Molecular Characterization of *Enterohemorrhagic Escherichia coli* (EHEC): O *rough* strains and the prevalence and importance of IS*629* in *E. coli* O157:H7

DISSERTATION

zur Erlangung des akademischen Grades

DOKTOR DER NATURWISSENSCHAFTEN

an der Universität Hamburg, Deutschland

Department Chemie
Institut für Lebensmittelchemie

In Partial Fulfillment of the Requirements for the

Degree of *Dr. rer. nat.*

at the University of Hamburg, Germany

Department of Chemistry
Institute for Food Chemistry

Division of Microbiology, U.S. Food and Drug Administration,
5100 Paint Branch Parkway, College Park, MD 20740, USA

Von
Lydia Vanessa Rump
aus Norderstedt

Hamburg 2011

© Lydia Vanessa Rump, 2011.
All rights reserved.

The present work was performed in the period between November 2007 till February 2011 under the supervision of **Prof. Dr. M. Fischer**, **Dr. E. Brown** and **T. Hammack** at the U.S. Food and Drug Administration, Division of Microbiology, College Park, USA in collaboration with the University of Hamburg, Hamburg, Germany.

Dissertation Committee: Prof. Dr. Fischer
Prof. Dr. Heisig

Disputation: 16. September 2011

Disputation committee: Prof. Dr. Fischer
Prof. Dr. Bisping
Prof. Dr. Hackl

Illustration du Chat Botté par Gustave Doré (1832-1883)

For My Parents and Narjol

ACKNOWLEDGEMENTS

There are lots of people I would like to thank for a huge variety of reasons. I owe my gratitude to all those people who have made this dissertation possible and because of whom my graduate experience has been one that I will look back at forever.

It is difficult to overstate my gratitude to my mentor, **Dr. Eric Brown**, and my supervisor **Thomas Hammack** for the opportunity to work in the Division of Microbiology and for their help with various applications for funding of the project. Their tremendous efforts in weal and woe are highly appreciated. Thank you also for managing to read the whole thesis so thoroughly and for helpful comments on the text. Without your knowledge, perceptiveness, cracking-of-the-whip and encouragement I would never have finished.

I would like to acknowledge the support of **Dr. Jeanne Rader**, **Dr. Paul Whittacker,** and **Dr. Mary Trucksess** who believed in me and fought to get me into the Ph.D. program in the first place.

I also gratefully acknowledge the supervision of **Prof. Dr. Marcus Fischer** during this work and his confidence in me. Your patience and support helped me overcome many crisis situations and finish this dissertation. Thanks for the many insightful conversations during the development of the ideas in this thesis.

I also want to thank **Prof. Dr. Bernward Bisping**, for instilling in me the qualities of being a good scientist. Your infectious enthusiasm and unlimited zeal have been major driving forces through my graduate career at the University of Hamburg. Additionally, I also wish to acknowledge the other members of the editorial and disputation panels **Prof. Dr. Heising** and **Prof. Dr. Haas.**

Thanks to all members of the **Division of Microbiology** for their friendship and supports. Special thanks to **Christine Keys, Guojie Cao, Dr. Insook Sun, Dr. Marc Allard, Dr. Peter Feng,** and **Dr. Steven Monday** for the help with experimental setup, general advice and numerous stimulating discussions.

For financial support, I thank the Center for Food Safety and Applied Nutrition (CFSAN) for providing the Oak Ridge Associated Universities (ORAU) scholarship in the Research Fellowship Program.

A big “thank you” to my friends **Dr. Angel Soler, Antonio DeJesus**, **Brenda Kroft**, **Cindy Kroon**, **Dr. Guodong Zhang**, **James Nides**, **Julie Halstead, Dr. Julie Kase, Dr. Karen Millians, Dr. Michael McLaughlin, Catharina Lüdeke, Nicole Schönekäß, Roberto D’Arminio, Sandra Lorenz, Shereen Tewfik, Dr. Steve Casper,** and **Petia Vladimirova.** You helped me staying sane through these difficult years and never stopped listening and supporting me during this significant part of my personal and professional education.

Finally, I am forever indebted to my **parents** and **Narjol** for their understanding, endless patience and encouragement when it was most required, my brother **Patrick**, for long-distance cheerleading and most of all **Mami** and **Papi** for their courage to send me off into the world to pursue my dreams.

ABSTRACT

Enterohemorrhagic *Escherichia coli* (EHEC) O157:H7 is the most frequent cause of hemorrhagic colitis (HC) and haemolytic-uremic syndrome (HUS) worldwide. The definitive identification of *E. coli* O157:H7 in clinical or food samples is done serologically by testing for the presence of somatic O157 and flagellar H7 antigens. One atypical stx_2 producing *E. coli* strain (MA6) was isolated from beef in Malaysia. It reacted positively for the H7 antigen, however negatively for the presence of the O157 antigen, yet carries the *rfbE* gene necessary for O157 biosynthesis. Therefore, MA6 is an *E. coli* O157:H7 strain genetically, however an O *rough*:H7 strain serologically. This trait makes it undetectable or unidentifiable with most serological assays used in clinical or food sample screening for *E. coli* O157:H7. The source of the lack of O157 antigen expression in MA6 was unknown. After PCR screening of genes involved in the O157 synthesis a 1,310 bp insertion, homologous to IS*629*, was observed within its *gne* gene, encoding an epimerase enzyme essential for the synthesis of an oligosaccharide subunit in the O157 antigen. *Trans*-complementation with a functional *gne* gene from O157:H7 restored O157 antigen expression in MA6. Shiga-toxigenic *E. coli* strains that are O *rough*:H7 due to inactivation of *gne* by IS*629* (*gne*::IS*629*) were thought to be rare and to have unknown pathogenic potential. However recently, another O *rough*:H7 strain caused by *gne*::IS*629* was isolated from a hemorrhagic colitis patient, suggesting that these strains are pathogenic and may not be as rare as previously thought. Insertion elements (IS) are known to play an important role in the evolution and genomic diversification of *Escherichia coli* O157:H7 lineages. In particular, IS*629* has been found in multiple copies in the *E. coli* O157:H7 genome and is one of the most prevalent ISs in this serotype. Numerous IS*629* insertion sites which are not uniformly distributed among strains were found in 4 *E. coli* O157:H7 genome and plasmid sequences. Although highly prevalent in *E. coli* O157:H7 genomes, IS*629* is absent in SFO157 which are on a divergent pathway in the emergence of O157:H7. Although IS*629* deficient, it permits IS*629* transposition with an excision frequency higher than ancestral O55:H7 strains but significantly lower than highly pathogenic O157:H7 strains. Thus, high IS*629* prevalence and high excision frequency in the O157:H7 genomes suggest that IS*629* might not only contribute to the appearance of atypical pathogenic strains like O *rough*:H7 (IS*629*::*gne* mutant), but also might play an integral role in divergence, genome plasticity, and possibly the pathogenicity of this important and dangerous bacterial pathogen.

KURZBESCHREIBUNG

Enterohemorrhagische *Escherichia coli* (EHEC) des Serotyps O157:H7 sind weltweit bekannt als die häufigste Ursache einer hämorrhagischen Colitis (HC) und Hauptauslöser des lebensbedrohlichen hämolytisch-urämischen Syndroms (HUS). *E. coli* O157:H7 wird in Routineuntersuchungen von Lebensmitteln und klinischen Proben mit Antiseren gegen das somatische O157 und das Geißel H7 Antigen nachgewiesen. Kürzlich wurde in Malaysia ein atypischer, Shiga-Toxin produzierender *E. coli* O157:H7 Stamm -MA6- aus Rindfleisch isoliert. Sowohl das H7 Antigen als auch das für die O157 Antigen Biosynthese spezifische *rfb*E Gen konnten in MA6 nachgewiesen werden, jedoch reagierte MA6 negativ für das O157 Antigen. Daher handelt es sich bei MA6 genetisch um einen O157:H7 Stamm, serologisch jedoch um einen O *rough*:H7 Stamm. Aufgrund dieser Eigenschaft ist MA6 mit den meisten serologischen *E. coli* O157:H7 Tests nicht identifizierbar oder nicht nachweisbar. Die Ursache für die Abwesenheit des O157 Antigens war jedoch unklar. Bei der Analyse aller in die O157 Biosynthese involvierten Gene wurde eine Einschiebung der Insertionssequenz (IS) IS*629* in das offene Leseraster des *gne* Gens, welches essenziell für die Synthese einer Oligosaccharid Unterheinheit des O157 Antigens ist, detektiert. Die *Trans*-Komplementierung von MA6 mit einem funktionellen *gne* Gen stellte die O157 Antigen Synthese in diesem Stamm wieder her. Atypische, Stx-produzierende O *rough*:H7 *E. Coli* Stämme, die aufgrund einer *gne*::IS*629* Mutation kein O157 Antigen aufweisen, sind bisher selten und von unbekannter Pathogenität. Allerdings wurde kürzlich ein O *rough*:H7 Stamm von einem HC-Patienten in Deutschland isoliert, was darauf schließen lässt, dass O *rough*:H7 Stämme pathogen und möglicherweise nicht so selten sind, wie zuvor angenommen, ISs sind dafür bekannt, eine wichtige Rolle in der Entwicklung und Diversifizierung von *E. coli* O157:H7 zu spielen. Das *E. coli* O157:H7 Genom trägt mehrere IS*629* Kopien und ist außerdem das am Häufigsten vorhandene IS in diesem Serotyp. Es wurden zahlreiche IS*629* Insertionsstellen in 4 *E. coli* O157:H7 Genom und Plasmiden identifiziert, die jedoch in den Stämmen ungleichmäßig verteilt sind. Darüber hinaus ist IS*629* einzigartig für Nicht-Sorbitol fermentierende (NSF O157) O157:H7 Klone und konnte in den nah verwandten Sorbitol fermentierenden (SF O157) O157:H$^-$ Klonen, die sich auf einem divergenten Weg in der Entstehung von den heutigen O157:H7 Klonen („*E. coli* O157:H7 stepwise model of evolution“) befinden, nicht nachgewiesen werden. Die Abwesenheit von IS*629* in den SF O157-Klonen ist aufgrund der nahen Verwandtschaft zu NSF O157 überraschend. Jedoch ist die Fähigkeit von IS*629* sich zu mobilisieren, in den SF O157 nicht eingeschränkt. Erstaunlicherweise mobilisiert sich IS*629* in den SF O157 Klonen sogar mit einer

größeren Häufigkeit als in den O55:H7 Vorfahren, jedoch mit einer deutlich geringeren Häufigkeit als in den pathogenen O157:H7 Klonen. Die hohe Prävalenz von IS*629* und die erhöhte IS*629*-Mobilisierung in O157:H7 könnte daher nicht nur eine wichtige Rolle in der Genom-Plastizität und der Divergenz von O157:H7 spielen, sondern auch zu der Entstehung von atypischen, pathogenen Stämmen wie O *rough*:H7 (*gne*::IS*629*) beitragen.

TABLE OF CONTENTS

I LIST OF ABBREVIATIONS

Amp	Ampicillin
BfR	Bundesinstitut für Risikobewertung
BLAST	Basic Local Alignment Search Tool
BVL	Bundesamt für Verbraucherschutz und Lebensmittelsicherheit
DNA	Deoxyribonucleic Acid
CBS	Cell Suspension Buffer
CC	Clonal complex
CCD	Charge-coupled device
CDC	Centers for Disease Control and Prevention
CDT	Cytolethal distending toxin
CHEF	Contour-clAmped homogeneous electric field
CLA	Cell Lysis Solution
CRA	Cell Resuspension Solution
CWB	Column Wash Solution
DHEC	diarrhea-associated hemolytic *E. coli*
DI	Distilled water
DR	Direct repeat
ECA	Enterobacterial common antigen
EAEC	enteroaggregative *E. Coli*
EFSA	European Food Safety Authority
EHEC	Enterohemorrhagic *Escherichia coli*
EIEC	enteroinvasive *E. Coli*
EPA	Environmental Protection Agency
EPEC	Enteropathogenic *E. Coli*
ETEC	Enterotoxigenic *E. Coli*
EtOH	Ethanol
Fuc	Fructose
Gal*N*Ac	*N*-acetylgalactosamin

Glc*N*Ac	*N*-acetylglucosamin
GDP	Guanine Diphosphate
Glu	Glucose
HC	Hemorrhagic Colitis
HUS	Hemolytic-uremic syndrome
H_2O	Water
IAA	Isoamyl alcohol
IC	Inner core
IEE	IS-excision enhancer
IM	Inner membrane
IS	Insertion Sequence
IR	Inerted repeat
LB	Luria-Bertani
LPS	Lipopolisaccharide
MCS	Multiple Cloning Site
NCBI	National Center for Biotechnology Information
NaOAc	Sodium Acetate
NSF	Non-sorbitol-fermenting
nt	Nucleotides
OC	Outer core
OD	Optical Density
OM	Outer membrane
ORF	Open Reading Frame
P	Phosphate
PCR	Polymerase chain reaction
PerNAc	N-acetylperosamine
RKI	Robert Koch-Intitute
RNA	Ribonucleic Acid
RT	Room temperature
SAP	Shrimp alkaline phosphatase

SF	Sorbitol-fermenting
SNP	Sugar nucleotide precursors
Stx	Shiga-toxin
SOD	Zinc/copper-type superoxide dismutase
STEC	Shigatoxigenic *Escherichia coli*
Tet	Tetracycline
TCL	Tissue and Cell Lysis
TCSMAC	Tellurite cefixime sorbitol MacConcey agar
UDP	Uridine Diphosphate
und-P	Undecaprenyl phosphate
UTI	Urinary tract infections
WHO	World Health Organization
X-gal	Bromo-chloro-indolyl-galactopyranoside

II LIST OF FIGURES

III LIST OF TABLES

1 INTRODUCTION

Shigatoxigenic *E. coli* (STEC) can cause varying degrees of illness including: diarrhea, hemorrhagic colitis (HC), and the life threatening hemolytic uremic syndrome (HUS) (Hayes et al. 1995). Various STEC serotypes have been implicated in foodborne illness worldwide. Enterohemorrhagic *E. coli* (EHEC) serotype O157 is most commonly associated with severe consequences like HUS, but also many other “non-O157 STEC” serogroups are gaining importance as foodborne pathogens. *E. coli* O26, O91, O103, O111, O118, O145, and O166, are the non-O157 serogroups that most often cause illness in the United States (EFSA, 2011). In May 2011 an unusual STEC serotype O104:H4 caused a major outbreak in Europe (mostly Germany) with 4,321 illnesses, and an unusually high number (885) of HUS cases and deaths (50) (07.25.2011) (EFSA, 2011). In the United States, 6 cases of O104:H4 infections linked to travel to Germany have been identified; 1 death has been reported (CDC, 2011).

EHEC O157:H7 was first recognized in the early 80’s as the cause of bloody diarrhea, utilizing a number of virulence factors, including attaching and effacing factors as well as the production of several cytotoxins (known as shiga toxins), responsible for its pathogenicity (Griffin and Tauxe 1991; Karmali 1989). All these factors however, do not appear to be essential for illness, as EHEC lacking some of the factors have caused severe diseases (Law, 2000). Strains of EHEC O157:H7 are detected by the absence of sorbitol fermentation and glucuronidase activity, and identified serologically by the O157 and H7 antigens. However, atypical O157:H7 are isolated from foods, animals and humans and some have caused illnesses and outbreaks worldwide. These strains do not exhibit typical traits and therefore, are not detected by routine assays used for O157:H7 analysis. Atypical strains of diverse serotypes are emerging and have been implicated in illness, however carrying different phenotypes and virulence factor patterns from those used to identify common STECs like O157:H7 (Feng, 1997; Feng et al., 1998b). Therefore, it is necessary to examine atypical strains using molecular characterization to determine the genetic causes for atypical traits and to identify unique genetic markers that may be used to detect these atypical strains in the food supply.

Previous research found that *E. coli* MA6, a shigatoxigenic O157:H7 strain does not react positively when tested for the presence of O157 antigen with most O157

commercially available kits. However, MA6 carries markers and virulence factors associated with EHEC O157:H7 and, therefore, is genotypically an O157:H7 strain. Furthermore, the strain tested positively by PCR for the *rfbE* gene, which is essential for O157 antigen biosynthesis. However, the O antigen side chain could not be detected by LPS sodium dodecyl sulfate polyacrylamide gel electrophoresis by Western blot analyses (Feng et al., 1998b). Hence, MA6 appears to be a rough (O-side strain deficient) O157:H7 (O *rough*:H7) strain that does not express O157 antigen and, consequently, is serologically undetectable and unidentifiable in routine and outbreak analysis for *E. coli* O157:H7. MA6-like O157:H7 strains that are O *rough* are thought to be rare, as MA6 was the only strain isolated thus far. Moreover, since MA6 was isolated only from a beef sample in Malaysia and was not implicated in illness, the pathogenic potential of this strain was also uncertain. A study in Germany characterized STEC isolates from patients over a 3-year period and found a few strains with the O *rough*:H7 phenotype (Beutin et al., 2004). One of these, CB7326, was isolated from a hemorrhagic colitis patient and found to carry Shiga toxin 1 (stx_1), Shiga toxin 2 (stx_2), and γ-intimin (γ-*eae*) genes, all of which are common in EHEC O157:H7 strains, suggesting that, like MA6, CB7326 may be an O *rough* variant of O157:H7.

1.1 Research Objectives

The detection of pathogens in foods, to ensure food safety, is a primary mission of the Food and Drug Administration (FDA). O157:H7 detection methods target specific phenotypes and genetic traits associated with EHEC pathogenicity. The principal concern for atypical O157:H7 and other EHEC is that they can elude detection, thus posing a health threat to the consumer. In this project, we examined the cause of the absence of O157 expression in the atypical *E. coli* O *rough*:H7 MA6 and CB7326 strains in order to determine the cause of the O *rough* phenotype. Both strains were also compared to determine whether these are analogous or related strains.

Biosynthesis and assembly of *E. coli* O157 antigen is highly complex. The *rfb* operon is comprised of 12 genes along with 3 ancillary genes. It is required for the biosynthesis of 4 sugar nucleotide precursors and the assembly of the O unit (Samuel

and Reeves, 2003; Wang and Reeves, 1998). The O unit is linked to the outer core antigen, which requires 3 other operons located in the *waa* gene cluster for biosynthesis and assembly (Raetz and Whitfield, 2002). Genetically defects in any of the O antigen biosynthesis and/or assembly genes, as well as in the outer core genes, could lead to the O antigen null phenotype. Therefore, the genes involved in O antigen and outer core antigen biosynthesis were systematically examined in MA6 to elucidate the cause of the absence of O157 expression. We found that the lack of O antigen synthesis in both MA6 and CB7326 was due to an inactivation by insertion sequence (IS) IS*629* of the *gne* gene which is essential for the O157 biosynthesis (Rump et al., 2010b; Rump et al., 2010a).

The following section of the project, due to the previous findings, focuses on the prevalence, distribution and evolutionary significance of the IS*629* insertion sequence (IS) in the emergence of EHEC O157:H7. IS are known to play an important role in the evolution and genomic diversification of *E. coli* O157:H7 lineages. Furthermore, IS insertion and IS-mediated deletions have been shown to generate phenotypic diversity among closely related O157 strains (Ooka et al., 2009a). Therefore, the prevalence of IS*629* in a panel of *E. coli* strains, including ancestral and atypical strains implicated in the stepwise emergence of *E. coli* O157:H7, was examined in order to determine its importance in the transitional steps that gave rise to current highly pathogenic *E. coli* O157:H7.

2 BACKGROUND

2.1 Foodborne illness

Foodborne illnesses are defined as diseases, either infectious or toxic in nature, caused by agents entering the body through ingestion of contaminated food or water. Foods and beverages which are produced, processed, and distributed to the consumer under good hygiene practices can generally be considered as “safe”. Improper handling, preparation or food storage however, can be the cause of contamination with viruses, bacteria, parasites, toxins, metals, and prions (Mead et al., 1999). Raw foods of animal origin (raw meat and poultry, raw eggs, unpasteurized milk, and raw shellfish) are most likely to be contaminated. Foods that are the product of many individual animals, such as pooled eggs, or ground beef are hazardous because a pathogen present in any one of the animals may contaminate the entire batch. Fruits and vegetable which are consumed raw are particularily insecure since washing may decrease but not eliminate contamination. The recent STEC O104:H4 outbreak was liked to raw vegetables (fenugreek seeds or sprouts) which are generally consumed raw or undercooked (BfR Pressemitteilung, 2011). Symptoms of an infection are mild gastroenteritis to life-threatening neurologic, hepatic, and renal syndromes. Important safety concerns are associated among population groups with greater susceptibility to foodborne infections. These include pregnant women (and their fetuses), young children, elderly persons, and persons with lowered immunity due to HIV/AIDS, or those on medications for cancer treatment or for organ transplantation (immune suppressors) (Centers for Disease Control and Prevention, 2010b).

Both in industrial and developing countries, foodborne diseases are a widespread and growing public health problem due to high mortality rates and consequential costs. In the USA alone, estimated costs from diseases caused by several major pathogens alone are close to $35 billion annually (1997) in medical treatment and loss of productivity (WHO (World Health Organization), 2007). The percentage of the population suffering from foodborne diseases each year is near to 30% in industrialized countries. In 2005, 1.8 million people died from diarrheal diseases worldwide. Around 76 million cases of foodborne diseases are estimated to occur every year in the United States of America (USA) resulting in 325,000 hospitalizations and 5,000 deaths. In Germany, over

1 million people get infected annually (Robert Koch Institut, 2000). Although it is less well-documented and difficult to estimate, developing countries bear the brunt of the problem albeit due to an even wider range of foodborne diseases including those caused by various parasites (WHO, 2010).

Ongoing changes in the food supply, the identification of new pathogens, as well as re-emerging pathogens with altered characteristics (e.g. multi-resistant *Salmonella Typhimurium* strains) comprise the main challenges for research and public health addressing foodborne diseases. Some pathogens are well recognized but are considered emerging because they have recently become more common. While technological advances, such as pasteurization and proper canning, have eliminated some diseases; new causes of foodborne illness have been identified. Some organisms were not recognized as causes of foodborne illness until as recently as 20 years ago. Moreover, globalization of the food industry changed traditional outbreak scenarios. Outbreaks are no longer only localized but can be rapidly widespread throughout a country (Robert Koch Institut, 2000). While most foodborne diseases remain sporadic and often not reported, foodborne disease outbreaks may take on massive proportions as seen for the *E. coli* O104:H4 outbreak in Germany, Spring 2011 (Bielaszewska et al., 2011).

The predominant foodborne pathogens are *Campylobacter jejuni*, *C. coli*, *Salmonella*, and shiga-toxigenic *Escherichia coli* (STEC) O157:H7 and related strains, parasites (*Toxoplasma gondii* and *Cryptosporidium parvum*) and caliciviruses (O'Brien et al., 1983). In the United States, foodborne disease outbreaks caused by produce contaminated with *E. coli* O157:H7 remain prominent. Additionally, *Salmonella* Enteritidis continues to be a major cause of illness and death especially for elderly persons and immunocompromised (Centers for Disease Control and Prevention, 2000).

2.2 Pathogenic *E. coli*

E. coli belongs to the family *Enterobacteriaceae*, which are facultatively anaerobic Gram-negative rods. Most *E. coli* are harmless commensal organisms that commonly inhabit the lower intestine of warm-blooded organisms. They are the predominant facultative organisms in the human gastrointestinal tract, underscoring their non-pathogenic and beneficial state in the gut. Some *E. coli*, however, are pathogenic to humans and are responsible for three types of infections: urinary tract infections (UTI), neonatal meningitis, and intestinal diseases (gastroenteritis). Pathogenic *E. coli* implicated in foodborne diarrheal diseases are furthermore classified on the basis of serological characteristics and virulence properties: enterotoxigenic *E. coli* (ETEC); enteroinvasive *E. coli* (EIEC); enterohemorrhagic *E. coli* (EHEC); enteropathogenic *E. coli* (EPEC); enteroaggregative *E. coli* (EAEC); diarrhea-associated hemolytic *E. coli* (DHEC) and cytolethal distending toxin (CDT)-producing *E. coli* (Clarke, 2001). Each class furthermore falls within a serological subgroup and manifests distinct features in pathogenesis. EPEC and ETEC are the most important in terms of total diarrheal episodes on a global scale, although in recent years EHEC has become more significant as the predominant cause of hemorrhagic colitis in humans (Feng, 1995).

2.2.1 Enterohemorrhagic *E. coli* (EHEC)

The EHEC clonal group is thought to have derived from EPEC and shares several virulence mechanisms (Whittam et al., 1993). However, it now is recognized as a distinct class of pathogenic *E. coli* due to two key epidemiologic observations. (1.) Riley et al. (1983) described a rarely isolated *E. coli* serotype, O157:H7, as a cause of two outbreaks in the United States. Both cases had distinctive gastrointestinal illness characterized by severe crampy abdominal pain, watery diarrhea, and little or no fever. These symptoms are common for hemorrhagic colitis (HC). (2.) Karmali et al. associated sporadic cases of hemolytic uremic syndrome (HUS) with fecal cytotoxin and cytotoxin-producing *E. coli* in stools. HUS is typically preceded by a bloody diarrheal illness indistinguishable from HC and furthermore can lead to acute renal failure, thrombocytopenia, and microangiopathic hemolytic

anemia (Nataro and Kaper, 1998; Karmali et al., 1983). These observations of enteric pathogenic *E. coli* causing diarrheal diseases distinguishable from EPEC spawned recognization of a new class, EHEC.

EHEC produces characteristic toxins which are closely related to Shiga toxins (Stx) of *Shigella dysenteriae* (O'Brien et al., 1983). The toxins have cytotoxic activity on Vero cells (also called Vero cytotoxin) whereby the organism is termed verocytotoxigenic *E. coli* (VTEC) or Shiga toxin-producing *E. coli* (STEC) (Levine et al., 1987). In addtition to Stx production, members of the EHEC group possess a number of other mechanisms involved in virulence. Most major enterohemmorrhagic *E. coli* contain a locus of enterocyte effacement (LEE) encoded by a pathogenicity island also found in EPEC strains. The LEE region encodes a 94- to 97-kDa outer-membrane protein called intimin. This important intestinal adherence factor encoded by the *eae* gene, mediates the attaching and effacing lesion on the intestinal brush border (Yu and Kaper, 1992). Intimin additionally binds to the intimin receptor (Tir) protein also encoded on the LEE region. Tir is translocated from the bacterium to the host cell via a LEE-encoded type III secretion system (Yoon and Hovde, 2008). Another characteristic trait for the Stx-producing EHEC group is a 60-MDa virulence plasmid which contains genes encoding a hemolysin (termed enterohemolysin) (Schmidt et al., 1995).

EHEC can be found in cattle and other ruminants and is transmitted to humans primarily through the consumption of contaminated foods. Cattle appear to be the main reservoir of EHEC and the pathogen has been isolated from healthy animals. Most cases are caused by consumption of raw or undercooked ground meat products or raw milk (Clarke, 2001). Fecal contamination of water and other foods, as well as cross-contamination during food preparation that also leads to infections. EHEC also has been isolated from bodies of water (ponds, streams) wells and water troughs and has been found to survive for months in manure and water-trough sediments (Nataro and Kaper, 1998). Additionally, visits to petting zoos, dairy farms, and camp grounds where cattle have previously gazed, as well as recreational water sources have all been implicated in infections (Yoon and Hovde, 2008). The broad spectrum of vehicles implicated in disease due to EHEC may be explained by the low infectious dose which has been estimated to be less than 100 organisms, as

determined from outbreak investigations (Nataro and Kaper, 1998). The incubation period varies between 1 to 6 days but may be up to 14 days. Most cases are self-resolving within a week, but disease sometimes processes to HC in one or two days. Patients develop watery diarrhea that typically is accompanied by abdominal pain. In 80% of cases, after a 2- to 4-day interval, bloody diarrhea occurs, which is the origin of the term hemorrhagic colitis (Table 2-1). Young children and elderly are especially at risk for the disease progressing to HUS, reported in 10-15% of HC patients (Yoon and Hovde, 2008). Furthermore, especially in children, HUS is responsible for up to 4.5 % of chronic renal replacement therapy and a common cause of acute renal failure and can lead to significant morbidity and mortality during the acute phase. In addition long-term renal and extrarenal complications can even occur years after the acute episode of HUS in a substantial number of children (Scheiring et al., 2008).

Table 2-1. ***Distribution of EHEC serotypes isolated from stool specimens of patients with HUS in Germany, 1996-2003 (Karch et al., 2009)***

Serotype	% of the isolates
O157:H7	47
O157:NM	17
O26:H11/NM	15
O145:H25/H28/NM	9
O11:H8/NM	4
O103:H2	3
Others[1]	5
Total	100

NM: non-motile. [1]Serotypes present in more than one isolate: O4:NM, O73:H18, O91,H21, O113:H21,O118:NM.

EHEC serotypes most frequently clinically associated with HC include O157:H7 and O157:H-, O26:H11, O103:H2, O111:H8, and O145:H28 (Yoon and Hovde, 2008). *E. coli* O157:H7 is the prototypic EHEC most often implicated in illness with bloody diarrhea and HUS worldwide (Tarr et al., 2005). and has been commonly identified as a cause of HUS in North America, Japan, and much of Europe (Karch et al., 2009).

In Germany *E. coli* O157:H7 has been found to be associated with HUS in almost half of all HUS cases (Table 2-1). In North America however, *E. coli* O157:H7 was isolated from >95% of all HUS cases (Tarr et al., 1990). In the United States, *E. coli* O157:H7 infects millions of people every year, accounting for 0.5%, 3%, and 2.9% of illnesses, hospitalizations, and deaths, respectively, of the total U.S. food-borne diseases caused by known food-borne pathogens (Mead et al., 1999; Feng et al., 2007).

Although the incidences seem to be relatively low, the severe and sometimes fatal health consequences, particularly among infants, children and the elderly, with the ability to cause large outbreaks, makes *E. coli* O157:H7 among the most serious foodborne infections.

2.2.2 *E. coli* O157:H7

E. coli O157:H7 is the prototypical EHEC strain. It expresses somatic (O) antigen O157 and flagellar (H) antigen H7. It is the serotype predominantly isolated from humans globally, and the serotype most often associated with HUS. Clinical data on HUS caused by non-O157 strains are rare. The first time it was recognized as a pathogen for humans as a cause of bloody diarrhea was in 1982, when two outbreaks in Michigan and Oregon were traced to ground beef contaminated with this organism (Riley et al., 1983). This time, the serotype was considered as rare, since the only known previous isolation in the United States of this serotype was from a sporadic case of HC in 1975 from a patient in California (Wells et al., 1983).

In recent years, *E. coli* O157:H7 has caused hundreds of outbreaks worldwide emerging rapidly as a major food pathogen. To date, the largest *E. coli* O157:H7 outbreak known has been in Sakai City, Japan, in 1996 (Watanabe et al., 1996). Thousands, mostly school children, were affected -probably infected by contaminated white radish sprouts served during school lunches. The largest outbreak in the USA caused by the consumption of poorly cooked hamburgers affected 732 people of which 55 developed HUS (mostly children), and 4 died in the States of Washington, Idaho, Nevada, and California (Bell et al., 1994; Rump et al., 2010a). In Scotland in 1996, one of the largest outbreaks in Europe occurred with 501 people affected, and 20 dying from the consequences. Surprisingly, the source of *E. coli* O157:H7 was identified as meat from a single butcher's shop (Dundas et al., 2001). In Germany O157 has been isolated from both humans and animals, however only a few outbreaks have been reported (Beutin, 1999). One of the lagest HUS outbreaks in Germany occurred due to the consumption of sausages in 1994-1995 with an estimated 300-400 affected persons, whereby 28 HUS cases (children), and no deaths (Ammon et al., 1999). An 1992 outbreak in a healthcare center

involved 39 children and 2 adults (3 HUS, 1 death) (Reida et al., 1994). Since 2006 8 multi-state outbreaks have been reported in the United States whereby most of them (5 of 8) have been associated with the consumption of beef products (Centers for Disease Control and Prevention, 2010a). However, other less likely vehicles of infection like cookie dough (2009), bagged spinach (2006), or fenugreek seeds and sprouts which were carrier of the STEC *E. coli* O104:H4 outbreak strain in Germany (BfR Pressemitteilung, 2011), have been the cause of recent outbreaks. The predominant vehicles of *E. coli* O157:H7 have been bovine products, however, other vehicles like raw milk, vegetables and even water have been implicated with foodborne illnesses caused by this pathogen. *E. coli* O157:H7 is an important example showing that unknown strains can emerge to major pathogens threatening the food supply. Also thought to be rare was serotype O104:H4, which has been implicated in one of the most deadly outbreaks worldwide concentrated in Germany with >4,000 illnesses and an unusually high death rate (50 deaths) (BfR Pressemitteilung, 2011). In order to diminish outbreak situations, it is crucial to rapidly identify *E. coli* O157:H7 and other STECs in foods and patients, since a prompt and accurate EHEC diagnostic can be substantial and live-saving. It is important to detect current and emerging pathogens accurately to prevent life-threatening outbreaks worldwide.

2.2.3 The Evolutionary Model of *E. coli* O157:H7

In Germany, *E. coli* O157:H7 is the most frequent EHEC serotype implicated in HUS, but is however, not the only relevant EHEC O157. Sorbitol-fermenting (SF) *E. coli* O157:H- (non-motile) strains cause ~20% of all cases of HUS (Karch et al., 2005). Unlike strains of serotype O157:H7, organisms of this clonal group can ferment sorbitol and until recently could, not be isolated from cattle and other domestic or wild animals, which is knowingly a key reservoir for *E. coli* O157:H7 (Karch et al., 2005; Bielaszewska et al., 2000). Several typing methods (Multilocus enzyme electrophoresis, multilocus sequence typing (MLST), analysis of bacteriophage insertion sites, and identification of additional inserted loci), indicated that SF O157 are closely related to EHEC O157:H7, yet clearly distinct from this

clonal group (Feng et al., 1998a). Point mutations at -10 and +92 in *uid*A confirm the close relatedness of the non-sorbitol-fermenting (NSF) O157 and SF O157 strains. Some of the characteristics and differences between the two pathogens are summarized in Table 2-2. In addition to their ability to ferment sorbitol, SF O157 produce β-*D*-glucuronidase, uniformly possess stx_2, lack the plasmid-encoded resistance to tellurium (*ter* genes) and are non-motile (Karch et al., 2005). However, both groups possess the virulence genes *eae*, *ehx*A, the rfb_{O157} operon encoding the O157 antigen biosynthesis, and a complete gene cluster encoding flagella. The loss of motility in SFO157 was found to be caused by a 12-bp in-frame deletion in *flh*C that is required for transcriptional activation of the flagellum biosynthesis (Monday et al., 2004).

E. coli O157:H7 is only distantly related to other stx-producing EHEC. The genetic distance between O157:H7 and other major O157 lineages exceed a level at which these strains cannot be assigned to a specific evolutionary branch within the O157 serogroup. Whittam and Wilson (Whittam and Wilson, 1988) suggested that O157:H7 did not recently derive from ancestral O157 but rather from another O serogroup. A clonal analysis based on multilocus enzyme electrophoresis revealed that O157:H7 and the SF non-motile O157:H- belong to a genetically distinct clone complex including EPEC clones of serotype O55:H7 (Feng et al., 1998a).

Table 2-2. Comparison of phenotypic and virulence characteristics of EHEC O157:H7 and EHEC O157:NM. Modified from (Karch et al., 2005)

Serotype	Phenotypic characteristics				Virulence factors				
	SF	GUD	Hly	Motility/ *flh*C	Stx	*eae*	*ehx A*	*ter* genes	plasmid
O157:H7	-	-	+	+/+	stx_1, stx_2, stx_{2c}, stx_{1+2}, stx_{2+2c}, stx_{1+2c}, stx_{2+2+2c}	γ	+	+	pO157 (~90 kb)
O157:NM	+	+	-	-/+ (12-bp-del.)	stx_2	γ	+	-	pSFO157 (~120 kb)

SF: sorbitol fermentation, GUD: production of β-D-glucuronidase, Hly: EHEC hemolysin production; -(+): only a portion of SF EHEC O157:NM display an enterohaemolytic phenotype

An evolutionary model postulates that *E. coli* O157:H7 might have evolved through a series of steps from closely related ancestral *E. coli* O55:H7 by acquisition or loss of virulence and phenotypic traits (Figure 2-1) (Feng et al., 1998a). The model begins with an ancestral O55:H7 (A1) which is able to express β-*D*-glucuronidase (GUD^+) and to ferment sorbitol (SOR^+). This ancestral cell (presented by TB182A and DEC5A) expresses both the somatic O55 and H7 antigen. In the next step, A1 → A2, gained the *stx*$_2$, probably by a bacteriophage, leading to the Stx_2-producing O55:H7 strain (presented by 3256-97 and USDA 5905). The lack of strains in clonal complexes (CC) A1 and A2 however, makes it harder to characterize and define those clonal groups. This is followed by transition from the O55 to the O157 antigen which might have occurred due to horizontal gene transfer of this region (Bilge SS et al., 1996).

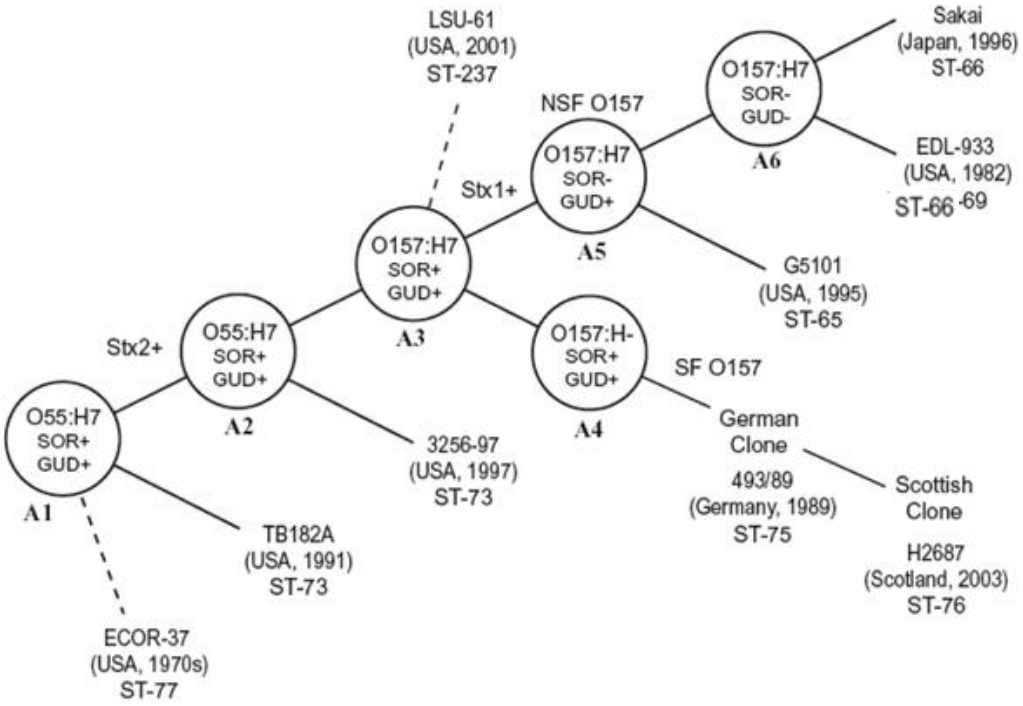


Figure 2-1. ***Evolution model for E. coli O157:H7.*** *Some strains, whose position on the model remain to be determined, are shown with dashed lines, from (Feng et al., 1998a).*

This ancestral O157 strain however, has not been isolated and remains a hypothetical "intermediate" strain (A3). This strain is thought to be able to ferment sorbitol (SOR^+) and produce β-*D*-glucuronidase (GUD^+) and possesses stx_2. At this point, two distinct clones evolved: A4 and A5. The clonal complex A4 is non-motile, carries stx_2, has the ability to ferment sorbitol and produces β-*D*-glucuronidase. This lineage is known as the German clone. The clonal complex A5 of SOR^-, GUD^+, stx_2 which acquired stx_1 (presented by G5101) furthermore gave rise to the immediate ancestor (A6) of the common O157:H7 clone belonging to sequence type (ST) 66 with multilocus sequence typing (MLST) analysis using 7 housekeeping genes (http://www.shigatox.net/ecmlst/cgi-bin/index). Strains from clonal complex A6 are readily available, whereas strains from other ancestral clonal complexes are rare (Feng et al., 1998a).

2.2.4 *E. coli* O *rough*:H7

E. coli O157:H7 isolates can be phenotypically separated from other *E. coli* using common analytical methods such as delayed *D*-sorbitol fermentation and the lack of β-*D*-glucuronidase activity. Therefore, identification of O157 strains relies on initial suspicion on selective media and confirmation of the presence of the O157 and H7 antigens. The emergence of phenotypic variations within *E. coli* O157:H7, which is known to be a highly clonal and phenotypically tight serotype, however is a recently arising concern (Feng, 1995). An increasing number of phenotypic variations in O157 isolates have been reported in Europe, which could potentially lead to

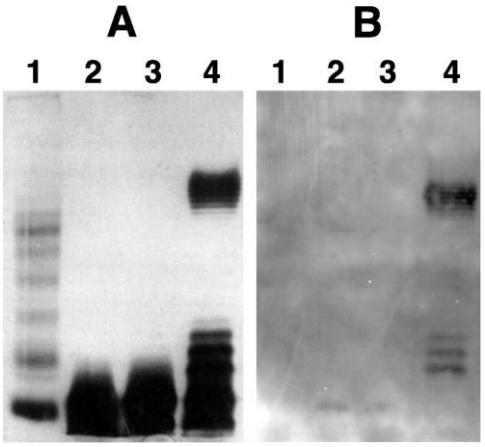


Figure 2-2. ***Sodium dodecyl sulfate-polyacrylamide gel (A) and Western blot (B) of bacterial LPS.***
Lanes: 1: prestained Kaleidoscope molecular weight standards (Bio-Rad, Hercules, CA), 2: DH5α λpir, 3: MA6, 4: TT12. Blot was probed with anti-O157 serum from the Escherichia coli Reference Center, Pennsylvania State University. From (Feng et al., 1998b)

misidentification of O157:H7.

Previous research found *E. coli* MA6, a shigatoxigenic O157:H7 strain isolated from beef in Malaysia, which is genotypically an O157:H7 strain and carries the typical virulence factors, associated with this organism (Table 2-3). However, it does not react positively when tested for the presence of the O157 antigen with most available kits. Consequently, MA6 is undetectable and unidentifiable in routine serological analysis for *E. coli* O157:H7. Furthermore, the strain tested positive for the *rfb*E gene by PCR, which is essential for O157 antigen expression (Bilge SS et al., 1996; Feng et al., 1998b). Moreover, the O antigen side chain could not be detected by LPS sodium dodecyl sulfate-polyacrylamide gel electrophoresis or Western blot analyses (Figure 2-2), and it was determined that the MA6 strain is a *rough* (O-side strain deficient) O157:H7 strain that does not synthesize O157 antigen (Feng et al., 1998b).

Table 2-3. ***Comparison of traits and markers among MA6, CB7326, and O157:H7 strains***

Trait or marker[a]	Result for strain: MA6	CB7326	O157:H7[b]
SOR	-	-	-
GUD	-	-	-
O157 antigen	-	-	+
wzx (O157)	+	+	+
H7 antigen	+	+	+
fliC (H7)	+	+	+
stx1	-	+	+
stx2	+	+	+
γ *-eae*	+	+	+
uidA	+	+	+
ehxA	+	+	+
MLST	ST-66	ST-66	ST-66

[a] SOR: sorbitol fermentation, GUD: ß-glucuronidase activity, O157 and H7: O157 and H7 antigens by latex agglutination, *wzx* (O157): *wzx* for O157, *fliC* (H7): *fliC* for H7, stx_1 and stx_2 : shigatoxin 1 and 2, γ-*eae:* γ intimin allele, *uidA:* +93 *uidA* SNP, *ehxA:* enterohemolysin, MLST: multilocus sequence typing, *gne*::IS*629*: insertion location. [b] EDL931

A study in Germany characterized patient isolates of STEC over a 3-year period and found a few strains with the O *rough*:H7 phenotype (Beutin et al., 2004). One of these, CB7326 (Table 2-3), was isolated from a hemorrhagic colitis patient and found to carry the characteristic traits common for O157:H7 strains, suggesting that, like MA6, CB7326 may be an O *rough* variant of O157:H7 (Rump et al., 2010a; Rump et al., 2010b).

2.3 The O Antigen

The O antigen is part of the cell envelope of gram-negative bacteria that consists of an inner (IM) and outer (OM) membrane, separated by periplasmic space (Raetz and Whitfield, 2002). The OM is an asymmetric bilayer which is composed by phospholipids and lipopolysaccharides (LPS) which face into the external environment. LPS are immunogenic glycolipids and essential components of bacterial endotoxins. This molecule is unique to Gram-negative bacteria and composed of three covalently linked domains: Lipid A, the core region and the O antigen (Schnaitman and Klena, 1993) (Figure 2-3) Lipid A functions as the hydrophobic anchor for LPS and is the bioactive component responsible for some pathophysiology associated with severe Gram-negative infections. The core region, a phosphorylated non-repeating oligosaccharide, is a barrier to antibiotics and can further be divided in the inner (IC) and outer core (OC) (Nikaido and Vaara, 1985). The OC provides the attachment site for O antigen (Kaniuk et al., 2002; , 2006), which is an immunogenic repeating oligosaccharide that varies greatly from strain to strain. Clinical isolates of enteric gram-negative bacteria usually possess an intact O antigen and are termed "*smooth*", whereby strains lacking the O antigen, including all laboratory strains of *E. coli* K-12, are described as "*rough*" (Stevenson et al., 1994). The LPS is essential for most gram-negative bacteria however in *E. coli*, the minimal part required for growth is Lipid A and the inner core. Strains with truncated cores

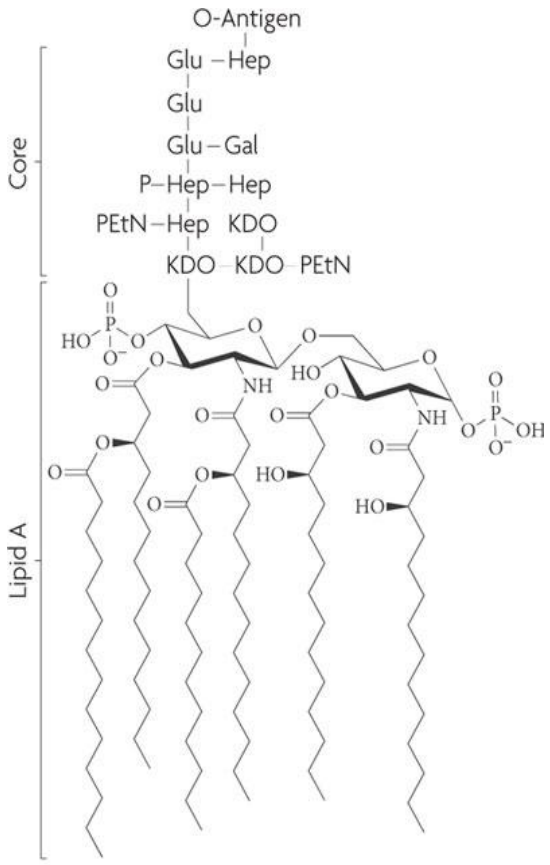


Figure 2-3. ***The structure of E. coli LPS (lipopolysaccharide) with a K-12 core region is shown. (Ruiz et al., 2009)***

("*deep-rough*") are able to grow but have been shown to be hypersensitive to antibiotics and detergents.

O157 antigen biosynthesis is mediated by the *rfb* operon (~15,000 bp) which is comprised of 12 genes and a remnant H-repeat unit (Figure 2-4) (Wang and Reeves, 1998). The genes fall into three general classes: genes for synthesis of nucleotide sugar precursors, genes for transfer of sugars to build the O unit and genes which carry out specific assembly or processing steps in conversion of the O unit to the O antigen as part of the complete lipopolysaccharide (Wang and Reeves, 1998;

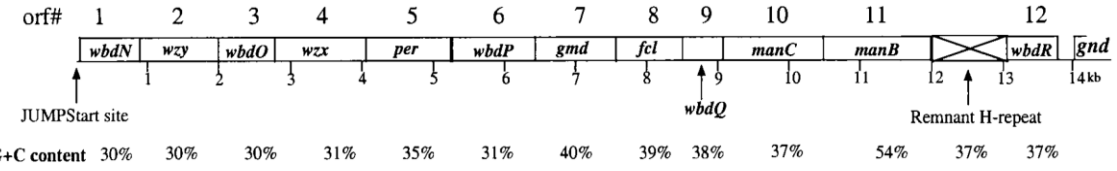


Figure 2-4. ***O antigen gene cluster of E. coli O157.***
Gene names and ORF numbers are given (Wang and Reeves, 1998).

Whitfield, 1995). They encode in particular: WbdN, WbdO, and WdbP, glycosyltransferases, Wzy, an O antigen polymerase, Per, a perosamine synthetase, Gmd, GDP-D-mannose dehydratase, FcI, fructose synthetase, WdbQ, GDP-mannose mannosylhydrolase, ManC, mannose-1-P guanosyltransferase, ManB, phosphomannomutase and WbdR, acetyl transferase (gene sizes and function see Table 0-1). In *E. coli* O157, the O-unit is composed of four different monosaccharides (Figure 2-5), which are synthesized from the respective sugar nucleotide precursors (SNPs), GDP-*D*-N-acetylperosamine (GDP-*D*-PerNAc), GDP-*L*-Fructose (GDP-Fuc), UDP-Glucose (UDP-Glu) and UDP-*N*-acetylgalactosamin (UDP-Gal*N*Ac) (Perry et al. 1986).

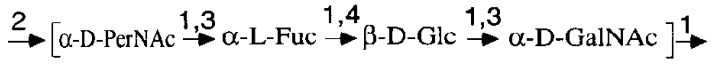


Figure 2-5. ***Structure of the O157 O antigen (Perry et al., 1986).***

The two precursors GPD-*L*-Fuc and GPD-*D*-PerNAc are synthesized in the same pathway (Figure 2-6) (Ginsburg, 1961; Wang and Reeves, 1998). Hereby, frucose-6-phosphate which is synthesized in the glycolysis, is isomerized to *D*-mannose-6-phosphate by phosphomannose isomerase (ManA) encoded by *man*A (Table 0-3). The *man*A gene is located outside the O antigen gene cluster. The following pathways for the synthesis of the final SNPs require genes located exclusively within the *rfb* operon. UDP-Glu is synthesized in the glycogenesis and plays an essential role in the assembly of the Lipid A and core polysaccharides, therefore is essential for cell viability (Galloway and Raetz, 1990). The fourth SNP UDP-Gal*N*Ac is epimerized from UDP-*N*-acetylglucosamine (UDP-Glc*N*Ac) by the *gne* gene enzyme product (Table 0-3), which is located upstream of the *rfb* operon. UDP-Glc*N*Ac is part of Lipid A and, consequently, also essential for the viability of the bacteria.

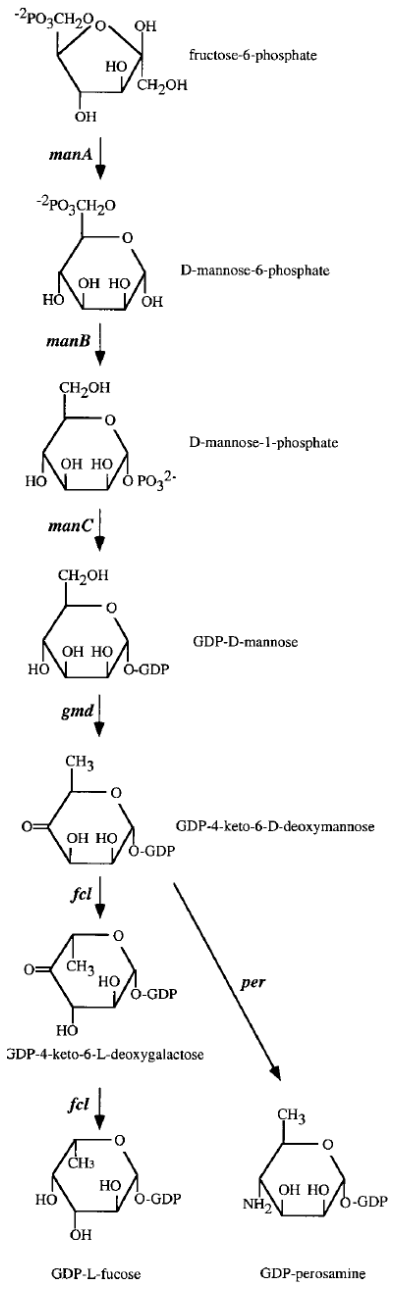


Figure 2-6. ***Biosynthetic pathway of GDP-L-fucose and GDP-perosamin with gene names (Wang and Reeves, 1998).***

Precursor polymerization takes place at the inner face of the cytoplasmic membrane. The process is initiated by the transfer of the first sugar unit to the membrane-bound carrier, undecaprenyl phosphate (*und*-P) catalyzed by WecA, encoded by the *wec*A gene located in the enterobacterial common antigen (ECA) operon (Table 0-3) (Raetz and Whitfield, 2002; Meier-Dieter et al., 1992). Subsequently, the linkages of the other three precursors are catalyzed by the enzymes expressed by the *wbd*O, *wbd*P, and *wbd*N genes, located in the *rfb* gene cluster. Following their assembly, the O antigen subunits are now assembled in a so called

Wzy-dependent pathway to the final O antigen chain. The *und*-PP-linked O units are exported to the site of polymerization at the periplasmic face of the plasma membrane by Wzx. In the periplasm the chain is transferred from its *und*-PP carrier to the "new" *und*-P linked subunit by Wzy. Both genes encoding the latter enzymes are located in the *rfb* operon. These steps will be repeated, until terminated by the Wzz chain length regulator-protein (Raetz and Whitfield, 2002).

The synthesis of the outer core region that serves as an attachment site for O antigen is mediated by the genes in the *waa* gene cluster (Table 0-2). The genes in this cluster are primarily involved in the synthesis and modification of the LPS core, but also in the attachment of the complete O antigen to the core. The *waa* locus consists of three operons (*gmh*D, *waaQ* and *waa*A) mapping between *cys*E and *pyr*E on the chromosome (Figure 2-7). The genes *gmbh*D-*waa*FC are required for biosynthesis of the IC which is essential for the outer-membrane stability. The longer *waa*Q operon *(waa*Q, *waa*G, *waa*P, *waa*I, *waa*Y, *waa*J, *waa*D, and *waa*L) contains the genes necessary for biosynthesis and modification of the OC including the "ligase" structural gene *waa*L whose product is required to link O polysaccharide to the complete core. Mutations in many glycosyl transferases encoded by this locus result in the production of LPS lacking O antigen since the O antigen polysaccharide can not be ligated to an incomplete lipid A-core acceptor molecule (Yethon et al., 1998).

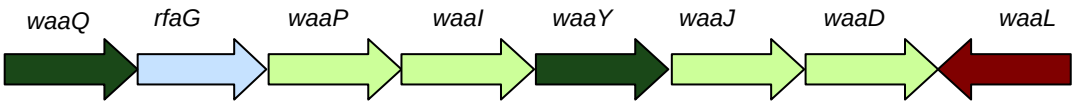


Figure 2-7. ***Organization of the waa locus****.*
Dark green: structure modifying genes, blue: outer core glycosyltransferase genes, red: ligase gene, light green: LPS biosynthesis genes (Raetz and Whitfield, 2002).

2.4 Insertion sequence IS*629*

Insertion sequences (IS) are widely distributed, self-mobilizing genetic elements that can introduce genetic changes (Mahillon and Chandler, 1998). Transposable insertion sequences are examples of selfish DNA whose short-term population dynamics are determined mainly by transposition and horizontal transmission among strains balanced against the regulation of transposition as a function of copy number, and negative effects on fitness (Hartl et al., 1992). Their integration into the chromosome mediates genetic rearrangement and facilitates horizontal gene transfer. Hence, they clearly play an important role in the evolution of prokaryotes (Mahillon, 1998; Szabo et al., 2008). The acquisition or loss of mobile genetic elements, like IS elements, may diverge those strains of particular bacterial species from each other (Arbeit, 1995). Twice as many insertions/deletions (indels) as single nucleotide polymorphisms (SNPs) influence the process of diversification in O157:H7 (Kudva et al., 2004). IS elements are distributed throughout the chromosome, inserted into numerous sites. Target site utilization is generally not random. Consistent with this is that it has been observed that several different elements were often clustered in “islands” (Bukhari AI et al., 1977) and, therefore, display some degree of target site specificity (Craig NL, 1997). The patterns of target site selection are characteristic of each element, such as a preferred target site and/or region or an avoided target site and/or region (Craig NL, 1997).

Over 1,600 bacterial ISs have been described to date, based on their similarities of their transposase, their genetic organization and features of their ends (Siguier et al., 2006). IS are very compact and usually only encode a transposase which is involved in their motility. The transposase is encoded by one or two open reading frames covering nearly the entire length of the element. Most IS exhibit short terminal inverted-repeat sequences (IR) (10 - 40 bp) which are involved in the Tpase binding, as well as the cleavage and strand transfer reactions leading to the transposition of the element (Mahillon and Chandler, 1998). On insertion, most ISs generate a short directly repeated sequence (DRs) of the target DNA flanking the IS. The length (2 - 14 bp) of the DR is characteristic for a given element. Transposition generally occurs at a very low level and is regulated through different mechanisms which depends on the type of

the IS element. IS are classified into three groups by the apparent strength of regulation: IS*1* and IS*5* (weakly regulated); IS*2*, IS*4*, and IS*30* (moderately regulated); and IS*3* (strongly regulated). Members of the IS*3* family generally have two overlapping open reading frames, *orf*A and *orf*B, and the transposase is a fusion protein, *orf*AB, generated by highly regulated and host-dependent translational frameshifting (Mahillon and Chandler, 1998).

The IS*629* is a member of the IS*3* family and was initially isolated from the chromosome of *Shigella sonnei* (Matsutani et al., 1987; Sambrook J and Russell RG, 2001). It has been detected in many other enteric bacteria in numerous copies, including S. *dysenteriae*, S. *flexneri*, S. *boydii*, *E. coli*, including O157:H7, *Enterobacter cloacae* MD36 and *Serratia marcescens* (Matsutani and Ohtsubo, 1993). Chen and Hu (2006) stated that the majority of IS*629* elements in nature might be transposable or able to produce the IS*629* transposase. IS*1203* and IS*4311* are two isoforms of IS*629* and are closely related with few differences at the nucleotide sequence level. IS*629* is 1,310 bp in length and has a pair of imperfect IRs at its termini. It carries two partially overlapping open reading frames, *orf*A and *orf*B, and two putative -1 translational frameshift signals, TTTTG and AAAAT, located near the 3'-end of the *orf*A. This suggests two frameshift products, *orf*AB', encoding the IS*629* transposase, and another transframe product, *orf*AB (Figure 2-8). The insertion site specificity of IS*629* remains unknown to date (Mahillon and Chandler, 1998; Craig NL, 1997).

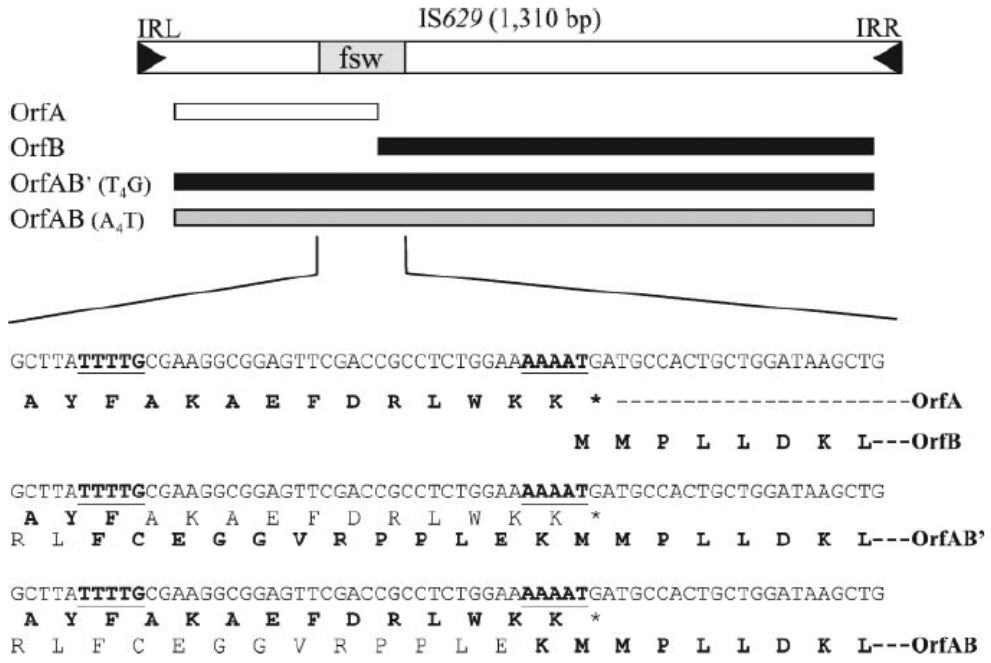


Figure 2-8. ***Organization of IS629***
Left (IRL) and right (IRR) terminal inverted repeats are indicated by triangles. The putative frameshift area (fsw) is shown as a gray box. The two partially overlapping open reading frames (OrfA and OrfB), as well as two putative frame-shift products (OrfAB' and OrfAB) including the nucleotide sequence of the fsw and the corresponding amino acid sequences are shown. The two putative frameshift signals (TTTTTG and AAAAT) are indicated in boldface and underlined (Chen CC and Hu ST, 2006).

