shiga toxin 2-converting bacteriophage changes host gene expression, enhances host acid resistance and motility." *Molecular biology* 44.1 (2010): 54-66.
- [59] Allison, Heather E., et al. "Immunity profiles of wild-type and recombinant Shiga-like toxin-encoding bacteriophages and characterization of novel double lysogens." *Infection and immunity* 71.6 (2003): 3409-3418.

- [60] Schaefer, KATHERINE L., and WILLIAM R. McCLURE. "Antisense RNA control of gene expression in bacteriophage P22. II. Kinetic mechanism and cation specificity of the pairing reaction." *RNA* 3.2 (1997): 157.
- [61] Fogg, P. C. M., et al. "Characterization of the relationship between integrase, excisionase and antirepressor activities associated with a superinfecting Shiga toxin encoding bacteriophage." *Nucleic acids research* (2010): gkq923.
- [62] Herold, Sylvia, et al. "Global expression of probacteriophage genes in *Escherichia coli* O157:H7 strain EDL933 in response to norfloxacin." *Antimicrobial agents and chemotherapy* 49.3 (2005): 931-944.
- [63] Dobrindt, Ulrich, et al. "Genomic islands in pathogenic and environmental microorganisms." *Nature Reviews Microbiology* 2.5 (2004): 414-424.
- [64] Ju, Wenting, et al. "Distribution of pathogenicity islands OI-122, OI-43/48, and OI-57 and a high-pathogenicity island in Shiga toxin-producing *Escherichia coli*." *Applied and environmental microbiology* 79.11 (2013): 3406-3412.
- [65] Coombes, Brian K. "Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria." *Trends in microbiology* 17.3 (2009): 89-94.
- [66] Croxen, Matthew A., and B. Brett Finlay. "Molecular mechanisms of *Escherichia coli* pathogenicity." *Nature Reviews Microbiology* 8.1 (2010): 26-38.
- [67] Coombes, Brian K., et al. "Molecular analysis as an aid to assess the public health risk of non-O157 Shiga toxin-producing *Escherichia coli* strains." *Applied and environmental microbiology* 74.7 (2008): 2153-2160.
- [68] Muniesa, Maite, Francisco Lucena, and Juan Jofre. "Comparative survival of free Shiga toxin 2-encoding bacteriophages and *Escherichia coli* strains outside the gut." *Applied and environmental microbiology* 65.12 (1999): 5615-5618.

- [69] Langsrud, Solveig, Even Heir, and Tone Mari Rode. "Survival of Shiga toxin-producing *Escherichia coli* and Stx bacteriophages in moisture enhanced beef." *Meat science* 97.3 (2014): 339-346.
- [70] Casas, Veronica, and Stanley Maloy. "Role of bacteriophage-encoded exotoxins in the evolution of bacterial pathogens." *Future microbiology* 6.12 (2011): 1461-1473.
- [71] Bielaszewska, Martina, et al. "Effects of antibiotics on Shiga toxin 2 production and bacteriophage induction by epidemic *Escherichia coli* O104: H4 strain." *Antimicrobial agents and chemotherapy* 56.6 (2012): 3277-3282.
- [72] Green, Michael R., and Joseph Sambrook. *Molecular cloning: a laboratory manual*. Vol. 1. New York: Cold Spring Harbor Laboratory Press, 2012.
- [73] Köhler, Bernd, Helge Karch, and Herbert Schmidt. "Antibacterials that are used as growth promoters in animal husbandry can affect the release of Shiga-toxin-2-converting bacteriophages and Shiga toxin 2 from *Escherichia coli* strains." *Microbiology* 146.5 (2000): 1085-1090.
- [74] Dowd, Scot E. "*Escherichia coli* O157: H7 gene expression in the presence of catecholamine norepinephrine." *FEMS microbiology letters* 273.2 (2007): 214-223.
- [75] Łoś, Joanna M., et al. "Hydrogen peroxide-mediated induction of the Shiga toxin-converting lambdoid probacteriophage ST2-8624 in *Escherichia coli* O157: H7." *FEMS Immunology & Medical Microbiology* 58.3 (2010): 322-329.
- [76] Wagner, Patrick L., David WK Acheson, and Matthew K. Waldor. "Human neutrophils and their products induce Shiga toxin production by enterohemorrhagic *Escherichia coli*." *Infection and immunity* 69.3 (2001): 1934-1937.
- [77] Tarr, Phillip I., Carrie A. Gordon, and Wayne L. Chandler. "Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome." *The Lancet* 365.9464 (2005): 1073-1086.
- [78] Sanger, Frederick, Steven Nicklen, and Alan R. Coulson. "DNA sequencing with chain-terminating inhibitors." *Proceedings of the National Academy of Sciences* 74.12 (1977): 5463-5467.

- [79] Prober, James M., et al. "A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides." *Science* 238.4825 (1987): 336-341.
- [80] Ronaghi, Mostafa, Mathias Uhlen, and Pål Nyren. "A sequencing method based on real-time pyrophosphate." *Science* 281.5375 (1998): 363-365.
- [81] Pfeifer, Gerd P., et al. "Genomic sequencing and methylation analysis by ligation mediated PCR." *Science* 246.4931 (1989): 810-813.
- [82] Dramanac, Radoje, et al. "Sequencing of megabase plus DNA by hybridization: theory of the method." *Genomics* 4.2 (1989): 114-128.
- [83] Liu, Lin, et al. "Comparison of next-generation sequencing systems." *BioMed Research International* 2012 (2012).
- [84] Koboldt, Daniel C., et al. "The next-generation sequencing revolution and its impact on genomics." *Cell* 155.1 (2013): 27-38.
- [85] Fullwood, Melissa J., et al. "Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses." *Genome research* 19.4 (2009): 521-532.
- [86] Cox, Murray P., Daniel A. Peterson, and Patrick J. Biggs. "SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data." *BMC bioinformatics* 11.1 (2010): 485.