3 MATERIALS

3.1 Bacteria

3.1.1 *E. coli* O157:H7 and related strains

The *E. coli* O157:H7, *E. coli* O157:H-, and *E. coli* O55:H7 strains used in this study are listed in Table 3-1 according to the clonal complex they belong to.

Table 3-1. E. coli strains used in this study

No.	Name	Other name	Serotype	stx	Special characteristics			ST	CC[a]	Source	Year	Reference
					GUD	SOR	plasmid					
1	BAA 460	Sakai	O157:H7	1, 2	-	-	pO157	66		Japan	1996	NC_002695
2	EDL 933	700927	O157:H7	1	-	-	pO157	66		USA	1982	AE005174
3	EC 4115		O157:H7	1, 2	-	-	pO157	66		USA	2006	NC_011353
4	TW 14359		O157:H7	1, 2	-	-	pO157	66		USA	2006	CP001368
5	EDL 931	35150	O157:H7	1, 2	-	-	pO157	66	A6			
6	MA6		O *rough*:H7	2	-	-	pO157	66		Malaysia	1998	(Feng et al., 1998b)
7	550654		O157:H7	2	-	-	pO157	66		USA	2009	(Feng et al., 2007)
8	FDA 413		O157:H7	2	-	-	pO157	66				(Feng et al., 2007)
9	G5101		O157:H7	1, 2	+	-	pO157	65		USA	1995	(Feng et al., 2007)
10	1628/98		O157:H7	1, 2	+	-	pO157	65		Norway	1998	
11	1659		O157:H7	1, 2	+	-	pO157	65				
12	EC 97144	TW 10707	O157:H7	1, 2	+	+	pO157	65	A5	Japan	1997	(Nagano et al., 2004)
13	EC 96038	TW 10201	O157:H7	1, 2	+	+	pO157	65		Japan	1996	(Feng et al., 2007)
14	EC 96012	TW 10189	O157:H7	1, 2	+	+	pO157	65		Japan	1996	(Nagano et al., 2002)
15	493/89		O157:H-	2	+	+	pSFO157	75		Germany	1989	(Feng et al., 2007)
16	5412/89		O157:H-	2	+	+	pSFO157	75		Germany	1989	(Feng et al., 2007)
17	IH56929	TW 09159	O157:H-	2	+	+	pSFO157	76		Finland	1999	(Feng et al., 2007)
18	IH56909	TW 09162	O157:H-	2	+	+	pSFO157	76	A4	Finland	1999	(Eklund et al., 2006)
19	H 1085c		O157:H-	2	+	+	pSFO157	76		Scotland	2003	(Feng et al., 2007)
20	H 2687		O157:H-	2	+	+	pSFO157	76		Scotland	2003	(Feng et al., 2007)
24	3256-97	TW 07815	O55:H7	2	+	+	-	73		USA	1997	(Feng et al., 2007)
22	USDA 5905		O55:H7	2	+	+	-	73	A2	USA	1994	(Feng et al., 2007)
23	TB 182A	TW 04062	O55:H7	-	+	+	-	73		USA	1991	(Feng et al., 2007)
21	DEC5A		O55:H7	-	+	+	-	73	A1			(Monday et al., 2004)
25	LSU-61		O157:H7	-	+	+	pO157	237	?	USA	2001	(Feng et al., 2007)
26	ECOR-37		Ont:Hnt				-	-	-	USA	1970s	(Feng et al., 2007)
27	Sakai	Sakai PF	O157:H7	1, 2	-	-	pO157	66	A6	Japan	1996	NC_002695
28	EDL 933	438-95	O157:H7	1	-	-	pO157	66	A6	USA	1982	AE005174

Stx: shigatoxin gene, GUD: β-glucuronidase activity, SOR: sorbitol fermentation, ST: sequence type as determined by MLST ([a]http://www.shigatox.net/ecmlst/cgi-bin/index) CC: clonal complex (Feng et al., 2007).

3.1.2 *E. coli* O *rough*:H7

Two atypical *E. coli* O *rough*:H7 strains which do not express O157 antigen but carry all the characteristic traits of the pathogenic O157:H7 serotype and belong to the most common O157:H7 clonal type (ST-66), were analyzed in this study. *E. coli* O157:H7 MA6, which produces Stx2, was initially isolated from beef in Malaysia (Radu et al., 1998) and obtained by the Food and Drug Administration (FDA) for this study. A study in Germany characterized patient isolates of STEC over a 3 year period and found a strain with the O *rough*:H7 phenotype (Beutin et al., 2004). Strain CB7326 was isolated from a hemorrhagic colitis patient and found to carry Stx1 and 2, as well as γ-intimin (*eae*) genes that are common in O157:H7, suggesting that it may also be an O *rough* variant of O157:H7.

Table 3-2. ***O rough:H7 E. coli strains used in this study***

No	Name	Serotype	*stx*	Special characteristics			ST	CC	Source	Year
				GUD	SOR	plasmid				
1	MA6	O *rough*:H7	2	-	-	pO157	66	A6	Malaysia	1998
2	CB7326	O *rough*:H7	1,2	-	-	pO157	66		Germany	2004

stx: shigatoxin gene, GUD: β-glucuronidase activity, SOR: sorbitol fermentation, ST: sequence type as determined by MLST (http://www.shigatox.net/ecmlst/cgi-bin/index), CC: Clonal complex (Feng et al., 2007),

3.1.3 Genomes and plasmids used for “in silico” analysis

Four E. coli O157:H7 and pO157 plasmids and O157:H- plasmid pSFO157 deposited at the National Center for Biotechnology Information (NCBI) database were used for “in silico” analysis.

Table 3-3. ***Genomes and plasmids investigated by “in silico” analysis.***

Strain	Serotype	Accession No.	
		Genomes	Plasmids
Sakai	O157:H7	NC_002695	AB011549 (pO157)
EDL933	O157:H7	AE005174	AF074613 (pO157)
EC4115	O157:H7	NC_011353	CP001163 (pO157)
TW14359	O157:H7	CP001368	CP001369 (pO157)
3072/96	SFO157	NA	NC_009602.1 (pSFO157)
CB9615	O55:H7	NC_013941	CP001847.1 (pO55)

NA: not available

3.1.4 Vector pTrc99A

The high copy number expression *E. coli* vector pTrc99A is 4,176 bp in length (Amann and Brosius, 1985). It contains the *bla* gene, conferring resistance to Ampicillin. Figure 3-1 shows the restriction map with the enzymes cutting pTrc99A DNA once with the coordinates of the first nucleotide in each recognition sequence provided.

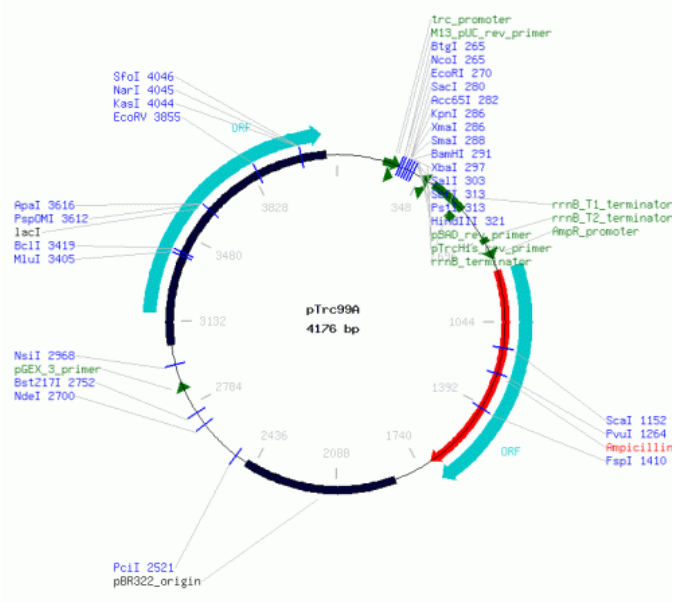


Figure 3-1. ***Expression vector pTrc99A (Accession No. M22744.1)*** *(http://www.lablife.org/p?a=vdb_view&id=g2.OXqiRm8r8oUFBLc87Ts.iJUB87w-), Pharmacia Biotech AB, Uppsala,. Sweden*

3.1.5 Vector pBR322

The *E. coli* multi-purpose cloning vector pBR322 is a commonly used cloning vector for the efficient cloning and selection of recombinant DNA molecules in *E. coli* (Balbas et al., 1986). This 4,361 bp DNA molecule contains the replicon *rep*; the *rop* gene, which can be used to decrease copy number; the *bla* gene, coding for beta-lactamase confering resistance to Ampicillin; and the *tet* gene, encoding tetracycline resistance protein. Figure 3-2 shows the restriction map of pBR322, including the coordinates.

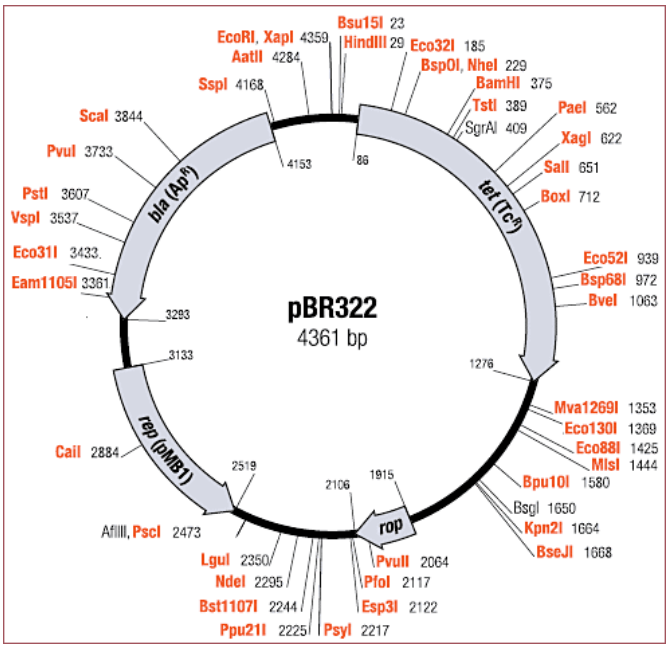


Figure 3-2. ***Expression vector pBR322 (Accession No. J01749)*** *(http://www.fermentas.com/en/products/all/molecular-cloning/vectors-phage/sd004-pbr322-dna).*

3.1.6 Vector pUC18

The high copy number expression *E. coli* vector pUC18 is 2,686 bp in length. It contains the replicon *rep*. The high copy number from the plasmid results of a single point mutation in *rep* and a lack of the *rop* gene (Balbas et al., 1986). It also contains the *bla* gene, coding for beta-lactamase, confering resistance to Ampicillin. Additionally, the presence of the *E. coli lac* operon with the P_{lac} promoter which allows α-complementation to form blue colonies on media with X-Gal and white colonies with insertion of DNA into MCS located within the *lacZ* gene. Figure 3-3 shows the restriction map with the enzymes cutting pUC18 DNA once. Coordinates refer to the position of the first nucleotide in each recognition sequence.

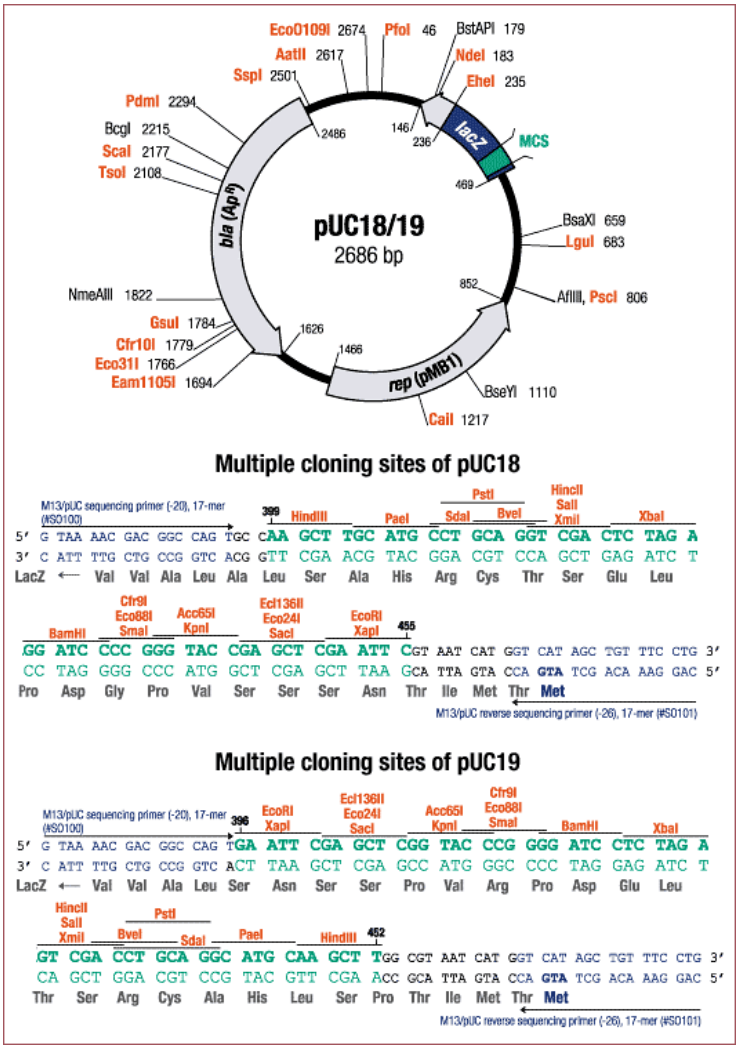


Figure 3-3. ***Expression vector pUC18 (Accession No. L09136).*** *(http://www.fermentas.com/en/products/all/molecular-cloning/sd005).*

4 METHODS

This paragraph portrays the theoretical background of the scientific methods used to achieve the research goal. Described are the principles of the individual methods as well as eventually performed modifications of those. Furthermore, the principles of each commercially used kit are elaborated.

4.1 Storage of Bacterial Cultures

To store bacterial cells over a longer period of time, 15% glycerol is a commonly used reagent to protect biological material. It protects the cells from freezing damage due to the fact that it reduces ice formation which could harm cells (Sambrook J and Russell RG, 2001).

4.2 Determining Bacterial Growth

Bacterial cells multiply rapidly by cell divison under favourable conditions. The growth of a population of bacteria occurs in several stages: Lag phase, Log (logarithmic or exponential) phase, Stationary phase and Decline (death) phase. After the adjustment to the new environment (Lag phase – no growth), the cells start replicating (Log phase), and cell reproduction exceeds cell death. This is followed by the Log phase, where the growth rate slowes down and the cell reproduction is equal to cell death (Stationary phase). The Decline phase occurs when nutrients become limited and cell deaths exceed the amount of new cells. In this experiment, bacteria were compared for their growth characteristics in the same media. The bacterial population in culture was estimated by turbidity (optical density, OD) measurement using a spectrophotometer. The OD is measured as the absorbance at 600 nm wavelength (Sambrook J and Russell RG, 2001).

4.3 *E. coli* O157:H7 Identification

To identify *E. coli* O157:H7, several selective isolation steps are conducted by using selective agars, such as cefixime tellurite sorbitol MacConkey agar (CT-SMAC) or

hemorrhagic coli agar. Additionally, detection and confirmation steps utilize enzyme-linked immunosorbent assay, agglutination, or DNA-methods (Cray et al., 2001).

One of the characteristic traits of *E. coli* O157:H7 is that strains of this serotype do not ferment *D*-sorbitol rapidly, unlike most human faecal *E. coli*. Therefore, the pathogen can be best detected by plating samples on sorbitol-Mac-Conkey agar (SMAC). Instead of lactose, on this selective media, sorbitol serves as a carbon source which cannot be fermented by *E. coli* O157:H7. Colonies of *E. coli* O157:H7 appear colorless, whereby other *E. coli* appear red (Figure 4-1) (March and Ratnam, 1986). A modification of the SMAC agar to improve the selectivity and differentiation is the inclusion of cefixime and tellurite (CT-SMAC agar) in the media. Both agents permit the growth of stx-producing *E. coli* O157:H7, but inhibit the growth of most other *E. coli* (Nataro and Kaper, 1998).

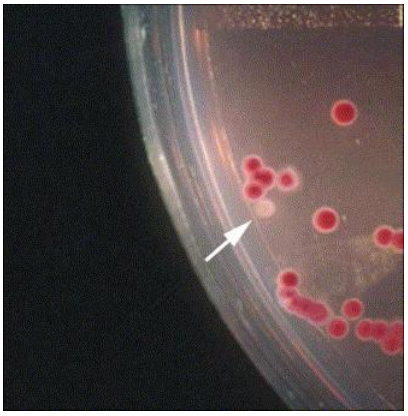


Figure 4-1. ***E. coli O157:H7 on a SMAC plate.***
Arrow indicates distinctive colorless E coli O157:H7 colony (, 2004).

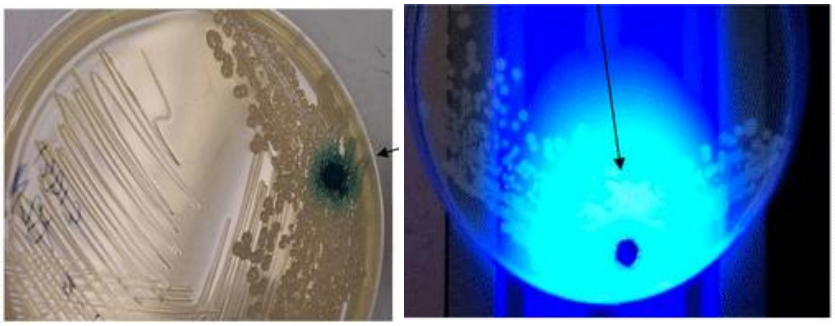


Figure 4-2. ***Result of ColiComplete disc for E. coli O157:H7.***
1, X-gal positive E. coli O157:H7 produce a blue color; 2, MUG positive E. coli O157:H7 show fluorescence under UV-light (365 nm). (BAM, www.fda.gov).

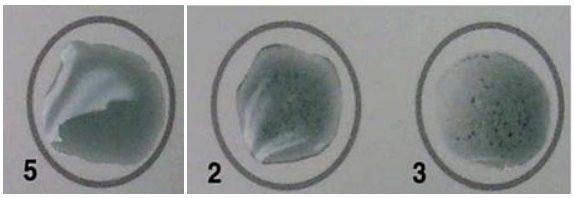


Figure 4-3. ***Typical E. coli O157:H7 latex agglutination result.***
1: Negative control, 2: O157 antigen, 3: H7 antigen (BAM, www.fda.gov).

Another characteristic trait for this pathogen is its inability to produce β-glucuronidase, which hydrolizes 4-methyl umbelliferyl-D-glycuronide (MUG) and related substrates (Thompson et al., 1990). *E. coli* O157:H7 are phenotypically selected on CT-SMAC agar with a ColiCompetent disc (Figure 4-2). Those presumptively identified *E. coli* O157:H7 furthermore can be screened on enterohemolysin agar, which detects enterohemolysin expressed by about 90% of Stx-producing *E. coli* isolates (Beutin et al., 1994). Suspected EHEC colonies additionally have to be screened for the presence of the O157 lipopolisaccharide (LPS) antigen. The most commonly used method to detect both the O and the H antigen or the Stx, are antiserum agglutination tests. The immunoassays for either the O or the H antigens are limited largely to detecting EHEC expressing O157 LPS and H7 flagellar antigen, whereas toxin immunoassays are suitable for the detection of both O157 as well as non-O157 serogroups of EHEC.

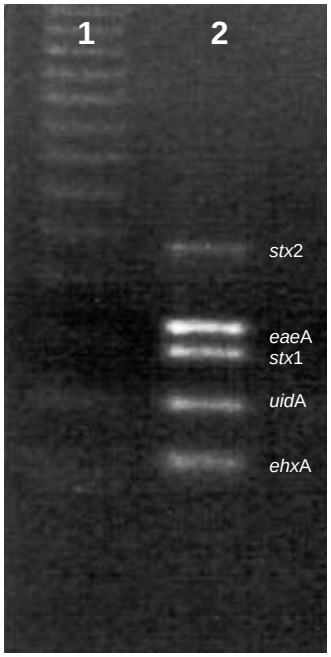


Figure 4-4. ***Agarose gel electrophoresis of DNA fragments amplified by 5P multiplex PCR.***
1: 123-bp ladder, 2: 35150 (O157:H7). (modified from (Feng and Monday, 2000)).

Besides culture techniques and immunoassays used to identify *E. coli*. O157:H7, the pathogen can be furthermore verified with PCR techniques. PCR has been extensively used to detect *stx* (*stx*1 and *stx*2) and other virulence genes (*eae*A (γ-intimin derivative gene), *ehx*A (pO157 plasmid/EHEC hemolysin gene), *uid*A (beta-*D*-glucuronidase gene), and *fli*C (flagellin gene) (Nataro and Kaper, 1998). To reliably detect those characteristic traits, a five gene product (5P) multiplex PCR simultaneously detecting five specific EHEC genes has been established (Figure 4-4) (Feng and Monday, 2000).

In the investigation of outbreaks, it is important to determine whether O157:H7 cases are linked to a common source or whether they represent sporadic and unrelated

cases. In order to discriminate closely related O157:H7 strains, different molecular epidemiological typing techniques have been established. The most important method of differentiation (the “gold-standard”) is pulsed-field gel electrophoresis (PFGE) which allows distinguishing among closely related clones due to individual restriction patterns (genetic “fingerprints”). To facilitate the early identification of common source outbreaks like *E. coli* O157:H7, the Center for Disease Control and Prevention (CDC) coordinates a national network of public health and food regulatory agency laboratories called “PulseNet” (Centers for Disease Control and Prevention, 2003). PulseNet stores PFGE patterns in an electronic database which allows for rapid comparison of new PFGE patterns of bacteria isolated from infected people to find out whether they are similar.

4.4 MasterPure Complete Nucleic Acid Purification

Nucleic acids were extracted from bacterial cells using the EPICENTRE® MasterPure™ Complete Purification kit. This kit uses a novel technology permitting the efficient purification of nucleic acids from every type of biological material. The kit utilizes a rapid desalting process used by MILLER (1988) to remove macromolecules with the advantage of avoidance of toxic organic solvents. In the first step, the usage of Proteinase K and Tissue and Cell Lysis Solution will lyse the cell membrane. In this procedure, lipid membrane layers will be disrupted (mainly by detergents) which exposes cell content, including the DNA. Salts stabilize the DNA to keep its helical formation. Proteinase K is able to digest proteins and therefore rapidly inactivates nucleases that might degrade DNA during the purification. Furthermore, Bovine Panceratic RNase A is added to remove RNA from the sample. RNase A works as an endonuclease which cleaves single-stranded RNA. The provided MPC Protein Precipitation Reagent precipitates the proteins but leaves DNA in solution. The addition of isopropanol in which DNA is insoluble to the recovered supernatant, precipitates the DNA which can then be resolubilized in a slightly alkaline buffer or ultra-pure water.

4.5 Nucleic Acid Quantification

The NanoDrop ND-1000 is a spectrophotometer which uses a sample retention system allowing to measure the nucleic acid content of a sample in volumes between

0.5 – 2 µL without the need of cuvettes or capillaries. Fiber optic technology and surface tension holds the sample in place between two optical surfaces that define the path length in a vertical orientation. The instrument uses a xenon flash lamp as source and a CCD (charge-coupled device) detector. DNA and RNA absorb ultraviolet light with an absorption maximum at 260 nm wavelength. To assess the DNA with respect to protein contamination, the absorption at 260 nm is compared to the absorption at 280 nm. The 260:280 ratio is a good indicator of protein contamination, whereby a 260/280 ratio of ~1.8 is considered “relatively pure” DNA, and ~2.0 is generally accepted as “pure” for RNA. Additionally, a secondary measure of nucleic acid purity is used to indicate the presence of contaminants absorbing at 230 nm. Expected 260/230 values are commonly in the range of 2.0-2.2.

4.6 Nucleic Acid Amplification (Polymerase Chain Reaction)

The Polymerase Chain Reaction (PCR) is used to amplify a specific region of a DNA by *in vitro* enzymatic replication. A strand of nucleic acids called primers, are used to target the DNA and therefore serve as the starting point of DNA replication. Then, the polymerase (Taq polymerase) enzymatically assembles a new DNA strand using desoxyribonucleotides (dNTPs), by using single-stranded DNA as template. To prevent unspecific amplification, a commonly used modified polymerase is HotStartTaq which remains in an inactive state until activated by 15-minutes incubation at 95°C. Each amplification cycle consists of different repeated temperature changes. Most cycles are composed of three temperature steps: denaturation (~95ºC), annealing of primers (50-65ºC) and extension (~72ºC). Those cycles are repeated 30 to 40 times on an automated PCR thermal cycler, generating millions of copies of the original DNA molecule (Sambrook J and Russell RG, 2001).

The PCR amplification conditions used in this study were as follows: a single enzyme activation step of 95 ºC for 15 min was followed by amplification using a “Touchdown PCR” technique in order to increase specificity and sensitifity of the PCR amplification (Korbie and Mattick, 2008). PCR usually relies on the accurate calculation of the melting temperature. As an incorrect estimation can lead to reduced hybridization performances the “Touchdown-PCR” technique eliminates this uncertainty and the amplification of nonspecific products by empirically addressing the approximations made

when calculating the melting temperature of PCR primers. The first 10 cycles each consisted of three steps: denaturation at 95ºC for 30 seconds (s), primer annealing starting at 68ºC for 20 s, decreasing in temperature one degree each cycle, and primer extension at 72 ºC for 1.5 min. This was followed by another 35 cycles consisting of 95ºC for 30 s, 60ºC for 20 s, and 72ºC for 1.5 min. The reaction was terminated with a final extension step consisting of a single cycle at 72ºC for 4 min and stored at 4°C. PCR amplicons were analyzed in 1 – 3% agarose gels containing ethidium bromide in Tris-Borate EDTA (TBE) buffer and examined for yield and specificity (see 4.8.1) and further used for sequencing or cloning reactions (see 0). Fragments were either purified directly after amplification or extracted from the gel in the presence of multiple amplification products and used for further downstream applications.

4.7 QIAquick PCR Purification

To further prepare DNA fragments for restriction and cloning reactions, PCR products were purified after the amplification. The Quiagen® QIAquick PCR Purification kit is designed for the fast cleanup of DNA fragments from enzymatic reactions. The system combines spin-column technology and selective binding properties of silica membranes. DNA absorbs to the silica membrane in the presence of high concentrations of salt. Contaminants do not bind to the silica column and therefore pass through the column and are later washed away. Afterwards, the DNA is eluted from the column with Tris buffer or ultra-pure water.

4.8 Gel Electrophoresis

Gel electrophoresis is commonly used for separation of molecules (nucleotides, proteins) according to their size or charge using an electric field applied to a matrix. The term “electrophoresis” is referring to the electromotive force which moves the molecules through the gel (Sambrook J and Russell RG, 2001).

4.8.1 Agarose Gel Electrophoresis

Agarose gel electrophoresis resolves fragments of DNA relative to their size using an electric field. The gel thereby allows the separation of negatively charged DNA macromolecules through a large mesh of channels formed during solidification. In a horizontal configuration in an electric field of constant strength and direction, the velocity of the DNA fragments decreases as their length increases and is proportional to the electric field strength. Ethidium bromide is commonly used to visualize double stranded DNA, since double stranded DNA fluoresces when exposed to ultraviolet (UV) light (Sambrook J and Russell RG, 2001). PCR products were analyzed on 1 – 3% agarose gels containing 0.5 µg/mL ethidium bromide. Depending on the targeted size of the amplification product, different molecular weight markers were included in each run (see 8.1.5).

4.8.2 Pulse-field Gel Electrophoresis

Pulse-field gel electrophoresis (PFGE) was conducted using a rapid standardized protocol for *E. coli* O157:H7 (Ribot et al., 2006) in order to confirm the identity of the strains used in this study and to determine changes due to storage conditions. Linear double stranded DNA molecules larger than a certain size (~40 kb) tend to migrate through agarose gels at the same rate. The velocity of the molecules becomes independent of their size and therefore, cannot be easily separated applying a constant electrical field to horizontal agarose gels. PFGE uses an electric field switched periodically between two different directions with pulse times ranging from 0.1 to 1000 seconds or more. This allows for separation of DNA molecules up to 5 MB in size. For this study, the contour-clamped homogeneous electric field (CHEF) device was used which is a hexagonal array of point electrodes in a voltage divider circuit. It produces homogenous fields, oriented at 120º, approximating those of pairs of infinitely long parallel electrodes (Sambrook J and Russell RG, 2001).

4.9 Standard Precipitation with Ethanol

The precipitation of DNA with ethanol is the standard method used for the recovery of DNA from aqueous solutions. The ethanol depletes the hydration shell from

nucleic acids and exposes the negatively charged phosphate backbone. Counter ions such as Na^+ (provided by the addition of Sodium Acetate (NaOAc) bind to the charged groups and therefore, reduce the repulsive forces between the polynucleotide chains to the point where a precipitate can form (Sambrook J and Russell RG, 2001).

4.10 DNA Extraction from Agarose Gels

The QIAEX II Gel extraction kit allows recovering DNA fragments from agarose gels using silica particles. Purified DNA can then furthermore be used for restriction reactions and cloning applications. High concentration of a chaotropic salt in the buffer disrupts hydrogen bonding between sugars in the agarose polymer. In addition, the high salt concentration helps to dissociate DNA binding proteins from DNA fragments. The agarose solubilizes and DNA molecules (40 bp to 50 kb) are absorbed at the QIAEX II particles. The DNA is forced to be absorbed to the silica particles due to the highly electrolytic environment with large anions modifying the structure of water. Non-nucleic acid impurities such as residual agarose, proteins, salts, and ethidium bromide are removed during the washing steps. The DNA elution is accomplished efficiently with water and can be used for further downstream applications.

4.11 Plasmid Vectors

4.11.1 Expression Vector pGNE

Vector pGNE was constructed in order to introduce a wild-type O157 *gne* gene in *E. coli* O *rough*:H7 strains MA6 and CB7326 (Figure 4-5). The *gne* gene appeared to be non-functional in those strains due to IS*629* insertions in the gene. The *trans*-complementation with pGNE allowed determining if the IS*629*::*gne* insertions caused the O *rough* phenotypes of these strains. The O157 wild- type *gne* gene has been inserted in the multiple cloning site (MCS) (*Bam*HI and *Sac*I) of the plasmid allowing the expression through the *trc* promoter. Transformants can be detected due to the exhibition of Ampicillin resistance (Amp^R).

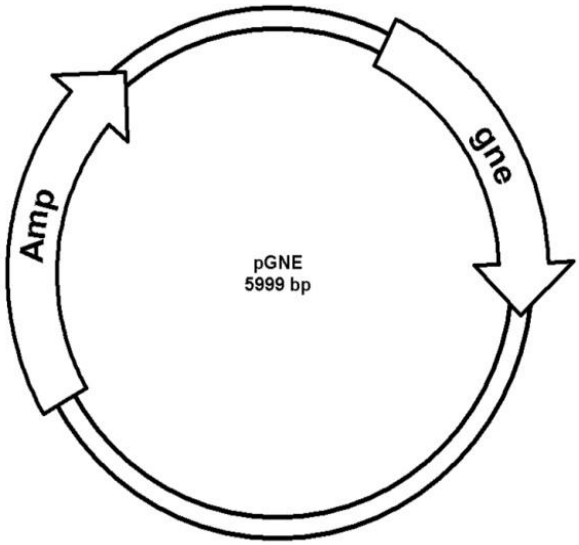


Figure 4-5. ***Schematic representations of the plasmid construct pGNE containing the wild-type gne gene***
AmpR: Ampicillin resistance, gne: wild-type gne gene

4.11.2 Vector pIS*629*AB-Tc

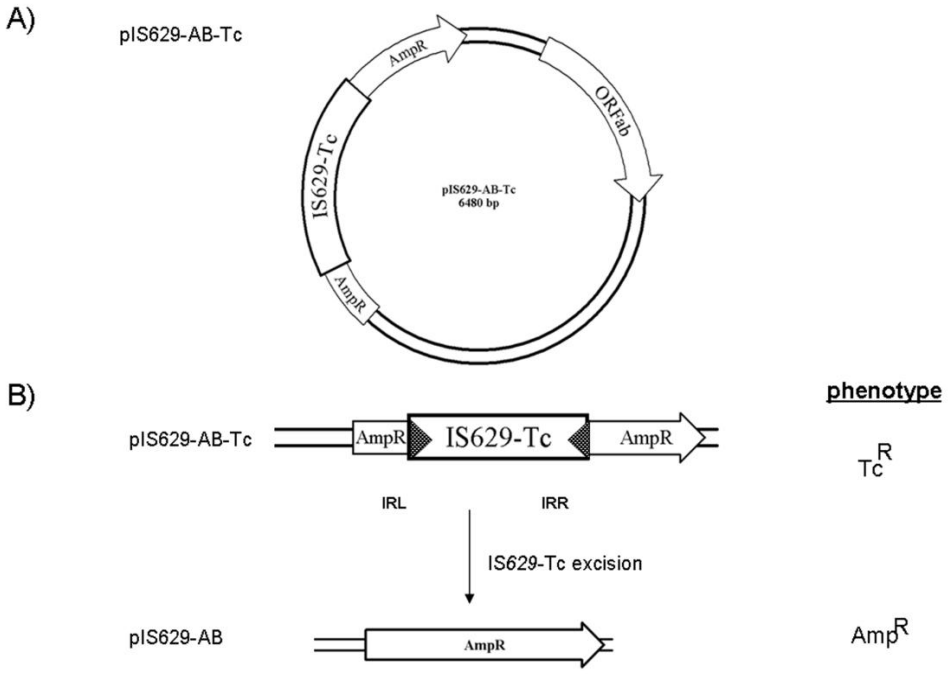
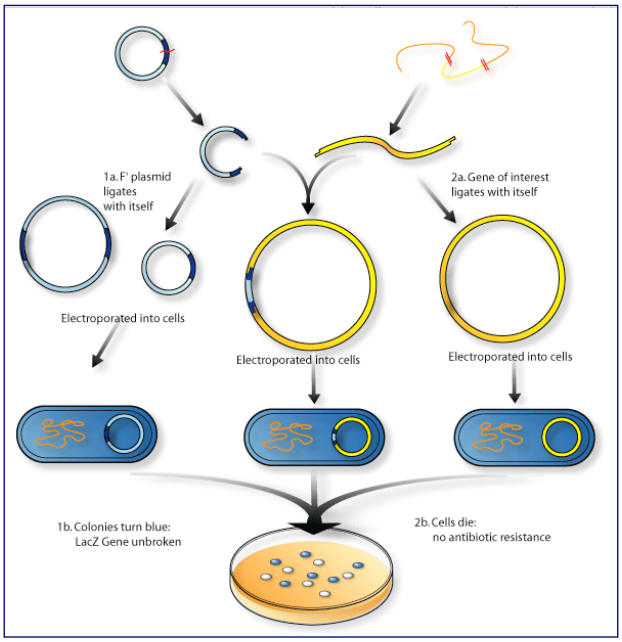


Figure 4-6. ***Schematic representations of the plasmid construct pIS629AB-Tc and determination of the transposition frequency***
A) Plasmid construct pIS629AB-Tc containing tetracycline resistance gene (tetC) disrupting the Ampicillin gene (Amp) and the IS629 transposase gene (ORFab). B) Successful transposition result in Ampicillin resistant colonies (transposition positive phenotype). Cells showing no transposition remain tetracycline resistant only (original phenotype). Excision frequency was calculated as follows: Amp^R/Tc^R cells. Tc^R: tetracycline resistance, Amp^R: Ampicillin resistance.

Vector pIS*629*AB-Tc was constructed in order to determine if IS*629* is able to transpose in IS*629*-deficient A4 clonal complex strains. The plasmid carries an IS*629* equivalent (designated IS*629*-Tc) consisting of both IS*629* IR flanking a *tet*C gene integrated in the *Amp* gene of the vector (*Sca*I site) (Figure 4-6). A single ORF encoding the IS*629* transposase (ORF*ab*) is introduced in the pUC18 MCS site (*Bam*HI and *Sac*I) and transcribed by the *lac*z promoter of the vector. Transformants can be detected due to the exhibition of tetracycline resistance (Tet^R). In the event of IS*629*-Tc transposition, the transformants regain Amp^R due to IS*629*-Tc excision from the *Amp* gene.

4.11.3 Directional Cloning into Plasmid Vectors

Cloning vectors are commonly used tools to introduce foreign DNA into a bacterial cell. To incorporate foreign DNA in a cloning vector *in vitro*, the closed circular plasmid will be cleaved with one or more restriction enzymes and ligated to the target DNA bearing compatible termini. The ligation product is then used to transform an appropriate cloning strain of *E. coli*. Since the vector contains a specific marker, such as antibiotic resistance, resistant colonies arise due to the presence of vector recombinants in antibiotic containing media (Figure 4-7). Cells able to grow on this media will have been transformed with the plasmid, as cells lacking the plasmid will be unable to grow. The transformants will afterwards be further analyzed for target DNA using PCR (Sambrook J and Russell RG, 2001).



Source: http://images.blog-u.net/view-450_Cloning_with_the_help_.png.html

Figure 4-7. Schematic Cloning principle
Figure showing different possible ligation events. 1a: Plasmid ligates with itself, 2a: Gene of interestligates with itself, middle: Gene of interest ligates into plasmid. 1b and middle: Colonies grow on media with resistancemarker, 2b: Cells die: no antibiotic resistance. Blue colonies: Plasmid without insert, White colonies: Plasmid with insert.

Plasmid vectors in common use contain multiple cloning sites (MCS) with recognition sequences for numerous restriction enzymes. To insert a DNA fragment in a cloning vector, both plasmid and target DNA will be digested with the same

restriction enzymes. To increase the yield of circular recombinants, a cloning strategy is used where the termini in the ligation reaction are not equivalent. In this case, the termini of the foreign DNA fragment will be non-complementary and are therefore, unable to ligate to each other. However, the foreign DNA will ligate preferably to a plasmid vector that has been prepared by cleavage with the same two enzymes. This will generate a high yield of desired circular recombinants containing a single insert in a predefined orientation.

4.11.4 Blunt-End Cloning

When using restriction enzymes generating blunt-ended termini for both, vector and foreign DNA fragment, the reaction is comparatively inefficient. Not only can the DNA fragments preferably ligate, but the vector can also re-ligate to its circular form. To minimize the latter, the vector is dephosphorylized by removal of the 5'-phosphate residue, suppressing self-ligation and recircularization of the linear plasmid. During the *in vitro* ligation, DNA ligase only catalyzes the formation of a phosphodiester bond between adjacent nucleotides if one nucleotide carries a 5'-phosphate residue and the other carries a 3'hydroxyl terminus. The removal of the 5'-phosphate residues of both termini of the cleaved plasmid can be performed with alkaline phosphatase (Ullrich et al., 1977; Seeburg et al., 1977). The foreign DNA with intact 5'-terminal phosphate residues can be ligated efficiently into the dephosphorylated plasmid DNA to generate a circular molecule.

4.11.5 Site Directed Mutation

A site directed mutatgenesis method by Mikaelian and Sergeant (1996) can be used for deletions, insertions and point mutations. This method is a rapid method derived from the overlap extension method described by HIGUCHI (Liu et al., 2008). The method consists of two successive rounds of PCR. In the first round, two simultaneous PCR reactions are performed (Figure 4-8). The first PCR is carried out with primer 1 and 2, whereby primer 2 contains a mismatched 3' end (~8 nucleotides (nt)). The second PCR is done with primer 3 and M, which contains the desired mutation (one ore more additional nt in the center of primer M). The amplified fragments will be purified and mixed together and amplified with primer 1 and M. Duing

this PCR parental strands and hybrid B cannot be amplified. Only the mutated strand A will be amplified. This resulting PCR product containing the desired mutation, can be used for further digestion steps.

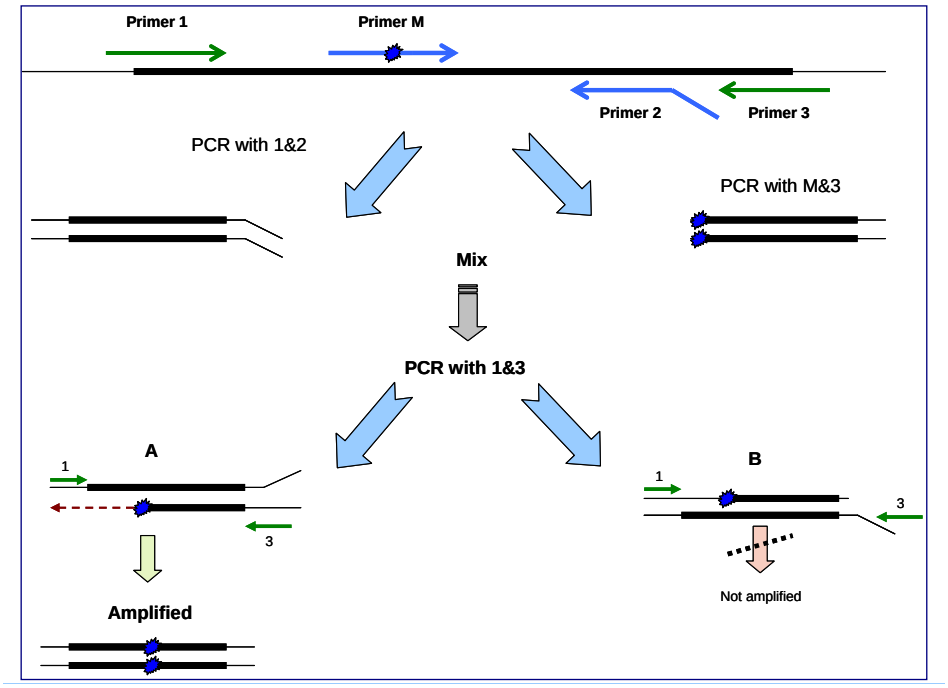


Figure 4-8. ***Principle of site directed mutation PCR (Mikaelian and Sergeant, 1996)***

4.11.6 Preparation and Transformation of Competent *E. coli* cells

The bacterial cell has to be in a state of competence during which DNA molecules may be admitted to the cell. In order to introduce foreign DNA into the bacterial cell molecules have to pass the outer and inner cell membranes until reaching an intracellular site where they can be expressed and replicated. For this purpose, chemical and physical methods are used to achieve the goals of competent cells.

The chemical methods are based on the observations of MANDEL and HIGA (1970), who showed that bacteria can be transfected with bacteriophage λ DNA if treated with ice-cold solutions of $CaCl_2$ and then briefly heated to 37ºC or 40ºC. The same method could afterwards be used to transform cells with plasmid DNA (Cohen et

al., 1972). Chemically induced competent cells are commercially available, yielding transformants at frequencies >10^8 colonies/µg of supercoiled plasmid DNA.

To physically transform cells, known as electroporation, the membrane is exposed to a short high-voltage electrical discharge, which destabilizes it and induces the formation of transient membrane pores (Neumann and Rosenheck, 1972). The vector DNA can easily pass from the medium into the cytoplasm. For electroporation, relatively small *E. coli* cells require very high field strengths (12.5 – 15 kV cm^{-1}) compared to the introduction of DNA into eukaryotic cells. An optimal efficiency can be archived by the use of dense slurry of bacteria (~ 2 x 10^{10}/mL). Although only 30-50% of the cells survive the procedure, under optimal conditions as many as 80% of those surviving cells may carry the desired vector DNA. Unlike chemical transformation, the number of transformants generated by electroporation is marker-dependent. If introducing a plasmid containing ampicillin as well as tetracycline resistance, the number of tetracycline-resistant transformants is ~100-fold less than the number of ampicillin-resistant (Amp^R) transformants (Steele et al., 1994). This effect has not been seen using chemical transformation methods. As for chemical transformation, the preparation of electrocompetent cells is essential for transformation with a vector. The bacteria will be grown till mid-log phase (OD: 0.3- 0.4), chilled, centrifuged and washed extensively with ice-cold 10% glycerol in order to reduce the ionic strength of the cell suspension. The cells are then resuspended in ice-cold buffer containing 10% glycerol and stored at –80°C.

4.11.7 Screening for Recombinant Plasmids

To identify whether a colony of transformed bacteria carries a recombinant plasmid or an empty wild-type plasmids, either optical methods can be used or the cells can be tested using PCR with vector specific primers. One optical method is the detection of β-galactosidase activity in transformed bacteria, called “blue/white screening”. Many vectors, like pUC18 and pTrc99A carry a short segment of *E. coli* DNA containing the regulatory sequences and the coding information for the first 146 amino acids of β-galactosidase (α-*lac*Z gene). Vectors of this type can be used in host cells that express the carboxy-terminal portion of β-galactosidase which is required to form a functional β-galactosidase enzyme. The MCS is embedded in the coding region

of the α-*lac*Z gene wherby inserting a DNA fragment in this region the production of β-galactosidase is disrupted (Davies and Jacob, 1968). To screen for β-galactosidase activity, X-gal (bromo-chloro-indolyl-galactopyranoside) is added to the media. Active β-galactosidase cleaves X-gal yielding galactose and 5-bromo-4-chloro-3-hydroxyindole. This product then oxidizes into an insoluble blue product (5,5'-dibromo-4,4'-dichloro-indigo) (Figure 4-9) (Horwitz et al., 1964).

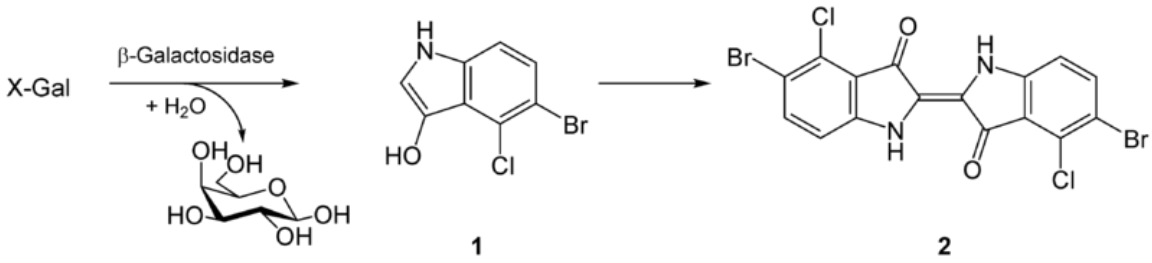


Figure 4-9. ***X-gal reaction in the presence of β-galactosidase***
1: bromo-chloro-indolyl-galactopyranoside, 2: 5,5'-dibromo-4,4'-dichloro-indigo

4.11.8 Plasmid DNA Purification

Plasmid DNA was extracted from bacterial cells using the Wizard® *Plus* Midiprep DNA Purification System. In this system, cellular structure is disrupted with chatotropic salts, detergents or alkaline denaturation to create a lysate to separate soluble DNA from cell debris and other insoluble materials. The resulting lysate is cleared by centrifugation. This is followed by purification of DNA of interest from soluble proteins and other nucleic acids on a silica matrix which binds DNA (Chen and Thomas, Jr., 1980; Marko et al., 1982). The lysate is combined with a resin slurry and then vacuum filtrated to wash the bound DNA. DNA elution is efficiently accomplished with water and can be used for further applications.

4.12 In silico analysis

Various *E. coli* O157:H7 and non-O157 genomes and pO157 plasmids deposited at the National Center for Biotechnology Information (NCBI) database were queried for IS*629* (accession number X51586) presence and insertion loci using Basic Local Alignment Search Tool (BLAST) analysis. Furthermore, approximately 400 bp up- and

downstream of the flanking regions of each new localized IS*629* in the genome and the plasmids were compared with each other. It was determined whether an IS*629* was also present in the other strains or appeared exclusively in either the genome or the plasmids.

5 RESULTS

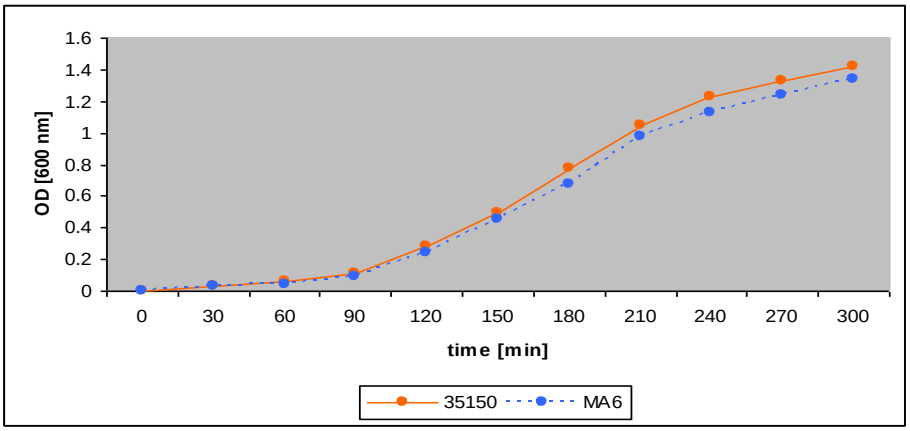
The first part of this chapter focuses on determining the cause(s) of the lack of O157 antigen expression in O *rough*:H7 strain MA6 by systematically examining the O antigen biosynthesis and assembly genes. Furthermore, strain CB7326, another O *rough*:H7 strain, was characterized and compared to MA6 to assess if the O *rough* phenotype was due to the same event and whether they were related strains. The second part deals with the prevalence and distribution of insertion sequence IS*629* among closely related *E. coli* O157:H7 strains as well as ancestral and atypical strains.

5.1 Genetic analysis for the absence of O157 expression in *E. coli* O *rough*:H7

5.1.1 Comparison of baterial growth

In order to determine if the lack of the O antigen in strain MA6 effects the viability of that strain in rich media, the growth rate under optimal conditions for the wild-type *E. coli* O157:H7 (EDL931) and *E. coli* O *rough*:H7 MA6 were compared (see 8.2.2). The growth curve of both strains was determined in Luria-Bertani (LB). Cell density was measured as optical density at 600 nm (OD600) (see 4.2). Comparison of these growth curves showed that MA6 and the wild-type strain had identical growth properties (Figure 5-1) and that the loss of the O antigen production in strain MA6 does not appear to have an impact on the its growth under optimal conditions. It remains unknown however noticable how the strain behaves in a more competitive environment such as a human host.

Figure 5-1. ***Growth curves for both E. coli strains (MA6 and ATCC 35150)***



5.1.2 O antigen operon

The O157 antigen is generally synthesized by gene products of the O antigen operon (*rfb*). Any defect in the operon could result in O157 deficient strains like MA6. The sequencing strategy was to generate a series of PCR amplicons that would span the entire *rfb* operon, including the promoter region in an over-lapping manner. The sequence data obtained from one amplicon could be overlapped to a sequence segment from another amplicon generated from an adjacent region in the operon. Significant changes would be detected by comparing the amplicon sizes of prototypic EDL931 wild-type strain and mutant strain MA6. Comparative sequence analysis of MA6 *rfb* operon sequences with that of EDL933 (GenBank accession no. AE005174) further allowed for identification of smaller scale mutations.

PCR of the *rfb* operon yielded 8 amplicons (~2,600 to 3,500 bp) that spanned all 12 genes of the operon including the promoter region (8.2.8.III). Agarose gel electrophoresis showed that amplicons derived from EDL931 and MA6 were identical in size, confirming the absence of major insertions or deletions within the MA6 *rfb* operon (8.5.2, Figure 0-1, Figure 0-2). Sequence analysis showed that there were no mutations in neither the promoter region, nor any ORF of the *rfb* MA6 operon that would affect O antigen biosynthesis. However, a one-base pair deletion between the genes *manB* and *wdbR* was identified. This mutation was not located inside an ORF, and unless this region has some other unknown function, it might not affect

functionality of the *rfb* operon. Furthermore, the promoter region did not show any mutations affecting functionality of the *rfb* operon either.

5.1.3 Core antigen biosynthesis gene cluster

Genes and operons (*waaA* and *gmhD*) that encode for inner core components of the LPS were not examined as these are essential for cell survival. PCR of the remaining *waa* cluster genes yielded 5 amplicons of 2,200 to 3,800 bp that spanned all 7 genes that encode the proteins necessary for the biosynthesis of the outer core (8.2.8.III). Comparison of the PCR amplicons of MA6 and EDL931 showed no apparent differences in sizes (Figure 0-3, Figure 0-4). Analysis of the gene sequences, including the predicted promoter regions, showed complete homology to that of EDL933, confirming the absence of insertions, deletions and base mutations in strain MA6. This finding indicates that the outer core biosynthesis gene cluster is fully functional since no mutations could be identified. Absence of the O antigen in MA6 must be due to other reasons than errors in the *waa* gene cluster.

5.1.4 Ancillary O antigen biosynthesis genes

Since no apparent defects were determined in neither the *rfb* nor the *waa* gene cluster, the lack of O157 antigen in MA6 could be due to a defect in other genes involved in O antigen synthesis. Other genes involved in O antigen biosynthesis but located outside the *rfb* and *waa* gene clusters are: *wec*A, m*an*A and *gne* (8.2.8.III). The amplicons of the *man*A gene as well as the *wec*A gene were identical in size to the EDL933 wild-type (Figure 0-5). Additionally, the sequences of each gene and the presumptive promoter region of both genes were identical to those of EDL933 in the GenBank database.

In contrast, the *gne* gene, encoding UDP-acetylgalactosamine (GalNAc)-4-epimerase, which is essential for the synthesis of one of the oligosaccharide subunits in the O157 antigen, showed a different result. When PCR primers that bound upstream of the putative promoter and downstream of the *gne* gene were used, an expected ~1,400 bp product was obtained for EDL931, but the MA6 PCR amplicon was ~1,300 bp larger (~2,700 bp) (Figure 0-5). PCR of other O157:H7 strains yielded the same ~1,400 bp product, while MA6 consistently produced the larger amplicons (Figure 0-6). Sequence comparison to that of EDL933 showed the presence of a 1,310 bp insertion within the MA6 *gne* ORF at position +385 (GenBank accession no. GU183138). The *gne* gene in MA6 seems to be non-functional due to this insertion. Consequently, the SNP UDP-Gal*N*Ac essential for the O157 antigen cannot be epimerized from UDP-Glc*N*Ac. Thus, the absence of the O157 antigen in MA6 appears to be due to an insertion in the *gne* gene.

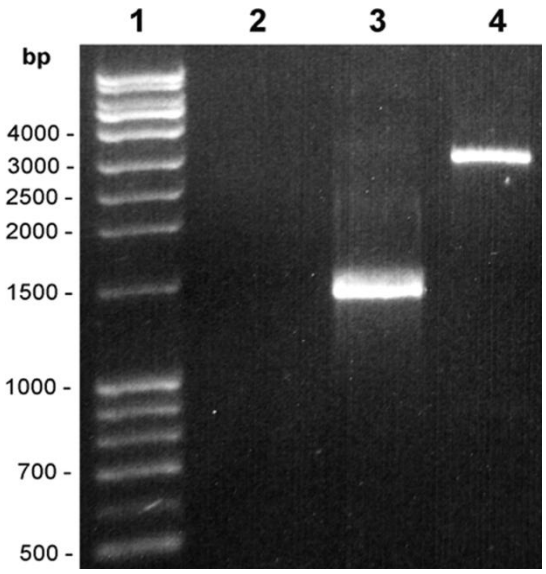


Figure 5-2. ***Agarose gel electrophoresis of gne amplicons derived from EDL931 (O157:H7) and MA6***
Lanes: 1: exACTGene (1 kb) plus molecular size ladder (Fisher BioReagents, Pittsburgh, PA), 2: negative control (reaction mix without DNA templat), 3: EDL933, 4: MA6.

5.1.5 Characterization of the insertion element in the MA6 *gne* gene

BLAST analysis of the 1,310 bp sequence showed that the insert located within the MA6 *gne* gene shared 96% homology to insertion sequence IS*629* (GenBank accession no. X51586). Comparison of the deduced protein sequence of the insert found in *gne* with that of IS*629* in GenBank, showed that the putative ORFA and ORFB proteins on the insert were 100% identical to that of IS*629* in O157:H7 Sakai and 99% homologous to those in O157:H7 EDL933 and O157:H7 EC4115 (GenBank acc. nos. see Table 3-3).

5.1.6 Trans-complementation with pGNE

IS*629* insertion in the *gne* gene of MA6 could be the reason for the O *rough* phenotype of this strain. If the absence of O157 antigen in MA6 is in fact due to this insertion resulting from a non-functional *gne* gene, the re-introduction of a wild-type *gne* gene in MA6 should recover O antigen expression. Therefore, MA6 was transformed with a vector (pGNE) which carries a functional *gne* gene and resistance to ampicillin (Amp^R) (see 4.11.1 and 8.2.14.I).

Amp^R MA6 (pGNE) transformants were PCR amplified with vector-specific primers to confirm that they carried the wild-type *gne* construct. The same transformant carrying the pGNE construct was tested by PCR to confirm possession of the *gne*::IS*629* locus. The PCR used primers which annealed to sequences outside *gne* and were not present on the vector or in the insert carried on pGNE, therefore only the MA6 possesing *gne*::IS*629* locus will produce an amplicon of ~2,700 bp. Positive results from both PCR reactions indicated that the MA6 (pGNE) transformant carried both the original *gne* gene harboring the IS*629* insertion and the wild-type *gne* gene carried on pGNE. Serological analysis of the MA6 (pGNE) strain with anti-O157 latex agglutination assay (see 4.3) confirmed that the O157 antigen was expressed. *Trans*-complementation of MA6 with wild-type *gne* restored O157 antigen expression confirming that the IS*629* insertion in the *gne* was the cause of the O *rough* phenotype.

5.1.7 Characterization of *E. coli* O *rough*:H7 from a hemorrhagic colitis patient

Another *E. coli* O *rough*:H7 strain, CB7326, was isolated in Germany from a hemorrhagic colitis patient harboring all characteristic O157:H7 traits. Previous findings of the IS*629*::*gne* insertion suggested that this strain might also be an O *rough* variant of O157:H7 due to a similar insertion event, like that observed for MA6. CB7326 did not ferment sorbitol or exhibit *ß*-glucuronidase activity, and serological analysis by latex agglutination confirmed presence of the H7 antigen but not the O157 antigen. Despite the absence of serological reactivity, PCR analysis for the *wzx* and *fliC* genes, which encode the O157 and H7 antigens, respectively, confirmed that CB7326 carried genetic sequences for both antigens. CB7326 was found to carry typical enterohemorrhagic *E. coli* (EHEC) virulence markers, including *stx*1, *stx*2, *ehxA* (enterohemolysin), the γ-*eae* allele, and the +93 *uidA* (ß-glucuronidase) single nucleotide polymorphism, which is unique to O157:H7 (Table 5-1). Except for the absence of the O157 antigen, these traits are consistent with those of O157:H7 (strain EDL931). Strain CB7326 had traits identical to those of strain MA6, except that MA6 did not carry *stx*1. To determine whether the cause of the O *rough* phenotype in CB7326 was also due to *gne*::IS*629*, the *gne* gene of CB7326 was amplified by PCR using *gne* specific primers (8.2.8.III). Analogous to the findings with MA6, CB7326 yielded a larger amplicon of ~2,700 bp in size. Sequencing of the amplicon

Table 5-1. ***Comparison of traits and markers among MA6, CB7326, and O157:H7 strains***

Trait or marker[1]	Result for strain		
	MA6	CB7326	O157:H7[2]
SOR	-	-	-
GUD	-	-	-
O157	-	-	+
Wzx (157)	+	+	+
H7	+	+	+
*fli*C (H7)	+	+	+
stx_1	-	+	+
stx_2	+	+	+
γ-*eae*	+	+	+
*uid*A	+	+	+
*ehx*A	+	+	+
MLST[a]	ST-66	S -66	ST-66
gne::IS*629*	+385	+711	-

[1]SOR: sorbitol fermentation, GUD: ß-glucuronidase activity, O157 and H7: O157 and H7 antigens, tested by latex agglutination, wzx (O157): wzx, encoding the O157 antigen, fliC (H7): fliC: encoding the H7 antigen, stx1 and stx2: Shiga toxin 1 and 2 genes, respectively, γ-eae: γ -intimin allele, uidA: +93 uidA single nucleotide polymorphism, ehxA: enterohemolysin, MLST: multilocus sequence typing, gne::IS629: insertion location. Strain EDL931 (http://www.shigatox.net/ecmlst/cgi-bin/index)

and BLAST analyses confirmed that the *gne* gene of CB7326 also had the IS*629* element; however, unlike in MA6, where *gne*::IS*629* was found at position +385, *gne*::IS*629* in CB7326 was located at position +711. *Trans*-complementation of CB7326 with the pGNE construct, carrying a wild-type *gne* insert, restored O157 antigen expression in CB7326 as determined by serological assays (8.2.3.II). These results confirm that, as for MA6, the O *rough* phenotype of CB7326 was due to *gne*::IS*629*.

5.2 Prevalence, distribution and evolutionary significance of the IS*629* insertion element in the stepwise emergence of *E. coli* O157:H7

5.2.1 PFGE comparison between the strains

PFGE patterns of the strains included in this study were compared and their identity queried to the PulseNet database (see 4.8.2 and 8.2.12). This comparison was conducted not only to confirm their identity but also in order to detect whether there were any noticeable changes between the PFGE patterns obtained from the strains in our study versus PFGE patterns recorded previously in the database. All strains were positively confirmed by comparing PFGE patterns obtained after *Xba*I digestion of genomic DNA from each strain with the PFGE patterns of the same strain recorded in the PulseNet database. This analysis was also conducted to determine if the same strain, although from different sources (FDA collection, ATCC) would show differences in their PFGE patterns. Indeed some strains analyzed showed some differences in PFGE patterns. In the ATCC, there are two deposited EDL933 strains: ATCC 43895 (CDC EDL933) and ATCC 700927. ATCC 700927 was derived from the original outbreak strain ATCC 43895. The PFGE pattern for ATCC 43895 was identical

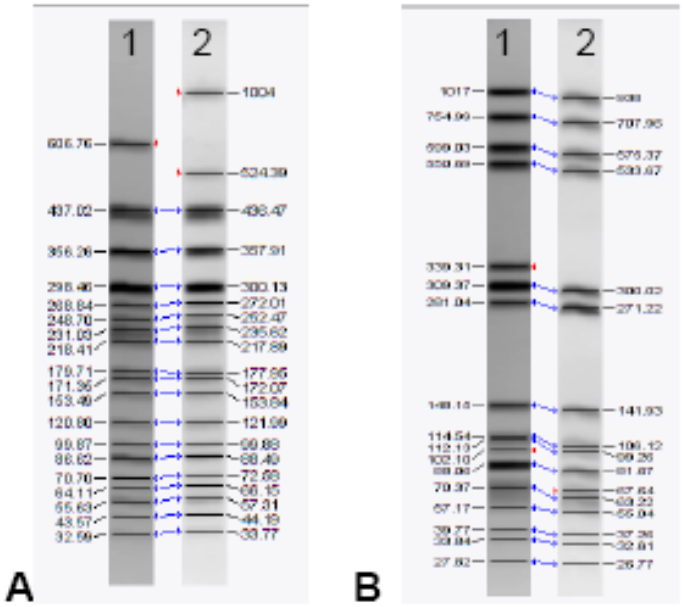


Figure 5-3. ***Pulsed-field gel electrophoresis results of EDL933 strains***
PFGE of XbaI- and BlnI-digested DNA from E. coli EDL933 strains. A: XbaI digested: 1: EDL933 ATCC 43895 - PulseNet, 2: EDL 933 ATCC 700927 (similarity 92.31%, (weight 39)); B: BlnI digested: 1: EDL933 ATCC 43895 - PulseNet, 2: EDL 933 ATCC 700927 (similarity 90.91%, (weight 33)).

to the PFGE pattern for EDL933 recorded in the PulseNet database (data not shown). However, the recently ordered strain ATCC 700927, although having a very similar PFGE pattern to that of ATCC 43895, showed differences in some bands. A similar result was observed when using another restriction enzyme (*Bln*I) (Figure 5-3).

Another strain which showed some alterations in PFGE pattern (*Xba*I and *Bln*I) when compared to its analogous PulseNet database pattern was the Sakai strain. The Sakai strain ordered from ATCC (ATCC BAA-460) was identical to that recorded in the PulseNet database for this strain, however, the Sakai strain from our collection (LVR) showed two band changes in the pattern with *Xba*I digestion and one band change with *Bln*I digestion compared to ATCC BAA-460 (Figure 5-4). These results showed that different storage conditions and multiple passages can induce changes in the PFGE patterns of strains. Albeit, the strains are remaining almost identical which allowed them to be identified sufficiently.

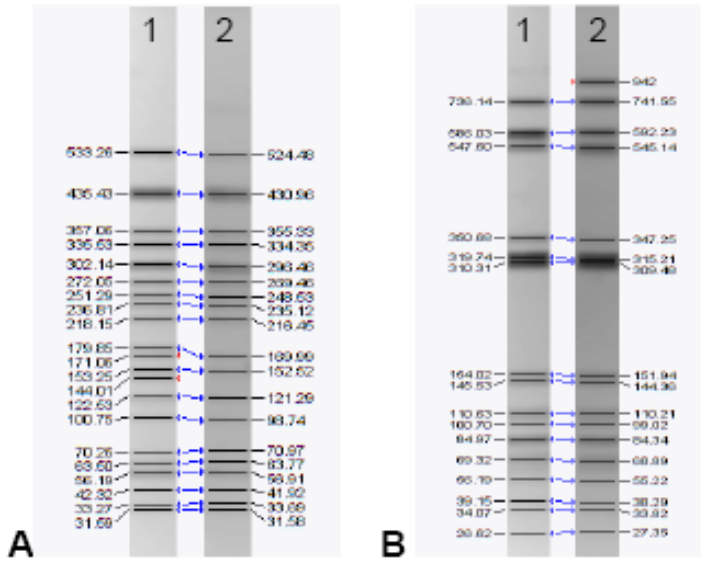


Figure 5-4. ***Pulsed-field gel electrophoresis results of Sakai strains***
PFGE of XbaI- and BlnI-digested DNA from E. coli O157:H7 str. Sakai. A: XbaI digested: 1: Sakai BAA-460 (ATCC), 2: Sakai (LVR) (similarity 96.97 %, (weight 39)); B: BlnI digested: 1: Sakai BAA-460 (ATCC), 2: Sakai (LVR) (similarity 95.00 %, (weight 33)).

5.2.2 IS*629* prevalence in *E. coli* O157:H7 genomes

The IS*629* sequence inserted into the *gne* gene in *E. coli* O *rough*:H7 strain MA6 (GenBank accession no. GU183138) was used for a BLAST analysis of the genomes of four *E. coli* O157:H7 strains (EDL933, Sakai, EC4115 and TW14359) and one O55:H7 strain (CB9615) (8.5.5; Table 0-5). The BLAST analysis for IS*629* showed the presence of between 22 and 25 copies in each strain along with their corresponding plasmid (Table 5-2). Strains Sakai and EDL933 shared 13 of those IS*629*s on the genome and three on their pO157 plasmids. Strains EC4115 and TW14359 had 17 IS*629*s on the genome and 4 on their pO157 plasmid in common.

Table 5-2. ***Prevalence of IS629 elements in each strain (genomes and plasmids) and number of shared IS629***

	Strain	Serotype	IS*629* sites	In common with strain			
				Sakai	EDL 933	EC 4115	TW 14359	CB 9615
Genomes								
	Sakai	O157:H7	19	-	13	9	9	0
	EDL933	O157:H7	21	13	-	6	6	0
	EC4115	O157:H7	19	9	6	-	17	0
	TW14359	O157:H7	21	9	6	17	-	0
	CB9615	O55:H7	**1 tr***	**0**	**0**	**0**	**0**	**-**
Plasmids				**pO157 Sakai**	**pO157 EDL 933**	**pO157 EC 4115**	**pO157 TW 14359**	**pO55 CB 9615**
	pO157 Sakai	O157:H7	3	-	3	3	3	0
	pO157 EDL933	O157:H7	3	3	-	3	3	0
	pO157 EC4115	O157:H7	4	3	3	-	4	0
	pO157 TW14359	O157:H7	4	3	3	4	-	0
	pSFO157	O157:H-	**0**	**0**	**0**	**0**	**0**	**0**
	pO55 CB9615	O55:H7	**1**	**0**	**0**	**0**	**0**	**-**

*tr: truncated

The analysis of the recently released *E. coli* O55:H7 genome strain CB9615 allowed for identification of one IS*629* with an internal 86 bp deletion and an IS*629* in its corresponding pO55 plasmid. Neither the O55 genomic nor the pO55 plasmid IS*629* insertion sites were present in other O157:H7 strains. The absence of the pO55 IS*629* insertion site in O157:H7 strains was expected since they do not carry the pO55 plasmid. However, lack of the genomic O55 IS*629* insertion site in O157:H7 strains is interesting as these strains are known to be closely related. Contrary to what was observed for plasmids pO157 and pO55, IS*629* was absent in plasmid pSFO157 (*E. coli* O157:H- strain 439-89). However, a 66 bp sequence identical to IS*629* was observed in the plasmid which could be a remnant of IS*629*. No genomic sequence is available for an O157:H- strain at this time, thus, this strain could not be investigated for the presence of IS*629*.

5.2.3 IS*629* target site specificity ("hot spots") on genomes and plasmids of four *E. coli* O157:H7 strains

The majority of IS*629* elements were located on prophages (Sp) or prophage-like elements (SpLE) (62%) ("strain-specific-loops", S-loops in Sakai (Mahillon and Chandler, 1998)). 28% of IS*629* locations were found on the in *E. coli* well-conserved genomic sequence widely regarded as the chromosome backbone (*E. coli* K-12 orthologous segment, 4.1-Mb) (Mahillon and Chandler, 1998), and 10 % were located on the pO157 plasmid. Since the majority of IS*629*s are preferably located on prophages or prophage-like elements, it seems likely that IS*629* was introduced into bacterial genomes through phages. In total, we observed 47 different IS*629* insertion sites (containing complete or partial IS*629*) in the 4 *E. coli* genomes and plasmids by computional analysis (8.5.7, Table 0-6). Seven of 47 IS*629* insertions were shared among the 4 diverged strains suggesting that they were also present in a common ancestor.

IS elements frequently generate short direct repeated sequences (DRs) of flanking target DNA upon insertion (Rump et al., 2010a) - this feature was also observed for IS*629* in the four O157:H7 strains. IS*629* duplicated between 3 to 4 base pairs at the insertion site and was observed for 21 of the 47 IS*629* insertion sites with matching identical base pairs up- and down-stream of IS*629*. A comparison

of 21 DRs created by IS*629* in the four strains analyzed did not reveal as many similarities as observed previously by OOKA et al (2009). The comparison of 25 bp up- and downstream of each insertion site did not show any similarities or patterns which would have suggested a target preference or "hot-spot" for IS*629* insertions. Hence, insertion site specificity for IS*629* remains unknown. However, IS*629* is frequently surrounded by other IS elements ('IS islands') and was found in the same gene (*gne*) but inserted in different sites (Mahillon and Chandler, 1998). These observations suggest that an insertion might occur preferentially in certain regions of the genome; however, these events may not be sequence specific.

5.2.4 IS*629* presence in strains belonging to the stepwise model of emergence of *E. coli* O157:H7

A total of 27 *E. coli* strains (Table 3-1) originally described in the stepwise model of Feng et al. (1998) were examined by PCR for the presence of IS*629* using IS*629* specific primers (8.2.8.IV). Every strain of clonal complex A6, A5, A2 and A1 carried IS*629*, except strain 3256-97 belonging to the ancestral CC A2. Strikingly, however, was the observation that IS*629* was absent in SFO157 strains belonging to closely related CC A4. Additionally, strains USDA5905 (A2) and TB182A (A1) as well as strain LSU-61 (CC unknown) appear to harbor a truncated IS*629* which could indicate the presence of genomic IS*629* found in the O55 strain CB9615. However, since no additional ancestral strains were available for analysis, the distribution of IS*629* in these groups is at present inconclusive.

5.3 IS*629* transposition in O157:H- IS*629*-deficient strains

It is striking that O157:H- strains, which are on a divergent evolutionary pathway in the stepwise emergence of O157:H7 (Feng et al., 2007), are IS*629*-deficient (Rump et al., 2011a). The absence of IS*629* among strains in the closely related clonal complex (CC) A4 (*E. coli* O157:H-) could be because an IS*629* carrying mobile element was excluded from infecting those strains. Another reason could be that A4 CC strains exhibit an IS*629* transposition inhibition mechanism disabling IS*629* transposition in A4 CC

strains. However these strains posses numerous other ISs belonging to the IS*3* family, like IS*629*. Thus, it is possible that other IS elements of the IS*3* family might interfere with its transposition.

To investigate if IS*629* transposition is being inhibited in the A4 CC strains, we constructed a vector pIS*629*AB-Tc and introduced it into various strains belonging to the stepwise evolutionary model for *E. coli* O157:H7 (Rump et al., 2011b; Feng et al., 2007). It carried an actively expressing IS*629* transposase gene (ORF*ab*) which has been shown to enhance IS*629* excision (Kusumoto et al., 2004; Kusumoto et al., 2011) but lacks IS*629* inverted repeats (IR) rendering the transposase unable to excise. It also carries an IS*629* analogue (IS*629*-Tc) in which a tetracycline resistance gene (*tetC*) replaces the IS*629* transposase gene embedded between both IS*629* IR truncating the vectors ampicillin resistance gene (*amp*). The IS*629*-Tc construct remains able to transpose if there is no inhibition of IS*629* transposition in the individual strain (Figure 1). In the event of precise IS*629* excision, amp resistant (amp^R), transformants will be observed. We introduced this vector into A2 CC strain (3265-97), A4 CC strains *E. coli* O157:H- (493-89, H56929c, H1085c), A5 CC strain (G5101), A6 CC strain (EDL933), and a possible evolutionary intermediary A3 CC strain (LSU-61) (Table 1). This vector allowed for determining IS*629* transposition and its excision frequency in those strains. We observed that IS*629*-Tc transposed in the tested strains. IS*629*-Tc successfully transposed in A4 strains signaling that absence of IS*629* in the A4 CC strains does not appear to be due to an IS*629* transposition inhibition mechanism. It is rather likely that after loss of IS*629* type II (until now for unknown reasons), these strains were not in contact with an IS*629* carrying mobile element after diverging from the hypothetical A3 CC.

5.4 Investigation of the IS*629* excision frequencies in different *E. coli* strains

We observed that IS*629*-Tc transposed in all strains tested although with notably different frequencies (Table 5-3). The excision frequency is the frequency of appearance of Amp^R cells caused by excision of IS*629*-Tc, and the values shown are the averages of

three independent analyses. As a result, the test strains were divided into frequent and infrequent groups whose frequencies were above 10^{-4} and below 10^{-7}, respectively.

Table 5-3. ***Serotype, sequence type and characteristics of E. coli strains used in this study and IS629-Tc excision frequencies from pIS629AB-Tc in each strain. In bold are the O157:H- strains lacking IS629***

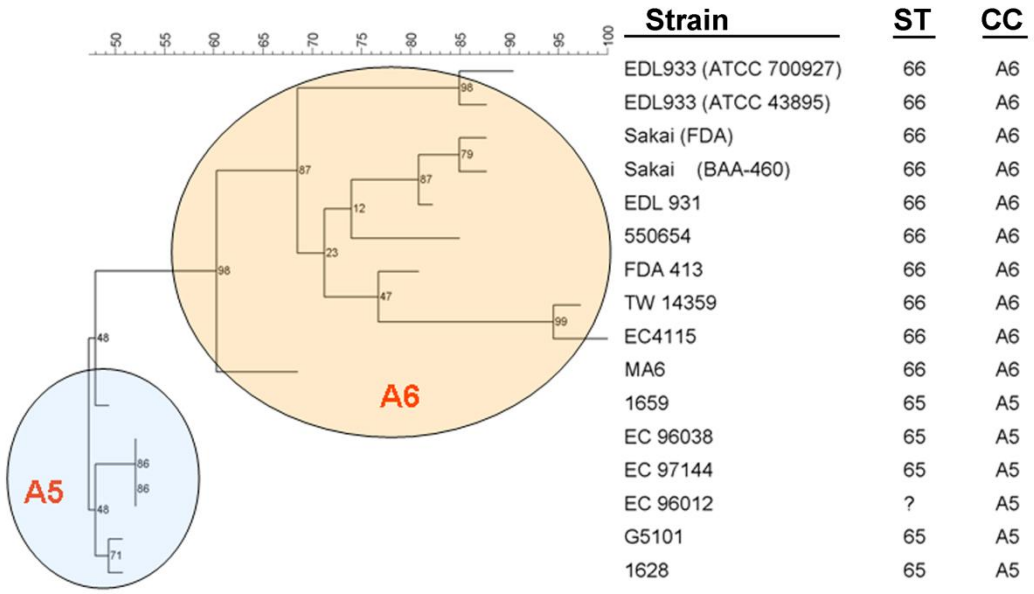
Name	Serotype	CC[a]	IS*629*[b]	Excision frequency[c]	
				Average	SD
EDL 933	O157:H7	A6	+	1.3×10^{-3}	$\pm 0.4 \times 10^{-5}$
G5101	O157:H7	A5	+	1.6×10^{-3}	$\pm 0.2 \times 10^{-3}$
LSU-61	O157:H7	A?	tr	0.6×10^{-3}	$\pm 0.2 \times 10^{-3}$
493-89	**O157:H-**	**A4**	**-**	**2.6×10^{-6}**	**$\pm 0.9 \times 10^{-6}$**
H56929	**O157:H-**		**-**	**2.2×10^{-6}**	**$\pm 0.3 \times 10^{-6}$**
H 1085c	**O157:H-**		**-**	**2.3×10^{-6}**	**$\pm 0.4 \times 10^{-6}$**
3256-97	O55:H7	A2	-	1.5×10^{-8}	$\pm 0.6 \times 10^{-8}$
DEC5A	O55:H7	A1	+	2.2×10^{-7}	$\pm 0.1 \times 10^{-7}$

[a] CC: clonal complex, Feng et al. (2001)
[b] IS*629:* IS*629* presence (+) or absence (-), tr: truncated IS*629*
[c] Three independent analysis for each strain were performed. The transposition frequency was calculated as follows: number of tet^R-amp^R cells/number of tet^R cells.

IS*629*-Tc excision frequencies in A5 and A6 CC strains were higher than in A1, A2, and A4 CC strains. These findings agree with previous results from Kusumoto et al. (2004), which noted that IS*629*-carrying strains have a higher IS*629* transposition frequency than IS*629*-deficient strains. The low excision frequency determined for IS*629*-deficient strain 3256-97 (1.5×10^{-8}) was similar to that of other strains lacking IS*629* of various serotypes and genotypes (Kusumoto et al., 2004). Strains from A1 and A2 CC (DEC5A and 3256-97), exhibited a low IS*629* excision frequency, regardless of the presence or absence of IS*629*. A4 CC strains exhibited a higher excision frequency (100-fold) compared to the tested A1 and A2 CC strains, although remaining lower than A5 and A6 CC strains. This intermediate excision frequency suggested that the presence of IS*629* alone might not enhance the transposition activity of IS*629*-Tc. Kusumoto and colleagues (2004) suggested that IS*629*-possessing strains use a "system to enhance IS*629* excision which might have been introduced by mobile genetic elements and may

be linked with IS*629* or other IS elements". The nature of this mechanism was recently described by Kusumoto et al. (2011) as a protein IS-excision enhancer (IEE) which promotes IS629 excision in O157. Sequence analysis of two A4 strains (493-89 acc. no. AETY00000000, H2687 acc. no. AETZ00000000) showed the absence of this specific gene. Hence, the elevated excision frequency in all A4 CC strains relative to A2 and A1 CC strains could indicate that these strains possess a different up-regulation mechanism than IEE. Conversely, A1 and A2 CC strains might lack IEE.

5.5 IS*629* distribution in strains belonging to the stepwise model of emergence of *E. coli* O157:H7



Strain	ST	CC
EDL933 (ATCC 700927)	66	A6
EDL933 (ATCC 43895)	66	A6
Sakai (FDA)	66	A6
Sakai (BAA-460)	66	A6
EDL 931	66	A6
550654	66	A6
FDA 413	66	A6
TW 14359	66	A6
EC4115	66	A6
MA6	66	A6
1659	65	A5
EC 96038	65	A5
EC 97144	65	A5
EC 96012	?	A5
G5101	65	A5
1628	65	A5

Figure 5-5. ***Maximum parsimony tree obtained using the distribution of IS629 and IS629 target sites in the 14 O157:H7 strains analyzed in the present study***
ST: sequence type, CC: clonal complex

Thirty eight of 47 observed IS*629* insertion sites could successfully be amplified by PCR in the 27 *E. coli* strains analyzed (8.2.8.V). The presence or absence of an IS*629* element as well as the IS*629* target site in each strain was determined. In accordance with the previous finding of total IS*629* absence in SFO157, none of the A4

CC strains harbored an IS*629* in any of the known IS*629* insertion sites (8.5.8 Table 0-7). Likewise, similar findings were observed for A1 and A2 CC strains, indicating that the previously detected IS*629* must be located in some other region of the genome. In A5 CC strains, only 3 of the 38 (7%) IS*629* insertion sites harbored an IS*629*. Those sites were located on prophage Sp12, the prophage-like element SpLE1, and on the chromosomal backbone (Sp and SpLE designations as by Hayashi et al (, 2004). Interestingly, one of the A5 CC strains (strain 1659) did not share any of known sites harboring IS*629*. The A6 CC strains shared between 6 (16 %) and 21 (55 %) IS*629* insertions in known sites, and 2 of them (IS.15: Sp14 and IS.41: pO157) were present in all A6 CC strains. IS*629* prevalence in A6 strains and its distribution amongst Sp, SpLE, backbone and pO157 plasmids did not show any specific pattern; however it appears that IS*629* transposes actively in O157 strains.

Figure 5-5 shows a maximum parsimony tree obtained for A5 and A6 CC strains using IS*629* presence/absence in the target site and presence/absence of IS*629* target site (chromosome or plasmid region). Strains belonging to A1, A2, and A4 CCs were not included in this analysis because they either lack IS*629* (A4) or IS*629* is located in other regions on the genome than the ones determined for O157:H7 strains. The parsimony tree allowed to separate A5 strains from A6 strains as proposed in the stepwise model (Feng et al., 2007). Furthermore, it showed the existence of high diversity among A5 and A6 CC strains. The validity of this analysis needs to be explored further using more O157:H7 strains belonging to A5 and A6 CCs. Besides using 25 different strains, two different EDL933 strains and two different Sakai strains were included in the analysis. EDL933 strains were provided by ATCC whereby strain EDL933 700927 was derived from EDL933 43895. Sakai strains were one from ATCC (BAA-460) and from a personal collection (PF FDA). PFGE analysis showed only minimal changes between those strains. The analysis based on IS*629* distribution also showed minimal changes in IS*629* distribution as well as among the Sakai and EDL933 strains. The use of IS*629* presence/absence in specific regions appears to be a promising and adequate technique to distinguish closely related O157:H7 strains. Furthermore, the presence/absence of a specific region in *E. coli* O157:H7 genomes, irrelevant of the presence of IS*629*, could provide additional information regarding an accurate phylogeny and relatedness among those strains.

5.6 IS*629* insertion site prevalence in strains belonging to the *E. coli* O157:H7 stepwise emergence model

PCR analysis for the presence of IS*629* insertion sites showed that sites located on the chromosomal backbone structure were present in all tested strains from the different clonal complexes (Table 5-4 and Table 0-8). However, neither A1, A2, nor A4 CC strains harbored any IS*629* in backbone IS*629* insertion sites. The backbone structure is highly conserved among *E. coli* O157:H7 and the presence of those sites in closely related strains suggest that it may be conserved in O55:H7 and O157:H- as well. Contrary to what was observed in the well-conserved backbone, IS*629* insertion sites in prophages and prophage-like elements in different strains were found to be highly variable (Table 5-5 and Table 0-8). As seen for the backbone IS*629* insertion sites, some of the phage associated IS*629* insertion sites were present in A1, A2 and A4 CC strains; however they lacked IS*629*. Many of the IS*629* sites on phages were unique to A6 CC strains (7 of 13) suggesting that they are strain-specific. This result underscores significant differences in the presence of phage-related sequences between strains belonging to the stepwise model of *E. coli* O157:H7. Phages are the major determinant in generating diversity, and since the majority of IS*629* is preferably located on prophages or prophage-like elements, it is reasonable to suggest that IS*629* could have been introduced into the genome through phages.

Table 5-4. ***Presence of IS629 target sites on the backbone***

IS629 target sites	A1	A2	A3	A4	A5	A6
IS.10	+/-	+	NA	+	+	+/-
IS.11	+	+	NA	+	+	+
IS.13	+	+	NA	+	+	+
IS.17	+	+	NA	+	+	+
IS.19	+	+	NA	+	+	+
IS.32	+	+	NA	+	+	+
IS.34	+	+	NA	+	+	+
IS.38	+	+	NA	+	+	+
IS.39	+	+	NA	+	+	+
IS.46	-	-	NA	+/-	+	+

NA: not applicable, "+": presence: "-": absence, "+/-": present in some strains.

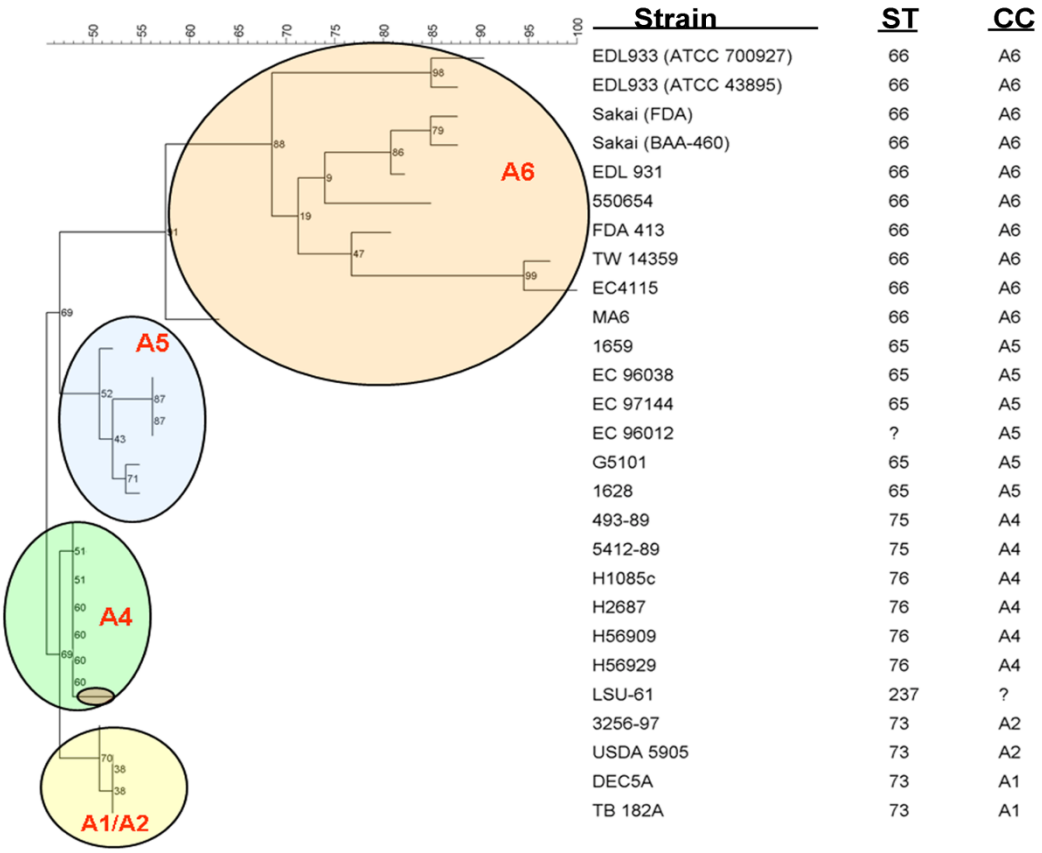
Table 5-5. ***Presence of phage or phage-like associated IS629 target sites***

IS*629* target sites	Sp SpLE	A1	A2	A3	A4	A5	A6
IS. 29	**Sp 1**	-	-	NA	-	-	+
IS. 45	**Sp 2**	+	+	NA	+	+	+
IS. 1, IS.2, IS.30	**Sp 4**	+	+	NA	+	+	+
IS. 3, IS.35	**Sp 5**	-	-	NA	-	-	+
IS.8/9	**Sp 8**	-	-	NA	-	-	+
IS.12	**Sp 12**	-	+	NA	+	+	+
IS. 14	**Sp 13**	-	-	NA	-	-	+
IS.15	**Sp 14**	-	-	NA	+	+	+
IS.20	**Sp 17**	-	-	NA	-	-	+
IS.4/5, IS.6/7, IS.24/25, IS.31	**SpLE 1**	-	-	NA	-	+	+
IS.27, IS.16	**SpLE 2**	-	-	NA	-	-	+
IS.21	**SpLE 3**	-	-	NA	-	-	+
IS.23	**SpLE 5**	-	-	NA	-	+	+

Sp: Phage, SpLE: Phage-like element, NA: not applicable, "+": presence, "-": absence.

The two IS*629* insertions in O55 and its corresponding plasmid pO55 were observed to be present in only one ancestral A2 and both A1 CC strains (Figure 0-9/Figure 0-18). A6, A5, and A4 CC strains as well as A2 CC strain 3256-97 (IS*629*-deficient) lacked the IS*629* insertion site in these regions. Interestingly, strain LSU-61 appeared to carry the truncated genomic IS*629* insertion, albeit, the exact location of this strain in the model of stepwise emergence of *E. coli* O157:H7 remains unclear.

Since strains belonging to the stepwise model share variable IS*629* insertion sites, we reconstructed their evolutionary path using this information. A parsimony tree using IS*629* target sites presence/absence produced a tree that was nearly analogous to the proposed model of stepwise evolution for O157:H7 from ancestral O55:H7 strains (Feng et al., 2007), with A1/A2 CC strains at the base of the tree, followed by A4 CC, A5 CC and A6 CC strains in that order (Figure 5-6). Additionally, MA6 appears to be more closely related to A5 CC strains than other A6 CC strains.



Strain	ST	CC
EDL933 (ATCC 700927)	66	A6
EDL933 (ATCC 43895)	66	A6
Sakai (FDA)	66	A6
Sakai (BAA-460)	66	A6
EDL 931	66	A6
550654	66	A6
FDA 413	66	A6
TW 14359	66	A6
EC4115	66	A6
MA6	66	A6
1659	65	A5
EC 96038	65	A5
EC 97144	65	A5
EC 96012	?	A5
G5101	65	A5
1628	65	A5
493-89	75	A4
5412-89	75	A4
H1085c	76	A4
H2687	76	A4
H56909	76	A4
H56929	76	A4
LSU-61	237	?
3256-97	73	A2
USDA 5905	73	A2
DEC5A	73	A1
TB 182A	73	A1

Figure 5-6. ***Maximum parsimony tree obtained using IS629 target sites for the 27 strains analyzed in the present study***

The colored ellipses mark the different CCs. CC: clonal complex (as by Feng et al., 2007), ST: sequence type (http://www.shigatox.net/ecmlst/cgi-bin/index).

5.7 Phylogenetic analysis of IS*629*-like elements in four *E. coli* O157:H7 and O55:H7 genomes

Phylogenetic analysis of IS*629* elements revealed that IS*629* in *E coli* O157:H7 can be divided into three different types (Figure 5-7). Evolutionary history was inferred using a Minimum Evolution method (Rzhetsky and Nei, 1992). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. IS*629* of *s*type I and II differed by an average of 4 % (> 55 nt) while type II and III differed by 5% (> 60 nt). It is notable that only four nucleotide differences were observed among seven housekeeping genes comprising a current MLST scheme (http://www.shigatox.net/ecmlst/cgi-bin/index) between A1 CC strain DEC5A and A6 CC strain Sakai. These two strains, in particular, are taken to represent the most ancestral and most derived *E. coli*, respectively, in the stepwise evolutionary model for this pathogen. If the IS*629* subsubtype I and III observed in A6 CC strains resulted from divergent evolution of IS*629* subtype II, the amount of changes observed among these IS subypes should be similar to those observed for the MLST loci examined above. However, the number of nucleotide substitutions between IS*629* subtype I and III in O157:H7 from subtype II in O55:H7 was 10-fold higher. Thus, the differences between IS*629* subtypes are more significant than those observed for housekeeping genes.

IS*629* subtype I appears to be most closely related to those of IS*1203* (IS*629* isoform) found in O111:H- (Paton and Paton, 1994), whereby IS*629* subtype II to IS*3411* are related to IS*629* found in *Shigella sonnei* (Matsutani and Ohtsubo, 1990). IS*629* subtype III seems to be related to IS*629* found in *E. coli* O26:H11 (Zhang et al., 2007). Analysis of all targeted IS*629*-like elements showed that strains from A6 CC seem to carry both IS*629* subtype I and subtype III whereby the ancestral O55:H7 strain carries IS*629* subtype II, hence IS*629* subtype II is no longer present in O157:H7 strains. This indicates that IS*629* subtype II was most likely lost and IS*629* subtype I and III were acquired independently in distinct *E. coli* O157:H7 lineages.

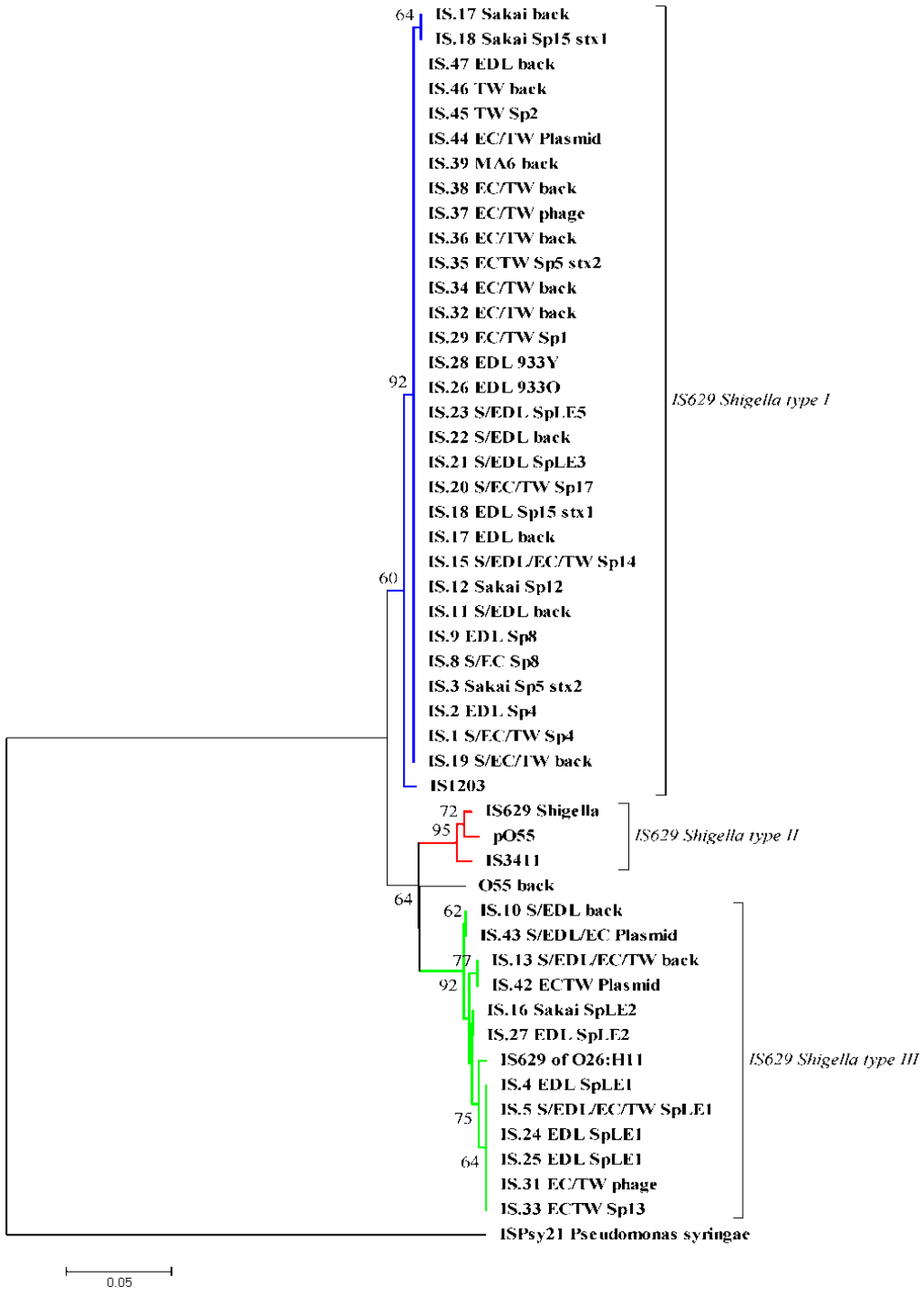


Figure 5-7. ***Phylogenetic tree of IS629 in E. coli O157:H7 and O55:H7 showing the three different IS629 subtypes present on those five genomes***

IS629 subtype I differed from subtype II by 4% (> 55 bp) and subtype II differed from subtype III by 5 % (>60 bp). IS629 subtype II was only present in O55:H7 genome (A1/A2 CC) while subtype I and III were present in all O157:H7 genomes (A6 CC). Bootstrap support when above 50% is shown at nodes. Sp- prophages; SpLE - prophage-like elements; and back – backbone.

5.8 IS*629* on phages and phage-like elements

The phage-like element SpLE1 carries multiple IS*629*s all belonging to IS*629* type III. In Sakai, this CP4-like cryptic prophage harbors one complete and one truncated IS*629*. EC4115 and TW14359 harbor an additional IS*629* as well as EDL933 which differs however at the insertion found in EC4115 and TW14359. Although SpLE1 has not been confirmed to be mobile, a second copy of this element, harboring the same IS*629*, is integrated into another site (tRNA serine gene) in EDL933 (Perna et al., 2001). Additionally, this specific mobile element (SpLE1) also can be found in other serotypes such as O26:H11 and O111:H- which harbor multiple IS*629*s, albeit in different locations (Figure 5-8). Some rearrangements between phages of O26 (Genbank acc. no. AP010953.1) and O111 (Genbank acc. no. AP010960.1) compared to Sakai and EDL933 appear to have happened. It is however unclear if IS*629* was the cause of the apparent rearrangements. One O111:H- strain carries two additional complete IS*629* s on SpLE1. This phage-like element appears to carry multiple IS*629*s in different locations, and some of them are unique to specific strains. No other phage in the four *E. coli* O157:H7 investigated carried multiple IS*629* elements, hence this seems to be exceptional for this specific phage-like element found in those strains. The presence of SpLE1 in other IS*629*-carrying serotypes makes it highly likely that phages and phage-like elements served as mobile vehicles for the introduction of IS*629* into O157:H7 genomes. Regrettably, it can not be determined if this specific phage-like element caused the IS*629* introduction into the genome of O157:H7 strains.

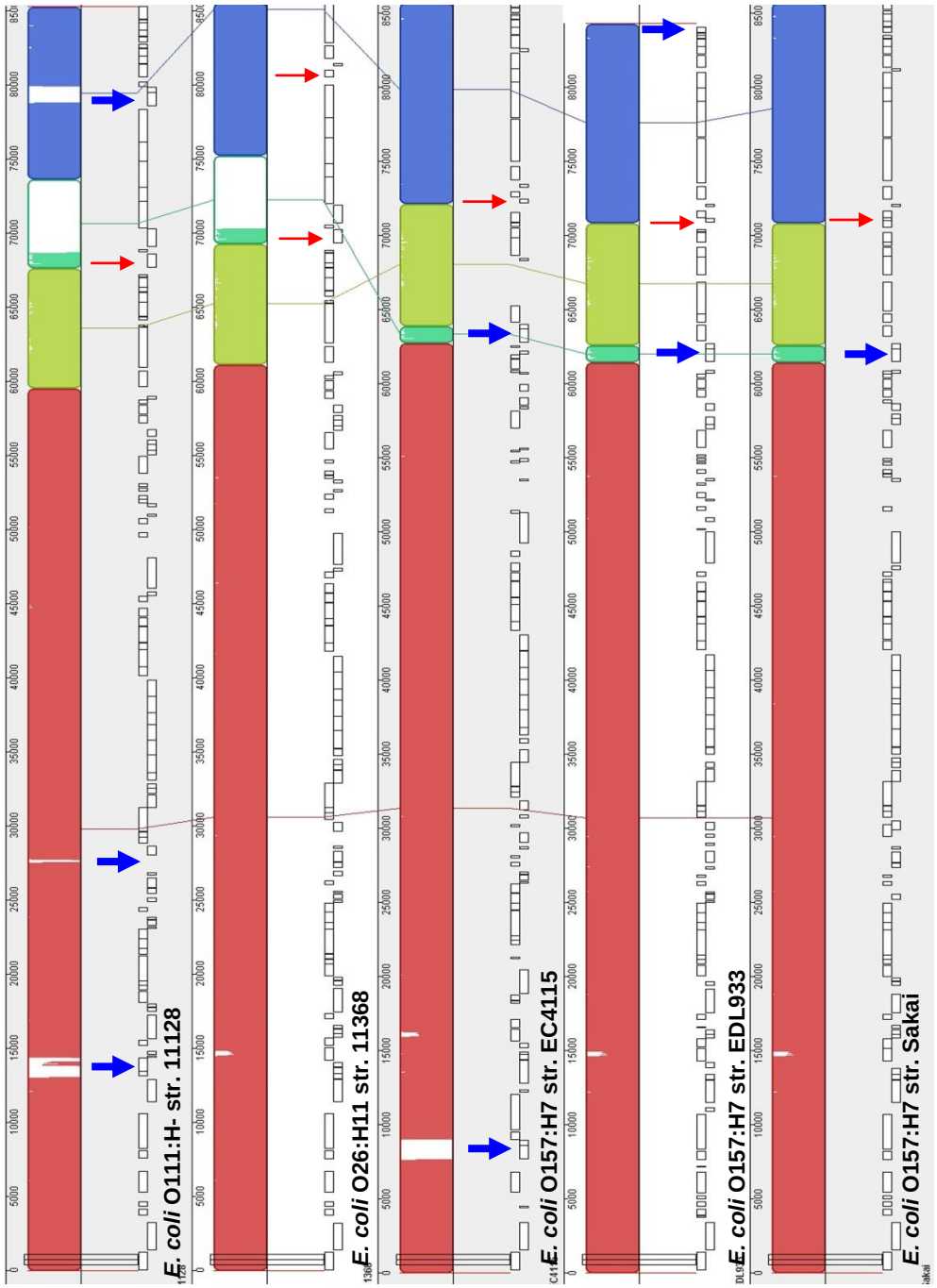


Figure 5-8. ***Comparison of the 5 SpLE1 elements in Sakai, EDL933, E. coli O26:H11 and E. coli O111:H- using*** *IS629 insertions are shown by arrows (blue: complete, red: truncated)*

5.9 IS elements on the pO55, pSFO157 and pO157 plasmids

Plasmid pO157 carries between 3 and 4 IS*629*s of type I and III however plasmid pO55 was shown to harbor one IS*629* of type II. Since CC A4 strains carrying pSFO157, are IS*629*-deficient, plasmid pSFO157 does not carry IS*629*. The 3 plasmids vary significantly in layout and size, whereby plasmid pSFO157 is the biggest (~120,000 kb), followed by pO157 (~90,000 kb) and the smaller pO55 (~66,000 kb) (Mahillon and Chandler, 1998). All plasmids carry multiple ISs which are mostly located in "IS-islands", flanking coherent gene clusters. The most prevalent IS in pO157 is IS*629* from the IS*3* family, which also is the predominant IS family in this plasmid and pSFO157. pSFO157 appears to carry more IS elements (14), however, considering the fact that this plasmid is significantly bigger than the two others, this finding appears to be more related to plasmid size.

Table 5-6. ***Insertion sequences present in E. coli plasmids pO55, pO157 and pSFO157***

IS family	IS element	Plasmid		
		pSFO157 (~120,00 kb)	pO157 (~90,00 kb)	pO55 (~66,00 kb)
IS*1*	isoIS*1*N	2	1	
IS*21*	IS*21*		1	
	IS*100*	1		1
IS*200*	IS*200*	1		
	IS*2*	1		
	IS*600*		1	
IS*3*	IS*629*		3-4	1
	IS*911*	3		
	IS*3*	2	2	2
IS*30* Tn	IS*30*	2		
IS*4*	IS*4*			1
IS*66*	ISEc*8*			2
Tn3	Tn3			2
	Tn2501	1		
IS*91*	IS*91*		1	
-	IS*697*	1		
Total		14	9-10	9

The sequence of four pO157 plasmids available in GenBank are highly similar, however the plasmids found in Sakai and EDL933, as well as those of EC4115 and TW14359, are more analogous, respectively. pO157 of EC4115 and TW14359 also

share one additional IS*629* located between the *ccd*A2 and a cytotoxic protein gene (*ccd*B) located in a region regarded as ancient transposon debris (Mahillon and Chandler, 1998). This IS*629* insertion (designated IS.42) is unique to those two strains (8.5.12, Figure 0-12), however, the site is present in all A6 CC strains that carry pO157 and all A4 CC strains carrying pSFO157. Interestingly, this region is absent in A5 CC strains which are closely related to the A6 CC strains.

IS.40, IS.44 and IS.48 are located between the genes *sop*A and *red*F (*rep*F), a region that contains a complete IS*629* sequence. EC4115 and TW14359 (IS.44) are analogous in this region, however these sequences in Sakai (IS.40) and EDL933 (IS.48) show significant differences in IS*629* flanking regions. Despite those differences, all A4 CC strains carry the IS.44 site (not containing IS*629*) yet none could be detected in A5 CC strains. These IS*629* sites are located on the beginning of a region related to F-like plasmids, whereby most of the genes carried in this region are the same as those in F-like plasmids (Mahillon and Chandler, 1998). IS.40/44/48 and IS.42 could be detected in A4 CC strains and LSU-61, but were absent in all A5 CC strains. Both sites (IS.40/44/48 and IS.42) appear to be absent in all A5 CC strains. This is surprising, since those strains are believed to be precursors of A6 CC strains and are thought to be closely related to A4 CC strains. Additionally, none of the ancestral A1 and A2 CC strains carry these regions. These findings suggest that the plasmid present in CC A5 strains is different from pO157 and pSFO157.

Two truncated IS*629*s are shared between all four pO157 plasmids. IS.41 is located between the *kat*P/*espP* (encoding a catalase-peroxidase and a serine protease, respectively) genes but is absent in pSFO157. Instead, the *sfp* gene cluster is inserted into this region in pSFO157 (Bielaszewska et al., 2009). The site was present in all A6 CC strains, however could not be detected in any strain of the other clonal complexes (8.5.12, Figure 0-13). According to the sequence date available in GenBank, this region is absent in pSFO157 and pO55, and thus, this result was expected for strains carrying those plasmids. However, it is surprising that this region is absent in A5 strains and in LSU-61, which were both thought to carry plasmid pO157 (Wick et al., 2005). Another IS*629* insertion site (IS.43) is located between the *tra*I/*tox*B genes. IS.43 could be detected in all but two A6 CC strains and, more importantly, seems to be present albeit ~200 bp smaller in all A5 CC strains. Since a truncated IS*3*

element (212 bp) is located in this region, it appears that the A5 CC strains might lack this element in that region, but still carry IS*629*, *tra*I and *tox*B.

5.10 Presence of pO157, pSFO157 and pO55

Table 5-7. ***Presence of plasmid specific regions in strains from different CC***

CC strains	pO157	pSFO157	pO55
	espP	gene 1	colicin gene
A6	+	-	-
A5	-	-	-
A? (LSU-61)	-	+	-
A4	-	+	-
A2 (3256-97)	-	-	-
A2 (USDA 5905)	-	-	+
A1	-	-	+

CC: clonal complex

Due to the fact that some regions in the pO157 as well as in the pO55 plasmids could not be detected in some of the A5 CC strains, new plasmid specific targets were chosen in order to determine which plasmid was carried by each strain (8.2.8.VI). In order to detect pO55, the colicin gene was used as a target. None of the A6, A5, A4 CC strains carried this region. Suprisingly, however, strain 3256-97 of CC A2 was lacking this region. As determined previously, this strain also lacks the IS*629* intertion site on this plasmid. This O55:H7 strain is thought to belong to CC A2 which all carry pO55, but the absence of those two regions might indicate that this strain lacks pO55.

The *esp*P gene was chosen as a specific target to detect pO157 (Brunder et al., 1999) which appears to be absent in pSFO157 and pO55. As expected, all A6 CC strains carried the *esp*P and subsequently harbor the pO157 plasmid. Additionally, the regions could neither be detected in A4 CC strains which carry pSFO157, nor in the ancestral A1 and A2 CC strains carrying pO55. However, none of the A5 CC strains as well as LSU-61, which all are O157:H7 strains, showed the presence of this region. This might indicate that those strains either lack the targeted region or carry entirely different plasmids. A5 CC strains were shown to carry *ehxA* which is carried by both pO157 and pSFO150 plasmids (Brunder et al., 2006) indicating the presence of a plasmid similar to

pSFO157 and pO157 in those strains. The absence of most of the pO157 associated IS*629* insertions in A5 CC strains might point to the presence of a plasmid other than pO157 or pSFO157 which has not been investigated to date.

Targeting the pSFO157 plasmid, as expected, all A6, A5 and ancestral A1 and A2 strains lacked the region, and all A4 CC strains showed the presence of the region. Most interestingly, also LSU-61 and the reference strain ECOR37 were positive for the site. Accordingly, LSU-61 might carry a pSFO157-like plasmid instead of pO157, contrary to what was expected for an O157:H7 strain. Additionally, LSU-61 shares specific characteristics with strains of the A5 CC strains like enterohemolysin activity and the H7 antigen (Table 5-8) (Feng et al., 2007), however, none of the A4 CC strains which carry pSFO157 show enterohemolysin activity. Therefore, LSU-61 appears to be an atypical H7 positive strain which carries pSFO157 but shows hemolysin activity.

Table 5-8. ***Characteristics of A4 CC strains and LSU-61***

Strains	CC	SOR	GUD	Te	H7	*stx*	*ehx*A	+776	BAP	plasmid
G5101	A5	-	+	R	+	1,2	+	+	+	pO157
498/89	A4	+	+	S	-	2	+	-	-	pSFO157
Germany 210-89	A4	+	+	S	-	2	+	-	-	pSFO157
Germany CB1009	A4	+	+	S	-	-	+	-	-	pSFO157
Finland IH56929	A4	+	+	S	-	2	+	-	-	pSFO157
LSU-61	A?	+	+	R	+	-	+	-	+	pSFO157-?

CC: clonal complex, SOR: sorbitol fermentation, GUD: β-glucuronidase activity, Te: tellurite resistant (R) or sensitive (S), H7: antigen, *stx:* shigatoxin gene, *ehx*A: enterohemolysin gene, +776: *uid*A SNP, BAP: enterohemolysin activity on blood agar plates (Feng et al., 2007).

6 DISCUSSION

6.1 Atypical *E. coli* O *rough*:H7 strains

In this project it was determined that two atypical O *rough*:H7 strains, which carry the characteristic traits of O157:H7 strains, did not produce O157 antigen due to IS*629* insertions in the *gne* gene. The *gne* gene encodes for UDP-*N*-acetylgalactosamine (GalNAc)-4-epimerase enzyme, which is essential for the synthesis of NSP UDP GalNAc, one of the oligosaccharide subunits necessary for O antigen biosynthesis (Wang et al., 2002). In *E. coli,* the genes involved in O antigen biosynthesis are usually clustered between *galF* and *gnd* (Hobbs and Reeves, 1994). However, a few genes, such as *gne*, can be located outside and upstream of the *galF* gene, as it occurs in *E. coli* O157:H7, O55 and a few other enteric bacteria (Samuel et al., 2004; Senchenkova et al., 2005). Interestingly, the *gne* gene has also been found inside the O antigen gene cluster, as is the case for *Pseudomonas aeruginosa* O6, *E. coli* (for example O113) and several *Shigella bodii* and *dysenteriae* strains (Creuzenet et al., 2000; Paton and Paton, 1999; Liu et al., 2008).

In the case for *Salmonella enterica* O50 and *Shigella boydii* O10, *gne* is located outside the cluster as in *E. coli* O157:H7 (Samuel et al., 2004; Senchenkova et al., 2005; Wang et al., 2002). Mutations at the level of nucleotide sugar synthesis, at some stage in the assembly or at the attachment of the O antigen side chains to the outer core region can give rise to O *rough* mutants (Schnaitman and Klena, 1993; Stevenson et al., 1994). It is not uncommon to find O *rough* strains of *E. coli* and other bacterial species (Keskimaki et al., 1999; Sheng et al., 2008), but O *rough* strains of O157:H7 seem to be rare. Sequence analysis of genes involved in O biosynthesis and transport in MA6 showed the presence of IS*629* within its *gne* gene. *Trans*-complementation with a wild type *gne* gene restored O157 antigen expression, confirming that the O *rough* phenotype of MA6 was caused by IS*629* insertion. IS*629* was first isolated from the chromosome of *Shigella sonnei* (Matsutani et al., 1987) and has since been found, often in multiple copies, in many other enteric bacteria, including O157:H7 (Matsutani and Ohtsubo, 1993). IS elements, as IS*629*, have been shown to inactivate genes. For example, an IS*629* insertion into the *fliC* gene that encodes flagella caused non-motility in an *E. coli* O111 strain (Zhang et al., 2007). Similarly, IS*629* inserted into *wbaM*, a ribosyltransferase gene in a strain of *Shigella boydii* O6, resulted in an O antigen

deficient strain (Wang et al., 2001). MA6 appeared to be the only naturally occurring O *rough* strain of O157:H7 that arose from the disruption of the *gne* gene by IS*629* thus far.

A study in Germany characterized patient isolates of STEC over a 3-year period and found a few strains with the O *rough*:H7 phenotype (Beutin et al., 2004). One of these, CB7326, was isolated from a hemorrhagic colitis patient and found to carry common O157:H7 virulence characteristics (Rump et al., 2010a). This lead to the suggestion that, like MA6, CB7326 may be an O *rough* variant of O157:H7 which has been implicated in illness. It was determined that the O *rough* phenotype of CB7326 was due to *gne*::IS*629,* however, unlike in MA6, where *gne*::IS*629* was found at position +385, *gne*::IS*629* in CB7326 was located at position +711. PFGE comparison of *Xba*I-digested genomic DNA showed that MA6 and CB7326 shared little homology with each other and, therefore, are not analogous strains. The IS*629* recognition site remains unknown (Mahillon and Chandler, 1998), hence, it is uncertain if there is an IS*629* hot-spot within the *gne* gene. While there are other enteric bacteria whose O antigen genes are structured similarly to those of O157:H7, where the *gne* gene lies upstream of the O antigen gene cluster (Liu et al., 2008; Senchenkova et al., 2005; Samuel et al., 2004), no O *rough* variants of these strains have been reported as a consequence of IS*629* insertion into *gne*.

These results suggest that the IS*629* insertion within the *gne* may have occurred as a result of a random mutation event and that IS insertions into *gne*, such as that which has occurred in MA6, are rare. The O antigen is not required for bacterial growth, but does confer protection from antibiotics, the mammalian complement system and other environmental and immunological responses, including serum-mediated killing, phagocytosis and killing by cationic peptides (Raetz and Whitfield, 2002). Consequently, the loss of the O antigen has been reported to make many pathogens serum sensitive or less virulent (Liu et al., 2008). For instance, mutants of *E. coli* serotypes O1, O7 and O18 that are devoid of O antigen were more sensitive to environmental factors and thought not survive to cause disease (Achtman and Pluschke, 1986; Pluschke et al., 1983). Similarly, *Shigella flexneri,* whose O antigen phenotype has been altered, has been found to be less virulent (Gemski et al., 1972). It has been postulated that in STEC, the lipopolysaccharide layer that contains the O antigen can stimulate the synthesis of Stx-

receptors on renal endothelial cells, thereby, promoting the progression of infection to HUS (Jackson et al., 1990). If so, we would expect O *rough*:H7 strains to be less pathogenic than its wild-type counterpart that has the O157 antigen because of its inability to resist the hosts' immune system and to induce Stx-receptor expression in renal endothelial cells. Consistent with those speculations, MA6 has only been isolated from a beef sample. Although, the virulence potential of MA6 remains undetermined, O *rough*:H7 CB7326 appears to be pathogenic since it was implicated in a HC case. Additionally, four other O *rough* STEC strains (O *rough*:H7:K1, O *rough*:H49:K-, O *rough*:H16, and O *rough*:H7) have been implicated in infections and cases of HUS (Keskimaki et al., 1999; Beutin et al., 1989; Bonnet et al., 1998; Vila et al., 1997).

Atypical O *rough* strains like MA6 were thought to be uncommon, still, the isolation of another O *rough*:H7 strain due to *gne*::IS*629* suggests that these strains may not be as rare as previously anticipated (Rump et al., 2010a). The insertion of IS*629* at two different *gne* locations may have been coincidental, or there may be multiple IS*629* insertion sites within *gne*. If so, however, it seems that such O *rough*:H7 strains would have been encountered more often. The preferred IS*629* insertion target sequence is unknown, so at this time, the existence of an IS*629* hot spot(s) within *gne* remains inconclusive. Although MA6 and CB7326 are not analogous strains, they are similar in that both are O *rough* due to *gne*::IS*629*. The fact that CB7326 was isolated from a hemorrhagic colitis patient also suggests that MA6 and other similar O *rough*:H7 strains, if found, may be pathogenic to humans. The presence of O *rough*:H7 strains, in food or clinical samples are of concern as they can not be detected serologically in the assays routinely used for fast detection of *E. coli* O157:H7. Consequently, should the occurrence of O *rough* variants of *E. coli* O157:H7 become more prevalent, they will create challenges to serological detection methods and could pose a potential health risk.

Highly pathogenic O157 and non-O157 Shiga-toxin producing *Escherichia coli* (STEC) (O26, O91, O103, O104, O111, O121, and O145) are emerging pathogens threatening the safety of our food supply. The reason for the emergence of those particular serotypes and more virulence genes carried in such serotypes than other serotypes are unknown. Also *E. coli* O26:H11 strain 11368, O111:H- strain 11128, O103:H2 strain 12009, as well as O104:H4 strain TY-2482 that caused a large outbreak

in Germany in 2011 were shown to carry various IS*629* copies (13 - 49) (Ogura et al., 2009). It is thereby worth mentioning, that O157:H- A4 strains (IS*629* deficient) were shown to be highly clonal and are geographically limited to Europe (Eklund et al., 2006). Changes in the genome related to IS*629* transposition thereby not only creates diversity, but also potentially increases the emergence of atypical, better adapted, and possibly more virulent strains such as the case of the new emerging O104:H4 strain.

6.2 Distribution of the IS*629* insertion element in *E. coli* O157:H7

IS elements are often considered selfish DNA sequences able to invade the genome of their host species and regarded as genetic factors that significantly contribute to genomic diversification and evolution (Ooka et al., 2009a). They have been shown previously to express species-specific distribution patterns, whereby *E. coli* strains showed similar distribution patterns to those of *Shigella* species (Matsutani and Ohtsubo, 1993). Furthermore, some IS elements might even be restricted to specific species (Matsutani and Ohtsubo, 1993). Their evolutive dynamics are complex, due to the interaction between their amplification capacity, selection at the host level, transposition regulation, and genetic drift (Le Rouzic et al., 2007). It was determined by Ooka et al. (2009) that IS elements IS*629* and IS*Ec8,* present in the *E. coli* O157:H7 lineage, serve as an important driving force behind genomic diversity. However, only a few genome-wide studies have been conducted to compare IS distributions in closely related genomes in order to determine to what extent IS elements actually contribute to genome diversification. In this work it has been determined that, IS*629* insertions in *E. coli* O157:H7 are widely distributed on the genome of this pathogen and differ significantly from strain to strain. Although ancestral *E. coli* O55:H7 strain carried only two IS*629*, with one on the genome and one on the pO55 plasmid, the four *E. coli* O157:H7 genomes still carried numerous IS*629*-like copies on the chromosome and corresponding pO157 plasmids.

IS*629* insertion sites in four *E. coli* O157:H7 genomes showed a predominant occurrence of IS*629* in prophages and prophage-like elements compared to the backbone structure. OOKA et al. (2009) determined that the genomic locations of other

types of IS elements in different *E. coli* O157 Sakai strains indicate that their distribution is also highly biased toward O157-specific regions, and thus this trend does not appear to be specific to IS*629*. This bias in distribution might be explained by the fact that the genome backbone carries essential genes, while genes located on phages are generally not essential for survival. IS*629* insertions in essential genes would be detrimental for the organism, whereby those on the phages might not affect their viability. Many ISs show specific target site selectivity, and their target site selection differs significantly from element to element (Mahillon and Chandler, 1998). Sequence-specific insertions are exhibited by several elements to some degree, though vary considerably in stringency. Other elements appear to exhibit regional preferences, which could reflect more global parameters, such as local DNA structure, which is less obvious to determine (Mahillon and Chandler, 1998). To identify the target specificity of an IS element, each individual insertion site has to be analyzed. IS*629* does not seem to specifically integrate in sequence-specific target sites, which explains the highly diverged flanking sites in the examined genomes. As mentioned above IS*629* has also been identified in the *gne* gene of two O *roug*h:H7 strains located in two different sites of the gene, indicating that this specific region could serve as a “hot-spot” for IS*629* insertions, even though IS*629* site specificity remains unknown. Although no specific “hot-spot” was observed for IS*629* insertions, the apparent preferred location in phages or phage-like elements suggest that these mobile elements could have functioned as vehicles for IS*629* introduction into *E. coli* O157:H7 genomes.

6.3 Prevalence of the IS*629* insertion element in the stepwise emergence of *E. coli* O157:H7

Comparison of IS*629*s found in *E. coli* O157:H7 and O55:H7 showed extensive divergence between these elements. At least three different IS*629* subtypes could be distinguished differing from 55 to 60 nt. *E. coli* O157:H7 strains carry IS*629*-like elements subtype I and III, whereby O55:H7 carries subtype II (Rump et al., 2011a). It is notable that just 4 nucleotide differences existed between A1 CC strain DEC5A and A6 CC strain Sakai among seven housekeeping genes sequences, comprising a current MLST scheme (http://www.shigatox.net/ecmlst/cgi-bin/index). These two strains, in particular,

are taken to represent the most ancestral and most derived *E. coli*, respectively, in the stepwise evolutionary model for this pathogen. If the IS*629* subsubtype I and III observed in A6 CC strains resulted from divergent evolution of IS*629* subsubtype II, the amount of changes observed among these IS subsubtypes should be similar to those observed for the MLST loci examined above. However, the number of nucleotide substitutions between IS*629* subsubtype I and III in *E. coli* O157:H7 from subsubtype II in O55:H7 was 10-fold higher. Thus, the differences between IS*629* subtypes are more significant than those observed for housekeeping genes. This indicates that IS*629*-subsubtype II was most likely lost and IS*629*-subsubtype I and III were acquired independently in distinct *E. coli* O157:H7 lineages. Further supporting this thesis was the fact that potentially functional IS*629* subsubtype II copy was found on the pO55 plasmid, which was subsequently lost during evolution towards O157:H7 strains. The other IS*629* copy in *E. coli* O55:H7, with a unique internal deletion (rendering a potentially non-functional IS*629*), is located in the genome. This truncated IS*629* appears to be part of a mobile region (Brunder et al., 2006) which is absent in *E. coli* O157:H7 strains.

Contrary to what was observed for the O157:H7 strains examined, none of the A4 CC strains, which are on a divergent evolutionary pathway, carried IS*629*. A reason for the absence of IS*629* in the A4 clonal complex could have been that the mobile vehicle carrying IS*629* was introduced but then IS*629* was incapable of transposing into the genome and further propagating as observed in A5 and A6 strains. However, since IS*629* was shown to transpose effectively in A4 strains, it is rather likely that the introduction of the IS*629* carrying mobile element never occurred in those strains. Interestingly, the ancestral IS*629*-deficient A2 O55:H7 strain 3256-97 is also lacking IS*629* and both IS*629* associated regions found in O55:H7 strains. Either IS*629* has been lost in strain 3256-97 as it seems for the hypothetical A3 precursor, or this strain never carried IS*629*. These results may indicate that a similar strain, lacking IS*629*, might have given rise to IS*629*-deficient A4 CC strains.

Since IS*629* is preferably located on prophages or prophage-like elements, IS*629* could have been introduced through phages in the O157:H7 strains. Therefore, the absence of IS*629* in the A4 CC strains might be explained through phage immunity, resulting in the exclusion of the phage carrying IS*629*. Consequently, the absence and presence of specific prophages and prophage-like elements in the different strains

belonging to each complex could provide an indication about the origin of IS*629* in the genomes of O157:H7 strains. The phage responsible for IS*629* type I and III introduction in SOR- O157:H7 strains must be present in A5 strains but absent in A4 strains. Elements exclusively present in A6 clonal strains, were probably acquired more recently, hence, those can be excluded from being a possible vehicle for IS*629* element, since both IS*629* types were shown to be already present in A5 strains. Additionally, prophages (Sp) and phage-like elements (SpLE) present in A4 strains were most likely present before IS*629* acquisition. There are 24 known Sp and SpLE in the genome of Sakai (Hayashi et al., 2001). Fourteen of those are found to harbor IS*629*. Only SpLE1 (IS*629* type III) and SpLE5 (IS*629* type I) were found to be unique to A5 and A6 CC strains and were absent in A4, A2 and A1 CC strains. SpLE1 can also be found in other *E. coli* serotypes carrying IS*629*, which makes this phage a potential IS*629* vehicle for their introduction into A5 strains. SpLE1, a CP4-like element, was shown to carry multiple complete and truncated IS*629*-type III copies in O157:H7, as well as in O26:H11 and O111:H- strains. The complete IS*629* found on this specific element, additionally inserted in different sites on SpLE1. Therefore, it seems likely that this phage either carries multiple “hot-spots” or IS*629* inserts specifically on this phage due to regional features that it exhibits. Conclusively, this specific phage-like element appears to be a possible vehicle for the IS*629* type III introduction into the genomes of O157:H7 strains. SpLE5 integrated in tandem with SpLE6 (Hayashi et al., 2001), however, BLAST analysis did not show any similarity with known phage-like elements in other serotypes. In this study, only IS*629* carrying phages were examined for their presence in related strains. Since IS*629* appears to use both replicative and non-replicative pathway of transposition (Ooka et al., 2009a; Kusumoto et al., 2004), it is likely that the mobile element initially carrying IS*629* does not harbor this element any longer in A6 CC strains. Although it seems possible that SpLE1 served as the vehicle, other elements also could have introduced IS*629* into the O157 lineage, however, not carrying IS*629* any longer. Lastly, it is intriguing to speculate which phage or phage-like element introduced IS*629* into O157 even though, it remains unknown which phage initially introduced IS*629* into the genome of O157:H7 strains.

Additionally, A6 CC strains carry pO157 which harbors 3 to 4 IS*629*, contrary to pSFO157 which lacks IS*629*. Plasmids are known to carry multiple IS elements and

transposons, making them a possible IS*629*s carrier (Mahillon and Chandler, 1998). Since pO157 carries multiple IS*629*, this plasmid could also have reintroduced IS*629* into the genome of O157:H7 strains. IS*629* seems to transpose actively in O157:H7 strains (Ooka et al., 2009a), and therefore, the original source which could have carried IS*629* remains difficult to identify.

6.4 Evolutionary significance of the IS*629* insertion element in the stepwise emergence of *E. coli* O157:H7

IS*629* insertion sites located on the backbone seem to be conserved in almost all of the strains studied, whereby sites located on phages and phage-like areas appear to differ between all strains. These findings confirm the presence of regions of genomic stability and regions of genomic variability existing within O157:H7 populations and closely related strains. It is noteworthy that sites associated with phages seem to be present predominantly in closely related strains. The majority of the phages present in the A6 CC strains appear to be unique to this complex. Since bacteriophages are known to contribute to the diversification of bacteria (Ohnishi et al., 2001), they seem to be a major determinant in generating diversity among O55:H7, O157:H- and O157:H7 strains.

The comparison of IS*629* prevalence in O157:H7 strains as well as IS*629* insertion sites allowed distinguishing strains similarly to what was observed by PFGE and by previous studies (Feng et al., 2007), into two clonal complexes A5 and A6. Adding the “same” strain from different collections, Sakai and EDL933, allowed the confirmation of IS*629* stability between closely related strains. Minimal changes in IS*629* presence / absence were observed and could have occurred due to different storage conditions and passages. On the parsimony tree strains grouped tightly together despite those subtle changes, therefore, this analysis can be used to further distinguish closely related O157:H7 strains. However, in order to distinguish strains not harboring IS*629*, as for A4 CC strains, it was crucial to include the presence / absence of specific IS*629* sites on the backbone, phages and plasmids in the analysis. This comparison allowed distinguishing all strains as it has been proposed in the evolutionary model for O157:H7

into different clonal complexes (Feng et al., 2007). Strains lacking IS*629* however were only grouped into each CC, but could not further be distinguished. Those strains lacked IS*629* insertions, which were used to differentiate A5 and A6 CC strains as well as multiple sites located on phages. The backbone structure appeared to be highly conserved among strains investigated and therefore only allowed distinguishing A1, and A2 CC strains from strains of the other clonal complexes due to one missing region. Sites on phages differed significantly between all the strains but only slightly between strains of the same CC. Consequently, this analysis grouped strains of the same CC together but cannot be used to further distinguish strains of CC A4, A2, and A1. Conclusively, this analysis could successfully differentiate closely related strains using mainly the presence of IS*629* insertion sites as well as the presence of specific phages. This finding shows, that bacteriophages seem to be a major determinant in generating diversity among O55:H7, O157:H- and O157:H7 strains allowing to group the strains in different CC of the evolutionary model of O157:H7. However the presence / absence of the insertion sequence IS*629* enables us to further distinguish highly pathogenic O157:H7 strains and determine their relatedness amongst each other.

Therefore, phages seem to play a major role in the diversification of *E. coli* genomes, whereby IS*629* significantly contributes to the diversification of closely related *E. coli* O157:H7 genomes. It also has been shown that IS elements play an important role in immobilizing newly incoming phages and plasmids upon insertion, causing defective phages, leading to their fixation in the O157 genome (Ooka et al., 2009a). Phages often encode virulence genes and therefore, contribute to pathogenicity of the host. The prophages and lambda-like elements found in O157 Sakai, contain various types of deletions and / or insertions in phage-essential regions and appear to be defective (Hayashi et al., 2001; Makino et al., 1999; Ohnishi et al., 2001). Those encode a variety of virulence-related proteins such as Stx, zinc / copper-type superoxide dismutase (SOD), TrcA-homologs, Bor proteins and many LOM homologs, which have been implicated in host serum resistance and cell adhesion, respectively (Hayashi et al., 2001; Vica et al., 1997; Barondess and Beckwith, 1995). tRNA genes (*ileZ, argN* and *argO*), which might contribute to the efficient expression of many O157-specific genes, can also be identified at O157-specific sites (Hayashi et al., 2001; Plunkett et al., 1999). This suggests that IS insertions as well as IS-mediated deletions causing immobilization

of phages or other mobile elements carrying virulence factors, could generate various phenotypic differences among O157 strains, including differences in virulence potential. Since IS*629* is the most prevalent IS element in O157:H7 strains, and appears to transpose actively in O157:H7 strains (Ooka et al., 2009a), it seems highly likely that IS*629* also contributes to the fixation of incoming phages in their genomes, indirectly contributing to their pathogenicity.

E. coli O157:H7 strains carry multiple IS*629* copies, while non-pathogenic K-12 strains lack IS*629* but carry other IS elements. Other pathogenic *E. coli* strains, amongst the top six non-O157 STEC O26:H11, O111:H- and O103:H2 (Fey et al., 2000), also harbor various copies of IS*629*-like elements in their genomes. Genome sequences for the other three most important pathogenic non-O157 STEC; O45, O145, and O121 are not available to date, thus, the presence of IS*629*-like elements in these is unknown. Interestingly, they also share the same reservoir with O157:H7 (*e.g.* cattle), shiga-toxins, the haemolysin gene cluster, other virulence factors and several phages and phage-like elements (Fey et al., 2000). OOKA et al. (2009) postulated that IS-related genomic rearrangements may have significantly altered virulence and other phenotypes in O157 strains. Therefore, IS*629* might not only have a great impact in O157:H7 genomic evolution but might also play an important role in its increase in pathogenicity. Moreover, the highly prevalent insertion sequence IS*629* might contribute to an increase in pathogenicity not only in O157:H7 but also in other non-O157 STEC strains.

6.5 Strains of *E. coli* O157:H7 stepwise evolutionary model exhibit different IS*629* transposition frequencies

To determine if the absence of IS*629* in A4 clonal complex strains is due to IS*629* transposition immunity or a similar mechanism, vector pIS*629*-AB-Tc was constructed and introduced in A2, A4, A5 and A6 strains. It was shown that IS*629*-Tc transposed in all tested strains, although with notably different frequencies. The absence of IS*629* in the A4 CC strains could have been due to an IS*629* transposition inhibition mechanism, whereby the mobile vehicle carrying IS*629* was introduced, but then IS*629* was incapable to transpose into the genome and further propagate like observed in A5 and

A6 strains. IS*629*-Tc successfully transposed in A4 strains therefore the absence of IS*629* in the A4 CC strains does not appear to be due to an IS*629* transposition inhibition mechanism. It is rather likely that a precursor strain lost IS*629* and did not get in contact with an IS*629* carrying mobile element after diverging from the hypothetical A3 CC.

IS*629*-Tc excision frequencies in A5 and A6 CC strains were higher than in A1, A2, and A4 CC strains. These findings agree with previous results from Kusumoto et al. (2004) which noted that IS*629*-carrying strains have a higher IS*629* transposition frequency than IS*629*-deficient strains. The low excision frequency determined for IS*629*-deficient strain 3256-97 (1.5×10^{-8}) was similar to that of other strains lacking IS*629* of various serotypes and genotypes (Kusumoto et al., 2004). Strains from A1 and A2 CC (DEC5A and 3256-97), exhibited a low IS*629* excision frequency, regardless of IS*629* presence or absence. A4 CC strains exhibited a higher excision frequency (100-fold) compared to the tested A1 and A2 CC strains although remaining lower than in A5 and A6 CC strain. This intermediate excision frequency suggests that the presence of IS*629* alone might not enhance transposition activity of IS*629*-Tc. Kusumoto et al. postulates that IS*629*-possessing strains use a currently unknown "system to enhance IS*629* excision which might have been introduced by mobile genetic elements that may be linked with IS*629* or other IS elements". The nature of this mechanism was recently described by Kusumoto et al. (2011) as a protein IS-excision enhancer (IEE) which promotes IS*629* excision in O157. The elevated excision frequency in all A4 CC strains relative to A2 and A1 CC strains could indicate that these strains possess a similar up-regulation mechanism as in A6 CC strains, while A1 and A2 CC strains might lack IEE. Le Rouzic et al. (2007) described that if an IS element has a constant transposition rate, it will disappear in most cases. Elements with low-transposition rates are frequently lost through genetic drift, while those with high-transposition rates may amplify, leading to sterility of their host. However, elements whose transposition rate is regulated are able to successfully invade populations, by an initial transposition burst, followed by a strong limitation of their own activity (Le Rouzic et al., 2007). In the case of IS*629*, the transposition is programmed by translational frameshifting (Chen CC and Hu ST, 2006). Thereby frameshifting frequency is critical in determining the overall transposition activity. Although it has not yet been explored in detail, frameshifting could be influenced by host physiology thus coupling transposition activity to the state of the host cell

(Mahillon and Chandler, 1998). Translational control of transposition by frameshifting has been demonstrated for many members of the IS*3* family (Polard et al., 1992), but might also occur in other elements. The frequency of frameshifting varies from element to element. It is approximately 50% for IS*150* (Vogele et al., 1991) and only 15% for IS*911* (Polard et al., 1991). Transposition activity is generally maintained at a low level. An often mentioned reason for this is that high activities and the accompanying mutagenic effect of genome rearrangements would be detrimental to the host cell (Doolittle et al., 1984). Transposition activity for various IS elements is frequently modulated by various host factors which are generally specific for each element (Mahillon and Chandler, 1998). Those factors include DNA chaperons (or histone-like proteins) integration host factors, replication initiator DnaA, protein chaperone / proteases ClpX, ClpP, and ClpA, SOS control protein LexA, and Dam DNA methylases (Mahillon and Chandler, 1998). In addition, proteins which govern DNA supercoiling in the cell might influence transposition. *E. coli* O157:H7 strains carry the common factors like DnaA, ClpX, ClpP, and ClpA, LexA which are known to positively influence transposition of IS elements (Mahillon and Chandler, 1998). The ancestral O55:H7 strains might lack the transposition up-regulation system IEE which explains why only a single complete IS*629* could be found in the genome. The O157:H7 strains carry an increased quantity of IS*629* in their genomes compared to the O55:H7 strain which highlights an obvious increase of IS*629* transposition promoted by IEE.

The elevated IS*629* transposition frequency observed among O157:H7 strains might explain the highly diverse distribution of IS*629* in O157:H7 genomes (Yokoyama et al., 2010). Ooka et al. (2009) postulated that IS-related genomic rearrangements may have significantly altered virulence and other phenotypes in O157 strains, which suggests that IS*629* might not only have a great impact in their genomic evolution, but possibly increase the pathogenicity of those strains. Other pathogenic *E. coli* strains, amongst the top six non-O157 STEC O26:H11, O111:H-, O103:H2, and O104:H4 strain TY-2482 also harbor various copies of IS*629*-like elements in their genomes and were also found to carry IEE (Kusumoto et al., 2011). Highly clonal O157:H- A4 strains lacking both IS*629* (Rump et al., 2011a) and IEE were shown to be highly clonal (Eklund et al., 2006). The possession of IS*629* together with IEE appears to generate diversity in O157:H7 strains most likely due to phage fixations, genome rearrangements and / or

other effects caused by IS*629* transposition (Rump et al., 2011b; Rump et al., 2011a). It might show that IS*629* could also effectively contribute to the emergence of pathogenic strains that new pose a risk for food supply. These events not only can create diversity, but also potentially increase the emergence of better adapted and possibly more virulent strains such as the case of the new emerging O104:H4 strain. Hence, IS*629* might contribute to the appearance of atypical pathogenic strains like O *rough*:H7 (IS*629*::*gne* mutant) (Rump et al., 2010b; Rump et al., 2010a). Consequently, should the rate of those atypical pathogenic STECs become more prevalent, they will create challenges to serological detection methods and could pose a potential health risk.

6.6 Plasmids in closely related strains of the *E. coli* O157:H7 stepwise evolutionary model

Virulence plasmids pO157 and pSFO157 as well as pO55, carry multiple IS elements and transposons. However, only pO157 carries several copies of IS*629* which makes this plasmid a possible carrier for introducing IS*629* into the genomes of O157:H7 strains. Genomes and plasmids naturally undergo constant rearrangements and reconfigurations and thereby it is not surprising that also closely related O157:H7 strains show differences in their pO157 plasmids. Strains EC4115 and TW14359 both carry an additional IS*629* type III (IS.42) in their corresponding pO157 plasmids. The other three IS*629* located on the plasmids present in all four strains are however either truncated (IS.41, IS.43) or belong to IS*629* type I (IS.40/44/48). Ooka et al. (2009) showed that IS*629* likely generates footprints upon excision or transposition which could not be identified for pO157 of Sakai and EDL933 in the IS.42 site. This suggests that the IS*629* found in those two strains probably transposed from the genome to the plasmid after the initial IS*629* type III introduction into the genome of O157:H7 strains. Interestingly, however, all A5 strains appear to lack this region which is regarded as the debris of an ancient transposon (Brunder et al., 1999). The IS.40/44/48 site could be detected, in all A6 CC strains as well as in all A4 CC strains, but was absent in all A5 CC strains. However, the IS.43 site containing a truncated IS*629* was found in A5 strains, although ~200 bp smaller. A5 CC strains are thought to carry pO157 and were positively tested for

the plasmid encoded *ehxA* for enterohemolysin (Feng et al., 2007). It is surprising that the A5 CC strains were found to lack several pO157 specific regions, and neither pO55 nor pSFO157 specific regions were detected. This could be an indication that the A5 CC strains have a plasmid other than pO157 or pSFO157, however it was generally assumed that A5 CC wich gave rise to A6 CC strains carry pO157 (Wick et al., 2005). It is possible that A5 CC strains carry a precursor plasmid similar to pO157 present in the hypothetical A3 CC strains. Preliminary sequencing data showed indeed the presence of a plasmid other than pO157 or pSFO157 (designated pG5101). Additionally, strain LSU-61, which was thought to be a strain most closely related to the A3 CC intermediate strain (Feng et al., 2007), lacks most regions found in pO157, however carries a site unique for pSFO15, hence probably carrying plasmid pSFO157. This strain, in contrary to other pSFO157 possessing A4 CC strains, does show hemolysin activity (Feng et al., 2007). The maximum parsimony tree obtained, using IS*629* target sites for the 27 strains analyzed in the present study, clustered strain LSU-61 closely to the A4 CC strains although being of serotype O157:H7. These findings could indicate that LSU-61 might be closely related to CC A4 strains, however being motile (H7 antigen) and hemolysin positive. LSU-61 had many of the traits proposed for the A3 intermediate; however its exact position on the evolutionary model remains to be determined. Further whole genome sequencing of this particular strain could help in understanding the definite positioning of LSU-61 in the evolutionary model for O157:H7, and reveal if this strain could possibly be the strain most closely related to the “thought to be extinct” intermediate A3 CC strain.

Ancestral O55:H7 strains were shown to possess plasmid pO55, which was presumably lost during emergence towards O157:H7 (Zhou et al., 2010). Wick et al. (2005) suggested that the virulence plasmid pO157 was gained leading to the hypothetical A3 CC strain. A4 CC strains however carry pSFO157 which is thought to be in evolutionary terms, older that pO157 (Brunder et al., 2006). The O55:H7 strain 3256-97 of CC A2 which did not harbor any IS*629* neither in the genome nor in the corresponding plasmid, also lacked the pO55 specific site used to determine the presence of pO55. Plasmid pO55 carries one complete IS*629* type II, which is naturally absent in this strain. In the maximum parsimony tree obtained, strain 3256-97 clustered in the A1/A2 CC group. This result is however based on minimal differences between the

strains. The apparent absence of pO55 in this strain could yet show that ancestral strains could have lost plasmid pO55 (or never possessed pO55) and the ancestral pO157/pSFO157 plasmid was acquired by the A3 intermediate. Brunder et al. (2006) suggested that both plasmids pO157 and pSFO157, evolved from a common ancestor plasmid present in A3 by acquisition and loss of various genes and involving, in particular, insertion/transposition events and possible plasmid fusions. During this process, IS*629* elements could have exclusively been gained by pO157, but not by pSFO157. This furthermore implies that the ancestral pO157/pSFO157 plasmid was also lacking insertion sequence IS*629*. A5 CC strains were shown to already carry IS*629* type I and III, but seem to lack the locations and IS*629* insertions on their corresponding plasmid (designated pO157/GUD). Accordingly, the IS*629* insertions found on plasmid pO157 might have been introduced to this plasmid after initial IS*629* introduction into the genomes of O157:H7 strains and indeed, phages or phage-like elements were most likely the responsible mobile vehicles introducing IS*629* into the O157:H7 clonal group. To date neither genome sequences nor plasmid sequences of A5 CC strains are available, which would allow further investigating these findings. Thus, it remains undetermined whether pO157/GUD carries IS*629* or not.

7 SUMMARY / ZUSAMMENFASSUNG

7.1 Summary

The main objective of this research was to determine the cause for the lack of O antigen expression in a rare *E. coli* O *rough*:H7 mutant, MA6, which is a stx2 producing O157:H7 strain also carrying other virulence factors. A 1,310 bp insertion, homologous to insertion sequence IS*629*, was observed within its *gne* gene rendering it inactive, encoding an epimerase enzyme essential for the synthesis of an oligosaccharide subunit in the O157 antigen. Although this O *rough*:H7 phenotype due to *gne*::IS*629* was thought to be uncommon, an O *rough*:H7 strain caused by *gne*::IS*629* was isolated from a hemorrhagic colitis patient, indicating that these strains are pathogenic and may not be as rare as anticipated. Interestingly, IS*629* has been found in multiple copies in the *E. coli* O157:H7 genome and plasmids and is one of the most prevalent IS in this serotype. Contrary to what was observed for O157:H7 strains, sorbitol fermenting O157:H- (SFO157) strains which are on a divergent evolutionary pathway, entirely lack IS*629*. Comparison of IS*629*s found in O157:H7 and O55:H7 showed the presence of at least three different IS*629* types: O157:H7 strains carry IS*629*-like elements type I and III whereby the ancestral O55:H7 carries type II. Numerous non-uniformly distributed IS*629* insertion sites with no obvious “hot-spot” were found in four *E. coli* O157:H7 genomes and plasmids. However a predominant occurrence of IS*629* in prophages and prophage-like elements was observed compared to the backbone structure, suggesting that those elements could serve as IS*629* carriers. The comparison of IS*629* prevalence in A5 and A6 CC strains, as well as the IS*629* insertion site prevalence in all other CC strains, allowed distinguishing strains from different complexes, as it has been proposed in the stepwise evolutionary model for O157:H7. Strain LSU-61, a motile, hemolysin positive O157:H7 strain grouped with A4 CC strains and seems to carry pSFO157, found usually in non-motile, hemolysin negative A4 CC strains. A5 CC strains, which are thought to have given rise to A6 CC strains, were found to carry a plasmid other than pO157 or pSFO157 (designated pO157/GUD). Hence, the prevalence of plasmids which were found to change

continuously among related strains might also help to understand relatedness and evolutionary changes amongst strains.

In agreement with previous findings of whole genome analysis O55:H7 and O157:H7 strains showed high conservation of backbone structure. The location of IS*629* insertion sites on the backbone structure in these strains were also found to be highly conserved. Furthermore, IS*629* insertion sites located on bacteriophages were non-uniformally distributed. This reinforces the observation by numerous researches that bacteriophages are a major determinant in generating diversity among O55:H7, O157:H- and O157:H7 strains. Hence, phage exclusion appears to be the most likely explanation for the IS*629* absence in SFO157 strains, since a possible IS*629* transposition inhibition mechanism was confirmed to be absent in those strains. An IS*629*-analogue actively transposed in SFO157 strains with an elevated frequency compared to ancestral O55:H7 strains, indicating the presence of an “up-regulating system” similar to that found in O157:H7. NSFO157 strains exhibited a significantly higher excision frequency than all the other strains which is due to the presence of IEE, enhancing IS*629* transposition exclusively in the O157:H7 lineage. The elevated IS*629* transposition frequency observed among O157:H7 strains might explain the highly diverse distribution of IS*629* across O157:H7 genomes. Also other non-O157 Shiga-toxin producing *E. coli* (STEC) carry various IS*629* copies and IEE in their genomes. It thereby appears that the possession of IS*629* together with IEE likely plays a significant role in genome plasticity and genetic diversity and might participate in pathogenicity increase of STECs and in the apparition of atypical and/or new pathotypes as was observed for the recent EAHEC O104:H4 strains.

7.2 Zusammenfassung

Das Hauptziel dieser Arbeit war, den Grund des nicht Vorhandenseins des O157 Antigens in MA6 zu ermitteln. Bei MA6 handelt es sich um eine seltene O *rough*:H7 Mutante, die genetisch ein Stamm des hochpathogenen Serotypen O157:H7 ist. Es wurde festgestellt, dass in MA6 die Funktion des *gne*-Gens durch die Einfügung der Insertionssequenz IS*629* in das offene Leseraster einschränkt ist. Das *gne* Gen kodiert ein Epimerase-Enzym das für die O157 Antigen Biosynthese essenziell ist. Demzufolge kann eine der 4 Oligosaccharid Untereinheiten des O157 Antigens in MA6 nicht synthetisiert werden. Obwohl O *rough*:H7 Mutanten aufgrund von *gne*::IS*629* Insertionen als selten gelten, wurde zwischen 1997-1999 ein ähnlicher Stamm von einem HC-Patienten in Deutschland isoliert. Die *gne*::IS*629* Insertion befindet sich in diesem Stamm jedoch an einer anderen Stelle als in MA6. Hierbei handelt es sich daher wahrscheinlich um eine individuelle IS*629* Insertion. Das lässt den Schluss zu, dass dieser Serotyp häufiger vorkommen könnte, als zuvor angenommen. Interessanterweise trägt *E. coli* O157:H7 zahlreiche IS Kopien, wobei IS*629* das häufigste IS in diesem Serotyp ist. IS*629* konnte außerdem in den Vorfahren (Serotyp O55:H7) nachgewiesen werden, jedoch in geringerer Anzahl. Im Gegensatz dazu, wurde IS*629* in O157:H- Stämmen nicht nachgewiesen. O157:H- sind nah verwandt zu O157:H7; befinden sich jedoch auf einem abweichenden evolutionären Weg.

Der Vergleich verschiedener IS*629* in O55:H7 und O157:H7 zeigte die Anwesenheit von mindestens drei verschiedenen IS*629* Subtypen: die O157:H7 Stämme tragen IS*629* Subtyp I und III; und der O55:H7 Stamm trägt IS*629* Subtyp II. Da IS*629* in O157:H- nicht vorkommt, scheint es wahrscheinlich, dass im Laufe der Entwicklung von O55:H7 zu O157:H7 der Subtyp II verloren gegangen und die Subtypen I und III ausschließlich in O157:H7 eingeführt wurden.

In 4 *E. coli* O157:H7 wurden zahlreiche, ungleichmäßig verteilte IS*629* Insertionsstellen gefunden. Dadurch war es nicht möglich einen IS*629* „Hot-Spot“ zu ermitteln. Allerdings wurde ein vermehrtes Auftreten von IS*629* in Prophagen und Prophagen-ähnlichen Elementen - im Vergleich zu der als in *E. coli* konserviert angesehenen „Backbone“-Struktur - beobachtet. Diese Beobachtung könnte unter anderem drauf hindeuten, dass Prophagen und Prophagen-ähnliche-Elemente als

IS*629* Träger dienen könnten. Ausserdem sind „Backbone“-Struktur Insertionen möglichwerweise seltener, da diese Struktur essenzielle Gene trägt.

Um das Vorhandensein und die Verteilung von IS*629* zu ermitteln, wurden für die Studie 27 Stämme verschiedener Serotypen (O157:H7, O157:H-, O55:H7) aus den einzelnen klonalen Komplexen des „evolutionary model” (Evolutionsmodells) für O157:H7 ausgewählt. Wie zuvor schon angenommen, war die „Backbone“-Struktur in allen Stämmen hoch konserviert. Die IS*629*-Seiten in Bakteriophagen und Phagen-ähnlichen Elementen hingegen, waren nicht uniform. Diese Vektoren stellen daher eine wichtige Determinante bei der Erzeugung der Diversität zwischen den Stämmen dar. Der Vergleich des Vorhandenseins der IS*629* Insertionen und den spezifischen IS*629* Insertionsstellen in den IS*629* tragenden O157:H7 Stämmen, ermöglichte die Gruppierung der Stämme in die jeweiligen klonalen Komplexe (A5 und A6) des Evolutionsmodells für O157. Außerdem konnten alle anderen in dieser Studie verwendeten Stämme in die einzelnen klonalen Komplexe gruppiert werden, wenn ausschließlich die Information über das Vorhandensein der IS*629* Insertionsstellen verwendet wurde. Überraschenderweise wurde der Stamm LSU-61 dabei zu den O157:H- (A4) Stämmen zugeordnet, obwohl es sich bei diesem Stamm um einen begeisselten, Hemolysin-positiven O157:H7 Stamm handelt. Außerdem scheint es, dass LSU-61 anstelle des für *E. coli* O157:H7 typischen Plasmids pO157, des in den A4 Stämmen vorhandenen Plasmid pSFO157 trägt. Die A5 Stämme, welche als Vorfahren der A6 Stämme angenommen werden, scheinen ein anderes Plasmid als pO157 oder pSFO157 zu tragen (namentlich pG5101). Infolgedessen könnten anhand des Vorkommens verschiedener Plasmide tiefere Einblicke in die evolutionäre Veränderung zwischen den Stämmen gewährt werden, da sich sogar die nahe verwandten Stämme kontinuierlich zu verändern scheinen.

Aufgrund der Beobachtungen, dass sich IS*629* bevorzugt auf Bakteriophagen oder Phagen-ähnlichen Elementen befindet und dass diese Vektoren hauptsächlich zu der Differenzierung nahe verwandter Stämme beitragen, ist das Fehlen von IS*629* in den O157:H- (A4) Stämmen, mit dem Ausschluss eines IS*629*-tragenden Vektors zu erklären. Da sich IS*629* in O157:H- Stämmen erfolgreich mobilisierte, konnte das Vorhandensein eines IS*629* Transpositionsinhibitions-Mechanismus ausgeschlossen werden. Erstaunlicherweise wiesen O157:H- Stämme eine höhere IS*629*-Transpositionsfrequenz auf, als es in den O55:H7 Stämmen beobachtet wurde.

Die Transpositionsfrequenz ist jedoch geringer als in O157:H7 Stämmen. O157:H7 Stämme wiesen die höchste IS*629*-Tranpostionsfrequenz auf. Dies ist auf das Vorhandensein eines „IS-excision enhancer (IEE)" zurückzuführen. IEE ist in O157:H- nicht nachweisbar, wodurch es daher möglich ist, dass die erhöhte Häufigkeit der IS*629*-Mobilisierung in den A4 klonalen Stämmen auf das Vorhandensein eines anderen „Hochregulierungssystems" zurückzuführen ist.

Die erhöhte IS*629*-Mobilisierung in den O157:H7 Stämmen aufgrund des Vorhandenseins von IEE, könnte daher die zahlreichen und ungleichmäßig verteilten IS*629*-Insertionsseiten in diesem Serotypen erklären. Dies könnte unter anderem bei der Entstehung von atypischen Stämmen, wie den beiden O *rough*:H7 Mutanten, beitragen. Es ist also möglich, dass IS*629* zusammen mit IEE eine signifikante Rolle in der Genom-Plastizität und genetischen Vielfalt spielt. Unter Umständen könnten diese beiden Elemente einen wichtigen Beitrag zu der Erhöhung der Pathogenität von O157:H7 leisten. Da auch andere STEC Serotypen wie O26:H11, O111:H-, O103:H2 und O104:H4 sowohl IS*629* als auch IEE in ihren Genomen tragen, erscheint es gut möglich, dass diese Kombination eine entscheidende Rolle in der Entstehung von atypischen und/oder neuen Pathogenen, wie dem neuartigen EAHEC O104:H4, spielen könnte.

7.3 Prospect

Current findings of the pathogenic potential of atypical O *rough*:H7 strains, which are due to IS*629*::*gne* are of concern, as those strains can not be detected serologically in assays routinely used to test for O157:H7. Since IS*629* actively transposes in the O157:H7 clonal group, the occurrence of O *rough* variants of O157:H7 and other pathogenic non-O157 STEC could become more prevalent, and create challenges to serological detection methods posing a potential health risk. Therefore, it would be of great interest to further investigate the prevalence of O *rough*:H7 and O *rough* STEC strains and determine if IS*629*::*gne* and other IS*629* instertions in genes involved in their O antigen synthesis are becoming more prevalent. Moreover, other detection methods that go beyond targeting the O antigens of pathogenic STECs, should be employed into routine analysis, in order to detect those atypical O *rough* strains.

Two different IS*629* types (IS*629* type I and III) have been detected in multiple copies in NSFO157 strains, but were absent in SFO157 strains. It remains unclear how

the O157:H7 clonal group gained two different IS*629* types, but it appears highly likely that phages were involved in the introduction. Highly valuable information could be gained by sequencing and comparing the genomes of A6, A5 and A4 strains. The comparison of phages and phage-like elements present in A5, A6 compared to A4 strains could help identify the IS*629* carrier. Since plasmids are also known to contribute in introducing transposons, ISs and other mobile elements to genomes, pG5101 found in A5 CC strains, could also have participated in the introduction of IS*629*. It is furthermore unclear what plasmid A5 CC strains possess and whether it harbors IS*629*. Sequencing of the corresponding plasmids would further enable a determination of whether pG5101 also contributed to IS*629* introduction in the O157:H7 clonal group.

The plasmid analysis gave unexpected results and showed that the ancestral A2 strain 3256-97 presumably lacks pO55, LSU-61 probably carries pSFO157, contrary to what was expected for an O157:H7 strain, and A5 CC strains appear to harbor a plasmid other than pO157 (pO157/GUD). Sequencing the corresponding plasmids of those particular strains, together with whole genome sequencing, would allow a better understanding of the evolutionary changes that happened during the emergence towards highly pathogenic O157:H7 lineage from its O55:H7 ancesor. Furthermore, should it be confirmed that strain LSU-61 carries pSFO157 it would be the first strain possessing pSFO157 but showing hemolysin activity. Thus far, the reason for the hemolysin negative phenotype of the A4 CC strains is unknown. The comparison of hemolysin positive pSFO157 with the pSFO157/LSU could allow potentially for determination of the origin of this hemolysin negative phenotype.

8 ATTACHMENTS

8.1 Materials

8.1.1 Equipment

Equipment	Model/Name	Manufacturer
Autoclave	n.s	Consolidated
Centrifuge 1 (Micro, refrigerated)	Model 5417R	Eppendorf North America, Hauppauge, NY, USA
Centrifuge 2 (Micro)	Model 5415	Eppendorf North America, Hauppauge, NY, USA
Centrifuge 3 (refrigerated)	Marathon 21000R	Fisher Scientific, Pittsburgh, PA, USA
Dry Bath	Dri-Bath Type 16500	Thermolyne, Thermo Scientific Inc., Waltham, MA, USA
Electroporation System	Gene Pulser II	Bio-Rad, Hercules CA, USA
Gel Electrophoresis Chamber (Vertical)	Mini-Protean®	Bio-Rad, Hercules CA, USA
Gel Electrophoresis Chamber (Horizontal)	MP-1015 Multipurpose	IBI, Peosta, IA, USA
Gel Electrophoresis Power Supply	n.s	LKB Bromma 2197,
Incubator (37ºC)	Isotemp Incubator	Fisher Scientific, Pittsburgh, PA, USA
Incubator Shaker	Innova 40	New Brunswick Scientific., Edison, NJ, USA

Microwave	Turntable Cooking	Emerson Radio Corp., Parsippany, NJ, USA
Nucleic Acid and Protein Quantitation	Micro-Volume UV-Vis Spectrophotometer for	Wilmington, DE, USA
pH-Meter	250 pH/Temp/mV Meter	Beckman Coulter
Scale 1	Sartorius analytic	Goettingen, Germany
Scale 2	Sartorius 1219MP	Goettingen, Germany
Sealer	n.s	Cleveland ,Equipment and Machinery Company, Memphis, TN, USA
Shaker Incubator (RT)		
Speed-Vac	SPD1010 Speed Vac System	Thermo Savant, Waltham, MA, USA
Thermocycler 1 (96-well)	GeneAmp PCR System 9700	Applied Biosystems, Foster City, CA, USA
Thermocycler (24-well)	Gene Amp PCR 2400	Perkin Elmer, Waltham, MA, USA
Transilluminator	White/UV Transilluminator,	Laboratory Imaging System, UVP Inc. Upland, CA, USA Software: Imagestore™
Transilluminator	UV Transilluminator	Spectoline, Westbury, NY, USA

PFGE Equipment	Electrophoretic Cell, Cooling Module, Chef Mapper™	Bio-Rad, Hercules CA, USA
PCR Workstation	PCR Workstation 600, Controller Model 300	AirClean Systems, Raleigh, NC, USA
UV-DNA Linker	UV Stratalinker 1800	Stratagene, La Jolla, CA, USA
UV/VIS-Spectrophotometer	Ultrospec 2000	Pharma Biotech, Chandigarh, India
Varistaltic Pump	Manostat Simon	Thermo Scientific, Waltham, MA, USA
Vacuum Manifold	n.s	Promega, Madison, WI, USA
Vortex	Scientific Vortex	VELP Scientific Vortex, Usmate, Italy

8.1.2 Software

Genome Alignment Visualization	MAUVE, Multiple Genome Alignment, Version 2.3.1, 2003-2009 Aaron Darlung, Paul Infield-Harm, and Anna Rissman
Primer design	NetPrime, Premier Biosoft, http://www.premierbiosoft.com/netprimer/
Sequence Alignment	BioEdit Sequence Alignment Editor 1997-2007 Tom Hall

8.1.3 Chemicals

Chemical	Safety instructions	Manufacturer
Agarose, UltraPure™		Invitrogen™, Carlsbad, CA, USA
Ampicillin	Xn R 36/37/38-42/43 S 22-26-36/37	Sigma-Aldrich, St. Louis, MO, USA
Chloroform, UltraPure™, Phenol:Chloroform:Isoamyl Alcohol (25:24:1 v/v)	T R 23/24/25-34-40-48/20/21/22-68 S 26-36/37/39-45	Invitrogen™, Carlsbad, CA, USA
EDTA, Ethylenediaminetetraacetic acid disodium salt dehydrate, 0.5M, pH 8		Invitrogen™, Carlsbad, CA, USA
Ethanol	F R 11 S 7-16	Acros Organics, Part of Fisher Scientific, Pittsburgh, PA, USA
Glycerol, 90-100%	S 36-36	Mallinckrodt Baker, Phillipsburg, NJ, USA
Phenol, UltraPure™-saturated Phenol Buffer	T R 23/24/25-34-40-48/20/21/22-68 S 26-36/37/39-45	Invitrogen™, Carlsbad, CA, USA
2-Propanol	F, Xi R 11-36-67 S 7-16-24/25-26	Mallinckrodt Baker, Phillipsburg, NJ, USA
Sodium-Acetate	Xi R 36/37/38 S 24/25-36/37	USB Corporation, Cleveland, OH, USA
SDS, Lauryl Sulfate Sodium Salt	F, Xn R 11-22-36/38 S 26-36/37/39	Sigma-Aldrich, St. Louis, MO, USA
Sodium Chloride	S 24/25	Mallinckrodt Baker,

		Phillipsburg, NJ, USA
Tetracycline hydrochloride	Xi R 36/37/38 S 26-36	Sigma-Aldrich, St. Louis, MO, USA
TRIS (Base), 2-Amino-2-(hydroxymethyl)-1,3-propanediol	Xi R 36 S 26	Mallinckrodt Baker, Phillipsburg, NJ, USA
X-Gal, 5-Brom-4-chlorindol-3-yl-ß-D-galactopyranosid		

8.1.4 Reagents

Description	Manufacturer
Desoxiribonuklease (DNase I)	Invitrogen™, Carlsbad, CA, USA
dNTPs	Invitrogen™, Carlsbad, CA, USA
Oligonucleotides	Integrated DNA Technologies, Coralville, IA, USA
Restriction Enzymes: BamHI (Buffer 4) SacI (Buffer 4) ScaI (Buffer 4) SAP, Shrimp Alkaline Phosphatase	BioLabs® Inc., Ipswich, MA USA
ColiComplete disc	BioControl, Bellevue, WA

8.1.5 DNA Ladders

1kb Plus DNA ladder

Invitrogen™, Carlsbad, CA, USA

Storage Buffer
10 mM Tris-HCl (pH 7.5) 50 mM NaCl 1 mM EDTA

12 bands ranging from 1,000 bp to 12,000 bp (1000 bp increments); 8 bands from 100 – 1650 bp.

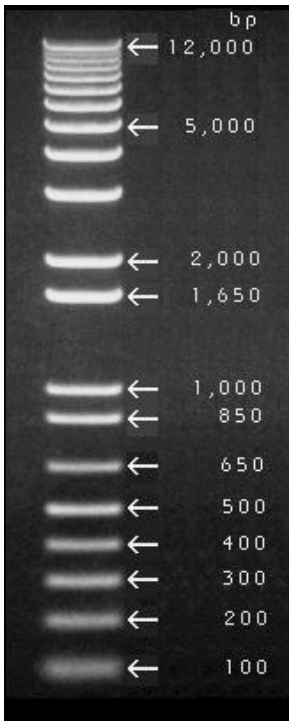


Figure 8-1. ***1 Kb Plus DNA Ladder.*** *0.7 µg/lane, 0.9% agarose gel. stained with ethidium bromide*

GeneRuler™ DNA Ladder Mix, O'GeneRuler™, 100-10,000 bp

Fermentas International Inc., Burlington, OT, Canada

Storage Buffer
10 mM Tris-HCl (pH 7.6) 1 mM EDTA

Loading dye
10 mM Tris-HCl (pH 7.6) 0.03 % bromphenol blue 0.03 % xylene cyanol FF 60 % glycerol 60 mM EDTA

21 bands ranging from 100 bp to 10,000 bp (10,000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1200, 1000, 900, 800, 700, 600, 500, 400, 300, 200, and 100). It contains three reference bands (3000, 1000 and 500 bp) for easy orientation.

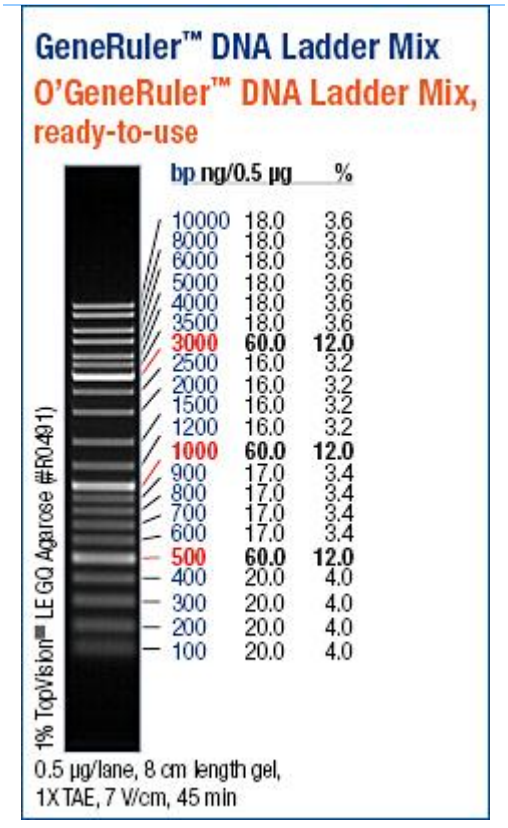


Figure 8-2. ***O'GeneRuler™***

HyperLadder™ II DNA Ladder, 50-20,000 bp
Bioline., Burlington, OT, Canada

HyperLadder II produces a pattern of 15 regularly spaced bands, ranging from 50bp to 2000bp. To allow easy identification and orientation, the 300bp, 1000bp and 2000bp bands have the highest intensity.

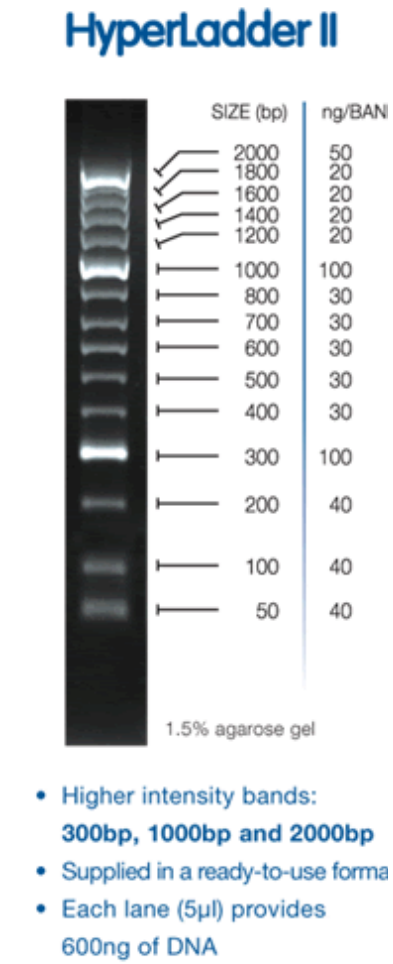


Figure 8-3. ***HyperLadder II.***

exACTGene 1kb Plus DNA Ladder, 10-10,000 bp
Fisher Scientific, Pittsburgh, PA, USA.

DNA Marker; Fisher BioReagents; exACTGene; 1kb Plus DNA ladder; 100 Lanes; Room temperature storage, Quantitative, Ready-to-use; 19 Bands, 100 bp-10kb; Ideal for analyzing digested DNA over a wide range of sizes.

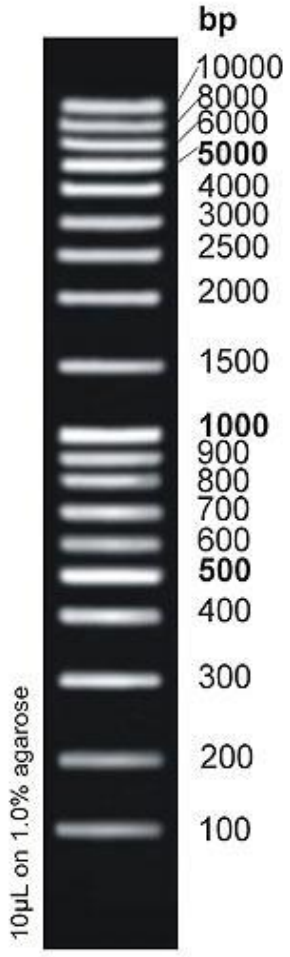


Figure 8-4. ***exACTGene 1kb Plus DNA Ladder***

8.1.6 Buffer and Solutions

Description	Composition	Manufacturer
TE-Buffer	10 mM Tris/HCl, pH 8,0 1 mM EDTA in Aqua dest.	-
Tris-Acetate-EDTA (TAE) Solution, 10 x		Promega Corporation, Madison, WI, USA
Tris/HCl		Promega Corporation, Madison, WI, USA
SSC 20x (pH 7.0)		Promega Corporation, Madison, WI, USA

8.1.7 Commercial Kits

Description	Name	Manufacturer
DNA Extraction Kit	MasterPure™ DNA Purification Kit	Epicentre® Biotechnologies, Madison, Wi, USA
Gel Extraction Kit	QUIAX II	QUIAGEN Inc., Valencia, CA, USA
HotStart Taq	HotStart Taq Polymerase kit	QUIAGEN Inc., Valencia, CA, USA
PCR Cleaning Kit	QIAquick PCR Purification Kit	QUIAGEN Inc., Valencia, CA, USA
Plasmid extraction	Wizard® *Plus* Midipreps DNA Purification System	Promega Corporation, Madison, WI, USA
Platinum Taq Polymerase	Platinum Taq Polymerase kit	QUIAGEN Inc., Valencia, CA, USA

RNA Extraction Kit	MasterPure™ RNA Purification Kit	Epicentre® Biotechnologies, Madison, Wi, USA
Serological O157:H7 latex agglutination kit	RIM O157:H7	Remel, Lenexa, KS
T4-DNA-Ligase and 5x Ligase - Buffer	T4-DNA-Ligase 5 U/µL	Invitrogen™, Carlsbad, CA, USA

8.1.8 Other Materials

Description	Model/Name	Manufacturer
Cuvettes (UV)	Ultra-Vu Disposable	Elkay, Shrewsbury, MA, USA
Cuvettes (Electroporation)	GenePulser® Cuvette 0.1 cm electrode gap	Bio-Rad, Hercules CA, USA
Microcentrifuge Tubes	0.5µL, 1.5µL, 2.0µL	Eppendorf North America, Hauppauge, NY, USA
PCR Tubes	0.2mL	MicroAmp®, ABI, Foster City, CA, USA
Petri Dish	stackable	BP Falcon™
Pipettes	Pipetman (1-10µL, 2-20µL, 20-100µL, 50-200µL, 200-1000µL)	Gibson, Middleton, WI, USA
Pipettes (aspirating)	5mL, 10mL, 25mL, 50mL	BP Falcon™
Pipette tips	10µL, 20 µL, 100 µL, 200 µL, 1000 µL	Eppendorf North America, Hauppauge, NY, USA

8.1.9 Media

All media was autoclaved (250ºC/25 min) before usage. To prepare the agar plates, 25 mL of media was divided in Petri-dishes and allow solidifying.

Name	Composition	Manufacturer
Difco™ Luria (LB)-Agar Base	Approximate Formula per liter 10 g Tryptone 5.0 g Yeast Extract 0.5 g Sodium Chloride 15.0 g Agar	Difco, Becton, Dickinson and Company, Franklin Lakes, NJ, USA
LB-Broth	Approximate Formula per liter 10 g Tryptone 5.0 g Yeast Extract 0.5 g Sodium Chloride	Difco, Becton, Dickinson and Company, Franklin Lakes, NJ, USA
MacConkey	Approximate Formula per liter 3 g Proteose peptone 17 g Peptone 10 g Lactose 1.5 g Bile salts No.3 5 g NaCl 0.03 g Neutral red 0.001 g Crystal violet 13.5 g Agar	BioControl, Bellevue, WA
Nutrient-Broth	0.5% Pancreatic Digest of Gelatin 0.3% Beef Extract	Becton, Dickinson and Company, Franklin Lakes, NJ, USA
SOC-Media	2% (w/v) Tryptone 0.5% (w/v) Yeast extract 8.6 mM NaCl 2.5 mM KCl 20 mM MgSO4 20 mM Glucose	Sigma-Aldrich, St. Louis, MO, USA

8.2 Methods

8.2.1 Storage of Bacterial Cultures

MATERIALS

Buffers and Solutions

Glycerol (90-100 %)
LB-Media

METHOD

1. Add 1 bacterial colony to 0.75 mL media (screw top tube)
2. Grow overnight, in shaker incubator at 37 ºC
3. Add 0.25 mL glycerol and mix
4. Store cultures at -20 ºC

8.2.2 Determining Bacterial Growth

MATERIALS

Buffers and Solutions

LB-Media

Equipment conditions:

UV/VIS-Spectrophotometer	Wavelength	600 nm

METHOD

1. Add 2.5 mL bacterial overnight culture to 250 mL media
2. Incubate in shaker incubator at 37 ºC
3. Every 30 minutes, take 1 mL and measure the OD_{600} in UV/VIS-Spectrophotometer
4. After 2 h incubation, measure OD every 1 h till reaching 5 hours total

8.2.3 Phenotypic Stain Characterization

8.2.3.I Sorbitol Fermentation and Glucoronidase Activity

MATERIALS

Buffers and Solutions

Cefixime-tellurite sorbitol MacConcey agar (BAM)
ColiComplete disc

Equipment conditions:

UV-light (365 nm)

METHOD

1. Plate colonies on specific media
2. Add the ColiComplete disc
3. Incubate at 37 ºC overnight
4. Sorbitol non-fermenting colonies appear colorless with grey center, sorbitol fermenting colonies appear red
5. Glucuronidase (MUG) positive bacteria show fluorescence under UV light

8.2.3.II Serological Antigen Analysis

MATERIALS

Buffers and Solutions

RIM O157:H7 test kit

METHOD

1. Add one drop of O157 or H7 specific antiserum on the provided paper
2. Use the spatula to transfer an overnight colony from the media to the paper
3. Suspend the sera with the bacteria
4. Examine for agglutination:
    - Positive: agglutination (clumping) of the suspension
    - Negative: suspension remains turbid

8.2.4 MasterPure ™ Complete DNA Purification

MATERIALS

Provided Buffers and Solutions

MasterPure™ DNA Purification Kit Contents

Tissue and Cell Lysis (TCL) Solution
2X T and C Lysis Solution
MPC Protein Precipitation Reagent
RNase A @ 5 µg/ µL
Proteinase K @ 50
TE Buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA)

The kit provides all reagents necessary for the extraction of genomic DNA from bacterial cells.

Buffers and Solutions

Ethanol (70%)	70 mL	EtOH (99 %)
	add 100 mL	bidest. H_2O
Isopropanol		

METHOD

1. Dilute 1 µL of Proteinase K into 300 µL of TCL Solution for each sample
2. Pellet cells by centrifugation of 0.5 mL of an overnight culture and discard the supernatant, leaving approximately 25 µL of liquid
3. Resuspend cell pellet by vortexing (10 sec)
4. Add 300 µL of TCL containing Proteinase K and mix thoroughly
5. Incubate at 65ºC for 15 minutes; vortex every 5 minutes
6. Cool sample to 37 ºC and add 1 µL of RNase A to the sample, vortex
7. Incubate at 37ºC for 30 minutes
8. Cool sample on ice for 3-5 minutes
9. Add 175 µL of MPC to the sample and vortex vigorously for 10 seconds
10. Pellet debris by centrifugation at 4 ºC for 10 minutes at ≥ 10,000 x g
11. Transfer the supernatant and discard the pellet
12. Add 500 µL isopropanol to the supernatant. Invert the the 30-40 times
13. Pellet DNA by centrifugation at 4 ºC for 10 minutes at ≥ 10,000 x g
14. Carefully pour off the isopropanol
15. Rinse twice with 70% ethanol an remove residual ethanol with a pipette
16. Resuspend the DNA in 35 µL of TE buffer
17. Analyze 1 µL of DNA on a 1% agarose gel (0).

8.2.5 QIAquick PCR Purification

MATERIALS

Provided Buffers and Solutions

QIAquick™ PCR Purification Kit Contents

QIAquick Spin Colum
Buffer PB*
Buffer PE
Buffer EB (10 mM Tris·Cl, pH 8.5)
pH indicator I
Collection tubes (2 mL)

The kit provides all reagents necessary for the silica-membrane-based purification of PCR products. *(contains chaotropic salts which are irritants. Take appropriate laboratory safety measures and ware gloves when handling)

Buffers and Solutions

Ethanol (70%)	70 mL	EtOH (99 %)
	add 100 mL	bidest. H_2O
Isopropanol		
Ethanol (70%)	70 mL	EtOH
	add 100 mL	H_2O bidest.
1 x TE buffer	10 mL	10x TE Buffer
	add 900 mL	H_2O bidest.

Specifications

	QIAquick kit
Max. binding capacity	10 µg
Max. weight of gel slice	-
Min. Elution volume	30 µL
Capacity of column reservoir	800 µL
Typical recoveries	
Recovery of DNA	90-95% (100 bp – 10 kb)
Recovery of oligonucleotides	0
Recovered	
Oligonucleotides	-
dsDNA	100 bp – 10 kb
Removed	
<10mers	YES
17-40mers	YES

METHOD

1. Dilute 1 µL of Proteinase K into 300 µL of TCL Solution for each sample Pellet
2. Add ethanol to Buffer PE
3. Add 1:250 volume pH indicator I to buffer PB. The yellow color indicates a pH of ≥ 7.5
4. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix
5. Check that the mixture is yellow
6. Place a QIAquick spin column in a provided 2 mL collection tube and apply the sample to the column
7. Centrifuge for 30-60 seconds
8. Discard flow-through and place column back in collection tube
9. Add 0.75 mL Buffer PE to the QIA quick column and centrifuge for 30-60 seconds
10. Discard flow-through and place column back in collection tube
11. Centrifuge the column for an additional 1 min
12. Place the column in a clean 1.5 mL tube
13. Add 30 µL Buffer EB to the center of the membrane, let the column stand for 1 minute, and then centrifuge for 1 minute
14. Analyze 1 µL of purified products on a 1% agarose gel (0).

8.2.6 Standard Precipitation with Ethanol

MATERIALS

Buffers and Solutions

Ethanol 100%		
Ethanol (70%)	70 mL	EtOH (100%)
	add to 100 mL	H_2O bidest.
1x TE buffer	10 mL	10x TE Buffer
	add 900 mL	H_2O bidest.

METHOD

1. Add TE buffer to the sample reach a final volume of 300 µL
2. Add 600 µL of ice-cold ethanol
3. Invert to mix 30-40 times
4. Place the solution in the -20ºC freezer for 1 hour (in case of small DNA fragments or small amounts, store at least 2h)
5. Pellet DNA by centrifugation at 4 ºC for 10 minutes at ≥ 10,000 x g
6. Carefully pour off the ethanol
7. Rinse twice with 70% ethanol an remove residual ethanol with a pipette
8. Resuspend the DNA in 5-15 µL of TE buffer
9. Analyze 1 µL of purified products on a 1% agarose gel (0).

8.2.7 Nucleic Acid Quantification

EQUIPMENT

Micro-Volume UV-Vis Spectrophotometer for Nucleic Acid and Protein Quantification

METHOD

1. Pipette 2 µL water on the pedestal to calibrate the instrument
2. Blank the instrument with 2 µL Buffer or water (depending on the sample)
3. Measure DNA/RNA concentration by loading 2 µL sample on the pedestal.

8.2.8 Nucleic Acid Amplification

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase* (5 units/µL) $MgCl_2$* (25 mM solution) 10 x PCR buffer* (15 mM $MgCl_2$) dNTPs** (10 mM Mix) Primer A + B (10x) DNA sample (0.5 ng/µL) H_2O	See individual requirements below

*QUIAGEN, HotStart Taq DNA polymerase kit **Invitrogen dNTP Mix

Thermal cycler program:

1 x	15 min	95ºC	(HotStart Polymerase activation)
10x	0:30 min	95ºC	(Denaturation)
	0:20 min	69ºC	-1ºC/cycle (Annealing)
	1 min	72ºC	(Extension)
35x	0:30 min	95ºC	(Denaturation)
	0:20 min	58ºC	(Annealing)
	1 min	72ºC	(Extension)
1 x	10 min	72ºC	(Final extension)
∞	--	4ºC	(Cooling)

METHOD

1. Prepare a volume of the Master mix 10 % greater than required for the total number of reactions to be performed
2. Mix the Master mixes thoroughly and dispense appropriate amounts into PCR tubes placed on ice
3. Add template DNA to the individual PCR tubes
4. Include negative control
5. Program the thermal cycler
6. Place the PCR tubes in the thermal cycler
7. After the Amplification, run 5 µL of the samples in agarose gel (0), or store at 4ºC.

8.2.8.I PCR Amplification for the O157 and H7 antigens

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase	2.5 U
$MgCl_2$	3 mM
10 x PCR buffer (50 nM KCl, 2.5 mM $MgCl_2$, 10 mM Tris-Cl, (pH 8.3))	1X
dNTPs	300 µM
Primer pairs	300 nM
DNA sample (0.5 ng/µL)	100 ng
H_2O	Add 50 µL

Primer

Primer	Sequences (5' - 3')	Product length [bp]	Gene	Reference
5'O157	GCTGCTTATGCAGATGCTC	133	*wzx*	*(Monday et al., 2007)*
3'O157	CGACTTCACTACCGAACACTA			
F-fliC1	CCATGGCACAAGTCATTAATACCAAC	1,739	*fliC*	*(Monday et al., 2004)*
R-fliC2	CTAACCCTGCAGCAGAGACA			

METHOD

1. Add the template DNA of each strain to the individual PCR tubes containing the master mix
2. Include negative control
3. Program the thermal cycler
4. Place the PCR tubes in the thermal cycler
5. After the Amplification, run 5 µL of the samples in agarose gel (0) to compare each Amplicons of the O rough:H7 strains to the wild-type strain
6. Store at 4ºC and use samples for further sequencing analysis.

8.2.8.II PCR Amplification for the O157:H7 virulence markers

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase	2.5 U
$MgCl_2$	-
10 x PCR buffer (50 mM KCl, 2.5 mM $MgCl_2$, 10 mM Tris-Cl, (pH 8.3))	1X
dNTPs	200 µM
Primer pairs	300 nM
DNA sample (0.5 ng/µL)	100 ng
H_2O	Add 50 µL

Primer

Primer	Sequences (5' - 3')	Product length [bp]	Genes	Reference
LP30 LP31	CAGTTAATGTGGTGGCGAAGG CACCAGACAATGTAACCGCTG	384	*stx*$_1$	*(Feng and Monday, 2000)*
LP43 LP44	ATCCTATTCCCGGGAGTTTACG GCGTCATCGTATACACAGGAGC	584	*stx*$_2$	
PT-2 PT-3	GCGAAAACTGTGGAATTGGG TGATGCTCCATCACTTCCTG	252	*uid*A	
AE22 AE20-2	ATTACCATCCACACAGACGGT ACAGCGTGGTTGGATCAACCT	397	*eae*A	
MFS1Fb MFS1R	GTTTATTCTGGGGCAGGCTC CTTCACGTCACCATACATAT	158	*ehx*A	

METHOD

1. Add the template DNA of each strain to the individual PCR tubes containing the master mix
2. Include negative control
3. Program the thermal cycler
4. Place the PCR tubes in the thermal cycler
5. After the Amplification, run 5 µL of the samples in agarose gel (0) to compare each Amplicons of the O rough:H7 strains to the wild-type strain
6. Store at 4ºC and use samples for further sequencing analysis.

8.2.8.III PCR Amplification of the O antigen, core antigen gene cluster, and the genes *man*A, *wec*A and *gne*

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase	2.5 U
$MgCl_2$	-
10 x PCR buffer (50 nM KCl, 2.5 mM $MgCl_2$, 10 mM Tris-Cl, (pH 8.3))	1X
dNTPs	200 µM
Primer (8.3.1; 8.3.2; 8.3.3; 8.3.4; 8.3.5; 8.3.6; 8.3.7; 8.3.8; 8.3.9; 8.3.10)	300 nM
DNA sample (0.5 ng/µL)	100 ng
H_2O	Add 50 µL

METHOD

1. Add the template DNA of each strain to the individual PCR tubes containing the master mix
2. Include negative control
3. Program the thermal cycler
4. Place the PCR tubes in the thermal cycler
5. After the Amplification, run 5 µL of the samples in agarose gel (0) to compare each Amplicons of the O rough:H7 strains to the wild-type strain
6. Store at 4ºC and use samples for further sequencing analysis.

8.2.8.IV PCR Amplification of IS*629*

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase	2.5 U
$MgCl_2$	1.5 mM
10 x PCR buffer	1X
dNTPs	400 µM
Primer pairs	300 nM
DNA sample (0.5 ng/µL)	3 ng
H_2O	Add 30 µL

Primer

Primer	Sequences (5' - 3')	Product length [bp]	Reference
IS*629*-inside-F	GAACGTCAGCGTCTGAAAGAGC	781	(Ooka et al., 2009b)
IS*629*-inside-R	GTACTCCCTGTTGATGCCAG		
SRM86	AGAAGCACCGGCTAACTC	562	(Monday et al., 2004)
SRM87	CGCATTTCACCGCTACAC		
VMP5	AGAAGCACCGGCTAACTC	204	(Monday et al., 2007)
VMP6	CGCATTTCACCGCTACAC		

METHOD

1. Add the template DNA of each strain (strain list 3.1.1) to the individual PCR tubes containing the master mix
2. Include negative control
3. Program the thermal cycler
4. Place the PCR tubes in the thermal cycler
5. After the Amplification, run 5 µL of the samples in agarose gel (0) to determine the presence/absence of the individual IS*629* insertion sites
6. Store at 4ºC.

8.2.8.V PCR Amplification of IS*629 and* IS*629* insertion sites

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase	2.5 U
$MgCl_2$	1.5 mM
10 x PCR buffer	1X
dNTPs	400 µM
Primer pairs (8.3.11)	300 nM
16S primer pairs (8.2.8.IVand 8.3.11)	300 nM
DNA sample (0.5 ng/µL)	3 ng
H_2O	Add 30 µL

PRINCIPLE

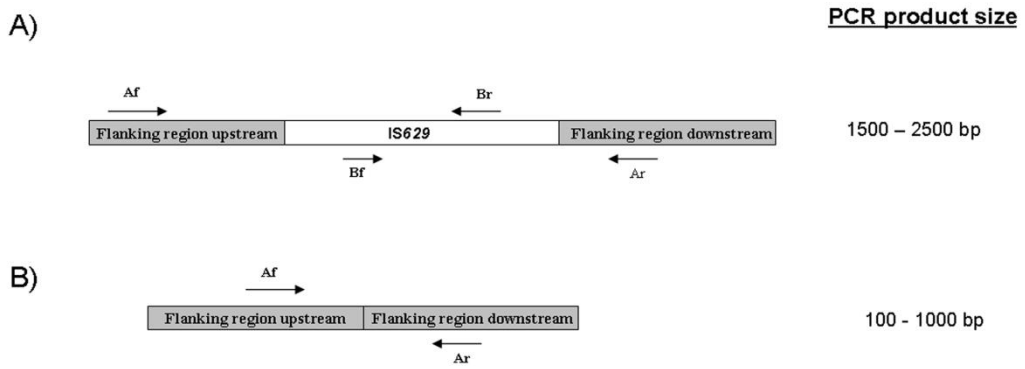


Figure 8-5. ***Schematic representation of the strategy used for primer design***
Primer pairs: A: presence/absence of IS629 at specific loci, B: IS629 internal primer. A) Amplification product for locations where the IS629 element is present; B) Amplification product for locations where the IS629 element is absent, although the up-and downstream flanking region is present in the genome but not carrying an insertion.

METHOD

1. Add the template DNA of each strain (strain list 3.1.1) to the individual PCR tubes containing the master mix
2. Include negative control
3. Program the thermal cycler
4. Place the PCR tubes in the thermal cycler
5. After the Amplification, run 5 µL of the samples in agarose gel (0) to determine the presence/absence of the individual IS*629* insertion sites
6. Store at 4ºC.

8.2.8.VI PCR Amplification of plasmid specific sites

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase	2.5 U
$MgCl_2$	1.5 mM
10 x PCR buffer	1X
dNTPs	400 µM
Primer pairs	300 nM
DNA sample (0.5 ng/µL)	3 ng
H_2O	Add 30 µL

Primer

Primer	Sequences (5' - 3')	Product length [bp]	target
pO157 F pO157 R	GCACTGTTCGTCTATGATAACCA AGTTCCGCCATTTTTCGC	642	*esp*P
pO157 katP F2 pO157 katP R2	GCATCAGGTTACAGCCCG GTAAACACTCCCAACGAAGAAT	454	*kat*P
pSFO157 F pSFO157 R	AATACGGGCTGCGTGAGTT TCAGATCAGAATACGGGTCCA	369	gene
pO55 F pO55 R	CCGCTGATATGCTTGCTGA TCACCGCTCACAACAATAGAAT	334	colicin gene
pO55 F3 pO55 R3	TTGAAAACCAGAACCTCGC AAACAGGGTACAAAACCAGTAACT	349	gene

METHOD

1. Add the template DNA of each strain (strain list 3.1.1) to the individual PCR tubes containing the master mix
2. Include negative control
3. Program the thermal cycler
4. Place the PCR tubes in the thermal cycler
5. After the Amplification, run 5 µL of the samples in agarose gel (0) to determine the presence/absence of the individual IS*629* insertion sites
6. Store at 4ºC.

8.2.9 Sequencing analysis

After the PCR samples were sequenced by Amplicon Express (Pullman, WA) using the 'Big Dye chemistry' technique in an Applied Biosystems 377 automated sequencer. Sequences were analyzed and assembled using BioEdit software and compared to the equivalent sequences in GenBank.

8.2.10 Agarose Gel Casting

MATERIALS

Buffers and Solutions

1x TE Gel Buffer	200 mL	5 x TE
	60µL	Ethidium Bromide
	add to 2000 mL	Water

Gels

1% Agarose Gel	1 g	Agarose
	100 mL	1 x TE buffer
2 % Agarose Gel	2 g	Agarose
	100 mL	1 x TE buffer
3 % Agarose Gel	3 g	Agarose
	100 mL	1 x TE buffer

METHOD

1. Add required amount of agarose to Erlenmeyer flask
2. Add 50 mL TE buffer
3. Heat for approximately 1:30 minutes in the microwave until agarose is dissolved
4. While stirring add the remaining 50 mL TE buffer
5. Choose the right comb size and pour gel mix in casting form
6. Let gel solidify for min. 30 minutes.

8.2.11 Agarose Gel Electrophoresis

MATERIALS

Buffers and Solutions

1x TE Gel Buffer	200 mL	5 x TE
	60µL	Ethidium Bromide
	add to 2000 mL	Water

Gels

1 % Agarose Gel
2 % Agarose Gel
3 % Agarose Gel

METHOD

1. Mount the required gel in the gel electrophoresis tank and remove the comb
2. Add buffer to the chamber and assure that the gel is just covered with buffer
3. Load 5µL DNA ladder and samples dissolved in gel-loading buffer to the wells

   Sample preparation:

Gel-Loading buffer	3 µL
Sample	1-30 µL

4. Close the chamber and start run. Running conditions:

Gel Electrophoresis Chamber (Horizontal)	Step 1	Voltage	95 V
		Amperage	100 – 115 mA
		Duration	5 min
	Step 2	Voltage	80 V
		Amperage	100 – 115 mA
		Duration	60 – 120 min

5. Continue electrophoresis until the dye has migrated approximately two thirds of the length of the gel
6. Capture the image from the fluorescent gel with the transilluminator.

8.2.12 Pulse-field Gel Electrophoresis (PFGE)

The PFGE was conducted using a rapid standardized protocol for *E. coli* O157:H7 (Centers for Disease Control and Prevention, 2003).

MATERIALS

Buffers and Solutions

Cell Suspension Buffer (CBS), pH 8.0	100 mM	TRIS
	100 mM	EDTA
		Sterile dest. H_2O
Cell Lysis Buffer (CL)	50 mM	TRIS
	50 mM	EDTA
	1 %	Sarcosyl
		Sterile dest. H_2O
TE Buffer, 1X		
TBE Buffer, 10X		
CLR Water (CLRW)		
1% Gold Agarose Gel	0.24 g	Gold Agarose
	20 mL	TE buffer
H Buffer		
Electrophoresis Running Buffer	110 mL	TBE
	2089 mL	Sterile dest. H_2O
Restriction enzymes	*Xba*I	
	*Bln*I	
Proteinase *K*		

METHOD

1. Resuspend a swap of overnight culture in 3 mL CBS (OD: 0.42-0.48)
2. Mix the following in a 1.5 mL tube and mix 3-4 times by pipetting:

Cell CBS solution	400 µL
Proteinase K	20 µL
Agarose solution	400 µL

3. Dispense solution in plug casting wells and let solidify (10-15 min)
4. Mix the following in a 50 mL tube and add the solid plugs:

CL buffer	5 mL
Proteinase K	25 µL

5. Incubate plugs at 54ºC for 1-2.5h in water bath (agitation)
6. Rinse plugs 1x with pre-heated Ultrapure H_2O
7. Wash plugs 2x with pre-heated Ultrapure H_2O at 54ºC for 10-15 min
8. Wash plugs 4x with pre-heated TE buffer at 54ºC for 10-15 min
9. Decant buffer and add 5-10 mL TE buffer (store at 4ºC)
10. For the further steps include min. 2 slices of *Salmonella Branderup* (M9812)
11. Mix the following in a 1.5 mL tube and add one 0.2x10mm slice of the desired bacteria:

CLRW buffer	180 µL
H Buffer	20 µL

12. Incubate slices at 37ºC for 10-15 min
13. For the digestion reaction mix the following in a 1.5 mL tube and add the pre-soaked slice:

CLRW buffer	175 µL
H Buffer	20 µL
Enzyme	5 µL

14. Incubate slices at 37ºC for 1.5-3 h
15. During the digestion, prepare the agarose gel (for a 21 cm wide gel) solution and equilibrate in water bath:

Agarose	1.5 g
TBE buffer	7.5 mL
Water	143 mL

16. After the digestion of the gel slices place each slice on one tooth of the gel-comb
17. Place comb in casting form and pour gel solution in the tray (let gel solidify for approx. 30 min)
18. Mount the gel in the gel electrophoresis tank and remove the comb
19. Add the running buffer to the chamber
20. Close the chamber and start run. Running conditions:

PFGE Gel Electrophoresis Chamber (Horizontal)	Step 1	Voltage	95 V
		Amperage	100 – 115 mA
		Duration	5 min
	Step 2	Voltage	80 V
		Amperage	100 – 115 mA
		Duration	60 – 120 min

21. After the run proceed with the staining of the gel
22. Wash gel with Ultrapure H_2O containing ethidiumbromide for 20-30 min
23. Wash gel with Ultrapure H_2O for 20-30 min
24. Capture the image from the fluorescent gel with the transilluminator.

8.2.13 DNA Extraction from Agarose Gels

MATERIALS

Buffers and Solutions

Ethanol (100%)
Tris·Cl

QIAEX II Gel Extraction Kit Contents

QIAEX II Suspension
Buffer QX1* (with pH indicator)
Buffer PE (concentrate)

The kit provides all reagents necessary for the silica-membrane-based purification of PCR products. *(contains chaotropic salts which are irritants. Take appropriate laboratory safety measures and ware gloves when handling)

METHOD

1. Add ethanol to Buffer PE (see bottle label for volume)
2. Excise the DNA band from the gel with a scalpel
3. Weight the gel slice and add 300 µL of buffer QX1 to each 100 mg of gel
4. Resuspend QIAEX II by vortexing for 30 s
5. Add 10 µL of QIAEX II to the sample and mix
6. Incubate at 50ºC for 10 minutes and vortex every 2 minutes
7. Centrifuge for 30-60 seconds and remove supernatant
8. Add 500 µL of Buffer QX1
9. Resuspend the pellet by vortexing
10. Centrifuge for 30-60 seconds and remove supernatant
11. Wash the pellet twice with 500 µL of Buffer PE
12. Air-dry the pellet for 10-15 min until the pellet becomes white
13. To elute DNA, add 20 µL of Tris·Cl and resuspend the pellet by vortexing
14. Incubate at room temperature for 5 minutes
15. Centrifuge for 30-60 seconds and pipet supernatant in clean tube
16. Analyze 1 µL of purified products on a 1% agarose gel (0).

8.2.14 Design of Plasmid Vectors

8.2.14.I Vector pGNE

8.2.14.I.A PCR Amplification of the gne gene for vector pGNE

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase	2.5 U
$MgCl_2$	1.5 mM
10 x PCR buffer	1X
dNTPs	400 µM
Primer pairs (gne.13 +gne.14)	300 nM
DNA sample *E. coli* EDL933 (0.5 ng/µL)	3 ng
H_2O	Add 30 µL

METHOD

1. Prepare a volume of the Master mix 10 % greater than required for the total number of reactions to be performed
2. Mix the Master mixes thoroughly and dispense appropriate amounts into PCR
3. Add template DNA to the individual PCR tubes
4. Include negative control for the PCR
5. Program the thermal cycler:

1 x	15 min	95ºC
35x	0:30 min	95ºC
	0:20 min	50ºC
	1 min	72ºC
1 x	10 min	72ºC
∞	--	4ºC

6. Place the PCR tubes in the thermal cycler
7. After the Amplification, run 2 µL of the sample in 1% agarose gel (0) and purify remaining sample
8. Analyze 1 µL of purified products on a 1% agarose gel (0)
9. Use purified product for cohesive-end restriction enzyme digestion (8.2.14.I.B).

8.2.14.I.B Cohesive-End Restriction Enzyme Digestion

MATERIALS

Buffers and Solutions

Restriction Enzymes:			
*Bam*HI (20 units/µL)	10x NEBuffer 4		
*Sac*I (20 units/µL)	10x NEBuffer 4		
Plasmid pTrc99A			
PCR product:			
gne gene Amplification product (8.2.14.I.A)			
BSA 10x	1	µL	100 x
	9	µL	Water

METHOD

1. Select the restriction enzymes required and determine appropriate reaction buffer
2. Prepare the following mixtures:

10x buffer		5 µL
10x BSA		5 µL
A	PCR product (100 ng/µL)	5 µL
B	Plasmid (100 ng/µL)	5 µL
Restriction enzyme		1 µL
Water		34 µL

3. Incubate mixtures for 1 hour at 37ºC
4. Purify digested DNA with standard Ethanol precipitation (8.2.6)
5. Resuspend in 6 µL buffer

Product	Size [bp]
gne gene	1,223
pTrc99A	4,176

6. Use purified products for second digestion with different enzyme
7. Run samples in 1% agarose gel (0) and excise the required bands from the gel (0) [Final volume 20 µL]
8. Analyze samples with NanoDrop (8.2.7)
9. Analyze 1 µL products on a 1% agarose gel (0) in comparison with the original PCR Product and undigested plasmid

Product	Size [bp]
gne gene	1,223
pTrc99A	4,155

10. Use double digested products for sticky end ligation reaction of the linear plasmid DNA with digested PCR product.

8.2.14.I.C Ligation Reaction

MATERIALS

Buffers and Solutions

T4-Ligase 5 U/µL
10x T4-Ligase buffer

METHOD

1. Select the restriction enzymes required and determine appropriate reaction buffer
2. Use purified double digested PCR product and vector (8.2.14.I.B)
3. Prepare the following mixtures:

10x ligase buffer		1 µL
A	PCR product (100 ng/µL)	1 µL
B	Plasmid (100 ng/µL)	1 µL
T4 ligase		1 µL
Water		5.5 µL

4. Incubate mixtures for at 15ºC overnight
5. Ligation product is ready for transformation with electrocompetent cells
6. Analyze ligation product on a 1% agarose gel (0) in comparison with the original double digested plasmid DNA

Product	Size [bp]
pGNE	5,999

7. Use ligation product for further application or store at -20ºC.

8.2.14.II Vector pIS*629*-Tc-AB

8.2.14.II.A Site directed mutation of IS629 ORFab

Method of Mikaelian and Sergerant (1992).

MATERIALS

Buffers and Solutions

Master Mix

High Fidelity HotStart Taq polymerase* (5 units/µL)	0.5 µL
$MgCl_2$* (25 mM solution)	1.2 µL
10 x PCR buffer* (15 mM $MgCl_2$)	3 µL
dNTPs** (10 mM Mix)	0.5 µL
Primer A + B (10x)	1.25 µL
DNA sample *E. coli* MA6 (0.5 ng/µL)	6 µL
H_2O	to 30µL

*Invitrogen, High Fidelity HotStart Taq DNA polymerase **Invitrogen dNTP Mix

Primer*

Primer	Sequences (5' - 3')	Product length [bp]
GNE.14	ATCTCTTTGGGAATCGCC	2,765
GNE.13	TACCTTCCCTGAGTAGTCAATGC	
IS*629*F-1	ATATAGAGCTCATGACTAAAAATACTCGTTTTTC	
IS*629*R-2	AATCGTCGATTGTTATACCAGTCC*ATACCTC*	
IS*629*R-3	CGGGATCCTCAGGCTGCCAGATCA	see below
IS*629*alter-M	TTCGACCGCCTCTGGAAA**A**AAATGATGCCACTGC TGGATA	

*underlined: restriction sites, **bold**: mutation, *italic*: mismatch

METHOD

1. Prepare a volume of the Master mix 10 % greater than required for the total number of reactions to be performed

Primer Mix 1	gne.14/gne.13	2,165 bp amplicon

2. Mix the Master mixes thoroughly and dispense appropriate amounts into PCR
3. Add template DNA (*E. coli* MA6) to the individual PCR tubes (Include negative control for the PCR)
4. Run PCR with the following conditions:

1 x	15 min	95ºC
35x	0:30 min	95ºC
	0:20 min	50ºC
	1 min	72ºC
1 x	10 min	72ºC
∞	--	4ºC

5. After the Amplification, run samples in 1% agarose gel (0) and excise the required bands from the gel (0)
6. Prepare two master mixes with primer combinations:

Primer Mix 2	IS*629*F-1 and IS*629*R-2	1,131 bp amplicon
Primer Mix 3	IS*629*alter-M and IS*629*R-3	914 bp amplicon

7. Mix the Master mixes thoroughly and dispense appropriate amounts into PCR
8. Add template DNA (purified Amplicons PCR No.1) to the individual PCR tubes (Include negative control)
9. Run PCR with the following conditions:

1 x	15 min	95ºC
35x	0:30 min	95ºC
	0:20 min	50ºC
	1 min	72ºC
1 x	10 min	72ºC
∞	--	4ºC

10. After the Amplification, run samples in 1% agarose gel (0) and excise the required bands from the gel (0)

11. Prepare master mix with primer combination:

Primer Mix 4	IS*629*F-1 and IS*629*R-3	1,213 bp Amplicon

12. Add both template DNA samples (purified Amplicons PCR No.2 and 3) to the same individual PCR tubes (Include negative control)
13. Run PCR with the following conditions:

1 x	15 min	95ºC
35x	0:30 min	95ºC
	0:20 min	58ºC
	1 min	72ºC
1 x	10 min	72ºC
∞	--	4ºC

14. After the Amplification, run samples in 1% agarose gel (0) and excise the required bands from the gel (0)
15. Analyze 1 µL of purified products on a 1% agarose gel (0)
16. Use purified product for cohesive-end restriction enzyme digestion (8.2.14.I).

8.2.14.II.B Cohesive-End Restriction Enzyme Digestion

MATERIALS

Buffers and Solutions

Restriction Enzymes:		
*Bam*HI (20 units/µL)	10x NEBuffer 4	
*Sac*I (20 units/µL)	10x NEBuffer 4	
BSA 10x	1 µL	100 x
	9 µL	Water

METHOD

1. Select the restriction enzymes required and determine appropriate reaction buffer
2. Prepare the following mixtures:

10x buffer	5 µL
10x BSA	5 µL
A PCR product (100 ng/µL) IS*629* ORF*ab* Amplification product (8.2.14.II.A)	5 µL
B Plasmid pUC18 (100 ng/µL)	5 µL
Restriction enzyme	1 µL
Water	34 µL

3. Incubate mixtures for 1 hour at 37ºC
4. Purify digested DNA with standard Ethanol precipitation (8.2.6)
5. Resuspend in 6 µL buffer

Product	Size [bp]
IS*629* ORF*ab*	1,236
pUC18	2,686

6. Run samples in 1% agarose gel (0) in comparison with the original PCR Product and undigested plasmid Use purified products for second digestion with different enzyme
7. Run samples in 1% agarose gel (0) in comparison with the original PCR Product and undigested plasmid and excise the required bands from the gel (0)

Product	Size [bp]
IS*629* ORF*ab*	1,223
pUC18	2,665

8. Analyze samples with NanoDrop (8.2.7)
9. Analyze 1 µL products on a 1% agarose gel (0).

8.2.14.II.C Ligation Reaction

MATERIALS

Buffers and Solutions

T4-Ligase 5 U/µL
10x T4-Ligase buffer

METHOD

1. Select the restriction enzymes required and determine appropriate reaction buffer
2. Use purified double digested PCR product and vector (8.2.14.I.B)
3. Prepare the following mixtures:

10x ligase buffer		1 µL
A	PCR product (100 ng/µL)	1 µL
B	Plasmid (100 ng/µL)	1 µL
T4 ligase		1 µL
Water		5.5 µL

4. Incubate mixtures for at 15ºC overnight
5. Analyze ligation product on a 1% agarose gel (0) in comparison with the original double digested plasmid DNA

Product	Size [bp]
pUC18 + IS*629* ORF*ab*	3,888

6. Ligation product is ready for transformation with electrocompetent cells
7. Use ligation product for further application or store at -20ºC.

8.2.14.II.D PCR Amplification of the tetA gene for vector pIS629-Tc-AB

MATERIALS

Buffers and Solutions

Master Mix

High Fidelity HotStart Taq polymerase* (5 units/µL)	0.5 µL
$MgCl_2$* (25 mM solution)	1.2 µL
10 x PCR buffer* (15 mM $MgCl_2$)	3 µL
dNTPs** (10 mM Mix)	0.5 µL
Primer A + B (10x)	1.25 µL
pBr322 vector (0.5 ng/µL)	6 µL
H_2O	to 30µL

Primer

Primer	Sequences (5’ - 3’)	Product length [bp]
IS-Tc IR-BclI FW	CTTAGTGATCATGAACCGCCCCGGAAATCCTGGAGACTAAACTCCCTGAGAAAGAGGTAAACAGGATGAAATCTAACAATGCGCTCATCGTC	1,309
IS-Tc IR-NdeI REV	TCTTGTATACTGAACCGCCCCGGGTTTCCTGGAGAGTGTTTTATCTGTGAACTCAGGTCGAGGTGGCCCGGCTCCATGC	

METHOD

1. Prepare a volume of the Master mix 10 % greater than required for the total number of reactions to be performed
2. Mix the Master mixes thoroughly and dispense appropriate amounts into PCR tubes
3. Add template DNA to the individual PCR tubes (Include negative control for the PCR)
4. Run PCR with the following conditions:

1 x	15 min	95ºC
35x	0:30 min	95ºC
	0:20 min	
58ºC		
	1 min	72ºC
1 x	10 min	72ºC
∞	--	4ºC

5. After the Amplification, run samples in 1% agarose gel (0) and excise the required bands from the gel (0)
6. Analyze 1 µL of purified products on a 1% agarose gel (0)
7. Use purified product for the blunt-end ligation reaction (8.2.14.II.G).

8.2.14.II.E Blunt-End Restriction Enzyme Digestion

MATERIALS

Buffers and Solutions

Restriction Enzymes:		
*Sca*I (20 units/µL)	10x NEBuffer 3	
BSA 10x	1 µL	100 x
	9 µL	Water

METHOD

1. Prepare the following mixtures:

10x buffer	1 µL
10x BSA	1 µL
Plasmid (1000 ng/µL)	5 µL
Restriction enzyme	2 µL
Water	1 µL

2. Incubate mixtures for 1 hour at 37ºC
3. Purify digested DNA with standard Ethanol precipitation (8.2.6)
4. Resuspend in 6 µL water
5. Analyze samples with NanoDrop (8.2.7)
6. Analyze 1 µL products on a 1% agarose gel (0) in comparison with the original PCR Product and undigested plasmid.

Product	Size [bp]
pIS*629*-AB	2,888

8.2.14.II.F Dephosphorylation of Plasmid DNA

MATERIALS

Buffers and Solutions

EDTA
Ethanol 100 %
SDS (10% w/v)
Sodium Acetate (3M, 5.2 and pH 7.0)
TE buffer (pH 8.0)
Tris-Cl (10 mM, pH 8.3)
Calf Intestine phosphatase (CIP)
10x SAP buffer

METHOD

1. Use purified digested vector (8.2.14.II.A)
2. To 90 µL plasmid DNA add 10 µL of 10x SAP Buffer and 0.2 units SAP
3. Prepare the following mixtures:

10x CIP buffer	2 µL
CIP	1 µL
Plasmid (100 ng/µL)	5 µL
Water	20 µL

4. Incubate for 60 minutes at 37°C
5. Purify DNA with the QIAquick PCR purification kit (4.5).
6. Analyze samples with NanoDrop (8.2.7)
7. Analyze dephosphorylated product on a 1% agarose gel (0).

8.2.14.II.G Blunt-End Ligation

MATERIALS

Buffers and Solutions

T4-Ligase
10x T4-Ligase buffer

METHOD

1. Use purified PCR product (8.2.14.II.D) and dephosphorylated vector (8.2.14.II.F)
2. Prepare the following mixtures:

10x ligase buffer	1 µL
Plasmid DNA (100 ng/µL)	1 µL
PCR Product (100 ng/µL)	1 µL
T4 ligase	1 µL
Water	5.5 µL

3. Incubate mixtures for at 15ºC overnight
4. Ligation product is ready for transformation with “ultra-competent” cells
5. Analyze ligation product on a 1% agarose gel (0) in comparison with the original double digested plasmid DNA.

Product	Size [bp]
pIS*629*-AB-Tc	5,214

8.2.15 Preparation of "ultra-competent" *E. coli* cells

MATERIALS

Buffers and Solutions

Glycerol (10%)	10 mL	Glycerol
	90 mL	Water

Media

LB-Agar
LB-Broth

METHOD

1. Inoculate desired bacteria on LB-Agar overnight at 37ºC
2. Pick one colony and grow overnight in 5 mL LB-Broth at 37ºC with agitation
3. Add 2.5mL of overnight culture to 50 mL LB-broth and incubate at 37ºC
4. Measure the OD_{600} until it reaches 0.35 – 0.4 (ca. 2.5 h)
5. It Is essential that the number of viable cells does not exceed 10^8 cells/mL, which for most *E. coli* strains is equivalent to an OD_{600} of~0.4
6. Transfer the bacterial cells to sterile, ice-cold 50-mL tube (25mL each tube)
7. Cool the cells to 4ºC
8. Pellet the cells by centrifugation for 25 min at 2,700 x g at 4ºC
9. Decant the medium from the pellet cells and stand the tubes in an inverted position on a pad of paper towels to remove last traces of media
10. Resuspend the pellet in 15 mL ice-cold 10% Glycerol solution
11. Combine both resuspended cells in one tube
12. Repeat the washing with 10% Glycerol solution two times
13. Pellet the cells by centrifugation for 25 min at 2,700 x g at 4ºC
14. Resuspend the cells in 400 µL glycerol solution
15. Transfer 40 µL in an ice-cold Electroporation cuvette and test whether arcing occurs. If this is the case, repeat the washing step
16. Dispense into 40 µL aliquots and freeze at -80ºC.

8.2.16 Electroporation

MATERIALS

Media

SOC-Media

EQUIPMENT

Electroporation System	Gene Pulser II	Voltage	1.8 kV/cm
		Resistance	200 Ω
		Capacitance	25 µFd

METHOD

1. Cool Gene Pulser cuvette in ice
2. Mix 40 µL electrocompetent *E. coli* DH5α with 1 µL vector
3. Transfer mixture to cuvette
4. Place cuvette in Gene Pulser and electroporate until sound
5. Immediately add 750 µL SOC media (RT), and transfer to a 1.5 mL tube
6. Incubate for 1 hour at 37ºC under agitation and use for further growth on selective media (8.2.17).

8.2.17 Growth of Transformants

MATERIALS

Buffers and Solutions

Ampicillin-Stock (50 mg/mL)	50 mg Dissolve in 500 µL Add 500 µL	Ampicillin Water Ethanol
Tetracycline-Stock (5 mg/mL)	5 mg 1000 µL	Tetracycline Ethanol

Media

LB Amp^{50} plates (containing 50 µg/mL Ampicillin)	250 µL 250 mL	Amp-Stock LB-Agar (55ºC)
LB Tet^{12} plates (containing 15 µg/mL tetracycline)	250 µL 250 mL	Tet-Stock LB-Agar (55ºC)

METHOD

1. Spread 50 µL and 100 µL of the transformed cells on the selective media until dry
2. Incubate plates at 37ºC overnight
3. Determine the amount of grown cells and test colonies for the presence of the required vector containing the insert (8.2.18).

8.2.18 Analyzing Transformants by PCR

MATERIALS

Buffers and Solutions

Master Mix

HotStart Taq polymerase* (5 units/µL)	0.5 µL
$MgCl_2$* (25 mM solution)	1.2 µL
10 x PCR buffer* (15 mM $MgCl_2$)	3 µL
dNTPs** (10 mM Mix)	0.5 µL
Primer A + B (10x)	1.25 µL
DNA	colony
H_2O	Add to 30µL

*QUIAGEN, HotStart Taq DNA polymerase kit **Invitrogen dNTP Mix

Primer

Primer	Sequences (5’ - 3’)
M13 rev	AGCGGATAACAATTTCACACAGG
TRCHis	GATTTAATCTGTATCAGG
M13 Forward (-20)	GTAAAACGACGGCCAG
M13 Reverse	CAGGAAACAGCTATGAC

METHOD

1. Prepare a volume of the Master mix 10 % greater than required for the total number of reactions to be performed
2. Mix the Master mixes thoroughly and dispense appropriate amounts into PCR
3. Use the appropriate primer combinations

pTrc99A in MCS	M13rev	TrcHis	1,350 bp
pUC18 in MCS	M13 FW	M13 REV	1,311 bp
pTrc99A in lac operon	pTrc BnlI	pTrc NdeI	1,565 bp
pUC18 in Amp gene	pUC ScaI FW	ScaI REV	1,639 bp

4. Add template DNA to the individual PCR tubes
5. Include negative control for the PCR
6. Program the thermal cycler (Touchdown-PCR:

1 x	15 min	95ºC
10x	0:30 min	95ºC
	0:20 min	69ºC
	1 min	72ºC
35x	0:30 min	95ºC
	0:20 min	58ºC
	1 min	72ºC
1 x	10 min	72ºC
∞	--	4ºC

7. Place the PCR tubes in the thermal cycler
8. After the Amplification, run samples in 1% agarose gel
9. When insert can be detected in the vector, send out purified plasmid DNA for sequencing.

8.2.19 Purification of plasmid DNA

MATERIALS

Buffers and Solutions

Wizard® *Plus* Midipreps DNA Purification System contents

Cell Resuspension Solution (CRA)
Cell Lysis Solution (CLA)
Neutralization Solution (NSA)
Wizard® Midipreps DNA Purification Resin
Column Wash Solution (CWB)
Midicolumns with Reservoirs

The kit provides all reagents necessary for the extraction of plasmid DNA from bacterial cells.

METHOD

1. Grow required cells in 50 mL LB broth over night at 37ºC
2. Pellet cells by centrifugation in 50 mL tubes (25 mL each tube) and discard the supernatant
3. Add 3mL of CRA and resuspend cell pellet by vortexing
4. Add 3mL of CLA and mix by inverting carefully for four times
5. Add 3mL of NSA and mix by inverting carefully for four times
6. Centrifuge for 15 min at ≥ 7,500 x g
7. Carefully decant the supernatant to a new centrifuge tube, avoiding the white precipitate
8. Thoroughly mix the provided Resin before removing an aliquot
9. Add 10mL of resin to the DNA solution and swirl to mix
10. Place a Midicolumn on the vacuum manifold port
11. Transfer the resin/DNA mixture into the column
12. Apply a vacuum for at least 15 inches of Hg to pull the mix into the column
13. When all of the sample has passed the column, break the vacuum at the source
14. Add 15mL of CWB to the column and apply vacuum again
15. When all of the wash solution has passed the column, break the vacuum at the source

16. And add another 15mL CWB and reapply a vacuum
17. When all of the wash solution has passed the column, dry the resin for additional 30 seconds
18. Do not dry longer than 30 seconds
19. Remove the Midicolumn from the manifold and separate the reservoir from the column by cutting with sharp scissors as shown in Figure 8-6.

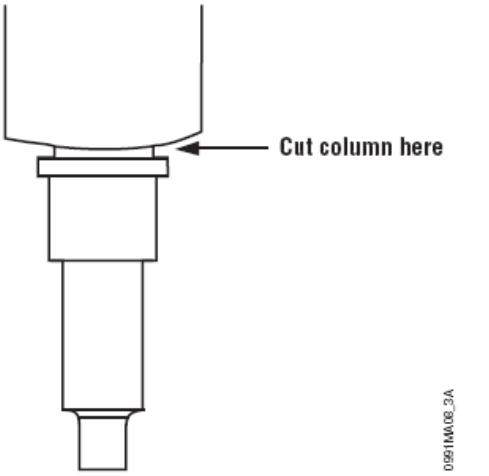


Figure 8-6. ***Location of the cut site for separating the reservoir from the Midicolumn***

20. Place a QIAquick spin column in a provided 1.5mL collection tube
21. Centrifuge for 2 minutes at ≥ 10,000 x g
22. Discard flow-through and place column in clean 1.5mL microcentrifuge tube
23. Add 0.75 mL Buffer PE to the QIA quick column and centrifuge for 30-60 seconds
24. Add 300 µL preheated water to the center of the membrane and let the column stand for 1 minute
25. Centrifuge for 30 seconds at ≥ 10,000 x g
26. Remove column and centrifuge for another 5 minutes at ≥ 10,000 x g
27. Carefully transfer supernatant to a clean microcentrifuge tube.

8.2.20 Determination of the Excision Frequency

MATERIALS

Buffers and Solutions

Ampicillin-Stock (50 mg/mL)	50 mg Dissolve in 500 µL Add 500 µL	Ampicillin Water Ethanol
Tetracycline-Stock (5 mg/mL)	5 mg 1000 µL	Tetracycline Ethanol

Media

LB Amp^{50} broth (containing 50 µg/mL Ampicillin)	250 µL 250 mL	Amp-Stock LB-broth (55ºC)
LB Tet^{12} broth (containing 15 µg/mL tetracycline)	250 µL 250 mL	Tet-Stock LB-broth (55ºC)
LB Amp^{50} plates (containing 50 µg/mL Ampicillin)	250 µL 250 mL	Amp-Stock LB-agar (55ºC)
LB Tet^{12} plates (containing 15 µg/mL tetracycline)	250 µL 250 mL	Tet-Stock LB-agar (55ºC)

METHOD

1. Inoculate transformants containing vector pIS629-Tc-AB in LB Tet12 broth overnight at 37ºC, with agitation
2. Dilute each sample serially and plated on selective LB plates containing either tetracycline or Ampicillin. (LB^{Tet}: dilution)
3. The appearance of Amp^{R} colonies is regarded as excision frequency of IS*629*-Tc

8.3 Primer

8.3.1 O Antigen Operon Amplification Primer

Primer	Sequences (5' - 3')	Product length [bp]	Genes
OAnt1 OAnt 13	CGTGAAGCGTCAACTACTGG ATTCGTTCCATCTGTAGAGCC	3,079	*wbdN, wzy, wbdO*
OAnt 8 OAnt 21	AGAAGGTATTCCTATATTTGATGATG GTTGAGTCCAGACATTCATTTACA	3,148	*wzy, wbdO, wzx*
OAnt 16 OAnt 29	GCTATATTACTTCTGGCATGATTG GGTTGGTTGTATTACAGCAATG	3,055	*wzx, per, wbdP*
OAnt 22 OAnt 35	TGGCATGACGTTATAGGCTAC CTACTATCCAACAAGTTCAATTCATC	2,954	*per, wbdP, gmd, fcl*
OAnt 30 OAnt 45	CTGACTGATTCATCTAACCTCACTA GCACGATAACCATATCTTCAAT	3,381	*gmd,fcl, wbdO, manC*
OAnt 40 OAnt 51	CGACTTTGCTACGACTTTCAG CGAAATCACGGAAGTAATGG	2,631	*manC, manB,*
OAnt 46 OAnt 61	GGCGACAAGGCTGAGTTC CACTGCCATACCGACTACG	3,488	*manB, wbdR, gnd*
OAnt 56 OAnt 69	AATGTATTCAAGGCAGGATTATC TACTGATTCTACACTGGATGTATTG	2,960	*wbdR, gnd*

8.3.2 O Antigen Operon Sequencing Primer

Primer	Sequences (5' - 3')	PCR Amplicons used for sequencing
OAnt 2	CTGACTGATGCCATTGCTG	OAnt 1+13
OAnt 3	CAGCAATGGCATCAGTCAG	OAnt 1+13
OAnt 4	GGATTATCATTTCGTACACTGGTAG	OAnt 1+13
OAnt 5	CTACCAGTGTACGAAATGATAATCC	OAnt 1+13
OAnt 6	TGACAGGAATCTATAATGCCAAC	OAnt 1+13
OAnt 7	GTTGGCATTATAGATTCCTGTCA	OAnt 1+13
OAnt 8	AGAAGGTATTCCTATATTTGATGATG	OAnt 1+13
OAnt 9	CATCATCAAATATAGGAATACCTTCT	OAnt 1+13
OAnt 79	GCGAAGAAGGCGTGGA	OAnt 1+13
OAnt 80	GATACCTTTGCGGAACTTCG	OAnt 1+13
OAnt 81	ATTGACGATTGTAGCACCG	OAnt 1+13
OAnt 82	CGGTGCTACAATCGTCAAT	OAnt 1+13
OAnt 83	GTGAAGTCAGCGGCTAAGTT	OAnt 1+13
OAnt 84	AACTTAGCCGCTGACTTCAC	OAnt 1+13
OAnt 10	GGAGAACGGTTACATAAACAAGT	OAnt 1+13 OAnt 8+21
OAnt 11	ACTTGTTTATGTAACCGTTCTCC	OAnt 1+13 OAnt 8+21
OAnt 12	GGCTCTACAGATGGAACGAAT	OAnt 1+13 OAnt 8+21
OAnt 85	AACCGTAATCCTAAAATAAAAAGA	OAnt 1+13 OAnt 8+21

OAnt 86	TCTTTTTATTTTAGGATTACGGTT	OAnt 1+13 OAnt 8+21
OAnt 87	TACTTAGTGGCTGGGAATGC	OAnt 1+13 OAnt 8+21
OAnt 88	GCATTCCCAGCCACTAAGTA	OAnt 1+13 OAnt 8+21
OAnt 89	AGGTGTTTCCTGAAGGGTTA	OAnt 8+21 OAnt 8+21
OAnt 13	ATTCGTTCCATCTGTAGAGCC	OAnt 8+21
OAnt 14	CGTTTAAGAAGAGTATTGCGG	OAnt 8+21
OAnt 15	CCGCAATACTCTTCTTAAACG	OAnt 8+21
OAnt 16	GCTATATTACTTCTGGCATGATTG	OAnt 8+21
OAnt 17	CAATCATGCCAGAAGTAATATAGC	OAnt 8+21
OAnt 90	TAACCCTTCAGGAAACACC	OAnt 8+21
OAnt 91	CTTTACCTTATTTAGGACAAGAGAG	OAnt 8+21
OAnt 92	CTCTCTTGTCCTAAATAAGGTAAAG	OAnt 8+21
OAnt 18	CCGTTATGGGCTGCTTATG	OAnt 8+21 OAnt 16+29
OAnt 19	CATAAGCAGCCCATAACGG	OAnt 8+21 OAnt 16+29
OAnt 20	TGTAAATGAATGTCTGGACTCAAC	OAnt 8+21 OAnt 16+29
OAnt 93	ATGCGGGACTACCAGTTTTA	OAnt 8+21 OAnt 16+29
OAnt 94	TAAAACTGGTAGTCCCGCAT	OAnt 8+21 OAnt 16+29
OAnt 95	CCAGCAAAATACATCATAGTTAG	OAnt 8+21 OAnt 16+29
OAnt 96	CTAACTATGATGTATTTTGCTGG	OAnt 8+21 OAnt 16+29
OAnt 21	GTTGAGTCCAGACATTCATTTACA	OAnt 16+29 OAnt 16+29
OAnt 22	TGGCATGACGTTATAGGCTAC	OAnt 16+29
OAnt 23	GTAGCCTATAACGTCATGCCA	OAnt 16+29
OAnt 24	GCTGAGGATCTTGGTTGGC	OAnt 16+29 OAnt 22+35
OAnt 25	GCCAACCAAGATCCTCAGC	OAnt 16+29 OAnt 22+35
OAnt 26	GATAGTAGCGATTGTGATGTTATTC	OAnt 16+29 OAnt 22+35
OAnt 27	GAATAACATCACAATCGCTACTATC	OAnt 16+29 OAnt 22+35
OAnt 28	CATTGCTGTAATACAACCAACC	OAnt 16+29 OAnt 22+35
OAnt 29	GGTTGGTTGTATTACAGCAATG	OAnt 22+35 OAnt 22+35
OAnt 97	CAAGTCCACAAGGAAAGTAAAG	OAnt 16+29 OAnt 22+35
OAnt 98	CTTTACTTTCCTTGTGGACTTG	OAnt 16+29 OAnt 22+35
OAnt 99	TTAGAAACGAAACCAGAAATA	OAnt 16+29 OAnt 22+35
OAnt 100	TATTTCTGGTTTCGTTTCTAA	OAnt 16+29 OAnt 22+35
OAnt 101	GCAAAACTACATTCTCTTCCA	OAnt 16+29 OAnt 22+35
OAnt 102	TGGAAGAGAATGTAGTTTTGC	OAnt 16+29 OAnt 22+35
OAnt 103	TTTATGAACCTACAACTCTGATAG	OAnt 22+35 OAnt 22+35
OAnt 30	CTGACTGATTCATCTAACCTCACTA	OAnt 22+35
OAnt 31	TAGTGAGGTTAGATGAATCAGTCAG	OAnt 22+35
OAnt 104	CTATCAGAGTTGTAGGTTCATAAA	OAnt 22+35
OAnt 32	GTCCAATACTCAGTCCGTCAG	OAnt 22+35 OAnt 30+45
OAnt 33	CTGACGGACTGAGTATTGGAC	OAnt 22+35 OAnt 30+45
OAnt 34	GATGAATTGAACTTGTTGGATAGTAG	OAnt 22+35 OAnt 30+45
OAnt 105	TTTATCCTCGTTCCCCTTATG	OAnt 22+35 OAnt 30+45
OAnt 106	CATAAGGGGAACGAGGATAAA	OAnt 22+35 OAnt 30+45
OAnt 107	AGTTGATACTTTGCTTGGAGAT	OAnt 22+35 OAnt 30+45

OAnt 108	ATCTCCAAGCAAAGTATCAACT	OAnt 22+35 OAnt 30+45
OAnt 35	CTACTATCCAACAAGTTCAATTCATC	OAnt 30+45 OAnt 30+45
OAnt 36	ATGTAGATGATATGGCTTCTGC	OAnt 30+45
OAnt 37	GCAGAAGCCATATCATCTACAT	OAnt 30+45
OAnt 38	TGGTATCTGGCAGCACTTCT	OAnt 30+45
OAnt 39	AGAAGTGCTGCCAGATACCA	OAnt 30+45
OAnt 40	CGACTTTGCTACGACTTTCAG	OAnt 30+45
OAnt 41	CTGAAAGTCGTAGCAAAGTCG	OAnt 30+45
OAnt 109	CACCAACCGATTATGGAAGA	OAnt 30+45
OAnt 110	TCTTCCATAATCGGTTGGTG	OAnt 30+45
OAnt 111	GGTCTTGAAAATACATACAACTGG	OAnt 30+45
OAnt 112	CCAGTTGTATGTATTTTCAAGACC	OAnt 30+45
OAnt 42	GAATATGCTGAAACTGGTTATGG	OAnt 30+45 OAnt 40+51
OAnt 43	CCATAACCAGTTTCAGCATATTC	OAnt 30+45 OAnt 40+51
OAnt 44	ATTGAAGATATGGTTATCGTGC	OAnt 30+45 OAnt 40+51
OAnt 45	GCACGATAACCATATCTTCAAT	OAnt 40+51 OAnt 40+51
OAnt 113	GCATCTGTTTATCTTGAGGAAT	OAnt 30+45 OAnt 40+51
OAnt 114	ATTCCTCAAGATAAACAGATGC	OAnt 30+45 OAnt 40+51
OAnt 46	GAAGAACTGAATGAAGATATTGCC	OAnt 40+51 OAnt 44+63
OAnt 47	GGCAATATCTTCATTCAGTTCTTC	OAnt 40+51 OAnt 44+63
OAnt 48	CCTCACGCCGCTCAAG	OAnt 40+51 OAnt 44+63
OAnt 49	CTTGAGCGGCGTGAGG	OAnt 40+51 OAnt 44+63
OAnt 50	CCATTACTTCCGTGATTTCG	OAnt 40+51 OAnt 44+63
OAnt 115	ATTGACCAAGGTGAGCGATA	OAnt 40+51 OAnt 44+63
OAnt 116	TATCGCTCACCTTGGTCAAT	OAnt 40+51 OAnt 44+63
OAnt 117	AAAGGTTTACAGGATGCGG	OAnt 40+51 OAnt 44+63
OAnt 118	CCGCATCCTGTAAACCTTT	OAnt 40+51 OAnt 44+63
OAnt 119	TTGACCGCTGTTTCCTGTT	OAnt 40+51 OAnt 44+63
OAnt 120	AACAGGAAACAGCGGTCTT	OAnt 40+51 OAnt 44+63
OAnt 51	CGAAATCACGGAAGTAATGG	OAnt 44+63 OAnt 44+63
OAnt 52	GCATTTCACGTCACGATTG	OAnt 44+63
OAnt 53	CAATCGTGACGTGAAATGC	OAnt 44+63
OAnt 54	GCTTGCCAGAAAGATATTGC	OAnt 44+63
OAnt 55	GCAATATCTTTCTGGCAAGC	OAnt 44+63
OAnt 56	AATGTATTCAAGGCAGGATTATC	OAnt 44+63
OAnt 57	GATAATCCTGCCTTGAATACATT	OAnt 44+63
OAnt 58	GTAATGGATATGTTGTTATTGAAGAC	OAnt 44+63 OAnt 56+69
OAnt 59	GTCTTCAATAACAACATATCCATTAC	OAnt 44+63 OAnt 56+69
OAnt 60	CGTAGTCGGTATGGCAGTG	OAnt 44+63 OAnt 56+69
OAnt 61	CACTGCCATACCGACTACG	OAnt 44+63 OAnt 56+69
OAnt 62	CTGAAGACGGTGAGCCAT	OAnt 44+63 OAnt 56+69
OAnt 121	GTGCGGTTGAATGTGGAAT	OAnt 44+63
OAnt 122	ATTCCACATTCAACCGCAC	OAnt 44+63
OAnt 123	GTCTTTGAGGAGATATTTACGC	OAnt 44+63

OAnt 124	GCGTAAATATCTCCTCAAAGAC	OAnt 44+63
OAnt 125	CTTTCAACCAACTGCTTTCTAA	OAnt 44+63 OAnt 56+69
OAnt 126	TTAGAAAGCAGTTGGTTGAAAG	OAnt 44+63 OAnt 56+69
OAnt 127	AGGCTTTAACTTCATCGGTAC	OAnt 44+63 OAnt 56+69
OAnt 128	GTACCGATGAAGTTAAAGCCT	OAnt 44+63 OAnt 56+69
OAnt 63	CTGCTCAGTTCACCGTTATTC	OAnt 56+69
OAnt 64	GGCGACAAGGCTGAGTTC	OAnt 56+69
OAnt 65	GCTGGTAGTCATCGGCAA	OAnt 56+69
OAnt 66	TGATAAAGAAGGTGTGTTCCATA	OAnt 56+69
OAnt 67	CGCTACGTTGTTCGATGGT	OAnt 56+69
OAnt 68	ATGTCATCATCGCCACTCC	OAnt 56+69

8.3.3 Core Antigen Gene Cluster Amplification Primer

Primer	Sequences (5' - 3')	Product length [bp]	Genes
Core 1	AGGATAACGATGACGCAGC	2,274	*waaQ, rfaG*
Core 13	GGTTATTCCATTCTTCTTACGG		
Core 8	AAATAACATGATTCAATTCTGGG	2,255	*rfaG, waaP, waaI*
Core 23	TAGATGCGATGGCTACACC		
Core 20	TTGGGCTTTATTTTTCTTCG	3,758	*waaP, waaI, waaY, waaJ, waaD*
Core 39	TTCCCTGACATTGCGTTGAT		
Core 28	CTGCTCGTGGACCTCTTGTA	3,136	*waaY, waaJ, waaD, waaL*
Core 43	TCCTGTTATTGGTTATGGTTACG		
Core 34	CACTGTCTTCATACATTATACTGGTA	3,737	*waaJ, waaD, waaL, rfaC*
Core 53	CTATCACAAAGTGCGTAAAGGTG		

8.3.4 Core Antigen Gene Cluster Sequencing Primer

Primer	Sequences (5' - 3')	PCR Amplicons used for sequencing
Core.2	CGCAGAGCAACAGTGAAAG	Core 1+13
Core.3	CTTTCACTGTTGCTCTGCG	Core 1+13
Core.4	TGCTGCTTTATCAGGACACC	Core 1+13
Core.5	GGTGTCCTGATAAAGCAGCA	Core 1+13
Core.6	GATTATCATCATCGGCAGTCT	Core 1+13
Core.7	AGACTGCCGATGATGATAATG	Core 1+13
Core.8	AAATAACATGATTCAATTCTGGG	Core 1+13
Core.54	AATCGGATAATCTCAACAAAAAG	Core 1+13
Core.55	CTTTTTGTTGAGATTATCCGATT	Core 1+13
Core.56	CAAGTTTTCCGCTGTGATT	Core 1+13
Core.57	AATCACAGCGGAAAACTTG	Core 1+13
Core.9	CCCAGAATTGAATCATGTTATTT	Core 1+13 Core 8+23
Core.10	GTTTATTATGCCGCTGATGTTT	Core 1+13 Core 8+23
Core.11	AAACATCAGCGGCATAATAAAC	Core 1+13 Core 8+23
Core.12	CCGTAAGAAGAATGGAATAACC	Core 1+13 Core 8+23
Core.13	GGTTATTCCATTCTTCTTACGG	Core 8+23
Core.14	TACCAGGAAGCAGCGGGAA	Core 8+23
Core.15	TTCCCGCTGCTTCCTGGTA	Core 8+23
Core.16	GGTGAGGTTTTTCGTGAACT	Core 8+23

Core.17	AGTTCACGAAAAACCTCAC	Core 8+23
Core.18	CTGACACCAACCATAAGTCTTGA	Core 8+23
Core.19	TCAAGACTTATGGTTGGTGTCAG	Core 8+23
Core.20	TTGGGCTTTATTTTTCTTCG	Core 8+23
Core.58	GCAGTAATGTTCACTTCTTCTC	Core 8+23
Core.59	GAGAAGAAGTGAACATTACTGC	Core 8+23
Core.60	TGCTACTGGAAGCGATTACTG	Core 8+23
Core.61	GCAGTAATCGCTTCCAGTAG	Core 8+23
Core.62	TGTTATATCTGTCATTTCCTGC	Core 8+23
Core.63	GCAGGAAATGACAGATATAACA	Core 8+23
Core.21	CGAAGAAAAATAAAGCCCAA	Core 8+23 Core 20+39
Core.22	GGTGTAGCCATCGCATCTA	Core 8+23 Core 20+39
Core.23	TAGATGCGATGGCTACACC	Core 20+39
Core.24	CTTCTGGATATTTCAATGCTGG	Core 20+39
Core.25	CCAGCATTGAAATATCCAGAAG	Core 20+39
Core.26	GTAAATAGCAATCAGTATCGGTA	Core 20+39
Core.27	TACCGATACTGATTGCTATTTAC	Core 20+39
Core.28	CTGCTCGTGGACCTCTTGTA	Core 20+39
Core.64	GACCTACAAAACCTTGGCAT	Core 20+39
Core.65	ATGCCAAGGTTTTGTAGGTC	Core 20+39
Core.29	TACAAGAGGTCCACGAGCAG	Core 20+39 Core 28+43
Core.30	TGATCTTGAACGGCACTATAATA	Core 20+39 Core 28+43
Core.31	TATTATAGTGCCGTTCAAGAT	Core 20+39 Core 28+43
Core.32	GCCAGTATCTATTTATTATCGTTT	Core 20+39 Core 28+43
Core.33	AAACGATAATAAATAGATACTGGC	Core 20+39 Core 28+43
Core.34	CACTGTCTTCATACATTATACTGGTA	Core 20+39 Core 28+43
Core.66	GTATTGATGAGAACTACCAGG	Core 20+39 Core 28+43
Core.67	CCTGGTAGTTCTCATCAATAC	Core 20+39 Core 28+43
Core.68	AAACACTTGCTTTACCAGAA	Core 20+39 Core 28+43 Core 34+53
Core.69	TTTCTGGTAAAGCAAGTGTT	Core 20+39 Core 28+43 Core 34+53
Core.70	AACGAAAGAAAGTGTTATTGTAGT	Core 20+39 Core 28+43 Core 34+53
Core.71	ACTACAATAACACTTTCTTTCGT	Core 20+39 Core 28+43 Core 34+53
Core.72	CCAATGACTAACACGGAAAG	Core 28+43 Core 28+43 Core 34+53
Core.35	TACCAGTATAATGTATGAAGACAGTG	Core 20+39 Core 28+43 Core 34+53
Core.36	TTGATCCTCTTCCATATTCACAGA	Core 20+39 Core 28+43 Core 34+53
Core.37	TCTGTGAATATGGAAGAGGAT	Core 20+39 Core 28+43 Core 34+53
Core.38	ATCAACGCAATGTCAGGGAA	Core 20+39 Core 28+43 Core 34+53
Core.39	TTCCCTGACATTGCGTTGAT	Core 28+43 Core 34+53
Core.40	AGCAAGGTCAGGACTTTGTG	Core 28+43 Core 34+53
Core.41	CACAAAGTCCTGACCTTGCT	Core 28+43 Core 34+53
Core.42	CGTAACCATAACCAATAACAGGA	Core 28+43 Core 34+53
Core.43	TCCTGTTATTGGTTATGGTTACG	Core 34+53 Core 34+53
Core.73	CTTTCCGTGTTAGTCATTGG	Core 34+53 Core 34+53
Core.44	TTATCCCTTGTTGATAGTCCTTAT	Core 34+53

Core.45	ATAAGGACTATCAACAAGGGA	Core 34+53
Core.46	AGGTGGTAATGAATAAGAATACG	Core 34+53
Core.47	CGTATTCTTATTCATTACCACCT	Core 34+53
Core.48	TAAGCAAATCCTTCCGCCA	Core 34+53
Core.49	TGGCGGAAGGATTTGCTTA	Core 34+53
Core.50	GTTTGCCAGTCCATGCCAT	Core 34+53
Core.51	ATGGCATGGACTGGCAAAC	Core 34+53
Core.52	CACCTTTACGCACTTTGTGATAG	Core 34+53
Core.74	GTGTCTATAATCTGCGAATGGC	Core 34+53
Core.75	GCCATTCGCAGATTATAGACAC	Core 34+53

8.3.5 *wec*A Amplification Primer

Primer	Sequences (5’ - 3’)	Product length [bp]	Genes
WecA 1 WecA 15	ATCGTCCGAAACGCTTCTT GGGGATAAACTGGATGTTGT	2,181	*wec*A

8.3.6 *wec*A Gene Cluster Sequencing Primer

Primer	Sequences (5’ - 3’)	PCR Amplicons used for sequencing
WecA.2	AAACTGGCAATGACCAAGAC	WecA1+15
WecA.3	GTCTTGGTCATTGCCAGTTT	WecA1+15
WecA.4	TCTGAATAAAGGTCTTCGTGGT	WecA1+15
WecA.5	ACCACGAAGACCTTTATTCAGA	WecA1+15
WecA.6	GCTGGATGACCGTTTTGAT	WecA1+15
WecA.7	ATCAAAACGGTCATCCAGC	WecA1+15
WecA.8	TCGGTATGATTTTGTGGTTC	WecA1+15
WecA.9	GAACCACAAAATCATACCGA	WecA1+15
WecA.10	ATGAGCCCATTCTCTCCTGA	WecA1+15
WecA.11	TCAGGAGAGAATGGGCTCAT	WecA1+15
WecA.12	CGTAGAAATCGTGGTGGC	WecA1+15
WecA.13	GCCACCACGATTTCTACG	WecA1+15
WecA.14	TATTCACCGTTGGTTGGAC	WecA1+15
WecA.15	GGGGATAAACTGGATGTTGT	WecA1+15
WecA16F	CGGTGTGCTTGTTTTCATTG	WecA1+15
WecA17R	CCGCATCACCCATAAAGACT	WecA1+15

8.3.7 *man*A Amplification Primer

Primer	Sequences (5’ - 3’)	Product length [bp]	Genes
ManA.1 ManA.14	ATGTCGGAGTTACGCAGGAAT ATGACGCCTACCGCTACCA	1,793	*man*A

8.3.8 *man*A Gene Cluster Sequencing Primer

Primer	Sequences (5’ - 3’)	PCR Amplicons used for sequencing
ManA.2	AGTGGAAAAGGAGCCTGATAAT	ManA1+14
ManA.3	ATTATCAGGCTCCTTTTCCACT	ManA1+14
ManA.4	GCGTTGACTGAACTTTATGGTA	ManA1+14
ManA.5	TACCATAAAGTTCAGTCAACGC	ManA1+14
ManA.6	CCAAACAAACACAAGTCTGAAAT	ManA1+14
ManA.7	ATTTCAGACTTGTGTTTGTTTGG	ManA1+14
ManA.8	CTCACTTTTTACAACAGCCTGA	ManA1+14
ManA.9	TCAGGCTGTTGTAAAAAGTGAG	ManA1+14
ManA.10	GGGTCTGACGCCTAAATACAT	ManA1+14
ManA.11	ATGTATTTAGGCGTCAGACCC	ManA1+14
ManA.12	ATGACCTTAGTGATAAAGAAACCA	ManA1+14
ManA.13	TGGTTTCTTTATCACTAAGGTCAT	ManA1+14
ManA.16R	TGTTGTAAACCCGTGCTAAAC	ManA1+14
ManA.17F	ACTGCCTTTCCTGTTCAAAGT	ManA1+14
ManA.15R	TCTTCACCCTGCATATTCAAC	ManA1+14

8.3.9 *gne* Amplification Primer

Primer	Sequences (5’ - 3’)	Product length [bp]	Genes
gne.13 gne.14	GCATTGACTACTCAGGGAAGGTA ATCTCTTTGGGAATCGCC	1,393	*gne*

8.3.10 *gne* Gene Cluster Sequencing Primer

Primer	Sequences (5’ - 3’)	PCR Amplicons used for sequencing
gne.2	AAACCTTACGATAACACTACACACC	gne13+14
gne.3	TGCCACAAAACCTGAAGAAT	gne13+14
gne.4	ATTCTTCAGGTTTTGTGGCA	gne13+14
gne.5	ATGTTCAGGTCTGGCTTATCAA	gne13+14
gne.6	TTGATAAGCCAGACCTGAACAT	gne13+14
gne.7	CGGTAGGACGGATGATAGTTA	gne13+14
gne.8	TAACTATCATCCGTCCTACCG	gne13+14
gne.9	GTTGACATCATAATAGAGAGAAGTAG	gne13+14
gne.10	CTACTTCTCTCTATTATGATGTCAAC	gne13+14
gne.11	TAAGTTTATGAATGGTCGCAA	gne13+14
gne.12	TTGCGACCATTCATAAACTTA	gne13+14
gne.15F	TGATAAACTCAACAATGTTTCCAAC	gne13+14
gne.16R	CGCAGGGACTAACTATAAGTC	gne13+14
gne.17F	ACGACGACTGGCTGAATAAGA	gne13+14
gne.18R	AGCACGAATGTCGTTTCCAT	gne13+14
IS*629*AF	TGAAAATCATCAGGTGTACGGT	gne13+14
IS*629*BR	GATAAACGCCACATAGACGAA	gne13+14

8.3.11 IS*629* Insertion Site Specific Primer

Primer sequences for the Amplification of each flanking IS*629* regions on the four *E. coli* genomes available. If IS absent size equal to 0 bp means that the primer pair was designed with one target region inside IS*629* therefore the IS*629* target site could not be observed.

Target	Primer	Sequences (5' - 3')	$MgCl_2$	16S rDNA Primers [µM]	Product length [bp]	
					IS present	IS absent
IS.1		primers that yielded specific amplicons could not be designed				
IS.2		primers that yielded specific amplicons could not be designed				
IS.3	Sak 3 IS FW	CCACAATGCGAAAGTGAACC	1.5 mM	VMP5 & VMP6 0.625	1,724	0
	Sak 3 REV	CCATCCTGACCATCAATAACC				
IS.4	Sak 4 FW	CTGGTTTGTTACTCAGCATTCT	1.5 mM	VMP5 & VMP6 0.625	1,730	339
	Sak 4 A REV	CAGTGACAGCAAGGACGGA				
IS.5	Sak 4 FW	Same as IS.4				
	Sak 4 A REV					
IS.6	Sak 5 FW	TCACCATCAACGAAAACCG	1.5 mM	VMP5 & VMP6 0.625	585	0
	Sak 5 REV	ATGCCAGTTCCACCATACG				
IS.7	Sak 5 FW	Same as IS.6				
	Sak 5 REV					
IS.8	EDL 8 /Sak 6 FW	ATCAGCAAAGTTGAACGAGG	2.5 mM	86 & 87 0.420	2027	717
	Sak 6 REV	GATCAAAACCGATATGTGAAT				
IS.9	EDL 8 /Sak 6 FW	Same as IS.8			1638	328
	Sak 6 REV					
IS.10	Sak 7 FW	ATCTACAAACTTGAGCAGGCAC	1.5 mM	VMP5 & VMP6 0.625	1,744	434
	Sak 7 REV	CCGATATTTCCGCATCTCC				
IS.11	Sak 8 FW	CCGCTAAGCCCTGAACTTT	1.5 mM	VMP5 & VMP6 0.625	1,778	454
	Sak 8 REV	CCTGATAACCGACAAAATCATC				
IS.12	Sak 9 FW	CCAAAATCAACAAAATGCCA	3.5 mM	VMP5 & VMP6 0.625	1,802	492
	Sak 9 REV	ATTAGCGAACAAACCACTCGTC				
IS.13	Sak 10 FW	CATCAGGAAGCGAAAGAAGA	2.5 mM	86 & 87 0.420	1,683	373
	Sak 10 REV	GAGGCACCCAATGGAACAA				

IS.14	Sak 11 FW	CAATGAACAAGGAGTAAACCAA	1.5 mM	VMP5 & VMP6 0.625	1,552	492
	Sak 11 REV	AAAACAGCCACGAAGCCAG				
IS.15	Sak 12 B FW	CCCAACTATCCTTCTTAGCCAGTA	2.5 mM	86 & 87 0.420	1,599	289
	Sak 12 B REV	CGGCGATTACGGTAACGAC				
IS.16	primers that yielded specific amplicons could not be designed					
IS.17	Sak 14 FW	TATTCAACTGCTCCATAACGG	2.5 mM	86 & 87 0.420	1,636	326
	Sak 14 REV	TAATCAAGGAACTGGTGACTCTC				
IS.18	Sak 15 FW	ATTGCCAAAATCAGAGGTGCT	3.5 mM	VMP5 & VMP6 0.625	1,664	356
	Sak 15 REV	GGTCGCTATCAGACGCTTCA				
IS.19	Sak 16 FW	GTGACCACCGACGCTGTAA	5.0 mM	VMP5 & VMP6 0.625	1,838	528
	Sak 16 REV	ATTTCTGATAGTTCGCACTCTG				
IS.20	Sak 17 FW	GCTGTGCTGTAAGTAACTCCCC	1.5 mM	VMP5 & VMP6 0.625	1,641	331
	Sak 17 REV	AAAGCGACTGTTGCCTGC				
IS.21	Sak 18 FW	TATTCGTGGGCTCAATGGAT	1.5 mM	VMP5 & VMP6 0.625	1,596	286
	Sak 18 REV	ATCATCGCTTCAGTTCCTGTATC				
IS.22	Sak 19 FW	TTACGGTGAACCAGACAGATAAC	1.5 mM	VMP5 & VMP6 0.625	1,559	245
	Sak 19 REV	GTAAACTTCTGACCATTGATGAATC				
IS.23	Sak 20 FW	AAGAAAGCGTGGCAAACAAA	4.5 mM	VMP5 & VMP6 0.833	1,640	330
	Sak 20 REV	TGACTGGCGGAGTGTGACTAA				
IS.24 /25	EDL 3&6 FW	TGAAAACATTATCCGACACACAT	4.5 mM	VMP5 & VMP6 0.625	1,834	530
	EDL 3&6 REV	CATCAACAACACGAAGGGAGTA				
Phage A	Phage B REV	GTCAATACTGCCATACGCTAAT	3.5 mM	VMP5 & VMP6 0.625	584	0
	Phage 1.1 FW	ACTTACGGATTCCAGAGTGC				
Phage B	Phage B REV	See above phage A			387	0
	Phage 2.1 FW	TAAGTGAGGAAGTGATAGGAAGTG				
IS.26	primers that yielded specific amplicons could not be designed					
IS.27	Sak 13 FW	TTCTTTGTCCTGATCTGCCC	4.5 mM	VMP5 & VMP6 0.833	1,698	388
	EDL 18 REV	ATGTCGGATGTTGTTGGTGA				
IS.28	primers that yielded specific amplicons could not be designed					

IS.29	EC 1 A FW	ATGCTTCGTTTCGTATCACACA	1.5 mM	VMP5 & VMP6 0.625	1,780	470
	EC 1 REV	ATTTACTTCTATGACTGTCCACCACTC				
IS.30	EC 3 FW	CTTACGACGGTCCTCTCTGATTT	3.5 mM	VMP6 & 86 0.625	1,525	215
	EC 3 REV	GCGGTGACGGAGACATACATC				
IS.31	EC 4.1 FW	AGTTTTATCAGGTTGTCTGTGCT	1.5 mM	VMP5 & VMP6 0.625	2,003	693
	EC 4.2 REV	TTCAGTTATGTCGTGGTCGG				
IS.32	EC 8 FW	AGAGCCAATCAACAGCACACT	2.5 mM	86 & 87 0.625	1,900	590
	EC 8 REV	GCCTTATCCAGACCAGAAAGTTT				
IS.33	EC11 IS FW	CGCTCAGACAGTGAACCG	1.5 mM	VMP5 & VMP6 0.420	1,335	0
	EC11 IS REV	CTCACCTGACACTGTGAACC				
IS.34	EC 13 FW	CCCAAATACGAAGTTGCTCAG	1.5 mM	VMP5 & VMP6 0.625	1,761	451
	EC 13 REV	CGGTAAGACTCATATCAGTATCAGGT				
IS.35	EC14 IS REV	TTCTGGCGTCGTTCTTGAA	3.5 mM	VMP5 & VMP6 0.625	1,697	0
	IS629 14 IS REV	GGCATCAACAGGGAGTACAG				
IS.36	primers that yielded specific amplicons could not be designed					
IS.37	EC 17 FW	GACGACTCACTGACTTCACCG	2.5 mM	86 & 87 0.420	2,030	720
	EC 17 REV	AGGACTTGGGGATGTGGATAT				
IS.38	EC 19 FW	CATCTTATCGGTCATTTTTATCG	5.0 mM	VMP5 & VMP6 0.420	2,221	636
	EC 19 REV	AGCCTCTTTCTTTGGCATCTT				
IS.39	gne 13	CTGAAAGAAGACGCTGGCTA	1.5 mM	VMP5 & VMP6 0.420	1,536	226
	gne 14	CGGGTATTCCGACGATATAAA				
IS.40	P1 A FW	CCTTGGGGTTATCCACTTATC	3.5 mM	VMP5 & VMP6 0.420	1,623	313
	P1 A REV	TCAGGCATACGCTGGTAACT				
IS.41	P2 A FW	AGAGATTATGCTGGCGAAAC	3.5 mM	VMP5 & VMP6 0.420	1,383	300
	P2 A REV	ACCCTCCTGTAATATGGCTGT				
IS.42	P3 A FW	ATGTTGTGTTGTATCTCTGGACTG	1.5 mM	VMP5 & VMP6 0.420	1,993	624
	P3 A REV	CACGGGAGACTTTATCTGACA				
IS.43	P4 A FW	CCTGACGGTTATCGGTGTAC	3.5 mM	VMP5 & VMP6 0.420	1,533	0
	P4 A REV	TTACTTTCGTGAACAAAATAGTGAT				

IS.44	P1 A FW	Same as IS.40	3.5 mM	VMP5 & VMP6	2,100	800
	P1 A REV			0.420		
IS.45	TW 3A FW	TCCCAAATAGCGTTATCCAG	1.5 mM	VMP5 & VMP6	1,880	570
	TW 3A REV	AAAACTTGTGGCAGAAGGAA		0.625		
IS.46	TW 15 FW	AGACCATTCTGACAATCCCA	1.5 mM	VMP5 & VMP6	1762	452
	TW 15 REV	GAAGATCAAACACAGGCGTAT		0.625		
IS.47	primers that yielded specific amplicons could not be designed					

8.4 Tables

8.4.1 Genes and their Function for the O Antigen Biosynthesis: *rfb* operon

Table 0-1. ***Genes and their functions of the O antigen operon***

O Antigen operon (*rfb* operon)			
Gene	**General Function**	**Specific Function**	**Size [bp]**
wbdN	Nucleotide Sugar Transfer	Glycosyl Transferase	783
wzy	O Unit Processing	O antigen Polymerase	394
wbdO	Nucleotide Sugar Transfer	Glycosyl Transferase	284
wzx	O Unit Processing	O antigen Flipase	463
per	Nucleotide Sugar synthesis	Perosamine Synthetase	364
wbdP	Nucleotide Sugar Transfer	Glycosyl Transferase	1,215
gmd	Nucleotide Sugar synthesis	GDP-Mannose Dehydrase	1,119
fcl	Nucleotide Sugar synthesis	Fucose Synthetase	966
wbdQ	Nucleotide Sugar Transfer	GDP-Mannose Mannosylhydrolase	510
manC	Nucleotide Sugar synthesis	Manose 1-P Guanosyltransferase	1,449
manB	Nucleotide Sugar synthesis	Phosphomannomutase	1,371
wbdR	Nucleotide Sugar Transfer	Acetyl Transferase	666

8.4.2 Genes and their Function for the O antigen Biosynthesis: *waa* Locus

Table 0-2. ***Genes and their functions of the inner and outer core antigen***

		***waa* locus**	
waaQ	Structure Modification	LPS core biosynthesis enzyme	1,059
rfaG	Nucleotide Sugar Transfer	Glycosyl Transferase	1,125
waaP	Structure Modification	LPS core biosynthesis enzyme	807
waaI		LPS core biosynthesis enzyme	1,008
waaY	Structure Modification	LPS core biosynthesis enzyme	708
waaJ		LPS core biosynthesis enzyme	1,014
waaD		LPS core biosynthesis enzyme	1,143
waaL	Ligase	LPS core biosynthesis enzyme	1,209
rfaC	Nucleotide Sugar Transfer	Heptosyl Tranferase	993

8.4.3 Genes and their Function for the O Antigen Biosynthesis: Ancillary Genes

Table 0-3. ***Genes and their functions of the three* ancillary genes *manA*, *wecA* and *gne***

		Ancillary genes	
manA	Nucleotide Sugar synthesis	Mannose-6-Phosphat Isomerase	1,176
wecA	O Unit Processing	GlcNAc-1-phosphate transferase	1,104
gne	Nucleotide Sugar synthesis	UDP-N-Acetylgalactosamine 4-Epimerase	996

8.5 Results: Tables and Figures

8.5.1 Bacterial growth

Table 0-4. Table for both E. coli strains comparing growth in LB-Media (MA6 and ATCC 35150)

Time [min]	OD [600 nm]	
	35150	MA6
0	0	0
30	0.033	0.026
60	0.057	0.045
90	0.107	0.093
120	0.281	0.237
150	0.496	0.454
180	0.778	0.678
210	1.043	0.978
240	1.229	1.149
270	1.329	1.236
300	1.420	1.341

OD: Optical density

8.5.2 O Antigen Operon

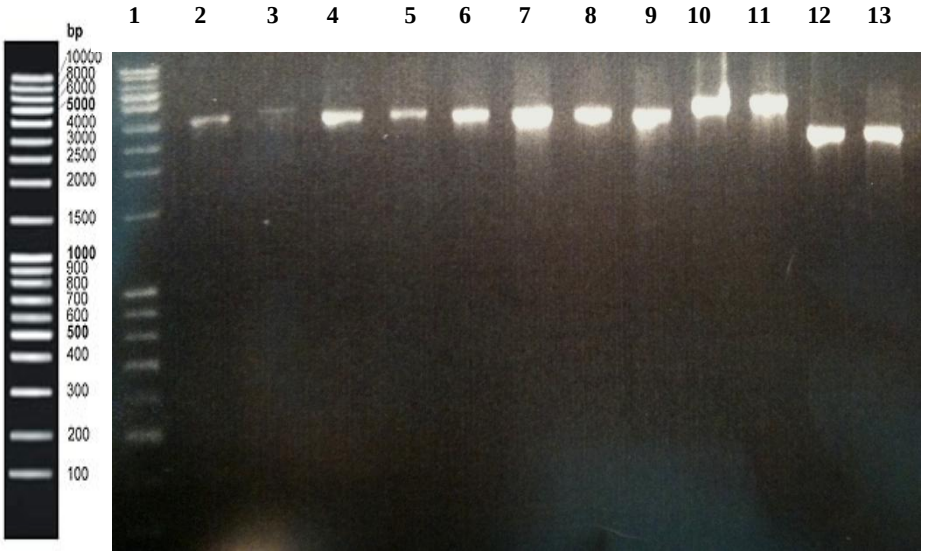


Figure 0-1. ***Gel-electrophoresis of the PCR amplification products of the O antigen operon (Part A)***

Lanes: 1: molecular weight ladder, 1kb Plus (exACTGene), 2: EDL931 and 3: MA6 (wbdN, wzy, wbdO), 4: EDL931 and 5: MA6 (wzy,wbdO, wzx), 6, EDL931 and 7: MA6 (wzx, per, wbdP), 8: EDL931 and 9: MA6 (wbdP, gmd, flc), 10: EDL931 and 11: MA6 (flc, wbdQ, manC), 12: EDL931 and 13: MA6 (manC, manB).;

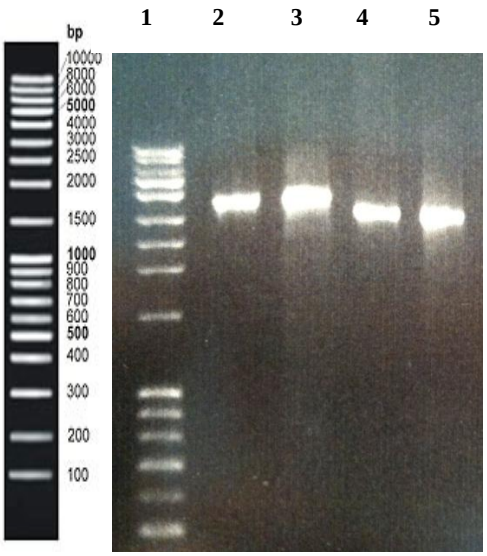


Figure 0-2. ***Gel-electrophoresis of the PCR amplification products of the O antigen operon (Part B)***

Lanes: 1: molecular weight ladder, 1kb Plus (exACTGene), 2: EDL931 and 3: MA6 (manB, wbdR, gnd), 4: EDL931 and 5: MA6 (wbdR, gnd).

8.5.3 Core Antigen Gene Cluster

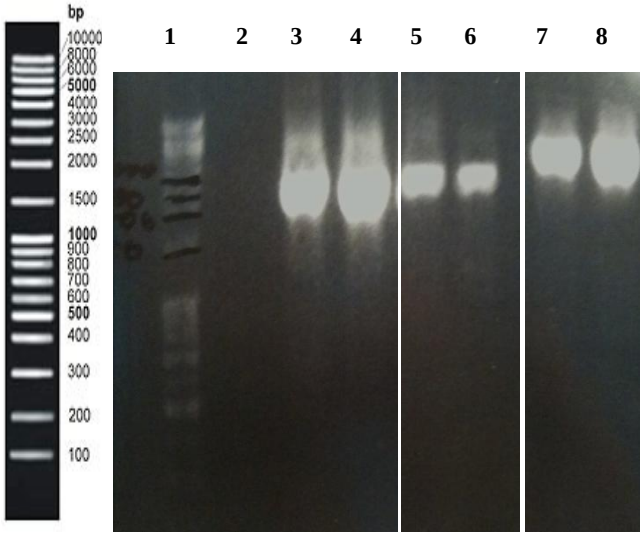


Figure 0-3. ***Gel-electrophoresis of the PCR amplification products of the core antigen operon (Part A)***
Lanes: 1: molecular weight ladder, 1kb Plus, (exACTGene), 2: Blank, 3: EDL931 and 4: MA6 (waaQ, rfaG), 5: EDL931 and 6: MA6 (rfaG, waaP, waaI), 7: EDL931 and 8: MA6 (waaY, waaJ, wad, waaL).

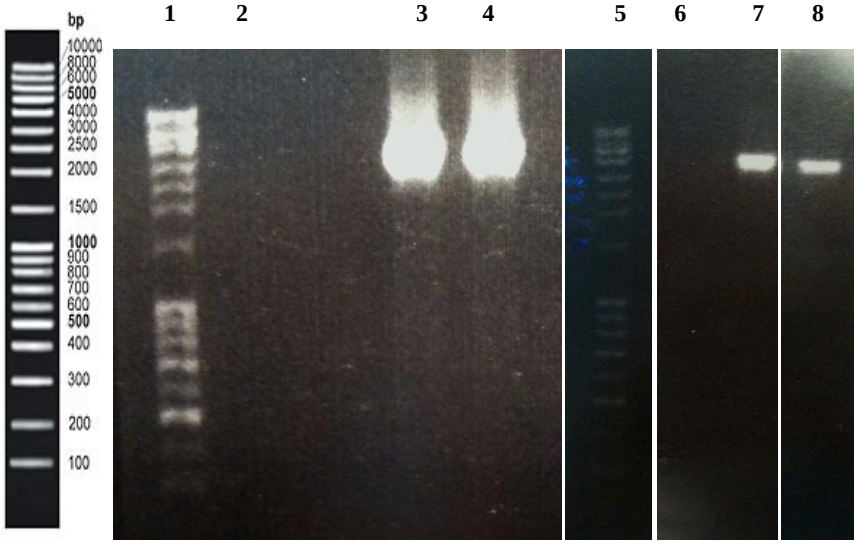


Figure 0-4. ***Gel-electrophoresis of the PCR amplification products of the core antigen operon (Part B)***
Lanes: 1: molecular weight ladder, 1kb Plus, (exACTGene), 2: Blank, 3: EDL931 and 4: MA6 (waaP, waaI, waaY, waaJ, wad), 5: molecular weight ladder, 1kb Plus, (exACTGene), 6: Blank, 7: EDL931 and 8: MA6 (waaJ, wad, waaL, rfaC).

8.5.4 Ancillary genes

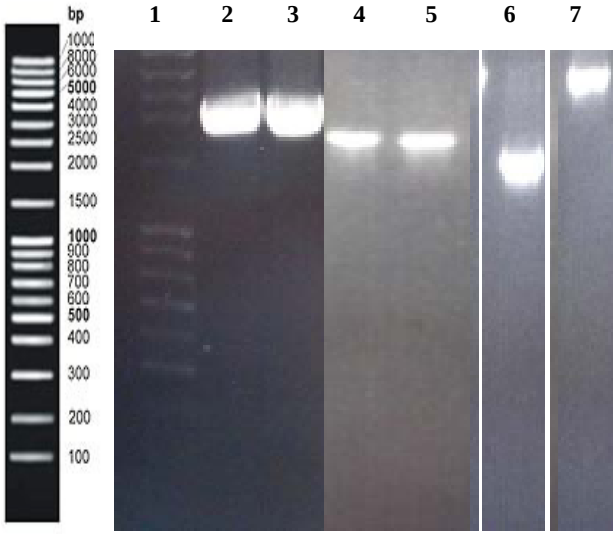


Figure 0-5 manA and wecA. ***Gel-electrophoresis of the PCR amplification products of the ancillary genes***

Lanes: 1: molecular weight ladder, 1kb Plus, (exACTGene), 2: EDL931 and 3: MA6 (manA), 4: EDL931 and 5: MA6 (wecA), 6: EDL931 and 7: MA6 (gne).

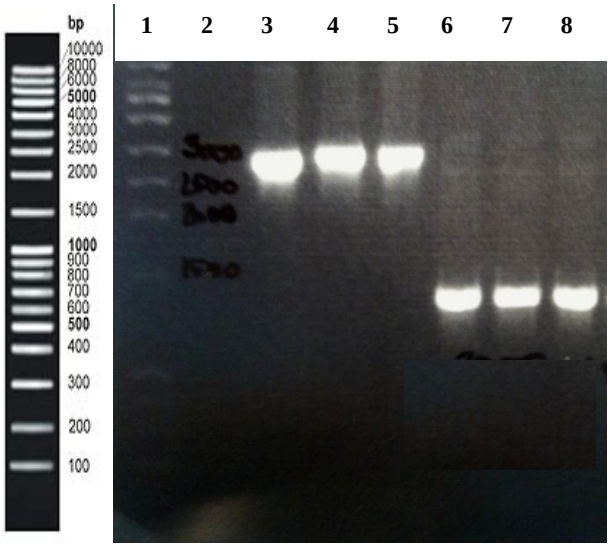


Figure 0-6 gne. ***Gel-electrophoresis of the PCR amplification products of gne in different strains.***

Lanes: 1: molecular weight ladder, 1kb Plus, (exACTGene), 2: Blank, 3: MA6, 4: MA6, 5: MA6, 6: EDL93, 7: EDL933, 8: 493/89.

8.5.5 PFGE Comparison of the Strains used in this Study with PulseNet

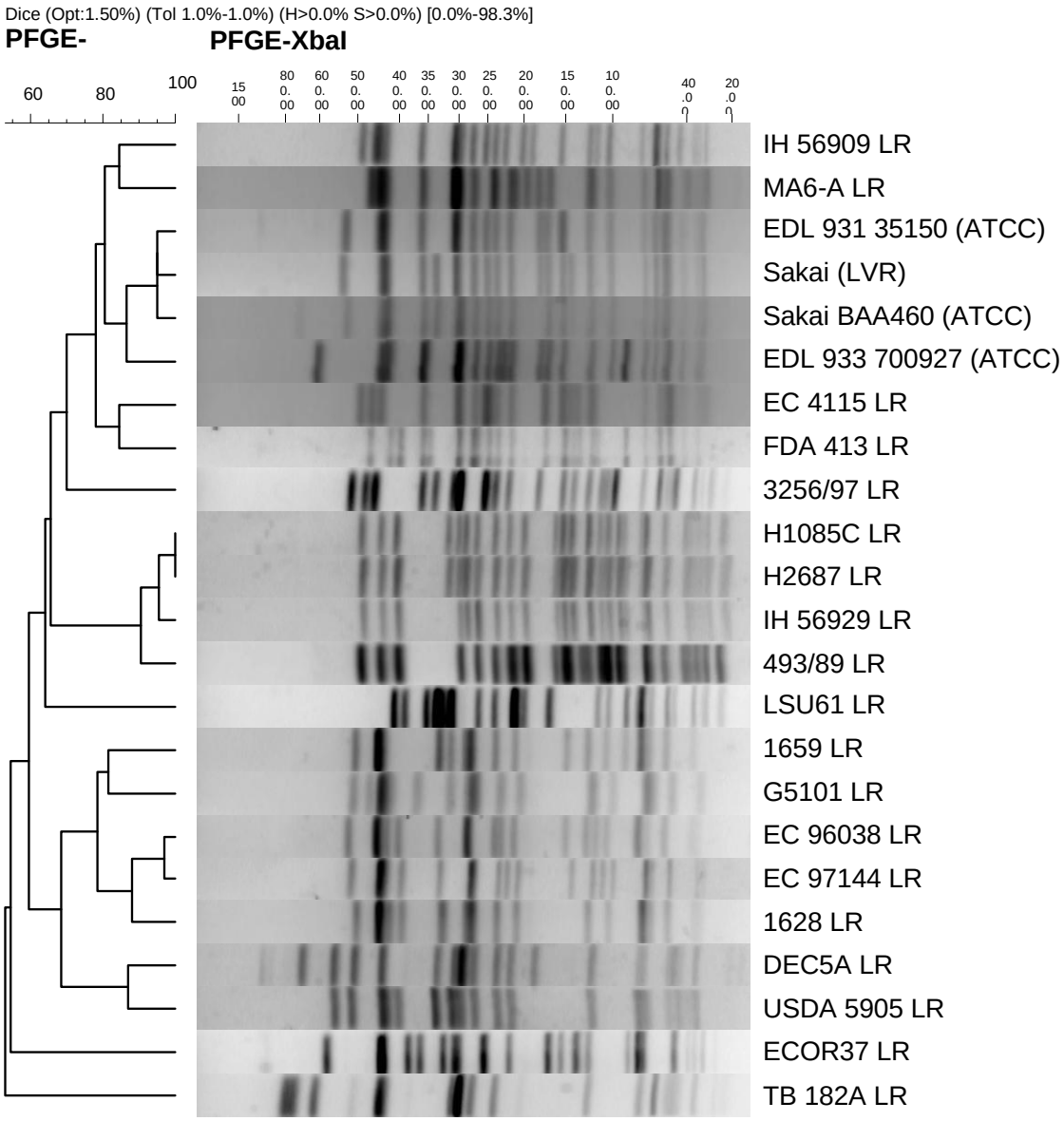


Figure 0-7. ***Pulse-field gel electrophoresis of XbaI digested DNA from the strains used in this study***

This unweighted pair-group method with arithmetic mean dendrogram was generated in BioNumerics software (Applied Maths, St-Martens-Latem, Belgium) by using Die coefficient with a 1.0% lane optimization and 1.0% band positiontolerance. The scale above the dendrogram indicates percent similarity.

8.5.6 IS*629* prevalence in *E. coli* O157:H7 genomes

Table 0-5. ***Genomes and plasmids investigated by "in silico" analysis***

Strain	Serotype	Accession No.	
		Genomes	Plasmids
Sakai	O157:H7	NC_002695	AB011549 (pO157)
EDL933	O157:H7	AE005174	AF074613 (pO157)
EC4115	O157:H7	NC_011353	CP001163 (pO157)
TW14359	O157:H7	CP001368	CP001369 (pO157)
3072/96	SFO157	NA	NC_009602.1 (pSFO157)
CB9615	O55:H7	NC_013941	CP001847.1 (pO55)

NA: not available.

8.5.7 IS*629* Target Site Specificity ("hot spots") on Genomes and Plasmids of Four *E. coli* O157:H7 Strains

Table 0-6. ***IS629 insertion sites in O157:H7 strains found in genomes available in Genbank***

NR.	Phage or backbone	Upstream sequence	Downsteam sequence	SAKAI location		EDL933 location		EC4115 location		TW14359 location	
IS.1	Sp 4	ATATGCGGAT	TTAATTAATG	1182129	1183201					1187070	1188379
IS.2	Sp 4	ATATGCGGAT	TTAATTAATG					1185471	1186780		
IS.3	Sp 5 stx2	AATGCGAAAG	AAAGTGGCGA	1270361	1271670						
IS.4	SpLE 1	AGGAGGCTGG	TGGTGATTGT			1120064	1121373				
IS.5	SpLE 1	AGGAGGCTGG	TGGTGATTGT	**1431898**	**1433207**	**1515671**	**1516980**	**1375483**	**1376792**	**1375770**	**1377079**
IS.6	SpLE 1	GAATCAGGTT	truncated			1129492	1129750				
IS.7	SpLE 1	GAATCAGGTT	truncated	**1441326**	**1441584**	**1525098**	**1525356**	**1384911**	**1385169**	**1385198**	**1385456**
IS.8	Sp 8	TGTTGATAAA	CCACTGATGC	1663333	1664642			1600004	1601313		
IS.9	Sp 8	CTGGTCTTTG	CCACTGATGC			1754739	1756048				
IS.10	backbone	TCTTGCCATC	ATCAATCACT	1685465	1686774	1776871	1778180				
IS.11	backbone	CGCTGACTTT	TTTCGGTCAA	1909167	1910476	2176690	2177999				
IS.12	Sp 12	ACGATGGCCT	CCTTTCAGTT	2205544	2204235						
IS.13	backbone	AAAAACCCTG	CTGTTAATAA	**2454183**	**2455492**	**2529386**	**2530695**	**2360477**	**2361786**	**2359458**	**2360767**
IS.14	Sp 13	truncated	ATCTGCTGGC	**2603612**	**2604668**	**2678823**	**2679879**	**2509906**	**2510962**	**2508887**	**2509943**
IS.15	Sp 14	TTAGCCAGTA	GTACTTCTCC	**2693477**	**2694786**	**2141966**	**2143277**	**2601091**	**2602400**	**2600071**	**2601380**
IS.16	SpLE 2	AAGAAGGTGG	GTCTCAGTCC	2741821	2743130						
IS.17	backbone	CAACAAGGAT	GATGAAATAG	2890776	2892085	2961007	2962316				
IS.18	Sp 15 stx1	CTCCTTGCCC	CCCATTTCAT	2912202	2913511	2761491	2762801				
IS.19	backbone	CCAGCCGGGC	GGATGCCCCA	3088838	3090147			3130200	3131509	3085071	3086380
IS.20	Sp 17	GTGCGCCATC	GTTCGGCGGT	3480689	3481998			3584797	3586106	3539674	3540983
IS.21	SpLE3	ACAAACTGAT	CACAGCCAGA	3869333	3870642	3936646	3937955			3928456	3929765
IS.22	backbone	ATTAAAACCA	GGGCGATGTT	4033206	4034515	4100422	4101731				
IS.23	SpLE 5	ATTAGCTCAG	CAGAATTGAT	5349269	5350578	5379272	5380581				

IS.24	SpLE 1	AAATACGCGC	CGCTGGTGCG			1142937	1144246				
IS.25	SpLE 1	AAATACGCGC	CGCTGGTGCG			1538543	1539852				
IS.26	933O	ATATGCGGAT	CCCATTTCAT			1896355	1897664				
IS.27	SpLE 2	AAGAAGGTGG	GTTTTCACTT			2818186	2819495				
IS.28	933Y	GATTACCCTG	GTACTTCTCC			3550226	3551537				
IS.29	Sp 1	CAATTGGTTA	TTATGGCTGT					304028	305337	304028	305337
IS.30	Sp 4	TGACGTGGTG	ATGACACACC					1189927	1191236		
IS.31	Phage	GAAATGGACT	GACTTCATGA					1321708	1320399	1321994	1320685
IS.32	backbone	TACCAGAAAC	AACTGATGCA					2157429	2158738	2156410	2157719
IS.33	Sp 13	GCTCAGACAG	CAGTGTCAGG					2519243	2520552	2519533	2518224
IS.34	backbone	ATCAGTAGAT	GATATATTCT					2646308	2647617	2645289	2646598
IS.35	Sp 5 stx 2	CAGTACGAGA	AGAACGACGC					2713108	2714417	2712089	2713398
IS.36	backbone	CTACGCGCAT	CATTCCTGCG					2971677	2972986	2926548	2927857
IS.37	Phage	ACGGGACTGG	TGGAGATAGT					3258306	3259615	3213177	3214486
IS.38	backbone	CGCTATAGCC	GCCAGCATCT					4259157	4260466	4215198	4216507
IS.40	pO157	TACGGTACTG	GTGACAATTT	**30694**	**32003**	**11103**	**10072**	**64171**	**63140**	**79197**	**80230**
IS.41	pO157	truncated	GTGTTTGTTA	80618	79587						
IS.42	pO157	ATGCAGTTTA	TTAAGGTTTA					13013	14322	29095	30404
IS.43	pO157	ATCCGGTGAT	truncated	**66033**	**66338**	**88594**	**88899**	**49585**	**49890**	**65642**	**65947**
IS.44	pO157	GTGAAGTGTG	GTGACAATTT					16109	17418	32189	33500
IS.45	pO157	TGCTCCGAAG	AAGTGTTTCG							315206	316515
IS.46	backbone	TGCCACTGTG	GTGGATATGT							2869454	2870763
IS.47	backbone	GATTACCCTG	GTTCGGCGGT			1871035	1872344				
IS.48	pO157	TACGGTACTG	CCACTCAGTT			53269	54578				

In bold are the locations shared by the four O157:H7 strains. The direct repeats (duplication are in red). NR, number of IS629, starting with all sites in Sakai, then all additional, unshared sites from EDL933, followed by additional unshared sites found in EC4115 and sites found in the plasmids and unshared IS629 insertion sites of strain TW1435. Truncated, partial IS629, with deletion in various regions of the IS element.

8.5.8 IS*629* presence in strains belonging to the stepwise model of emergence of *E. coli* O157:H7

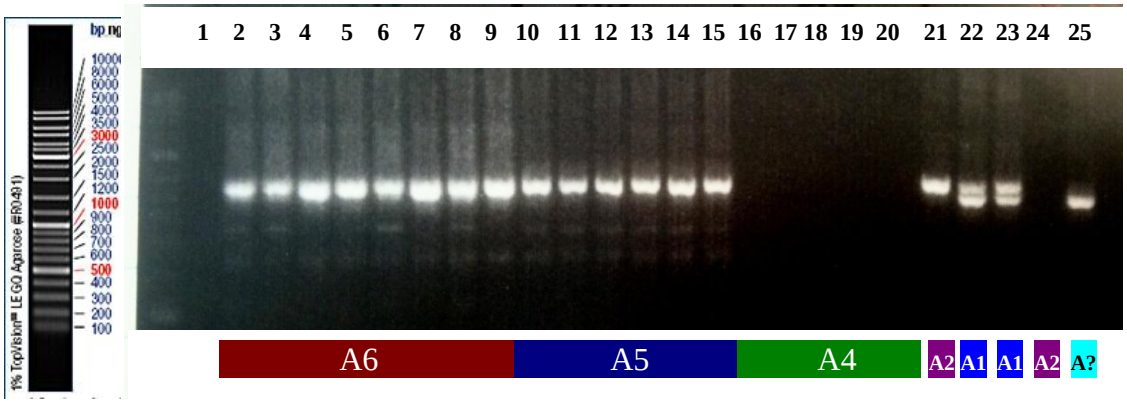


***Figure 0-8*. Gel-electrophoresis of the PCR amplification products for the presence of IS629 in different CC strains**

Lanes: 1: molecular weight ladder, (Gene Ruler), 2: Blank, 3-25- strains Table 3-1. A1-A6: Clonal complexes, A?: CC unknown.

8.5.9 IS*629* distribution in strains belonging to the stepwise model of emergence of *E. coli* O157:H7

Table 0-7. IS629 presence/absence in CC strains from the O157:H7 stepwise evolutionary model

Site	A6								A5						A4						A2		A1		A?	A6	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
IS. 3	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 4	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 5	+	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 6	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 7	+	+	+	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 8	+	-	+	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 9	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 10	+	+	-	-	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 11	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 12	+	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 13	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 14	+	+	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 15	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 17	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 18	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 19	+	-	+	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 20	+	-	+	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 21	+	+	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 23	+	+	-	-	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 24	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 25	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 27	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 29	-	-	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 30	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 31	-	-	+	+	-	-	-	-	+	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 32	-	-	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 33	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 34	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 35	-	-	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 37	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 38	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 39	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 40	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 41	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 42	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 44	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 45	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 46	-	-	-	+	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 0-8. ***IS629 target site presence/absence in CC strains from the O157:H7 stepwise evolutionary model***

	A6								A5						A4						A2		A1		A?	A6	
PRIMER	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
IS. 3	Site not determined – forward primer located inside IS*629*																										
IS. 4	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 5	+	+	+	+	+	-	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 6	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 7	+	+	+	+	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 8	+	+	+	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 9	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 10	+	+	-	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 11	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 12	+	+	+	+	+	-	+	-	+	-	+	+	+	+	+	-	+	+	+	+	+	-	-	-	+	+	+
IS. 13	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 14	+	+	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 15	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	+	+	+
IS. 17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 18	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 19	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 20	+	-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 21	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 23	+	+	-	-	+	-	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 24	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 25	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 27	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
IS. 29	+	+	+	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 30	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 31	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+	+	+
IS. 32	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 33	Site not determined – forward primer located inside IS*629*																										
IS. 34	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 35	Site not determined – forward primer located inside IS*629*																										
IS. 37	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IS. 38	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 39	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 40	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
IS. 41	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
IS. 42	+	+	+	+	+	+	+	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+	+	+
IS. 44	-	-	+	+	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+	-	-
IS. 45	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
IS. 46	+	+	+	+	+	+	+	+	+	+	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+	+	+

8.5.10 IS*629* insertion site in *E. coli* O55:H7

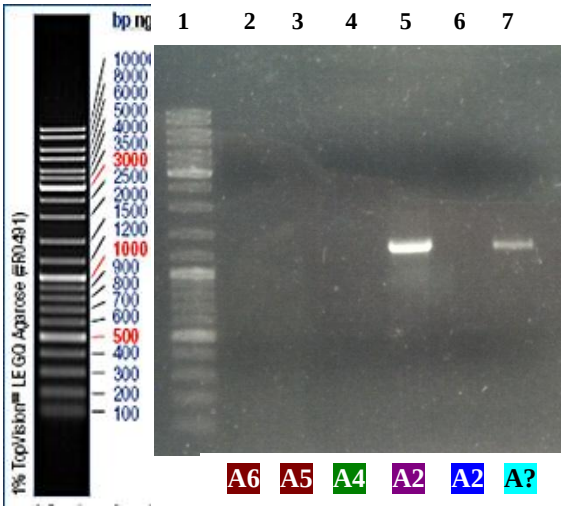


Figure 0-9. ***O55: Gel-electrophoresis of the PCR amplification products for IS629 located on the O55 genome in different CC strains.***
Lanes: 1: molecular weight ladder, (GeneRuler), 2: Sakai, 3: 493-89, 4: DEC5a, 5: USDA 5905, 6: TB182A, 7: 3256-9, 8: LSU-61.

8.5.11 Plasmid comparison

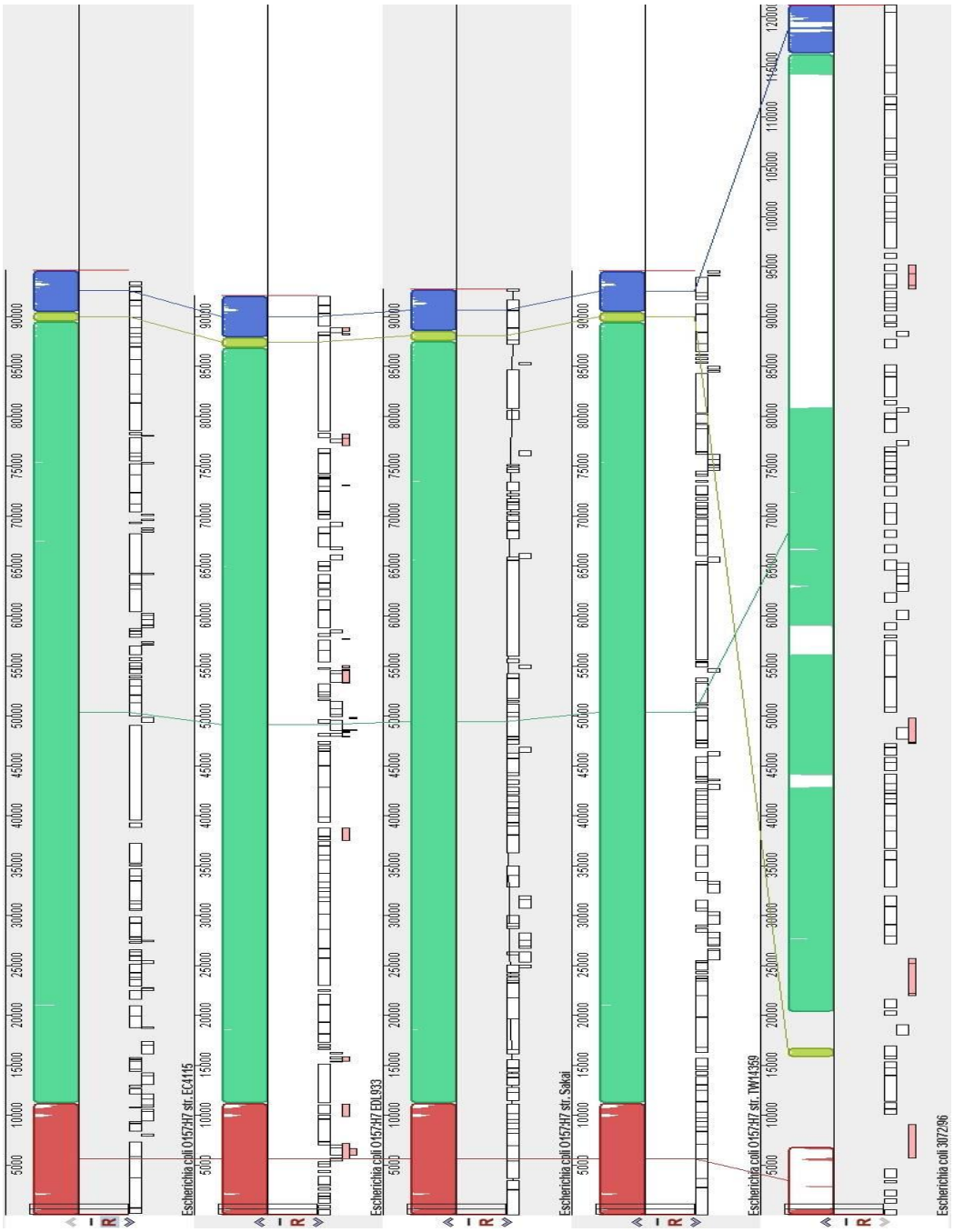


Figure 0-10. ***Comparison of the four pO157 plasmids with the pSFO157 plasmid using MAUVE***

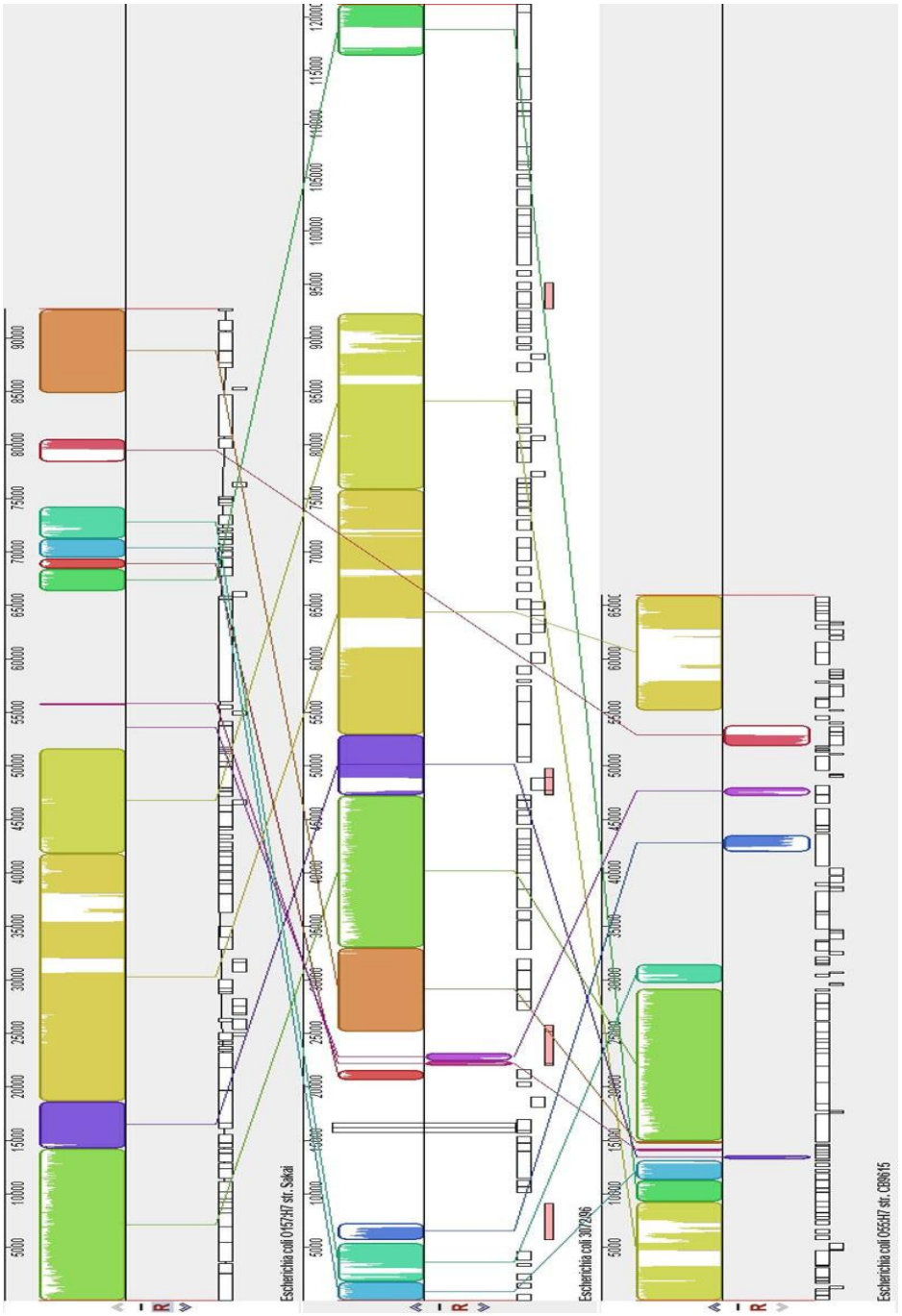


Figure 0-11. ***Comparison of pO157, pSFO157 and pO55 using MAUVE.***

8.5.12 IS*629* insertions in pO157

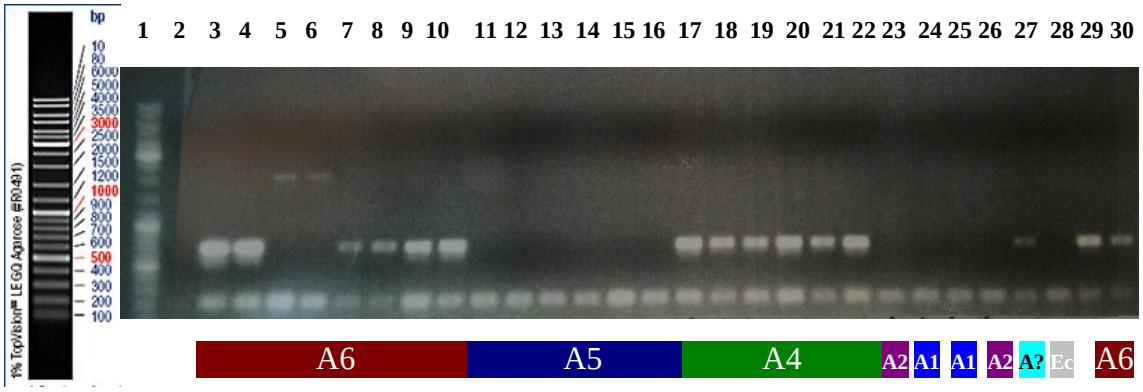


***Figure 0-12*. IS.42: Gel-electrophoresis of the PCR amplification products for IS629 insertion site IS.42 located on pO157 in different CC strains.**
Lanes: 1: molecular weight ladder, (Gene Ruler), 2: Blank, 3-30- strains Table 3-1. A1-A6: Clonal complexes, A?: CC unknown, Ec: ECOR.

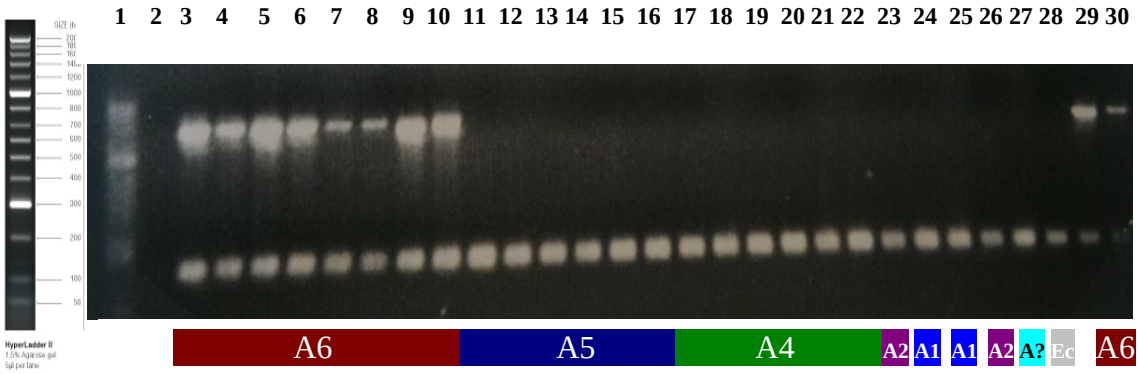


***Figure 0-13*. IS.41: Gel-electrophoresis of the PCR amplification products for IS629 insertion site IS.41 located on pO157 in different CC strains.**
Lanes: 1: molecular weight ladder, (HyperLadder II); 2: Blank, 3-30- strains Table 3-1. A1-A6: Clonal complexes, A?: CC unknown, Ec: ECOR.

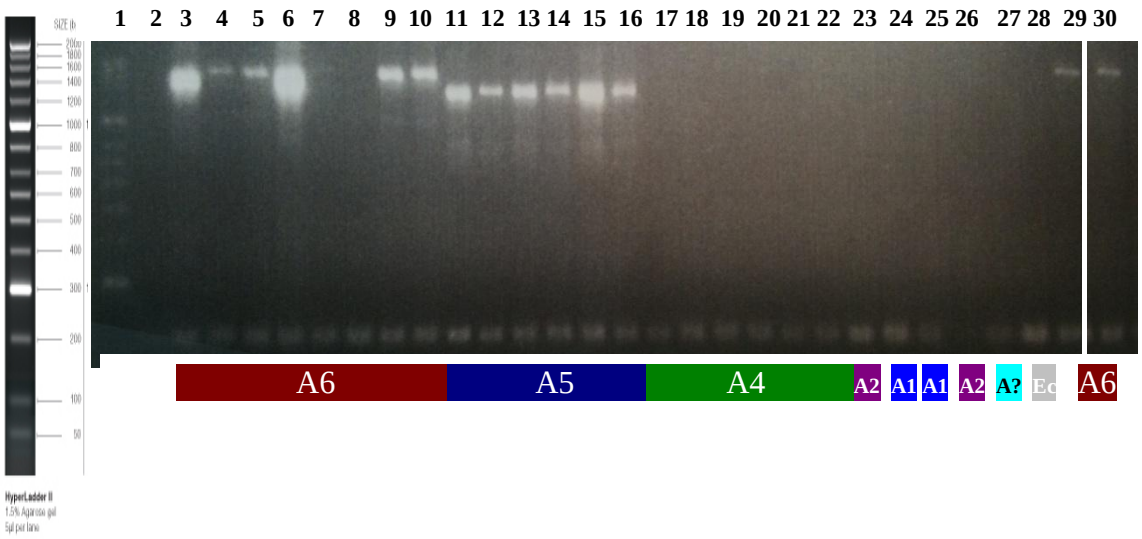


***Figure 0-14*. IS.43: Gel-electrophoresis of the PCR amplification products for IS629 insertion site IS.43 located on pO157 in different CC strains.**
Lanes: 1: molecular weight ladder, (GeneRuler), 2: Blank, 3-30- strains Table 3-1. A1-A6: Clonal complexes, A?: CC unknown, Ec: ECOR.

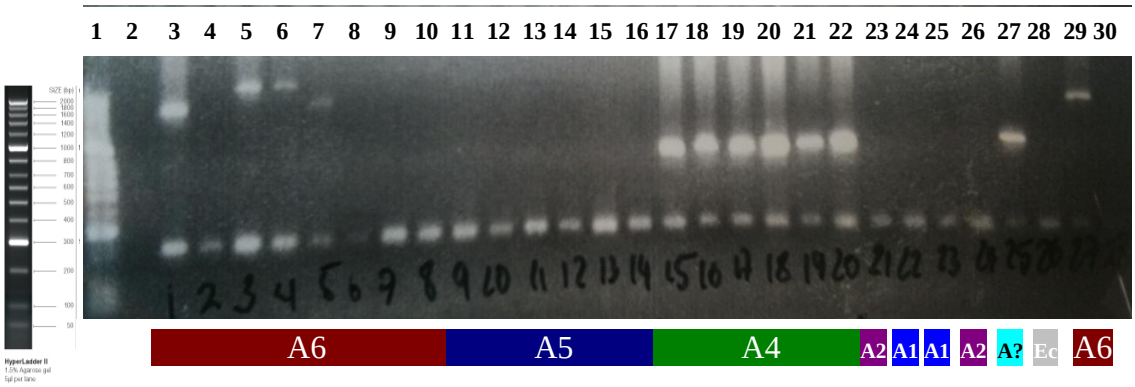


***Figure 0-15*. IS.40/IS.44: Gel-electrophoresis of the PCR amplification products for IS629 insertion site IS.40/IS.44 located on pO157 in different CC strains**
Lanes: 1: molecular weight ladder, (GeneRuler), 2: Blank, 3-30- strains Table 3-1. A1-A6: Clonal complexes, A?: CC unknown, Ec: ECOR.

8.5.13 Presence of pO157, pSFO157 and pO55

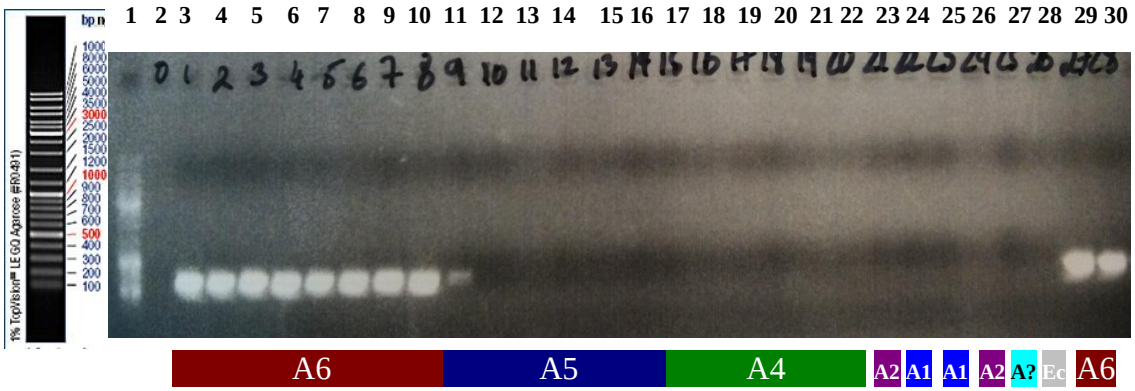


Figure 0-16. ***pO157: Gel-electrophoresis of the PCR amplification products with pO157 specific primer (pO157 F/R) in different CC strains.***
Lanes: 1: molecular weight ladder, (GeneRuler), 2: Blank, 3-30- strains Table 3-1. A1-A6: Clonal complexes, A?: CC unknown, Ec: ECOR.

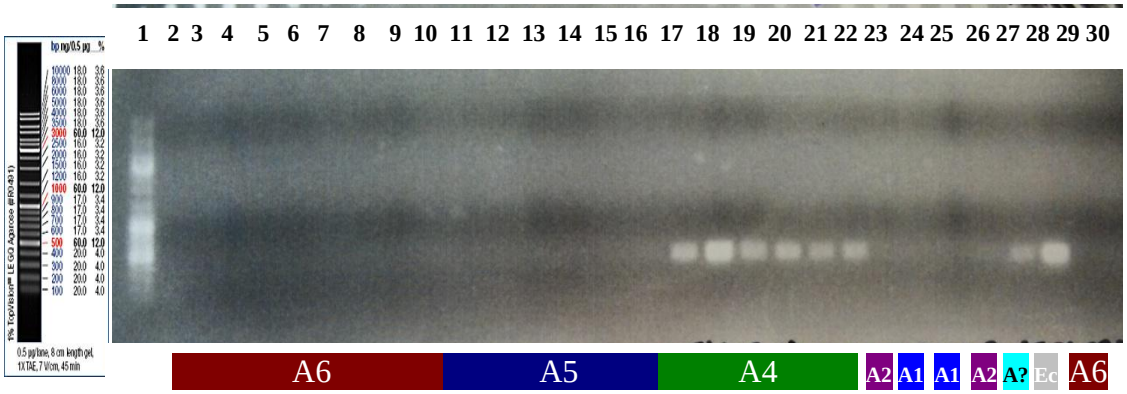


Figure 0-17. ***pSFO157: Gel-electrophoresis of the PCR amplification products with pSFO157 specific primer (pSFO157 F/R) in different CC strains.***
Lanes: 1: molecular weight ladder, (GeneRuler), 2: Blank, 3-30- strains Table 3-1. A1-A6: Clonal complexes, A?: CC unknown, Ec: ECOR.

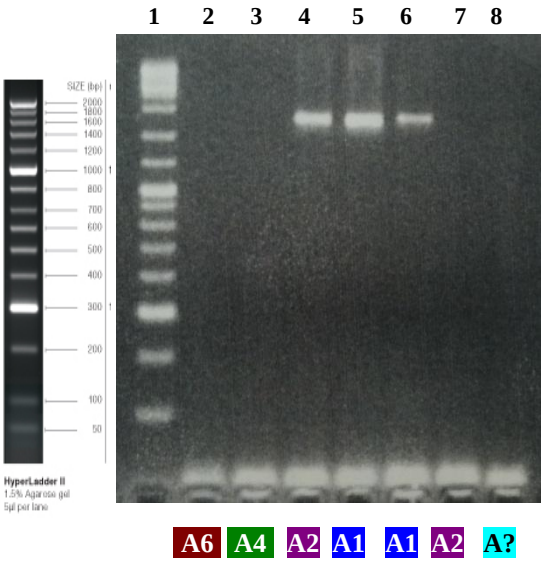


Figure 0-18. ***pO55: Gel-electrophoresis of the of the PCR amplification products for IS629 insertion located on the pO55 plasmid in different CC strains..***

Lanes: 1: molecular weight ladder, (GeneRuler), 2: Sakai, 3: 493-89, 4: DEC5a, 5: USDA 5905, 6: TB182A, 7: 3256-9, 8: LSU-6.

9 REFERENCES

Achtman M., Pluschke G.; Clonal analysis of descent and virulence among selected *Escherichia coli*; *Annu Rev Microbiol;* (40): 185-210; 1986.

Amann E., Brosius J.; ATG vectors' for regulated high-level expression of cloned genes in *Escherichia coli*; *Gene;* (40): 183-190; 1985.

Ammon A., Petersen L.R., Karch H.; A large outbreak of hemolytic uremic syndrome caused by an unusual sorbitol-fermenting strain of Escherichia coli O157:H-; *J Infect Dis;* (179): 1274-1277; 1999.

Arbeit R.D.; Laboratory procedures for the epidemiologic analysis of microorganisms; In: Manual of clinical microbiology; (Eds.Murray PJ, Baron EJ, Pfaller MA, Tenover FC, Yolken RH); Washington, D.C.: *ASM Press*; 6: 190-208;1995.

Balbas P., Soberon X., Merino E., Zurita M., Lomeli H., Valle F., Flores N., Bolivar F.; Plasmid vector pBR322 and its special-purpose derivatives--a review; *Gene;* (50): 3-40; 1986.

Barondess J.J., Beckwith J.; bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein; *J Bacteriol;* (177): 1247-1253; 1995.

Bell B.P., Goldoft M., Griffin P.M., Davis M.A., Gordon D.C., Tarr P.I., Bartleson C.A., Lewis J.H., Barrett T.J., Wells J.G., .; A multistate outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience; *JAMA;* (272): 1349-1353; 1994.

Beutin L.; *Escherichia coli* O157 and other types of verocytotoxigenic *E. coli* (VTEC) isolated from humans, animals and food in Germany; In: *Escherichia coli* O157 in Farm Animals; (Eds.Stewart CS, Flint HJ); Wallingford, Oxon, UK: *CABI Publishing*; 121-145;1999.

Beutin L., Aleksic S., Zimmermann S., Gleier K.; Virulence factors and phenotypical traits of verotoxigenic strains of *Escherichia coli* isolated from human patients in Germany; *Med Microbiol Immunol;* (183): 13-21; 1994.

Beutin L., Krause G., Zimmermann S., Kaulfuss S., Gleier K.; Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period; *J Clin Microbiol;* (42): 1099-1108; 2004.

Beutin L., Montenegro M.A., Orskov I., Orskov F., Prada J., Zimmermann S., Stephan R.; Close association of verotoxin (Shiga-like toxin) production with enterohemolysin production in strains of *Escherichia coli*; *J Clin Microbiol;* (27): 2559-2564; 1989.

BfR, BVL, RKI; EHEC: BfR, BVL und RKI konkretisieren Verzehrsempfehlung zu rohen Sprossen und Keimlingen; *BfR Presseinformation;* 2011.

Bielaszewska M., Mellmann A., Zhang W., Kock R., Fruth A., Bauwens A., Peters G., Karch H.; Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study; *Lancet Infect Dis;* 11:671-6; 2011.

Bielaszewska M., Prager R., Vandivinit L., Musken A., Mellmann A., Holt N.J., Tarr P.I., Karch H., Zhang W.; Detection and characterization of the fimbrial *sfp* cluster in enterohemorrhagic *Escherichia coli* O165:H25/NM isolates from humans and cattle; *Appl Environ Microbiol;* (75): 64-71; 2009.

Bielaszewska M., Schmidt H., Liesegang A., Prager R., Rabsch W., Tschape H., Cizek A., Janda J., Blahova K., Karch H.; Cattle can be a reservoir of sorbitol-fermenting shiga toxin-producing Escherichia coli O157:H(-) strains and a source of human diseases; *J Clin Microbiol;* (38): 3470-3473; 2000.

Bilge SS, Vary JC Jr., Dowell SF, Tarr PI; Role of the *Escherichia coli* O157:H7 O side chain in adherence and analysis of an rfb locus; *Infect Immun;* (64): 4795-4801; 1996.

Bonnet R., Souweine B., Gauthier G., Rich C., Livrelli V., Sirot J., Joly B., Forestier C.; Non-O157:H7 Stx2-producing *Escherichia coli* strains associated with sporadic cases of hemolytic-uremic syndrome in adults; *J Clin Microbiol;* (36): 1777-1780; 1998.

Brunder W., Karch H., Schmidt H.; Complete sequence of the large virulence plasmid pSFO157 of the sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H-strain 3072/96; *Int J Med Microbiol;* (296): 467-474; 2006.

Brunder W., Schmidt H., Frosch M., Karch H.; The large plasmids of Shiga-toxin-producing *Escherichia coli* (STEC) are highly variable genetic elements; *Microbiology;* (145 (Pt 5)): 1005-1014; 1999.

Bukhari AI, Shapiro JA, Adhya SL; DNA: Insertion Elements, Plasmids, and Episomes; *Cold Spring Harbor Laboratory*; 1977.

CDC; Surveillance for Foodborne-Disease Outbreaks - United States, 1993 - 1997; 2000.

CDC; Standardized molecular subtyping of foodborne bacterial pathogens by pulsed-field gel electrophoresis: CDC training manual.; *Atlanta: The Centers;* 2003.

CDC; *E. coli* Outbreak investigation; 2010a.

CDC; National Center for Zoonotic,Vector-Borne, and Enteric Diseases; 2010b.

CDC (Centers for Disease Control and Prevention); Investigation Update: Outbreak of Shiga toxin-producing *E. coli* O104 (STEC O104:H4) Infections Associated with Travel to Germany; 2011.

Chen CC, Hu ST; Two frameshift products involved in the transposition of bacterial insertion sequence IS*629*; *J Biol Chem;* (281): 21617-21628; 2006.

Chen C.W., Thomas C.A., Jr.; Recovery of DNA segments from agarose gels; *Anal Biochem;* (101): 339-341; 1980.

Clarke S.C.; Diarrhoeagenic *Escherichia coli*--an emerging problem?; *Diagn Microbiol Infect Dis;* (41): 93-98; 2001.

Cohen S.N., Chang A.C., Hsu L.; Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA; *Proc Natl Acad Sci U S A;* (69): 2110-2114; 1972.

Craig NL; Target site selection in transposition; *Annu Rev Biochem;* (66): 437-474; 1997.

Cray W.C., Abbott D.O., Beacorn F.J., Benson S.T.; Detection, isolation and identification of *Escherichia coli* O157:H7 and O157:NM (nonmotile) from meat products; In: In USDA/FSIS microbiology laboratory guidebook; Washington, D.C.: *Food Safety and Inspection Service, U.S. Department of Agriculture*; [Online]: 5.1-5.13;2001.

Creuzenet C., Belanger M., Wakarchuk W.W., Lam J.S.; Expression, purification, and biochemical characterization of WbpP, a new UDP-GlcNAc C4 epimerase from *Pseudomonas aeruginosa* serotype O6; *J Biol Chem;* (275): 19060-19067; 2000.

Davies J., Jacob F.; Genetic mapping of the regulator and operator genes of the lac operon; *J Mol Biol;* (36): 413-417; 1968.

Doolittle W.F., Kirkwood T.B., Dempster M.A.; Selfish DNAs with self-restraint; *Nature;* (307): 501-502; 1984.

Dundas S., Todd W.T., Stewart A.I., Murdoch P.S., Chaudhuri A.K., Hutchinson S.J.; The central Scotland *Escherichia coli* O157:H7 outbreak: risk factors for the hemolytic uremic syndrome and death among hospitalized patients; *Clin Infect Dis;* (33): 923-931; 2001.

Eklund M., Bielaszewska M., Nakari U.M., Karch H., Siitonen A.; Molecular and phenotypic profiling of sorbitol-fermenting *Escherichia coli* O157:H- human isolates from Finland; *Clin Microbiol Infect;* (12): 634-641; 2006.

European Food Safety Authority (EFSA); The community summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2008; *EFSA Journal 2010;* (8): 1498; 2011.

Feng P.; *Escherichia coli* serotype O157:H7: novel vehicles of infection and emergence of phenotypic variants; *Emerg Infect Dis;* (1): 47-52; 1995.

Feng P.; Impact of molecular biology on the detection of foodborne pathogens; *Mol Biotechnol;* (7): 267-278; 1997.

Feng P., Lampel K.A., Karch H., Whittam T.S.; Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7; *J Infect Dis;* (177): 1750-1753; 1998a.

Feng P., Monday S.R.; Multiplex PCR for detection of trait and virulence factors in enterohemorrhagic *Escherichia coli* serotypes; *Mol Cell Probes;* (14): 333-337; 2000.

Feng P., Sandlin R.C., Park C.H., Wilson R.A., Nishibuchi M.; Identification of a rough strain of *Escherichia coli* O157:H7 that produces no detectable O157 antigen; *J Clin Microbiol;* (36): 2339-2341; 1998b.

Feng P.C., Monday S.R., Lacher D.W., Allison L., Siitonen A., Keys C., Eklund M., Nagano H., Karch H., Keen J., Whittam T.S.; Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model.; Emerg. Infect. Dis.; 13:(11); 1701-1706; 2007.

Fey P.D., Wickert R.S., Rupp M.E., Safranek T.J., Hinrichs S.H.; Prevalence of non-O157:H7 shiga toxin-producing *Escherichia coli* in diarrheal stool samples from Nebraska; *Emerg Infect Dis;* (6): 530-533; 2000.

Galloway S.M., Raetz C.R.; A mutant of *Escherichia coli* defective in the first step of endotoxin biosynthesis; *J Biol Chem;* (265): 6394-6402; 1990.

Gemski P.Jr., Sheahan D.G., Washington O., Formal S.B.; Virulence of *Shigella flexneri* hybrids expressing *Escherichia coli* somatic antigens; *Infect Immun;* (6): 104-111; 1972.

Ginsburg V.; Studies on the biosynthesis of guanosine diphosphate L-fucose; *J Biol Chem;* (236): 2389-2393; 1961.

Hartl D.L., Lozovskaya E.R., Lawrence J.G.; Nonautonomous transposable elements in prokaryotes and eukaryotes; *Genetica;* (86): 47-53; 1992.

Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H.; Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12; *DNA Res;* (8): 11-22; 2001.

Hobbs M., Reeves P.R.; The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters; *Mol Microbiol;* (12): 855-856; 1994.

Horwitz J., Ch, J., Cu, y R., To, on A., Da, oge M., Fi, er B., Ma, icio J., Kl, dt I.; Substrates For Cytochemical Demonstration Of Enzyme Activity. I. Some Substituted 3-Indolyl-Beta-D-Glycopyranosides Results; *J Med Chem;* (7): 574-575; 1964.

Jackson M.P., Wadolkowski E.A., Weinstein D.L., Holmes R.K., O'Brien A.D.; Functional analysis of the Shiga toxin and Shiga-like toxin type II variant binding subunits by using site-directed mutagenesis; *J Bacteriol;* (172): 653-658; 1990.

Kaniuk N.A., Monteiro M.A., Parker C.T., Whitfield C.; Molecular diversity of the genetic loci responsible for lipopolysaccharide core oligosaccharide assembly within the genus *Salmonella*; *Mol Microbiol;* (46): 1305-1318; 2002.

Karch H., Mellmann A., Bielaszewska M.; Epidemiology and pathogenesis of enterohaemorrhagic *Escherichia coli*; *Berl Munch Tierarztl Wochenschr;* (122): 417-424; 2009.

Karch H., Tarr P.I., Bielaszewska M.; Enterohaemorrhagic *Escherichia coli* in human medicine; *Int J Med Microbiol;* (295): 405-418; 2005.

Karmali M.A., Petric M., Lim C., Fleming P.C., Steele B.T.; Escherichia coli cytotoxin, haemolytic-uraemic syndrome, and haemorrhagic colitis; *Lancet;* (2): 1299-1300; 1983.

Karmali M.A., Steele B.T., Petric M., Lim C.; Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools; *Lancet;* (1): 619-620; 1983.

Keskimaki M., Ratiner Y., Oinonen S., Leijala E., Nurminen M., Saari M., Siitonen A.; Haemolytic-uraemic syndrome caused by vero toxin-producing *Escherichia coli* serotype Rough: K-: H49; *Scand J Infect Dis;* (31): 141-144; 1999.

Korbie D.J., Mattick J.S.; Touchdown PCR for increased specificity and sensitivity in PCR amplification; *Nat Protoc;* (3): 1452-1456; 2008.

Kudva I.T., Griffin R.W., Murray M., John M., Perna N.T., Barrett T.J., Calderwood S.B.; Insertions, Deletions, and Single-Nucleotide Polymorphisms at Rare Restriction Enzyme Sites Enhance Discriminatory Power of Polymorphic Amplified Typing Sequences, a Novel Strain Typing System for *Escherichia coli* O157:H7; *J Clin Microbiol;* (42): 2388-2397; 2004.

Kusumoto M., Ooka T., Nishiya Y., Ogura Y., Saito T., Sekine Y., Iwata T., Akiba M., Hayashi T.; Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions; *Nat Commun;* (2): 152; 2011.

Kusumoto M., Suzuki R., Nishiya Y., Okitsu T., Oka M.; Host-dependent activation of IS1203v excision in Shiga toxin-producing *Escherichia coli*; *J Biosci Bioeng;* (97): 406-411; 2004.

Law D.; Virulence factors of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli*; *J Appl Microbiol;* (88): 729-745; 2000.

Le Rouzic A., Boutin T.S., Capy P.; Long-term evolution of transposable elements; *Proc Natl Acad Sci U S A;* (104): 19375-19380; 2007.

Levine M.M., Xu J.G., Kaper J.B., Lior H., Prado V., Tall B., Nataro J., Karch H., Wachsmuth K.; A DNA probe to identify enterohemorrhagic *Escherichia coli* of O157:H7 and other serotypes that cause hemorrhagic colitis and hemolytic uremic syndrome; *J Infect Dis;* (156): 175-182; 1987.

Liu B., Knirel Y.A., Feng L., Perepelov A.V., Senchenkova S.N., Wang Q., Reeves P.R., Wang L.; Structure and genetics of *Shigella* O antigens; *FEMS Microbiol Rev;* (32): 627-653; 2008.

Mahillon J.; Transposons as gene haulers; *APMIS Suppl;* (84): 29-36; 1998.

Mahillon J., Chandler M.; Insertion sequences; *Microbiol Mol Biol Rev;* (62): 725-774; 1998.

Makino K., Yokoyama K., Kubota Y., Yutsudo C.H., Kimura S., Kurokawa K., Ishii K., Hattori M., Tatsuno I., Abe H., Iida T., Yamamoto K., Onishi M., Hayashi T., Yasunaga T., Honda T., Sasakawa C., Shinagawa H.; Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak; *Genes Genet Syst;* (74): 227-239; 1999.

Mandel M., Higa A.; Calcium-dependent bacteriophage DNA infection; *J Mol Biol;* (53): 159-162; 1970.

March S.B., Ratnam S.; Sorbitol-MacConkey medium for detection of Escherichia coli O157:H7 associated with hemorrhagic colitis; *J Clin Microbiol;* (23): 869-872; 1986.

Marko M.A., Chipperfield R., Birnboim H.C.; A procedure for the large-scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder; *Anal Biochem;* (121): 382-387; 1982.

Matsutani S., Ohtsubo E.; Complete sequence of IS629; *Nucleic Acids Res;* (18): 1899; 1990.

Matsutani S., Ohtsubo E.; Distribution of the *Shigella sonnei* insertion elements in *Enterobacteriaceae*; *Gene;* (127): 111-115; 1993.

Matsutani S., Ohtsubo H., Maeda Y., Ohtsubo E.; Isolation and characterization of IS elements repeated in the bacterial chromosome; *J Mol Biol;* (196): 445-455; 1987.

Mead P.S., Slutsker L., Dietz V., McCaig L.F., Bresee J.S., Shapiro C., Griffin P.M., Tauxe R.V.; Food-related illness and death in the United States; *Emerg Infect Dis;* (5): 607-625; 1999.

Meier-Dieter U., Barr K., Starman R., Hatch L., Rick P.D.; Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. Molecular cloning of the rfe-rff gene cluster; *J Biol Chem;* (267): 746-753; 1992.

Mikaelian I., Sergeant A.; Modification of the overlap extension method for extensive mutagenesis on the same template; *Methods Mol Biol;* (57): 193-202; 1996.

Miller S.A., Dykes D.D., Polesky H.F.; A simple salting out procedure for extracting DNA from human nucleated cells; *Nucleic Acids Res;* (16): 1215; 1988.

Monday S.R., Beisaw A., Feng P.C.; Identification of Shiga toxigenic *Escherichia coli* seropathotypes A and B by multiplex PCR; *Mol Cell Probes;* (21): 308-311; 2007.

Monday S.R., Minnich S.A., Feng P.C.; A 12-base-pair deletion in the flagellar master control gene flhC causes nonmotility of the pathogenic German sorbitol-fermenting *Escherichia coli* O157:H- strains; *J Bacteriol;* (186): 2319-2327; 2004.

Nagano H., Hirochi T., Fujita K., Wakamori Y., Takeshi K., Yano S.; Phenotypic and genotypic characterization of beta-D-glucuronidase-positive Shiga toxin-producing *Escherichia coli* O157:H7 isolates from deer; *J Med Microbiol;* (53): 1037-1043; 2004.

Nagano H., Okui T., Fujiwara O., Uchiyama Y., Tamate N., Kumada H., Morimoto Y., Yano S.; Clonal structure of Shiga toxin (Stx)-producing and beta-D-glucuronidase-positive *Escherichia coli* O157:H7 strains isolated from outbreaks and sporadic cases in Hokkaido, Japan; *J Med Microbiol;* (51): 405-416; 2002.

Nataro J.P., Kaper J.B.; Diarrheagenic *Escherichia coli*; *Clin Microbiol Rev;* (11): 142-201; 1998.

Neumann E., Rosenheck K.; Permeability changes induced by electric impulses in vesicular membranes; *J Membr Biol;* (10): 279-290; 1972.

Nikaido H., Vaara M.; Molecular basis of bacterial outer membrane permeability; *Microbiol Rev;* (49): 1-32; 1985.

O'Brien A.O., Lively T.A., Chen M.E., Rothman S.W., Formal S.B.; *Escherichia coli* O157:H7 strains associated with haemorrhagic colitis in the United States produce a *Shigella dysenteriae* 1 (SHIGA) like cytotoxin; *Lancet;* (1): 702; 1983.

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

Ohnishi M., Kurokawa K., Hayashi T.; Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors?; *Trends Microbiol;* (9): 481-485; 2001.

Ooka T., Ogura Y., Asadulghani M., Ohnishi M., Nakayama K., Terajima J., Watanabe H., Hayashi T.; Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes; *Genome Res;* (19): 1809-1816; 2009a.

Ooka T., Terajima J., Kusumoto M., Iguchi A., Kurokawa K., Ogura Y., Asadulghani M., Nakayama K., Murase K., Ohnishi M., Iyoda S., Watanabe H., Hayashi T.; Development of a multiplex PCR-based rapid typing method for enterohemorrhagic *Escherichia coli* O157 strains; *J Clin Microbiol;* (47): 2888-2894; 2009b.

Paton A.W., Paton J.C.; Characterization of IS1203, an insertion sequence in *Escherichia coli* O111:H-; *Gene;* (150): 67-70; 1994.

Paton A.W., Paton J.C.; Molecular characterization of the locus encoding biosynthesis of the lipopolysaccharide O antigen of *Escherichia coli* serotype O113; *Infect Immun;* (67): 5930-5937; 1999.

Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.D., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.; Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7; *Nature;* (409): 529-533; 2001.

Perry M.B., Maclean L., Griffith D.W.; Structure of the O-chain polysaccharide of the phenol-phase soluble lipopolysaccharide of *Escherichia coli* O157:H7; *Biochem Cell Biol;* (64): 21-28; 1986.

Plunkett G.I., Rose D.J., Durfee T.J., Blattner F.R.; Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product; *J Bacteriol;* (181): 1767-1778; 1999.

Pluschke G., Mayden J., Achtman M., Levine R.P.; Role of the capsule and the O antigen in resistance of O18:K1 *Escherichia coli* to complement-mediated killing; *Infect Immun;* (42): 907-913; 1983.

Polard P., Prere M.F., Chandler M., Fayet O.; Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS911; *J Mol Biol;* (222): 465-477; 1991.

Polard P., Prere M.F., Fayet O., Chandler M.; Transposase-induced excision and circularization of the bacterial insertion sequence IS911; *EMBO J;* (11): 5079-5090; 1992.

Radu S., Abdul M.S., Rusul G., Ahmad Z., Morigaki T., Asai N., Kim Y.B., Okuda J., Nishibuchi M.; Detection of *Escherichia coli* O157:H7 in the beef marketed in Malaysia; *Appl Environ Microbiol;* (64): 1153-1156; 1998.

Raetz C.R., Whitfield C.; Lipopolysaccharide endotoxins; *Annu Rev Biochem;* (71): 635-700; 2002.

Reida P., Wolff M., Pohls H.W., Kuhlmann W., Lehmacher A., Aleksic S., Karch H., Bockemuhl J.; An outbreak due to enterohaemorrhagic *Escherichia coli* O157:H7 in a children day care centre characterized by person-to-person transmission and environmental contamination; *Zentralbl Bakteriol;* (281): 534-543; 1994.

Ribot E.M., Fair M.A., Gautom R., Cameron D.N., Hunter S.B., Swaminathan B., Barrett T.J.; Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet; *Foodborne Pathog Dis;* (3): 59-67; 2006.

Riley L.W., Remis R.S., Helgerson S.D., McGee H.B., Wells J.G., Davis B.R., Hebert R.J., Olcott E.S., Johnson L.M., Hargrett N.T., Blake P.A., Cohen M.L.; Hemorrhagic colitis associated with a rare *Escherichia coli* serotype; *N Engl J Med;* (308): 681-685; 1983.

Robert Koch Institut; Epidemiologisches Bulletin; 2000.

Ruiz N., Kahne D., Silhavy T.J.; Transport of lipopolysaccharide across the cell envelope: the long road of discovery; *Nat Rev Microbiol;* (7): 677-683; 2009.

Rump L.V., Beutin L., Fischer M., Feng P.C.; Characterization of a *gne*::IS629 O rough:H7 *Escherichia coli* strain from a hemorrhagic colitis patient; *Appl Environ Microbiol;* (76): 5290-5291; 2010a.

Rump L.V., Feng P.C., Fischer M., Monday S.R.; Genetic analysis for the lack of expression of the O157 antigen in an O Rough:H7 *Escherichia coli* strain; *Appl Environ Microbiol;* (76): 945-947; 2010b.

Rump L.V., Fischer M., Gonzalez-Escalona N.; Prevalence, distribution and evolutionary significance of the IS*629* insertion element in the stepwise emergence of *Escherichia coli* O157:H7; *BMC Microbiol;* (11): 133; 2011a.

Rump L.V., Fischer M., Gonzalez-Escalona N.; Strains of *Escherichia coli* O157:H7 stepwise evolutionary model exhibit different IS*629* transposition frequencies; *Appl Environ Microbiol;* 2011b.

Rzhetsky A., Nei M.; Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference; *J Mol Evol;* (35): 367-375; 1992.

Sambrook J, Russell RG; In: Molecular Cloning. A Laboratory Manual; Cold Spring Harbor: *Cold Spring Harbor Laboratory Press*;2001.

Samuel G., Hogbin J.P., Wang L., Reeves P.R.; Relationships of the *Escherichia coli* O157, O111, and O55 O-antigen gene clusters with those of *Salmonella enterica* and *Citrobacter freundii,* which express identical O antigens; *J Bacteriol;* (186): 6536-6543; 2004.

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

Scheiring J., Andreoli S.P., Zimmerhackl L.B.; Treatment and outcome of Shiga-toxin-associated hemolytic uremic syndrome (HUS); *Pediatr Nephrol;* (23): 1749-1760; 2008.

Schmidt H., Beutin L., Karch H.; Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157:H7 strain EDL 933; *Infect Immun;* (63): 1055-1061; 1995.

Schnaitman C.A., Klena J.D.; Genetics of lipopolysaccharide biosynthesis in enteric bacteria; *Microbiol Rev;* (57): 655-682; 1993.

Seeburg P.H., Shine J., Martial J.A., Baxter J.D., Goodman H.M.; Nucleotide sequence and amplification in bacteria of structural gene for rat growth hormone; *Nature;* (270): 486-494; 1977.

Senchenkova S.N., Feng L., Yang J., Shashkov A.S., Cheng J., Liu D., Knirel Y.A., Reeves P.R., Jin Q., Ye Q., Wang L.; Structural and genetic characterization of the *Shigella boydii* type 10 and type 6 O antigens; *J Bacteriol;* (187): 2551-2554; 2005.

Sheng H., Lim J.Y., Watkins M.K., Minnich S.A., Hovde C.J.; Characterization of an *Escherichia coli* O157:H7 O-antigen deletion mutant and effect of the deletion on bacterial persistence in the mouse intestine and colonization at the bovine terminal rectal mucosa; *Appl Environ Microbiol;* (74): 5015-5022; 2008.

Siguier P., Perochon J., Lestrade L., Mahillon J., Chandler M.; ISfinder: the reference centre for bacterial insertion sequences; *Nucleic Acids Res;* (34): D32-D36; 2006.

Steele C., Zhang S., Shillitoe E.J.; Effect of different antibiotics on efficiency of transformation of bacteria by electroporation; *Biotechniques;* (17): 360-365; 1994.

Stevenson G., Neal B., Liu D., Hobbs M., Packer N.H., Batley M., Redmond J.W., Lindquist L., Reeves P.; Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster; *J Bacteriol;* (176): 4144-4156; 1994.

Szabo M., Kiss J., Nagy Z., Chandler M., Olasz F.; Sub-terminal sequences modulating IS*30* transposition in vivo and in vitro; *J Mol Biol;* (375): 337-352; 2008.

Tarr P.I., Gordon C.A., Chandler W.L.; Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome; *Lancet;* (365): 1073-1086; 2005.

Tarr P.I., Neill M.A., Clausen C.R., Watkins S.L., Christie D.L., Hickman R.O.; *Escherichia coli* O157:H7 and the hemolytic uremic syndrome: importance of early cultures in establishing the etiology; *J Infect Dis;* (162): 553-556; 1990.

Thompson J.S., Hodge D.S., Borczyk A.A.; Rapid biochemical test to identify verocytotoxin-positive strains of *Escherichia coli* serotype O157; *J Clin Microbiol;* (28): 2165-2168; 1990.

Ullrich A., Shine J., Chirgwin J., Pictet R., Tischer E., Rutter W.J., Goodman H.M.; Rat insulin genes: construction of plasmids containing the coding sequences; *Science;* (196): 1313-1319; 1977.

Vica P.S., Garcia G.O., Paniagua Contreras G.L.; The lom gene of bacteriophage lambda is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells; *FEMS Microbiol Lett;* (156): 129-132; 1997.

Vila J., Vargas M., Ruiz J., Gallardo F., Jimenez de Anta M.T., Gascon J.; Isolation of verotoxin-producing *Escherichia coli* O-rough:K1:H7 from two patients with traveler's diarrhea; *J Clin Microbiol;* (35): 2279-2282; 1997.

Vogele K., Schwartz E., Welz C., Schiltz E., Rak B.; High-level ribosomal frameshifting directs the synthesis of IS*150* gene products; *Nucleic Acids Res;* (19): 4377-4385; 1991.

Wang L., Huskic S., Cisterne A., Rothemund D., Reeves P.R.; The O-antigen gene cluster of *Escherichia coli* O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene; *J Bacteriol;* (184): 2620-2625; 2002.

Wang L., Qu W., Reeves P.R.; Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships; *Infect Immun;* (69): 6923-6930; 2001.

Wang L., Reeves P.R.; Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes; *Infect Immun;* (66): 3545-3551; 1998.

Watanabe H., Wada A., Inagaki Y., Itoh K., Tamura K.; Outbreaks of enterohaemorrhagic *Escherichia coli* O157:H7 infection by two different genotype strains in Japan, 1996; *Lancet;* (348): 831-832; 1996.

Wells J.G., Davis B.R., Wachsmuth I.K., Riley L.W., Remis R.S., Sokolow R., Morris G.K.; Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype; *J Clin Microbiol;* (18): 512-520; 1983.

Whitfield C.; Biosynthesis of lipopolysaccharide O antigens; *Trends Microbiol;* (3): 178-185; 1995.

Whittam T.S., Wilson R.A.; Genetic relationships among pathogenic *Escherichia coli* of serogroup O157; *Infect Immun;* (56): 2467-2473; 1988.

Whittam T.S., Wolfe M.L., Wachsmuth I.K., Orskov F., Orskov I., Wilson R.A.; Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea; *Infect Immun;* (61): 1619-1629; 1993.

WHO (World Health Organization); Food safety and foodborne illness; 2007.

WHO (World Health Organization); WHO Initiative to Estimate the Global Burden of Foodborne Diseases; 2010.

Wick L.M., Qi W., Lacher D.W., Whittam T.S.; Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7; *J Bacteriol;* (187): 1783-1791; 2005.

Yethon J.A., Heinrichs D.E., Monteiro M.A., Perry M.B., Whitfield C.; Involvement of *waaY*, *waaQ*, and *waaP* in the modification of *Escherichia coli* lipopolysaccharide and their role in the formation of a stable outer membrane; *J Biol Chem;* (273): 26310-26316; 1998.

Yokoyama E., Hashimoto R., Etoh Y., Ichihara S., Horikawa K., Uchimura M.; Biased distribution of IS*629* among strains in different lineages of enterohemorrhagic *Escherichia coli* serovar O157; *Infect Genet Evol;* 2010.

Yoon J.W., Hovde C.J.; All blood, no stool: enterohemorrhagic *Escherichia coli* O157:H7 infection; *J Vet Sci;* (9): 219-231; 2008.

Yu J., Kaper J.B.; Cloning and characterization of the eae gene of enterohaemorrhagic *Escherichia coli* O157:H7; *Mol Microbiol;* (6): 411-417; 1992.

Zhang W., Mellmann A., Sonntag A.K., Wieler L., Bielaszewska M., Tschape H., Karch H., Friedrich A.W.; Structural and functional differences between disease-associated genes of enterohaemorrhagic *Escherichia coli* O111; *Int J Med Microbiol;* (297): 17-26; 2007.

Zhou Z., Li X., Liu B., Beutin L., Xu J., Ren Y., Feng L., Lan R., Reeves P.R., Wang L.; Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor; *PLoS One;* (5): e8700; 2010.

10 PUBLICATIONS

10.1 Refereed Journal Articles

10.1.1 From this research

[1] Rump L.V., Fischer M., Gonzalez-Escalona N.; Prevalence, distribution and evolutionary significance of the IS*629* insertion element in the stepwise emergence of *Escherichia coli* O157:H7; *BMC Microbiol;* (11): 133; 2011.

[2] Rump L.V., Fischer M., Gonzalez-Escalona N.; Strains of *Escherichia coli* O157:H7 stepwise evolutionary model exhibit different IS*629* transposition frequencies; *Appl Environ Microbiol;* 2011.

[3] Rump L. V., Strain E. A., Cao G., Allard M. W., Fischer M., Brown E. W. and Gonzalez-Escalona N.; Draft genome sequences of six *Escherichia coli* isolates from the stepwise model emergence of *Escherichia coli* O157:H7; *J. Bacteriol*.; 193: 8; 2011.

[4] Rump L.V., Beutin L., Fischer M., Feng P.C.; Characterization of a *gne*::IS629 O rough:H7 *Escherichia coli* strain from a hemorrhagic colitis patient; *Appl Environ Microbiol;* (76): 5290-5291; 2010.

[5] Rump L.V., Feng P.C., Fischer M., Monday S.R.; Genetic analysis for the lack of expression of the O157 antigen in an O Rough:H7 *Escherichia coli* strain; *Appl Environ Microbiol;* (76): 945-947; 2010.

10.1.2 Other publications

[1] Rump L. V., Asamoah, B., Gonzalez-Escalona, N.; Comparison of commercial RNA extraction kits for preparation of DNA-free total RNA from *Salmonella* cells, *BMC Research Notes*; 3:211; 2010.

[2] Trucksess M. W., Weaver C. M., Oles C. J.; Rump L. V., White K. D., Betz J. M., Rader J. I.; Use of multitoxin immunoaffinity columns for determination of aflatoxins and ochratoxin A in ginseng and ginger; *J AOAC Int*; 90(4):1042-9; 2007.

10.2 Refereed Conference Publications

[1] Rump L. V., Feng P. C. H., Fischer M., Monday S.R.; Genetic analysis for the absence of O157 antigen expression in MA6, a shigatoxigenic *Escherichia coli* strain of O rough:H7 serotype; *American Society for Microbiology*; 109th General Meeting; 2009.

[2] Rump L. V., Feng P. C. H., Fischer M., Monday S.R., Prevalence and distribution of the IS629 insertion element in *Escherichia coli* O157:H7 and its impact on the emergence of atypical O157:H7 pathogens; *American Society for Microbiology*; 110th General Meeting; 2010.

[3] Rump L. V., Fischer M., Gonzalez-Escalona, N.; Strains of the *Escherichia coli* O157:H7 stepwise evolutionary model exhibit different IS*629* transposition frequencies; *American Society for Microbiology*; 111th General Meeting; 2011.

10.3 Presentations

[1] Rump L. V.; Der Mutant *E. coli* MA6: Genetische Analyse der Abwesendheit des O157 Antigens; *University of Hamburg*, Germany, 2009.

[2] Rump L. V.; Molecular characterization of *E. coli* (EHEC): O *rough* strains and the prevalence and importance of IS*629* in *E. coli* O157:H7; *University of Maryland,* College Park*,* USA*,* 2011.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Feb. 2010, p. 945–947 Vol. 76, No. 3
0099-2240/10/$12.00 doi:10.1128/AEM.02046-09
Copyright © 2010, American Society for Microbiology. All Rights Reserved.

Genetic Analysis for the Lack of Expression of the O157 Antigen in an O Rough:H7 *Escherichia coli* Strain[▽]

Lydia V. Rump,[1,2] Peter C. H. Feng,[1]* Markus Fischer,[2] and Steven R. Monday[1]

Division of Microbiology, United States Food and Drug Administration, College Park, Maryland 20740,[1] and Institute of Food Chemistry, University of Hamburg, Hamburg, Germany[2]

Received 25 August 2009/Accepted 16 November 2009

The O-antigen (*rfb*) operon and related genes of MA6, an O rough:H7 Shiga-toxigenic *Escherichia coli* strain, were examined to determine the cause of the lack of O157 expression. A 1,310-bp insertion, homologous to IS*629*, was observed within its *gne* gene. *trans* complementation with a functional *gne* gene from O157:H7 restored O157 antigen expression in MA6.

Shiga-toxigenic *Escherichia coli* (STEC) serotype O157:H7 carries O157 and H7 antigens, so these traits are extensively used in identification (1). Strain MA6, isolated from beef in Malaysia (8), carries the O157:H7 virulence factor genes, including the Shiga toxin 2 gene (stx_2), the γ intimin allele (γ-*eae*), the enterohemolysin gene (*ehxA*), and the +93 *uidA* single nucleotide polymorphism (SNP) found only in O157:H7 strains (1). Multilocus sequence typing also showed MA6 to have the most common sequence type (ST-66) for O157:H7 strains. However, and in spite the fact that MA6 had *per* gene sequences essential for O157 antigen synthesis (2), no O157 antigen is expressed (O rough), and therefore, it is undetectable with serological assays used in O157:H7 analysis.

The biosynthesis and assembly of *E. coli* O antigen are highly complex (9). The *rfb* operon (12 genes) (16), along with 3 ancillary genes outside of the *rfb*, is required for the biosynthesis of the 4 sugar nucleotide precursors and the assembly of the O unit (11). This is then linked to the core antigen, comprising an inner and an outer component, which require 3 other operons for biosynthesis and assembly (9). As defects in any of these genes could produce the O-null phenotype (13), we systematically examined these genes (Table 1) to elucidate the cause of the absence of O157 expression in MA6.

PCR and sequencing primers for the individual genes were designed from sequences for the O157:H7 strain EDL933 (GenBank accession no. AE005174). The 50-μl PCR mix contained 5 U of HotStar *Taq* (Qiagen, Valencia, CA), 1× polymerase buffer, 2.5 to 3.5 mM $MgCl_2$, 400 μM each dNTP, 300 nM of each primer, and ~100 ng of template DNA from either MA6 or the EDL931 reference strain. The "touchdown" PCR (10) consisted of 95°C for 15 min and 10 cycles of 95°C for 30 s, 69 to 60°C (−1°C/cycle) for 20 s, and 72°C for 1.5 min, followed by 35 cycles of 95°C for 30 s, 60°C for 20 s, and 72°C for 1.5 min, with a single step of 72°C for 1 min for final extension. Products were examined on a 1% agarose gel in Tris-borate-EDTA (TBE) buffer. Comparison of amplicons from respective genes from MA6 and EDL931 showed that no gross differences in size were observed for any of the *rfb* or related genes, suggesting the absence of major insertions or deletions. Consistently, contigs assembled from the MA6 amplicon were identical in sequence to those of EDL933 in GenBank, indicating the absence of base mutations in either the promoter or any of the open reading frames (ORF). One exception was the *gne* gene, encoding UDP-acetylgalactosamine (GalNAc)-4-epimerase, which is essential for the synthesis of one of the oligosaccharide subunits in the O antigen (14). When PCR primers that bound upstream of the putative promoter and downstream of the *gne* gene were used, an expected ~1,400-bp product was obtained from EDL931 (Fig. 1, lane 3), but the MA6 amplicon was ~2,700 bp (Fig. 1, lane 4). PCR of other O157:H7 strains all yielded the ~1,400-bp product, while MA6 consistently produced the larger amplicon. Comparison of sequences to that of EDL933 showed the presence of a 1,310-bp insertion within the MA6 *gne* ORF at +385 that shared 96% homology to the insertion sequence *629* (IS*629*) (accession no. X51586) element. Furthermore, the deduced protein sequences for the putative *orfA* and *orfB* genes on the insert were 100% and 99% identical to those of the IS*629* transposase in O157:H7 strains Sakai (accession no. NC_002695), and EDL933 and EC4115 (accession no. NC_011353), respectively.

To determine whether *gne*::IS*629* (accession no. GU183138) caused the absence of O157 expression in MA6, the wild-type EDL931 *gne* ORF was amplified using primers that added BamHI and SacI restriction sites at the 5′ and 3′ termini, respectively. The purified amplicon was digested accordingly, ligated into pTrc99A vector (Stratagene, La Jolla, CA), and electroporated into *E. coli* DH5α (10). Transformants were selected on LB plates with 100 μg/ml ampicillin (Amp). Colonies that were Amp resistant (Amp^r) were PCR amplified with vector-specific primers, and those carrying the insert were sequenced to confirm the presence of the wild-type *gne* insert in the construct (pGNE). For *trans*-complementation studies, pGNE was electroporated into MA6. Amp^r transformants were PCR amplified with vector-specific primers as well as primers that annealed to sequences outside the *gne* gene and also not present on the vector, to confirm that they carried both pGNE and the

* Corresponding author. Mailing address: HFS-711, FDA, 5100 Paint Branch Parkway, College Park, MD 20740. Phone: (301) 436-1650. Fax: (301) 436-2644. E-mail: peter.feng@fda.hhs.gov.

▽ Published ahead of print on 30 November 2009.

Downloaded from aem.asm.org at FDA BIOSCIENCES LIBRARY on February 1, 2010

TABLE 1. *rfb* operon genes, ancillary genes, and *waa* cluster genes examined in this study

Category	General function[a]	Gene(s)
O-antigen (*rfb*) operon	Nucleotide sugar transfer	*wbdN*, *wbdO*, *wbdP*, *wbdQ*, *wbdR*
	O-unit processing	*wzy*, *wzx*
	Nucleotide sugar synthesis	*per*, *gmd*, *fcl*, *manC*, *manB*
waa core gene clusters	Structure modification	*waaQ*, *waaP*, *waaY*
	Nucleotide sugar transfer	*rfaG*, *rfaC*
	LPS core biosynthesis enzyme	*waaI*, *waaJ*, *waaD*, *waaL*
Ancillary genes	Nucleotide sugar synthesis	*manA*
	O-unit processing	*wecA*
	Nucleotide sugar synthesis	*gne*

[a] LPS, lipopolysaccharide.

gne::IS*629* locus. Serological testing with the RIM O157:H7 latex kit (Remel, Lenexa, KS) confirmed that the Amp^r MA6 transformants expressed O157 antigen.

These results confirmed that *gne*::IS*629* caused the O rough phenotype of MA6. Originally isolated from *Shigella sonnei* (7), IS*629* has since been found, often in multiple copies, to cause gene disruptions in other enteric bacteria (6). *fliC*::IS*629* caused nonmotility of an *E. coli* O111 strain (17), and *wbaM*::IS*629* resulted in an O rough *Shigella boydii* strain (15). The IS*629* recognition site remains unknown (5), so it is uncertain that there is an IS*629* hot spot within the O157:H7 *gne* ORF. Other bacteria, like O157:H7, also have the *gne* gene positioned upstream of the *rfb* operon (12), but no *gne*::IS*629* rough strains of these have been reported. This suggests that the IS*629* insertion site within the *gne* of MA6 may have occurred as a result of a random mutation and that MA6 appears to be the only naturally occurring O rough O157:H7 strain that resulted from the *gne*::IS*629* insertion.

The O antigen is not required for growth but does confer protection (9), so the loss of the O antigen has been reported to make pathogens serum sensitive or less virulent (4). If that is so, we would expect MA6 to be less pathogenic than O157:H7; consistent with that speculation, MA6 has not been implicated in illness. Even so, while no O rough O157:H7 strains have caused disease, other O rough STEC strains have caused illnesses (3); hence, the virulence potential of MA6 remains undetermined.

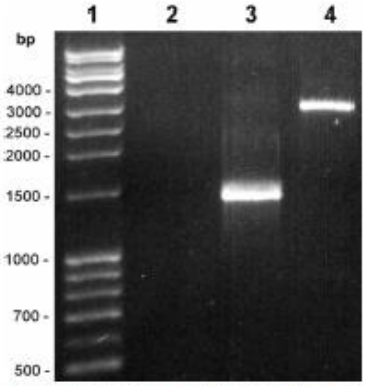


FIG. 1. Agarose gel electrophoresis of *gne* amplicons derived from EDL931 (O157:H7) and MA6. Lanes: 1, exACTGene (1 kb) plus molecular size ladder (Fisher BioReagents, Pittsburgh, PA); 2, negative control (reaction mix without DNA template); 3, EDL931; 4, MA6.

In conclusion, the absence of O157 antigen expression by MA6 is caused by *gne*::IS*629*. Occurrence of O rough:H7 strains like MA6 in food or clinical samples is of concern, as they are undetectable by the serological assays used to identify O157:H7. However, the IS*629* insertion site within the O157:H7 *gne* ORF appears to have been due to a random mutational event, and therefore, MA6-like O rough mutants of O157:H7 are thus far uncommon.

This project was supported by an appointment to the Research Fellowship Program for the Center for Food Safety and Applied Nutrition administered by the Oak Ridge Associated Universities through a contract with the FDA.

REFERENCES

1. **Feng, P.** 1995. *Escherichia coli* serotype O157:H7: novel vehicles of infection and emergence of phenotypic variants. Emerg. Infect. Dis. **1**:47–52.
2. **Feng, P., R. C. Sandlin, C. H. Park, R. A. Wilson, and M. Nishibuchi.** 1998. Identification of a rough strain of *Escherichia coli* O157:H7 that produces no detectable O157 antigen. J. Clin. Microbiol. **36**:2339–2341.
3. **Keskimaki, M., Y. Ratiner, S. Oinonen, E. Leijala, M. Nurminen, M. Saari, and A. Siitonen.** 1999. Haemolytic-uraemic syndrome caused by vero toxin-producing *Escherichia coli* serotype Rough: K-: H49. Scand. J. Infect. Dis. **31**:141–144.
4. **Liu, B., Y. A. Knirel, L. Feng, A. V. Perepelov, S. N. Senchenkova, Q. Wang, P. R. Reeves, and L. Wang.** 2008. Structure and genetics of *Shigella* O antigens. FEMS Microbiol. Rev. **32**:627–653.
5. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62**:725–774.
6. **Matsutani, S., and E. Ohtsubo.** 1993. Distribution of the *Shigella sonnei* insertion elements in *Enterobacteriaceae*. Gene **127**:111–115.
7. **Matsutani, S., H. Ohtsubo, Y. Maeda, and E. Ohtsubo.** 1987. Isolation and characterization of IS elements repeated in the bacterial chromosome. J. Mol. Biol. **196**:445–455.
8. **Radu, S., M. S. Abdul, G. Rusul, Z. Ahmad, T. Morigaki, N. Asai, Y. B. Kim, J. Okuda, and M. Nishibuchi.** 1998. Detection of *Escherichia coli* O157:H7 in the beef marketed in Malaysia. Appl. Environ. Microbiol. **64**:1153–1156.
9. **Raetz, C. R., and C. Whitfield.** 2002. Lipopolysaccharide endotoxins. Annu. Rev. Biochem. **71**:635–700.
10. **Sambrook, J., and R. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11. **Samuel, G., and P. R. Reeves.** 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr. Res. **338**:2503–2519.
12. **Samuel, G., J. P. Hogbin, L. Wang, and P. R. Reeves.** 2004. Relationships of the *Escherichia coli* O157, O111, and O55 O-antigen gene clusters with those of *Salmonella enterica* and *Citrobacter freundii*, which express identical O antigens. J. Bacteriol. **186**:6536–6543.
13. **Schnaitman, C. A., and J. D. Klena.** 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol. Rev. **57**:655–682.
14. **Wang, L., S. Huskic, A. Cisterne, D. Rothemund, and P. R. Reeves.** 2002. The

Downloaded from aem.asm.org at FDA BIOSCIENCES LIBRARY on February 1, 2010

O-antigen gene cluster of *Escherichia coli* O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. J. Bacteriol. **184:**2620–2625.
15. **Wang, L., W. Qu, and P. R. Reeves.** 2001. Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships. Infect. Immun. **69:**6923–6930.
16. **Wang, L., and P. R. Reeves.** 1998. Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. Infect. Immun. **66:**3545–3551.
17. **Zhang, W., A. Mellmann, A. K. Sonntag, L. Wieler, M. Bielaszewska, H. Tschape, H. Karch, and A. W. Friedrich.** 2007. Structural and functional differences between disease-associated genes of enterohaemorrhagic *Escherichia coli* O111. Int. J. Med. Microbiol. **297:**17–26.

Downloaded from aem.asm.org at FDA BIOSCIENCES LIBRARY on February 1, 2010

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Aug. 2010, p. 5290–5291 Vol. 76, No. 15
0099-2240/10/$12.00 doi:10.1128/AEM.00740-10
Copyright © 2010, American Society for Microbiology. All Rights Reserved.

Characterization of a *gne*::IS*629* O Rough:H7 *Escherichia coli* Strain from a Hemorrhagic Colitis Patient[▽]

Lydia V. Rump,[1,2] Lothar Beutin,[3] Markus Fischer,[2] and Peter C. H. Feng[1]*

Division of Microbiology, Food and Drug Administration, College Park, Maryland 20740[1]; Institute of Food Chemistry, University of Hamburg, Hamburg, Germany[2]; and National Reference Laboratory for Escherichia coli, Federal Institute for Risk Assessment, Berlin, Germany[3]

Received 24 March 2010/Accepted 20 May 2010

Shiga-toxigenic *Escherichia coli* strains that are O rough:H7 due to *gne*::IS*629* were thought to be rare and to have unknown pathogenic potential. Recently, an O rough:H7 strain caused by *gne*::IS*629* was isolated from a hemorrhagic colitis patient, suggesting that these strains are pathogenic and may not be as rare as anticipated.

Serotype O157:H7 is the prototypic Shiga-toxigenic *Escherichia coli* (STEC) strain that causes food-borne infections worldwide, and it is identified by the presence of the somatic (O)157 and the flagellar (H)7 antigens. Previously, we characterized MA6, an O rough:H7 strain that did not express the O157 antigen (5), and found that it was an O157:H7 strain that belonged in the most common O157:H7 clonal type. Despite the absence of O157 antigen expression, MA6 had all of the genes and operons essential for O157 antigen synthesis, but it exhibited the O rough phenotype due to an IS*629* insertion in the *gne* gene (*gne*::IS*629*) (8), which encodes an epimerase enzyme essential for the synthesis of an oligosaccharide subunit in the O antigen. The IS*629* element is commonly present (and in multiple copies) in enteric bacteria, including O157:H7 strains, and has been reported to cause gene disruptions (6). However, MA6-like O157:H7 strains that are O rough due to *gne*::IS*629* were thought to be rare, as MA6 was the only strain isolated thus far. Moreover, since MA6 was isolated only from a beef sample in Malaysia and was not implicated in illness, the pathogenic potential of this strain was also uncertain. A study in Germany characterized patient isolates of STEC over a 3-year period and found a few strains with the O rough:H7 phenotype (1). One of these, CB7326, was isolated from a hemorrhagic colitis patient and found to carry Shiga toxin 1 (stx_1), Shiga toxin 2 (stx_2), and γ-intimin (γ-*eae*) genes, which are common in O157:H7 strains, suggesting that, like MA6, CB7326 may be an O rough variant of O157:H7. In this study, we characterized strain CB7326 to determine the cause of its O rough phenotype and compared it to MA6 to determine whether these are analogous or related strains.

Strain CB7326 was plated on sorbitol MacConkey agar with ColiComplete (BioControl, Bellevue, WA), and it was determined that the strain did not ferment sorbitol or exhibit ß-glucuronidase activity. Serological analysis by latex agglutination (RIM O157:H7; Remel, Lenexa, KS) confirmed the presence of the H7 antigen but not the O157 antigen. Despite the absence of serological reactivity, however, PCR analysis (7) for the *wzx* and *fliC* genes, which encode the O157 and H7 antigens, respectively, confirmed that CB7326 carried genetic sequences for both antigens. By use of a multiplex PCR (4), CB7326 was found to carry typical enterohemorrhagic *E. coli* (EHEC) virulence markers, including stx_1, stx_2, *ehxA* (enterohemolysin), the γ-*eae* allele, and the +93 *uidA* (ß-glucuronidase) single nucleotide polymorphism, which is unique to O157:H7. Except for the absence of the O157 antigen, these traits are consistent with those of O157:H7 (strain EDL931). Strain CB7326 had traits identical to those of strain MA6, except that MA6 did not carry stx_1 (Table 1).

To determine whether the cause of the O rough phenotype in CB7326 was also due to *gne*::IS*629*, the *gne* gene of CB7326 was amplified by PCR using primers and parameters described previously (8). Analogous to the findings with MA6, CB7326 yielded a larger amplicon (~2,700 bp in size) than did O157:H7 (strain EDL931) (1,400-bp product) (data not shown). Sequence and BLAST analyses of the 2,700-bp amplicon confirmed that the *gne* gene of CB7326 also had the IS*629* element; however, unlike in MA6, where *gne*::IS*629* was found at position +385, *gne*::IS*629* in CB7326 was located at position +711. *trans*-Complementation of CB7326 with the pGNE construct, which carried a wild-type *gne* insert (8), restored O157 antigen expression to that for CB7326 as determined by a serological assay (data not shown). These results confirm that, as for MA6, the O rough phenotype of CB7326 was due to *gne*::IS*629*.

Multilocus sequence typing analysis using 7 specific housekeeping genes (http://www.shigatox.net/ecmlst/cgi-bin/index) confirmed that CB7326 was of sequence type 66 (ST-66) and, therefore, belongs in the most common clonal type of O157:H7 strains. However, a pulsed-field gel electrophoresis (PFGE) comparison of XbaI-digested genomic DNA showed that MA6 and CB7326 shared little homology with EDL931 or with each other and, therefore, are not analogous strains (data not shown). Still, the isolation of another O rough:H7 strain due to *gne*::IS*629* suggests that these strains may not be as rare as previously anticipated (8). The insertion of IS*629* at two different *gne* locations may have been coincidental, or there may be multiple IS*629* insertion sites within *gne*. If so, however, it

* Corresponding author. Mailing address: HFS-711, FDA, 5100 Paint Branch Parkway, College Park, MD 20740. Phone: (301) 436-1650. Fax: (301) 436-2644. E-mail: peter.feng@fda.hhs.gov.

▽ Published ahead of print on 11 June 2010.

TABLE 1. Comparison of traits and markers among MA6, CB7326, and O157:H7 strains

Trait or marker[a]	Result for strain:		
	MA6	CB7326	O157:H7[b]
SOR	−	−	−
GUD	−	−	−
O157	−	−	+
wzx (O157)	+	+	+
H7	+	+	+
fliC (H7)	+	+	+
stx_1	−	+	+
stx_2	+	+	+
γ-*eae*	+	+	+
uidA	+	+	+
ehxA	+	+	+
MLST	ST-66	ST-66	ST-66
gnr::IS*629*	+385	+711	−

[a] SOR, sorbitol fermentation; GUD, ß-glucuronidase activity; O157 and H7, O157 and H7 antigens, tested by latex agglutination; *wzx* (O157), *wzx*, encoding the O157 antigen; *fliC* (H7), *fliC*, encoding the H7 antigen; stx_1 and stx_2, Shiga toxin 1 and 2 genes, respectively; γ-*eae*, γ-intimin allele; *uidA*, +93 *uidA* single nucleotide polymorphism; *ehxA*, enterohemolysin; MLST, multilocus sequence typing; *gnr*::IS*629*, insertion location.

[b] Strain EDL931.

would seem that such O rough:H7 strains would have been encountered more often. The IS*629* insertion site sequence is unknown, so at this time, the existence of an IS*629* hot spot(s) within *gnr* remains inconclusive. Lastly, O rough strains of other STEC serotypes have been implicated in infections and cases of hemolytic-uremic syndrome (2, 3), but the pathogenic potential of MA6 is uncertain, as it has been isolated only from foods. Although MA6 and CB7326 are not analogous strains, they are similar in that both are O rough due to *gnr*::IS*629*, and the fact that CB7326 was isolated from a hemorrhagic colitis patient suggests that MA6 and other similar O rough:H7 strains, if found, may also be pathogenic to humans.

This project was supported by an appointment to the Research Fellowship Program for the Center for Food Safety and Applied Nutrition administered by the Oak Ridge Associated Universities through a contract with the FDA.

REFERENCES

1. **Beutin, L., G. Krause, S. Zimmermann, S. Kaulfuss, and K. Gleier.** 2004. Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period. J. Clin. Microbiol. **42:**1099–1108.
2. **Beutin, L., M. A. Montenegro, I. Orskov, F. Orskov, J. Prada, S. Zimmermann, and R. Stephan.** 1989. Close association of verotoxin (Shiga-like toxin) production with enterohemolysin production in strains of *Escherichia coli*. J. Clin. Microbiol. **27:**2559–2564.
3. **Bonnet, R., B. Souweine, G. Gauthier, C. Rich, V. Livrelli, J. Sirot, B. Joly, and C. Forestier.** 1998. Non-O157:H7 Stx2-producing *Escherichia coli* strains associated with sporadic cases of hemolytic-uremic syndrome in adults. J. Clin. Microbiol. **36:**1777–1780.
4. **Feng, P., and S. R. Monday.** 2000. Multiplex PCR for the detection of trait and virulence factors in enterohemorrhagic *Escherichia coli* serotypes. Mol. Cell. Probes **14:**333–337.
5. **Feng, P., R. C. Sandlin, C. H. Park, R. A. Wilson, and M. Nishibuchi.** 1998. Identification of a rough strain of *Escherichia coli* O157:H7 that produces no detectable O157 antigen. J. Clin. Microbiol. **36:**2339–2341.
6. **Matsutani, S., and E. Ohtsubo.** 1993. Distribution of the *Shigella sonnei* insertion elements in *Enterobacteriaceae*. Gene **127:**111–115.
7. **Monday, S. R., A. Beisaw, and P. C. H. Feng.** 2007. Identification of Shiga toxigenic *Escherichia coli* seropathotypes A and B by multiplex PCR. Mol. Cell. Probes **21:**308–311.
8. **Rump, L. V., P. Feng, M. Fischer, and S. Monday.** 2010. Genetic analysis for the lack of O157 antigen expression in an O rough:H7 *Escherichia coli* strain. Appl. Environ. Microbiol. **76:**945–947.

JOURNAL OF BACTERIOLOGY, Apr. 2011, p. 2058–2059 Vol. 193, No. 8
0021-9193/11/$12.00 doi:10.1128/JB.00118-11
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

GENOME ANNOUNCEMENTS

Draft Genome Sequences of Six *Escherichia coli* Isolates from the Stepwise Model of Emergence of *Escherichia coli* O157:H7▽

L. V. Rump,[1,4] E. A. Strain,[2] G. Cao,[3] M. W. Allard,[1] M. Fischer,[4] E. W. Brown,[1] and N. Gonzalez-Escalona[1]*

Division of Microbiology[1] and Biostatistics Branch,[2] Center for Food Safety and Applied Nutrition, Food and Drug Administration, 5100 Paint Branch Parkway, College Park, Maryland 20740; Joint Institute for Food Safety and Applied Nutrition (JIFSAN), University of Maryland, College Park, Maryland 20742[3]; and Institute of Food Chemistry, University of Hamburg, Hamburg, Germany[4]

Received 25 January 2011/Accepted 31 January 2011

Enterohemorrhagic *Escherichia coli* (EHEC) of serotype O157:H7 has been implicated in food-borne illnesses worldwide. An evolutionary model was proposed in which the highly pathogenic EHEC O157:H7 serotype arose from its ancestor, enteropathogenic *E. coli* (EPEC) O55:H7 (sorbitol fermenting [SOR$^+$] and β-glucuronidase positive [GUD$^+$]), through sequential gain of virulence, phenotypic traits, and serotype change. Here we report six draft genomes of strains belonging to this evolutionary model: two EPEC O55:H7 (SOR$^+$ GUD$^+$) strains, two nonmotile EHEC O157:H$^-$ strains (SOR$^+$ GUD$^+$) containing plasmid pSFO157, one EHEC O157:H7 (SOR$^-$ GUD$^+$) strain, and one O157:H7 strain containing plasmid pSFO157 (SOR$^+$ GUD$^+$).

Enterohemorrhagic *Escherichia coli* (EHEC) O157:H7 has become a significant worldwide cause of food-borne illness since its discovery about 20 years ago. It frequently causes large outbreaks of severe enteric infections, including bloody diarrhea, hemorrhagic colitis (HC), and hemolytic-uremic syndrome (HUS) (1, 4). This serotype expresses the somatic (O) 157 and flagellar (H) 7 antigens, so these traits are extensively used in clinical analysis to identify this highly pathogenic serotype (1). A stepwise evolutionary model has been proposed in which the highly pathogenic enterohemorrhagic *E. coli* (EHEC) serotype O157:H7 arose from its ancestor, enteropathogenic *E. coli* (EPEC) O55:H7, belonging to clonal complex (CC) A1/A2 (sorbitol fermenting and β-glucuronidase positive [SOR$^+$ GUD$^+$]), through sequential acquisition of virulence and phenotypic traits and serotype change (2, 3, 9). After the somatic antigen change from O55 to O157 gave rise to a probably extinct intermediary (CC A3), two separate O157 CCs evolved, splitting into two diverging clonal groups. One was composed of sorbitol-fermenting (SF) nonmotile O157:H$^-$ strains containing plasmid pSFO157 (CC A4) (SOR$^+$ GUD$^+$). The other was composed of non-sorbitol-fermenting (NSF) O157:H7 strains containing plasmid pO157 (CC A5) (SOR$^-$ GUD$^+$). The latter by a mutational inactivation of the *uidA* gene lost its β-glucuronidase activity, which is the most typical O157:H7 phenotype at present (CC A6) (3). These CC A6 strains expanded and spread geographically and account for most of the diseases caused by EHEC (10).

* Corresponding author. Mailing address: FDA, CFSAN, 5100 Paint Branch Parkway HFS-712, College Park, MD 20740. Phone: (301) 436-1937. Fax: (301) 436-2644. E-mail: narjol.gonzalez-escalona@fda.hhs.gov.

▽ Published ahead of print on 11 February 2011.

So far, only four complete genome sequences for pathogenic *E. coli* O157:H7 belonging to the CC A6 have been reported and deposited in GenBank: Sakai (NC_002695) (5), EDL933 (AE005174) (8), TW14359 (CP001368) (7), and EC4115 (NC_011353). Recently, the genome of an ancestral O55:H7 strain, CB9615 (NC_013941), was made available (11). In the present publication, we announce the availability of six draft genome sequences for other enteropathogenic *E. coli* strains belonging to the stepwise model of emergence of *E. coli* O157:H7 (2). The strains sequenced are EPEC O55:H7 3256-97 (CC A2) and USDA 5905 (CC A2), O157:H7 LSU-61 (CC A unknown), EHEC O157:H$^-$ 493/89 (CC A4) and H2687 (CC A4), and EHEC O157:H7 G5101 (CC A5) (3).

Genomic DNA from each strain was isolated from overnight cultures with a DNeasy blood and tissue kit (Qiagen). The genomes were sequenced by 454 Titanium pyrosequencing (Roche), according to the manufacturer's instructions at 20× coverage. Genomic sequence contigs for strains 3256-97 and USDA 5905 were assembled with the 454 Life Sciences Newbler software package version 2.3 (Roche) using the complete *E. coli* O55:H7 strain CB9615 genome (see above for accession number) as a reference. Genomic sequence contigs for strains LSU-61, 493/89, H2687, and G5101, were assembled with the complete *E. coli* O157:H7 strain Sakai genome as a reference. Sequences were annotated with the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html) (6).

A detailed report of a full comparative analysis between the genomes of these six isolates will be included in a future publication.

Nucleotide sequence accession numbers. The draft genome sequences for these six *E. coli* strains are available in GenBank under accession no. AEUB00000000, AEUA00000000,

AETY00000000, AETZ00000000, AEUC0000000, and AETX00000000.

This project was supported by an appointment to the Research Fellowship Program for the Center for Food Safety and Applied Nutrition administered by the Oak Ridge Associated Universities through a contract with the FDA.

REFERENCES

1. **Feng, P.** 1995. *Escherichia coli* serotype O157:H7: novel vehicles of infection and emergence of phenotypic variants. Emerg. Infect. Dis. **1:**47–52.
2. **Feng, P., K. A. Lampel, H. Karch, and T. S. Whittam.** 1998. Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. J. Infect. Dis. **177:**1750–1753.
3. **Feng, P. C., et al.** 2007. Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. Emerg. Infect. Dis. **13:**1701–1706.
4. **Griffin, P. M., and R. V. Tauxe.** 1991. The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome. Epidemiol. Rev. **13:**60–98.
5. **Hayashi, T., et al.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
6. **Klimke, W., et al.** 2009. The National Center for Biotechnology Information's Protein Clusters Database. Nucleic Acids Res. **37:**D216–D223.
7. **Kulasekara, B. R., et al.** 2009. Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. Infect. Immun. **77:**3713–3721.
8. **Perna, N. T., et al.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.
9. **Whittam, T. S., et al.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect. Immun. **61:**1619–1629.
10. **Wick, L. M., W. Qi, D. W. Lacher, and T. S. Whittam.** 2005. Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. J. Bacteriol. **187:**1783–1791.
11. **Zhou, Z., et al.** 2010. Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. PLoS One **5:**e8700.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, July 2011, p. 5030–5033 Vol. 77, No. 14
0099-2240/11/$12.00 doi:10.1128/AEM.00249-11
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

Different IS*629* Transposition Frequencies Exhibited by *Escherichia coli* O157:H7 Strains in the Stepwise Evolutionary Model[▽]

Lydia V. Rump,[1,2] Markus Fischer,[2] and Narjol González-Escalona[1]*

Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, 5100 Paint Branch Parkway, College Park, Maryland 20740,[1] and Institute of Food Chemistry, University of Hamburg, Hamburg, Germany[2]

Received 3 February 2011/Accepted 17 May 2011

The insertion sequence IS*629*, which is highly prevalent in *Escherichia coli* O157:H7 genomes, was found to be absent in O157:H− strains, which are on a divergent pathway in the emergence of O157:H7. Although O157:H− is deficient in IS*629*, it permits IS*629* transposition, with an excision frequency higher than that of ancestral O55:H7 strains but significantly lower than that of pathogenic O157:H7 strains.

Insertion elements (IS) play an important role in the evolution and genomic diversification of *Escherichia coli* O157:H7 (somatic [O] 157 and flagellar [H] 7 antigen) lineages and have been confirmed to actively transpose in O157 genomes (8). In particular, IS*629* has been found in multiple copies in the *E. coli* O157:H7 genome and is the most prevalent IS in this serotype (8). On the other hand, the ancestral O55:H7 strain (GenBank accession no. CP001846) carries only two IS*629* copies (15). It is striking that O157:H− strains, which are on a divergent evolutionary pathway in the stepwise emergence of O157:H7 (4), are IS*629* deficient (Fig. 1). IS*629* presence/absence was determined in the strains analyzed by using IS*629*-specific primers (IS*629*-insideF, GAACGTCAGCGTCTGAAAGAGC; IS*629*-insideR, GTACTCCCTGTTGATGCCAG), targeting conserved regions of the insertion element previously described by Ooka et al. (9). The reason for the absence of IS*629* among strains in the closely related clonal complex (CC) A4 could be either that an IS*629*-carrying mobile element was excluded from infecting those strains or that CC A4 strains exhibit an IS*629* transposition inhibition mechanism, disabling IS*629* transposition. These strains possess numerous other IS that belong to the IS*3* family, as IS*629* does, and it is thereby possible that these might interfere with its transposition, as has been observed for Tn*5* transposition (10). The mechanism of IS*629* transposition is unknown. However, IS*911*, which is another member of the IS*3* family, transposes replicatively, suggesting that IS*629* could also transpose by the copy-paste mechanism (1, 2).

To investigate if IS*629* transposition is inhibited in the CC A4 strains, we constructed vector pIS*629*AB-Tc and introduced it into various strains belonging to the stepwise evolutionary model for *E. coli* O157:H7 (3). pIS*629*AB-Tc carries an actively expressed IS*629* transposase gene (ORFab), which has been shown to enhance IS*629* excision (5, 6) but which lacks the IS*629* inverted repeats (IR), rendering the transposase unable to excise. It also carries an IS*629* analogue (IS*629*-Tc) in which a tetracycline resistance gene (*tetC*) replaces the IS*629* transposase gene, embedded between both IS*629* IR, truncating the vector's ampicillin resistance (Amp^r) gene (Fig. 2). The IS*629*-Tc construct remains able to transpose if there is no inhibition of IS*629* transposition in the individual strain. In the event of precise IS*629* excision, Amp^r transformants are observed. We introduced this vector into a CC A1 strain (DEC5A), a CC A2 strain (3265-97), CC A4 *E. coli* O157:H− strains (493-89, H56929c, and H1085c), a CC A5 strain (G5101), a CC A6 strain (EDL933), and a possible evolutionary intermediary CC A3 strain (LSU-61) (Table 1). The LSU-61 strain possesses various characteristics from strains belonging to clonal complexes A4 and A5 (4). This vector allowed for determining IS*629* transposition and excision frequency in those strains.

Vector pIS*629*AB-Tc was constructed in two stages. First, ORFab was constructed by site-directed mutagenesis and ligated into vector pUC18, creating pIS*629*AB. Second, IS*629*-Tc was inserted into the *amp* gene of the vector (pIS*629*AB-Tc). In detail, IS*629* ORFab was generated by a 1-bp insertion in the overlapping region of IS*629* ORFa and ORFb by the overlap extension method (7). The *gne* gene of *E. coli* O rough:H7 MA6 containing an IS*629* element was amplified by PCR as described previously by Rump et al. (12), using Platinum *Taq* DNA polymerase high fidelity (Invitrogen, Carlsbad, CA). The ~2,700-bp amplicon was gel purified using a Qiaex II agarose gel extraction kit (Qiagen, Valencia, CA) following the manufacturer's instructions and used to generate two fragments needed to obtain the 1-bp insertion. The amplicon was derived using primers IS*629*F-1 (5′-ATATAGAGCTCATGACTAAAAATACTCGTTTTTC-3′) (restriction site SacI underlined) and IS*629*R-2 (5′-AACATCGTATCGTCGATTGTTATACCAGTCC-3′). The second PCR was conducted using primers IS*629*R-3 (5′-CGGGATCCTCAGGCTGCCAGATCA-3′) (restriction site BamHI underlined) and IS*629*alter-M (5′-TTCGACCGCCTCTGGAAAAAAATGATGCCACTGCTGGATAA-3′), which contained the 1-bp insertion (underlined). Three nanograms of each gel-purified fragment was combined with the others, followed by PCR amplification with the primers IS*629*F-1 and IS*629*R-3 using conditions described previously (7). The purified amplicon was digested accordingly and ligated into pUC18 (Stratagene, La

* Corresponding author. Mailing address: FDA, CFSAN, 5100 Paint Branch Parkway, 21HFS-712, College Park, MD 20740. Phone: (301) 436-1937. Fax: (301) 436-2644. E-mail: narjol.gonzalez-escalona@fda.hhs.gov.

▽ Published ahead of print on 27 May 2011.

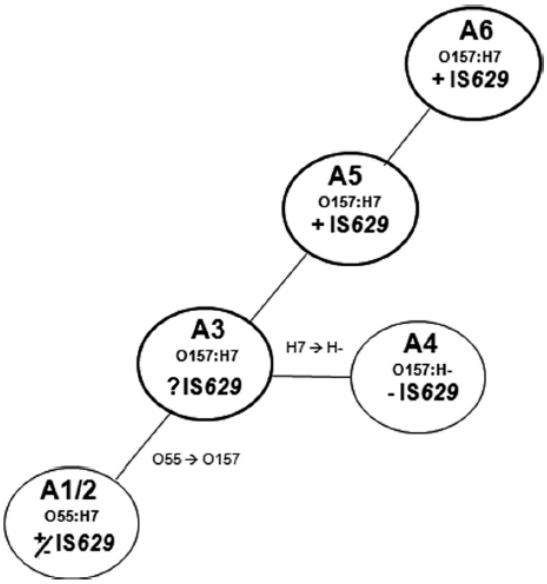


FIG. 1. IS*629* presence/absence in the stepwise evolutionary model of *E. coli* O157:H7 from ancestral O55:H7 strains (representation modified from reference 4). The circles represent the different clonal complexes. IS*629* presence and absence is indicated in bold by +IS*629* and −IS*629*, respectively. Strains belonging to hypothetical CC A3 have not yet been isolated.

Jolla, CA), and the vector construct was electroporated into *E. coli* DH5α (13). Transformants were selected on Luria-Bertani (LB) plates containing 100 μg/ml ampicillin. Amp^r colonies were PCR amplified with vector-specific primers, and those carrying the insert were sequenced in both directions by MCLAB (South San Francisco, CA) to confirm the presence of the ORFab insert in the construct (pIS*629*AB).

IS*629*-Tc was prepared by PCR using primers specific for the 5′ and 3′ ends of the *tetC* gene of the pBR322 plasmid containing both IS*629* IR (IS*629*-Tc-ScaIF [5′-ATCTGAACCGCCCCGGAAATCCTGGAGACTAAACTCCCTGAGAAAGAGGTAAACAGGATGAAATCTAACAATGCGCTCATCGTC-3′] and IS*629*-Tc-ScaIR [5′-GATTGAACCGCCCCGGGTTTCCTGGAGAGTGTTTTATCTGTGAACTCAGGTCGAGGTGGCCCGGCTCCATGC-3′]; the terminal IS*629* seq uences are underlined). The purified amplicon was ligated into the ScaI site of vector pIS*629*-AB. The new vector, pIS*629*A B-Tc, was electroporated into *E. coli* DH5α (13), and tra nsformants were selected on LB plates with 12.5 μg/ml tetrac ycline. Tetracycline-resistant (Tet^r) colonies were PCR amplified with vector-specific primers, and those carrying the insert were sequenced in both directions by MCLAB to confirm the presence of the pIS*629*-Tc insert in the construct (pIS*629*AB-Tc). For transposition frequency studies, pIS*629*A B-Tc was electroporated into the different strains, and 100 CFU of each of the transformants was grown at 37°C overnight in 100 ml of LB broth containing tetracycline (12.5 μg/ml). The transposition occurred during the overnight incubation. Each overnight culture was serially diluted and plated in triplicate on selective LB plates containing either 100 μg/ml ampicillin or 12.5 μg/ml tetracycline. The appearance of Amp^r colonies in comparison to Tet^r colonies is regarded as excision frequency.

We observed that IS*629*-Tc transposed in all strains tested, although with notably different frequencies correlating with grouping by CC, irrespective of the presence or absence of H7 antigen (Table 1). IS*629*-Tc successfully transposed in CC A4 strains, signaling that the absence of IS*629* in the CC A4 strains does not appear to be due to an IS*629* transposition inhibition mechanism. Rather, it is likely that the IS*629*s found in O55:H7 strains were lost and that the CC A4 strains were not in contact with an IS*629*-carrying mobile element after diverging from the hypothetical CC A3 (Fig. 1).

IS*629*-Tc excision frequencies in CC A5 and A6 strains were higher than those in CC A1, A2, and A4 strains. These findings agree partially with previous results from Kusumoto et al. (6), which noted that IS*629*-carrying strains have a higher IS*629* transposition frequency than IS*629*-deficient strains. However, strains from CC A1 and A2 (DEC5A and 3256-97) exhibited a low IS*629* excision frequency, regardless of the presence or absence of IS*629*. The low excision frequency determined for IS*629*-deficient strain 3256-97 (1.5×10^{-8}) was similar to that of other strains lacking IS*629* of various serotypes and genotypes (6). CC A4 strains exhibited a 100-fold-higher excision frequency than the tested CC A1 and A2 strains, although the frequency remained lower than that in CC A5 and A6 strains. This intermediate excision frequency suggested that the presence of IS*629* alone might not enhance the transposition activity of IS*629*-Tc. Kusumoto et al. (6) suggested that IS*629*-possessing strains use a "system to enhance IS*629* excision which might have been introduced by mobile genetic elements that may be linked with IS*629* or other IS elements." This mechanism was recently described by Kusumoto et al. (5) as a protein IS excision enhancer (IEE) which promotes IS*629* excision in O157:H7. Analysis of genome sequences for strains EDL933 (GenBank accession no. AE005174), G5101 (GenBank accession no. AETX01000000), and LSU-61 (GenBank accession no. AEUC01000000) showed the presence of the IEE, explaining the higher excision frequencies. Strain 3256-97 (GenBank accession no. AEUA01000000) lacks the IEE, explaining the low observed IS*629* excision frequency. On the other hand, sequence analysis of two A4 strains (493-89 [GenBank accession no. AETY01000000] and H2687 [GenBank accession no. AETZ01000000]) showed the absence of this specific gene. Hence, the elevated excision frequency in all CC A4 strains relative to that in CC A2 and A1 strains could indicate that these strains possess an upregulation mechanism different than the IEE.

Ooka et al. (8) postulated that IS-related genomic rearrangements may have significantly altered virulence and other phenotypes in O157 strains. However, the elevated IS*629* transposition frequency observed among O157:H7 strains might explain the highly diverse distribution of IS*629* in the O157:H7 genomes (14) and suggests that, in addition to impacting genomic evolution, IS*629* might increase pathogenicity in those strains. Additionally, it might contribute to the appearance of atypical pathogenic strains, like O rough:H7 (*gne*::IS*629* mutant) strains (11, 12). Consequently, should atypical O157:H7 strains become more prevalent, they will create challenges to serological detection methods and could pose a potential health risk. Regardless of the final explana-

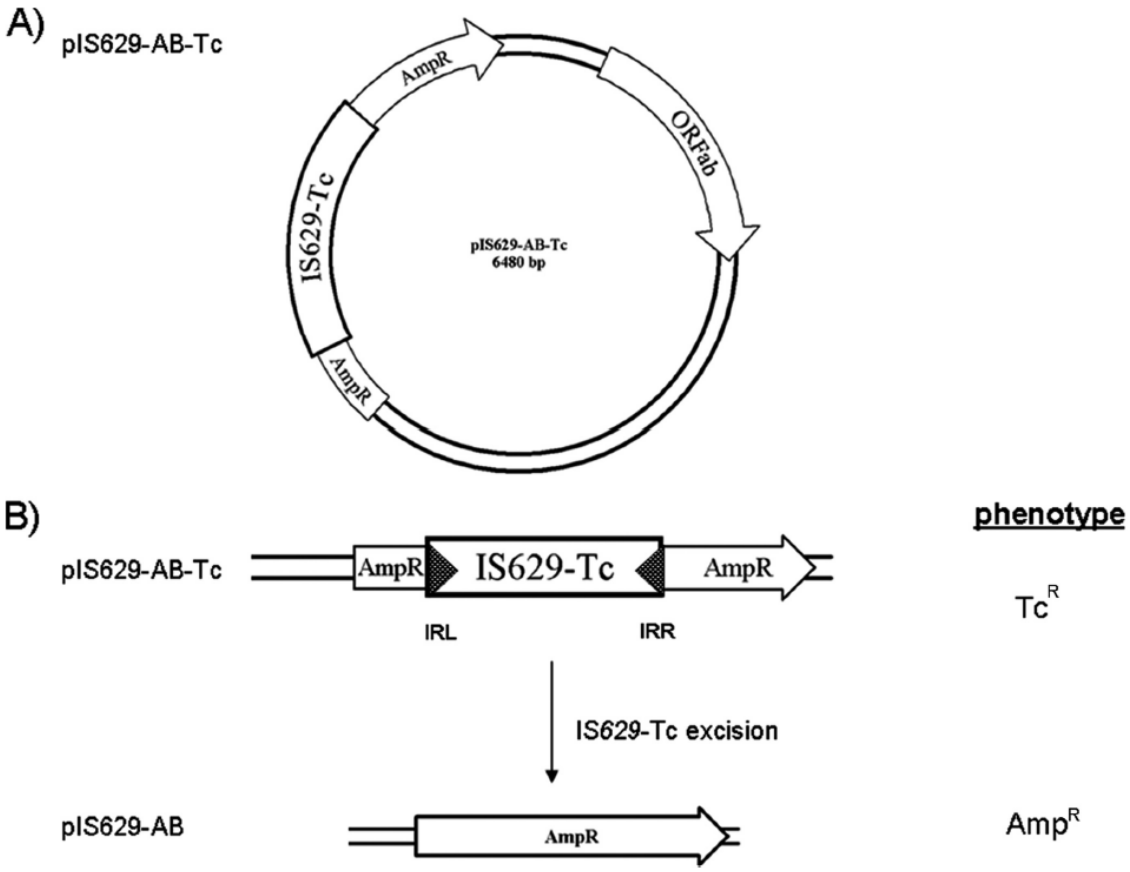


FIG. 2. Schematic representations of the plasmid construct pIS*629*AB-Tc and determination of the transposition frequency. (A) Plasmid construct pIS*629*AB-Tc, containing the IS*629* transposase gene (ORFab) and a tetracycline resistance gene (*tetC*) disrupting the ampicillin (Amp^r) gene. (B) Successful transposition results in ampicillin-resistant colonies (transposition-positive phenotype). Cells showing no transposition remain tetracycline resistant only (original phenotype). Excision frequency was calculated as follows: Amp^r/Tet^r cells.

TABLE 1. Characteristics of *E. coli* strains used in this study and IS629-Tc excision frequencies from pIS*629*AB-Tc in each strain, grouped according to CC, from recent O157 to ancestral O55 serotypes[d]

Strain name	Serotype	CC[a]	IS629[b]	Excision frequency[c]	
				Avg	SD
EDL933	O157:H7	A6	+	1.3×10^{-3}	$\pm 0.4 \times 10^{-3}$
G5101	O157:H7	A5	+	1.6×10^{-3}	$\pm 0.2 \times 10^{-3}$
LSU-61	O157:H7	A?	tr	0.6×10^{-3}	$\pm 0.2 \times 10^{-3}$
493-89	**O157:H−**	**A4**	**−**	$\mathbf{2.6 \times 10^{-6}}$	$\mathbf{\pm 0.9 \times 10^{-6}}$
H56929c	**O157:H−**	**A4**	**−**	$\mathbf{2.2 \times 10^{-6}}$	$\mathbf{\pm 0.3 \times 10^{-6}}$
H1085c	**O157:H−**	**A4**	**−**	$\mathbf{2.3 \times 10^{-6}}$	$\mathbf{\pm 0.4 \times 10^{-6}}$
3256-97	O55:H7	A2	−	1.5×10^{-8}	$\pm 0.6 \times 10^{-8}$
DEC5A	O55:H7	A1	+	2.2×10^{-7}	$\pm 0.1 \times 10^{-7}$

[a] CCs are defined in reference 4.
[b] +, presence; −, absence; tr, truncated IS629.
[c] Three independent experiments were performed for each strain. Excision frequency was calculated as follows: number of Amp^r cells/number of Tet^r cells.
[d] In bold are results for the O157:H− strains lacking IS629.

tion, it is clear that IS*629* has played an integral role in generating genetic diversity across lineages of this important and dangerous bacterial pathogen.

We thank Eric W. Brown for his helpful comments.
This project was supported by an appointment to L.V.R. through the Research Fellowship Program for the Center for Food Safety and Applied Nutrition administered by the Oak Ridge Associated Universities through a contract with the FDA.

REFERENCES

1. **Chandler, M., and J. Mahillon.** 2002. Insertion sequences revisited, p. 305–366. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, DC.
2. **Duval-Valentin, G., B. Marty-Cointin, and M. Chandler.** 2004. Requirement of IS911 replication before integration defines a new bacterial transposition pathway. EMBO J. **23**:3897–3906.
3. **Feng, P., K. A. Lampel, H. Karch, and T. S. Whittam.** 1998. Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. J. Infect. Dis. **177**:1750–1753.
4. **Feng, P. C., et al.** 2007. Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model. Emerg. Infect. Dis. **13**:1701–1706.
5. **Kusumoto, M., et al.** 2011. Insertion sequence-excision enhancer removes

transposable elements from bacterial genomes and induces various genomic deletions. Nat. Commun. **2**:152.
6. **Kusumoto, M., R. Suzuki, Y. Nishiya, T. Okitsu, and M. Oka.** 2004. Host-dependent activation of IS1203v excision in Shiga toxin-producing *Escherichia coli*. J. Biosci. Bioeng. **97**:406–411.
7. **Mikaelian, I., and A. Sergeant.** 1992. A general and fast method to generate multiple site directed mutations. Nucleic Acids Res. **20**:376.
8. **Ooka, T., et al.** 2009. Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. Genome Res. **19**:1809–1816.
9. **Ooka, T., et al.** 2009. Development of a multiplex PCR-based rapid typing method for enterohemorrhagic *Escherichia coli* O157 strains. J. Clin. Microbiol. **47**:2888–2894.
10. **Reznikoff, W. S.** 2008. Transposon Tn5. Annu. Rev. Genet. **42**:269–286.
11. **Rump, L. V., L. Beutin, M. Fischer, and P. C. Feng.** 2010. Characterization of a *gne*::IS*629* O rough:H7 *Escherichia coli* strain from a hemorrhagic colitis patient. Appl. Environ. Microbiol. **76**:5290–5291.
12. **Rump, L. V., P. C. Feng, M. Fischer, and S. R. Monday.** 2010. Genetic analysis for the lack of expression of the O157 antigen in an O rough:H7 *Escherichia coli* strain. Appl. Environ. Microbiol. **76**:945–947.
13. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14. **Yokoyama, E., et al.** 2011. Biased distribution of IS629 among strains in different lineages of enterohemorrhagic *Escherichia coli* serovar O157. Infect. Genet. Evol. **11**:78–82.
15. **Zhou, Z., et al.** 2010. Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor. PLoS One **5**:e8700.

Rump *et al. BMC Microbiology* 2011, **11**:133
http://www.biomedcentral.com/1471-2180/11/133

BMC Microbiology

RESEARCH ARTICLE Open Access

Prevalence, distribution and evolutionary significance of the IS*629* insertion element in the stepwise emergence of *Escherichia coli* O157:H7

Lydia V Rump[1,2*], Markus Fischer[2] and Narjol Gonzalez-Escalona[1]

Abstract

Background: Insertion elements (IS) are known to play an important role in the evolution and genomic diversification of *Escherichia coli* O157:H7 lineages. In particular, IS*629* has been found in multiple copies in the *E. coli* O157:H7 genome and is one of the most prevalent IS in this serotype. It was recently shown that the lack of O157 antigen expression in two O rough *E. coli* O157:H7 strains was due to IS*629* insertions at 2 different locations in the *gne* gene that is essential for the O antigen biosynthesis.

Results: The comparison of 4 *E. coli* O157:H7 genome and plasmid sequences showed numerous IS*629* insertion sites, although not uniformly distributed among strains. Comparison of IS*629*s found in O157:H7 and O55:H7 showed the presence of at least three different IS*629* sub-types. O157:H7 strains carry IS*629* elements sub-type I and III whereby the ancestral O55:H7 carries sub-type II. Analysis of strains selected from various clonal groups defined on the *E. coli* O157:H7 stepwise evolution model showed that IS*629* was not observed in sorbitol fermenting O157 (SFO157) clones that are on a divergent pathway in the emergence of O157:H7. This suggests that the absence of IS*629* in SFO157 strains probably occurred during the divergence of this lineage, albeit it remains uncertain if it contributed, in part, to their divergence from other closely related strains.

Conclusions: The highly variable genomic locations of IS*629* in O157:H7 strains of the A6 clonal complex indicates that this insertion element probably played an important role in genome plasticity and in the divergence of O157:H7 lineages.

Background

Enterohemorrhagic *Escherichia coli* (EHEC) of serotype O157:H7 has been implicated in foodborne illnesses worldwide. It frequently causes large outbreaks of severe enteric infections including bloody diarrhoea, hemorrhagic colitis (HC) and haemolytic uremic syndrome (HUS) [1,2]. This serotype constitutively expresses the somatic (O) 157 and flagellar (H) 7 antigens, thus, these traits are used extensively in clinical settings to identify this highly pathogenic serotype [1]. However some O157:H7 strains, although being genotipically O157 or H7 do not express either of those antigens [3,4]. According to the latest CDC report, *E. coli* O157:H7 infections affect thousands of people every year accounting for 0.7%, 4% and 1.5%, of illnesses, hospitalizations and deaths, respectively of the total U.S. foodborne diseases caused by all known foodborne pathogens [5].

Previously, we characterized two potentially pathogenic O *rough*:H7 strains that did not express the O157 antigen [4,6] but belonged to the most common O157:H7 clonal type. The O *rough* phenotype was found to be due to two independent IS*629* insertions in the *gne* gene that encodes for an epimerase enzyme essential for synthesis of an oligosaccharide subunit in the O antigen. Of the IS elements identified in O157 strains, IS*629* elements are the most prevalent in this serotype and have been confirmed to very actively transpose in O157 genomes [7]. The presence of O-rough strains of this serotype in food and clinical samples is of concern as they cannot be detected serologically in assays routinely used to test for O157:H7 [3].

* Correspondence: Lydiarump@gmail.com
[1]Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, 5100 Paint Branch Parkway, College Park, MD 20740, USA
Full list of author information is available at the end of the article

BioMed Central

© 2011 Rump et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rump et al. BMC Microbiology 2011, 11:133
http://www.biomedcentral.com/1471-2180/11/133

The occurrence of other atypical O157:H7 strains due to IS*629* insertions therefore, might be more common than anticipated. It is generally assumed that IS elements play important roles in bacterial genome evolution and in some cases are known contributors to adaptation and improved fitness [7]. The acquisition or loss of mobile genetic elements, like IS elements, may differ between strains of a particular bacterial species [8]. IS insertion and IS-mediated deletions have been shown to generate phenotypic diversity among closely related O157 strains [7]. It has been shown that O157 is a highly diverse group and a major factor that effects this diversity are prophages [7]. However, in addition to prophages, IS*629* also appears to be a major contributor to genomic diversification of O157 strains. Therefore, it is questionable how much influence IS*629* had on the evolution of O157:H7, or how much importance IS*629* has to changes in virulence in this bacterium.

It has been proposed in an evolutionary model previously that highly pathogenic enterohemorrhagic *E. coli* (EHEC) O157:H7 arose from its ancestor enteropathogenic *E. coli* (EPEC) O55:H7 (SOR+ and GUD+) through sequential acquisition of virulence, phenotypic traits, and serotypic change (A1(*stx*$^-$)/A2(*stx*2) in Figure 1A) [9-11]. After the somatic antigen change from O55 to O157 gave rise to an intermediary (A3) which has not yet been isolated, two separate O157 clonal complexes evolved, splitting into two diverged clonal groups. One of these groups was composed of sorbitol fermenting (SF) non-motile O157:NM strains containing plasmid pSFO157 (A4) (SOR+, GUD+). The other was composed of non-sorbitol fermenting (NSF) O157:H7 strains containing plasmid pO157 (A5) (SOR-, GUD+). The latter, by a mutational inactivation of the *uidA* gene, lost its β-glucuronidase activity which is the most typical O157:H7 phenotype at present (A6) [11]. These A6 strains have spread geographically into disparate locales and now account for most of the diseases caused by EHEC [12].

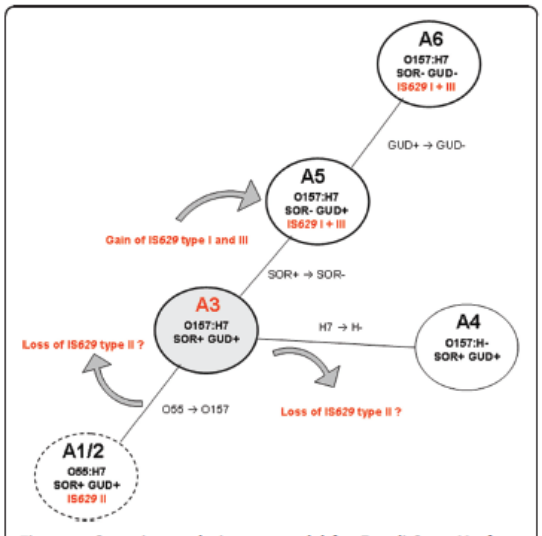


Figure 1 Stepwise evolutionary model for *E. coli* O157:H7 from ancestral O55:H7 [11]. In red letters are the possible events happening and where they occurred during the stepwise evolution. The circle in gray represents an intermediary A3 CC, which has not yet been isolated. SOR - sorbitol fermentation [if (+) fermenting, if (-) non-fermenting or slow fermenting]. GUD - β-D-glucuronidase activity.

IS*629* seems to play an important role in the diversification of closely related strains, specifically O157:H7 [7]. In the present study, we examined the prevalence of IS*629* in a panel of *E. coli* strains, including ancestral and atypical strains associated with the stepwise emergence of *E. coli* O157:H7 to determine the prevalence of IS*629* and its impact on the transitional steps that gave rise to today's highly pathogenic *E. coli* O157:H7.

Results

IS*629* prevalence in *E. coli* O157:H7 genomes

The IS*629* sequence, recently found to be inserted into the *gne* gene in *E. coli* O *rough*:H7 (MA6 and CB7326) [4,13], was used for a BLAST analysis of the genomes of 4 *E. coli* O157:H7 strains belonging to A6 CC (EDL933, Sakai, EC4115 and TW14359) and one O55:H7 strain (CB9615) (Additional file 1, Table S1). The BLAST analysis for IS*629* showed the presence of between 22 and 25 copies in each strain along with their corresponding plasmid (Table 1). Strains Sakai and EDL933 shared 13 of those IS*629* on the chromosome and three on their pO157 plasmids. Strains EC4115 and TW14359 had 17 IS*629* on the chromosome and four on their pO157 plasmid in common. The analysis of the recently released *E. coli* O55:H7 genome strain CB9615 [14] allowed for identification of one IS*629* with an internal 86 bp deletion on the chromosome and an IS*629* in its corresponding pO55 plasmid. Neither the O55 genomic (located on the chromosome backbone) nor the pO55 plasmid IS*629* insertion sites were present in other O157:H7 strains. The absence of the pO55 IS*629* insertion site in O157:H7 strains was expected since they do not carry the pO55 plasmid. However, lack of the genomic O55 IS*629* insertion site in O157:H7 strains is interesting as these strains are known to be closely related [14]. Contrary to what was observed for plasmids pO157 and pO55, IS*629* was absent in plasmid pSFO157 (*E. coli* O157:H- strain 439-89). However, a 66 bp sequence identical to IS*629* was observed in the plasmid which could be a remnant of IS*629*. No genomic sequence is available for an O157:H- strain at this time, thus, this

Rump *et al. BMC Microbiology* 2011, **11**:133
http://www.biomedcentral.com/1471-2180/11/133

Table 1 Prevalence of IS*629* elements in each strain (chromosomes and plasmids) and number of shared IS*629*

	Strain	Serotype	IS*629* Sites	In common with strain				
				Sakai	EDL933	EC4115	TW14359	CB9615
Chromosomes								
	Sakai	O157:H7	19	-	13	9	9	0
	EDL933	O157:H7	21	13	-	6	6	0
	EC4115	O157:H7	19	9	6	-	17	0
	TW14359	O157:H7	21	9	6	17	-	0
	CB9615	O55:H7	1	0	0	0	0	-
Plasmids				pO157 Sakai	pO157 EDL933	pO157 EC4115	pO157 TW14359	pO55 CB9615
	pO157 Sakai	O157:H7	3	-	3	3	3	0
	pO157 EDL933	O157:H7	3	3	-	3	3	0
	pO157 EC4115	O157:H7	4	3	3	-	4	0
	pO157 TW14359	O157:H7	4	3	3	4	-	0
	pSFO157	O157:H-	0	0	0	0	0	0
	pO55 CB9615	O55:H7	1 tr*	0	0	0	0	-

tr* - truncated.

strain could not be investigated for the presence of IS*629*.

IS*629* target site specificity ("hot spots") on chromosomes and plasmids of four *E. coli* O157:H7 strains

The majority of IS*629* elements were located on prophages or prophage-like elements (62%) ("strain-specific-loops", S-loops in Sakai [15]). 28% of IS*629* locations were found on the well-conserved 4.1-Mb sequence widely regarded as the *E. coli* chromosome backbone (*E. coli* K-12 orthologous segment) [15] and 10% were located on the pO157 plasmid. In total, we observed 47 different IS*629* insertion sites (containing complete or partial IS*629*) in the four *E. coli* chromosomes and plasmids by "in silico" analysis (Additional file 2, Table S2). Seven of 47 IS*629* insertion were shared among the 4 diverged strains which suggest that they were also present in a common ancestor.

IS*629* presence in strains belonging to the stepwise model of emergence of *E. coli* O157:H7

A total of 27 *E. coli* strains (Table 2) belonging to the stepwise model proposed by Feng et al. (1998) were examined by PCR for the presence of IS*629* using specific primers [16]. Every strain of clonal complex (CC) A6, A5, A2 and A1 carried IS*629*, except strain 3256-97 belonging to the ancestral CC A2 (Figure 1). Strikingly, however, was the observation that IS*629* was absent in the SFO157 strains belonging to the closely related CC A4 (Figure 2). Whole genome analysis of two A4 strains (493-89 accession no. AETY00000000 and H2687 accession no. AETZ00000000) confirmed the absence of this specific IS element in SFO157 strains [17]. On the other hand, O55:H7 strain 3256-97 (AEUA00000000) carried a truncated IS*629* version missing the target area for the reverse primer (IS*629*-insideR) located in ORFB, explaining the lack of IS*629* by PCR [17]. Additionally, strains USDA5905 (A2) and TB182A (A1) as well as strain LSU-61 (A?) appear to harbor a truncated IS*629* which could indicate the presence of genomic IS*629* found in the O55 strain CB9615. However, since no additional ancestral strains were available for analysis, the distribution of IS*629* in these groups is at present inconclusive.

IS*629* distribution in strains belonging to the stepwise model of emergence of *E. coli* O157:H7

We successfully PCR amplified 38 of the 47 observed IS*629* insertion sites in the 27 O157:H7 strains analyzed (Additional file 3, Table S2). We determined presence or absence of an IS*629* element as well as the IS*629* target site in each strain (Additional file 1, Figure S1). In accordance with the previous finding of total absence of IS*629* in SFO157, none of the A4 CC strains harbored an IS*629* in any of the known IS*629* insertion sites. Likewise, it was observed for A1 and A2 CC strains, indicating that the previously detected IS*629* must be located in some other region of the chromosome. In A5 CC strains, only 3 of the 38 (7%) IS*629* insertion sites harbored an IS*629* (Table 3). Those sites were located on the prophage Sp12, the prophage-like element SpLE1, and on the chromosomal backbone. Interestingly one of the A5 CC strains (strain 1659) did not share any of the known sites harboring IS*629*. The A6 CC strains shared between 6 (16%) and 21 (55%) IS*629* insertions in the known sites and two of them (IS.15: Sp14 and IS.41: pO157) were present in all A6 CC strains. IS*629* prevalence in the A6 strains and the distribution amongst Sp, SpLE, backbone and the pO157 plasmids did not show any

Rump et al. BMC Microbiology 2011, 11:133
http://www.biomedcentral.com/1471-2180/11/133

Table 2 Serotype, sequence type, characteristics and isolation information of strains of *E. coli* used in this study

No.	Name	Other name	Serotype	*stx*	Special characteristics			ST	CC	Source	Year	Reference
					GUD	SOR	plasmid					
1	Sakai	BAA 460	O157:H7	1, 2	-	-	pO157	66	A6	Japan	1996	NC_002695
2	EDL 933	700927	O157:H7	1,2	-	-	pO157	66		USA	1982	AE005174
3	EC 4115		O157:H7	1, 2	-	-	pO157	66		USA	2006	NC_011353
4	TW 14359		O157:H7	1, 2	-	-	pO157	66		USA	2006	CP001368
5	EDL 931	35150	O157:H7	1, 2	-	-	pO157	66				[26]
6	MA6		O157:H7	2	-	-	pO157	66		Malaysia	1998	[6]
7	550654		O157:H7	2	-	-	pO157	66		USA	2009	
8	FDA 413		O157:H7	2	-	-	pO157	66				[10]
9	G5101		O157:H7	1, 2	+	-	pO157	65	A5	USA	1995	[11]
10	1628		O157:H7	1, 2	+	-	pO157	65				[32]
11	1659		O157:H7	1, 2	+	-	pO157	65				[11]
12	EC 97144	TW 10707	O157:H7	1, 2	+	+	pO157	65		Japan	1997	[33]
13	EC 96038	TW 10201	O157:H7	1, 2	+	+	pO157	65				[11]
14	EC 96012	TW 10189	O157:H7	1, 2	+	+	pO157	65				[11]
15	493-89		O157:H-	2	+	+	pSFO157	75	A4	Germany	1989	[11]
16	5412-89		O157:H-	2	+	+	pSFO157	75		Germany	1989	[34]
17	H56929	TW 09159	O157:H-	2	+	+	pSFO157	76		Finland	1999	[11]
18	H56909	TW 09162	O157:H-	2	+	+	pSFO157	76		Finland	1999	[11]
19	H 1085c		O157:H-	2	+	+	pSFO157	76		Scotland	2003	[11]
20	H 2687		O157:H-	2	+	+	pSFO157	76		Scotland	2003	[11]
21	3256-97	TW 07815	O55:H7	2	+	+	?	73	A2	USA	1997	[11]
22	USDA 5905		O55:H7	2	+	+	?	73		USA	1994	[26]
23	TB 182A	TW 04062	O55:H7	-	+	+	?	73	A1	USA	1991	[11]
24	DEC5A		O55:H7	-	+	+	?	73				[11]
25	LSU-61		O157:H7	-	+	+	?	237	?	USA	2001	NC_002695
26	Sakai PF		O157:H7	1, 2	-	-	pO157	66	A6	Japan	1996	AE005174
27	43895	CDC EDL 933	O157:H7	1,2	-	-	pO157	69	A6	USA	1982	

stx - shiga toxin gene, GUD - β-glucuronidase activity, SOR - sorbitol fermentation, ST - sequence type as determined by a combination of seven genes http://www.shigatox.net/stec/cgi-bin/index, CC - Clonal complex [11], ? - Unknown. Sakai PF and 43895 are strain derived after numerous subculture passages from the original Sakai and EDL933 strains, respectively.

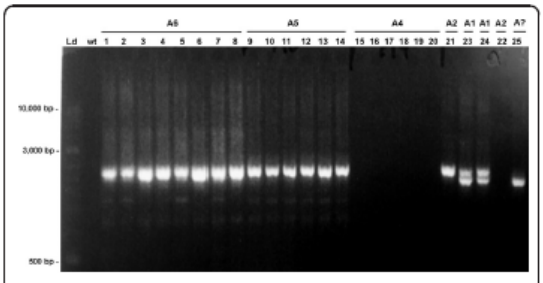


Figure 2 Gel-electrophoresis of the PCR products for IS*629* presence in strains belonging to the stepwise model of emergence of *E. coli* O157:H7. Lanes: Ld, molecular weight ladder (Gene Ruler); wt, Blank; 1 - 25, strains numbered according to Table 2. A1-A6, Clonal complexes, A?, CC unknown.

specific pattern, however it appears that IS*629* transposes actively in the A6 CC.

Figure 1B shows a maximum parsimony tree obtained for A5 and A6 CC strains using IS*629* presence/absence in the target site and presence/absence of IS*629* target site (chromosome or plasmid region) (Table 3 and Additional file 4, Table S3). Strains belonging to A1, A2, and A4 CCs were not included in this analysis because they either lack IS*629* (A4) or IS*629* is located in other regions on the chromosome than the ones determined for O157:H7 strains. The parsimony tree allowed to separate strains belonging to A5 from A6 strains as proposed in the stepwise model (Figure 1 and 3A) [10,12]. Furthermore, it showed the existence of high diversity among A5 and A6 CC strains similar to what has been shown by PFGE [11]. The validity of this analysis needs

esence/absence in CC strains from the O157:H7 stepwise evolutionary model *(Continued)*

+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
-	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	+	+	-	-	-	-	+	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
-	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	+	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
D	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND

- 47 (NR) starting with all sites in Sakai, followed by all additional, unshared sites from EDL933, EC4115, the sites found in the plasmids and unshared sites of strain T
ough:H7 strain MA6 was numbered IS.39 [4]. A1 - A6 are strains belonging to the different clonal complexes.
ıent; back - backbone; ND -Not determined, primers failed to amplify the region.

Rump *et al. BMC Microbiology* 2011, **11**:133
http://www.biomedcentral.com/1471-2180/11/133

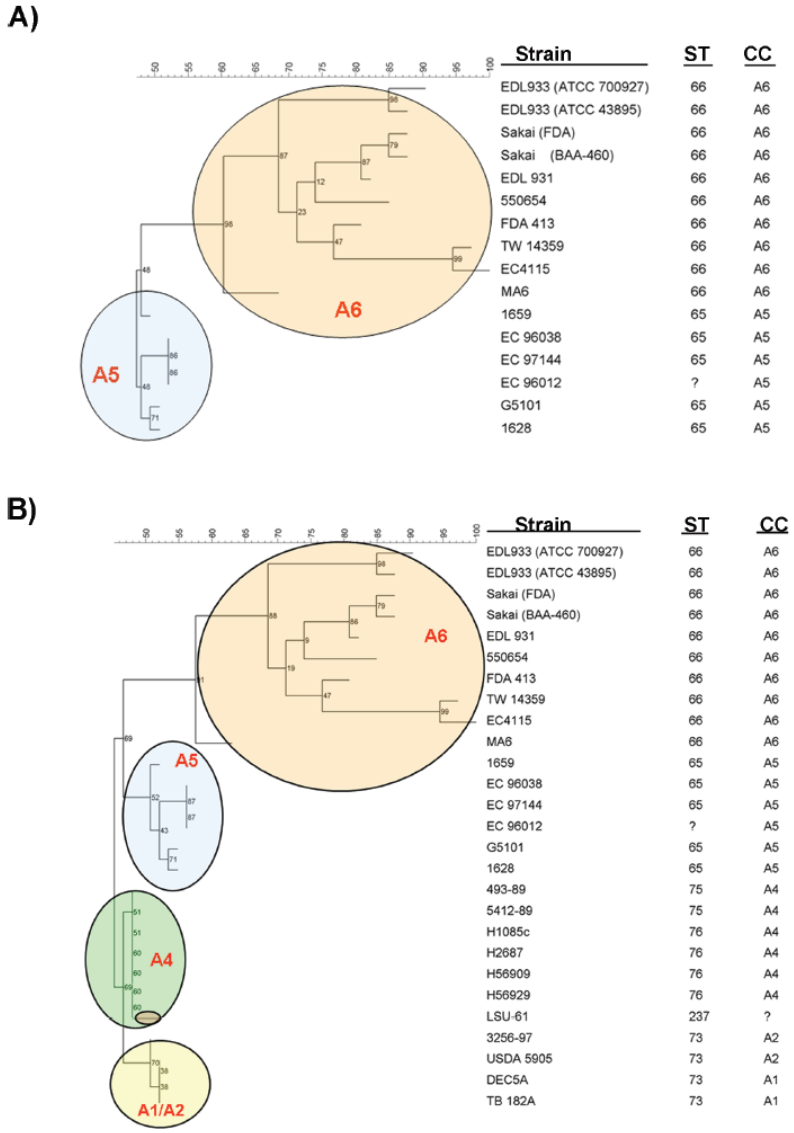


Figure 3 Evolutionary significance of IS*629* in the emergence of *E. coli* O157:H7. A) Maximum parsimony tree obtained using the distribution of IS*629* and IS*629* target sites in the 14 O157:H7 strains analyzed in the present study (Table 3 and Additional file 4, Table S3). B) Maximum parsimony tree obtained using IS*629* target sites for the 27 strains analyzed in the present study (Additional file 4, Table S3). The colored ellipses mark the different CCs. CC - clonal complex; ST - sequence type.

to be explored further using more O157:H7 strains belonging to either A5 or A6 CCs. Besides using 25 different strains for the analysis, we also included additional Sakai and EDL933 strains. Sakai strains were one from ATCC (BAA-460) and the other from a personal collection (FDA). EDL933 strains were provided by ATCC whereby strain EDL933 700927 derived from EDL933 43895. PFGE analysis showed only minimal changes between the original (ATCC) and the derived ones confirming their identity (data not shown). The analysis using the IS*629* distribution also showed minimal changes in the IS*629* distribution as well among the Sakai and EDL933 strains. The use of IS*629* presence/absence in specific regions has been used before to help

Rump *et al. BMC Microbiology* 2011, **11**:133
http://www.biomedcentral.com/1471-2180/11/133

detecting outbreak related strains as described by Ooka, et al (2009) and for population genetics analysis as described by Yokoyama, et al (2011), and appears to be a promising and adequate technique to distinguish closely related O157:H7 strains. However, both methodologies takes no notice of the information about the presence of the region where IS*629* was inserted into. The presence/absence of a specific region in *E. coli* O157:H7 chromosomes, irrelevant of the presence of IS*629*, could provide additional information regarding relatedness among those strains.

IS*629* insertion site prevalence in the strains belonging to the stepwise model of emergence of *E. coli* O157:H7

PCR analysis for the presence of IS*629* insertion sites showed that sites located on the chromosomal backbone structure were present in all tested strains from the different clonal complexes (Table 4 and Additional file 4). However, neither A1, A2, nor A4 CC strains harbored any IS*629* in backbone IS*629* insertion sites.

Contrary to what was observed in the well-conserved backbone, IS*629* insertion sites in prophages and prophage-like elements in different strains were found to be highly variable (Table 5 and Additional file 4, Table S3). As seen for the backbone IS*629* insertion sites, some of the phage associated IS*629* insertions sites were present in A1, A2 and A4 CC strains; however they lacked IS*629*. Many of the IS*629* sites on phages were unique to the A6 CC strains (7 of 13) suggesting that they are strain-specific. This result underscores significant differences in the presence of phage-related sequences between the strains belonging to the stepwise model of *E. coli* O157:H7.

The two IS*629* insertions in O55 and its corresponding plasmid pO55 were observed to be present in only one ancestral A2 and both A1 CC strains (data not shown). A6, A5, and A4 CC strains as well as A2 CC strain 3256-97 (IS*629*-deficient) lacked the IS*629* insertion site in these regions. Interestingly, strain LSU-61 which carries multiple characteristics for O157:H7 and is thought to be ancestral to A5 CC strains (Feng et al 2007), appeared to carry the truncated genomic IS*629* insertion.

Table 4 Presence of IS*629* target sites on the backbone

IS*629* target sites	A1	A2	A3	A4	A5	A6
IS.10	+/-	+	NA	+	+	+/-
IS.11	+	+	NA	+	+	+
IS.13	+	+	NA	+	+	+
IS.17	+	+	NA	+	+	+
IS.19	+	+	NA	+	+	+
IS.32	+	+	NA	+	+	+
IS.34	+	+	NA	+	+	+
IS.38	+	+	NA	+	+	+
IS.39	+	+	NA	+	+	+
IS.46	-	-	NA	+/-	+	+

NA, not applicable; + presence; - absence; +/- present in some strains.

Table 5 Presence of phage or phage-like associated IS*629* target sites

IS*629* target sites	A1	A2	A3	A4	A5	A6
Sp 1	_	_	NA	_	_	+
Sp 2	+	+	NA	+	+	+
Sp 4	+	+	NA	+	+	+
Sp 5	_	_	NA	_	_	+
Sp 8	_	_	NA	_	_	+
Sp 12	_	+	NA	+	+	+
Sp 13	_	_	NA	_	_	+
Sp 14	_	_	NA	+	+	+
Sp 17	_	_	NA	_	_	+
SpLE 1	_	_	NA	_	+	+
SpLE 2	_	_	NA	_	_	+
SpLE 3	_	_	NA	_	_	+
SpLE 5	_	_	NA	_	+	+

Sp - Phage; SpLE - Phage-like element; NA - not applicable; + presence; - absence.

Since the strains belonging to the stepwise model share variable IS*629* insertion sites we reconstructed their evolutionary path using this information. A parsimony tree using the IS*629* target sites presence/absence produced a tree that was nearly analogous to the proposed model of stepwise evolution for O157:H7 from ancestral O55:H7 strains [10], with A1/A2 CC strains at the base of the tree, followed by A4 CC, A5 CC and A6 CC strains in that order (Figure 3B).

Phylogenetic analysis of IS*629* elements in the four *E. coli* O157:H7 and O55:H7 genomes

The phylogenetic analysis of IS*629* elements revealed that IS*629* in *E coli* O157:H7 can be divided into three different sub-types (Figure 4). That is, IS*629* of sub-type I and II differ in average 4% (> 55 bp) while sub-type II and III differed by 5% (> 60 bp). Sub-type I appears to be most closely related to those of IS*1203* (IS*629* isoform) found in O111:H- [18]. IS*629* sub-type II appears to be most closely related to those of IS*629* found in *Shigella* [19]. IS*629* sub-type III appears to be most closely related to those of IS*629* found in *E. coli* O26:H11 [20]. Therefore, analysis of all targeted IS*629* elements showed that strains from A6 CC seem to carry both IS*1203* (sub-type I) and IS*629* (sub-type III) whereby the ancestral O55:H7 strain carries IS*629* (sub-type II). Since IS*629* sub-type II found in the ancestral O55:H7 strain is significantly different from the other two IS*629* sub-types (O157:H7 strains) and sub-type II is no longer present in certain O157:H7 strains (A6 CC), these data imply that IS*629* sub-type I and III were recently

Rump et al. BMC Microbiology 2011, 11:133
http://www.biomedcentral.com/1471-2180/11/133

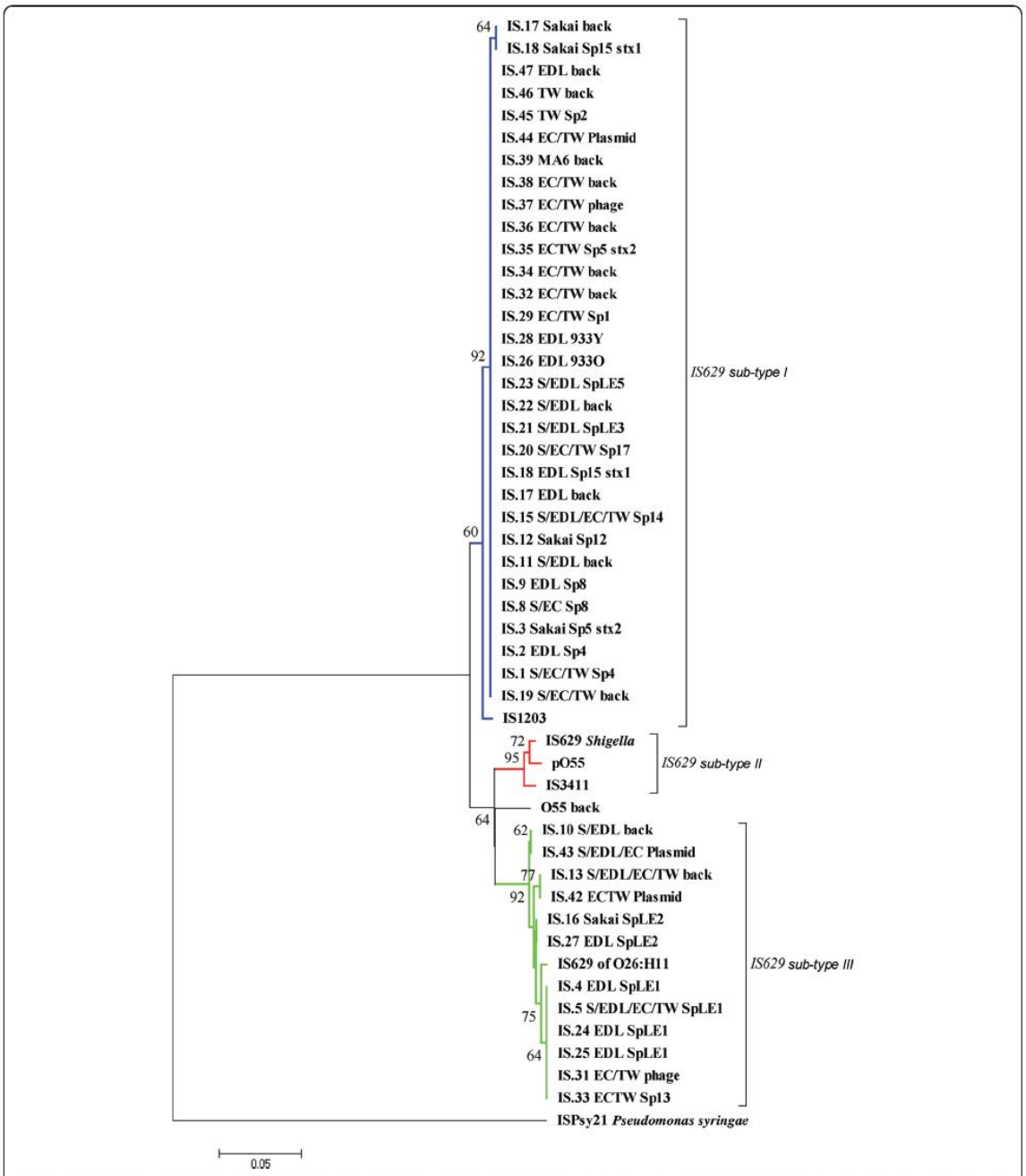


Figure 4 Phylogenetic tree of IS*629* in *E. coli* O157:H7 and O55:H7 showing the three different IS*629* sub-types present on those five genomes. IS*629* sub-type I differed from sub-type II by 4% (> 55 bp) and sub-type II differed from sub-type III by 5% (> 60 bp). IS*629* sub-type II was only present in O55:H7 genome (A1/A2 CC) while sub-type I and III were present in all O157:H7 genomes (A6 CC). The evolutionary history was inferred using the Minimum Evolution method [31]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Bootstrap support when above 50% is shown at nodes. Sp- prophages; SpLE - prophage-like elements; and back - backbone.

acquired by *E. coli* O157:H7 strains after the separation from the sub-lineage leading to the A4 CC strains therefore not carrying IS*629*.

Discussion

IS elements are in general regarded as genetic factors that significantly contribute to genomic diversification and evolution [7]. It was determined by Ooka et al (2009) that IS elements IS*629* and IS*Ec8*, found in the O157:H7 lineage, serve as an important driving force behind the genomic diversity. However, only a few genome-wide studies have been conducted to compare IS distributions in closely related genomes. In our study we determined that IS*629* insertions in *E. coli* O157:H7 are widespread distributed on the genome and differ significantly from strain to strain. Although the ancestral O55:H7 strain carried only two IS*629* with one on the chromosome and one on the pO55 plasmid, the four O157:H7 genomes carried between 22 and 25 IS*629* copies on the chromosome and the corresponding pO157 plasmid.

IS*629* does not seem to specifically integrate in sequence-based target sites, which explains the highly diverged flanking sites found in the genomes we examined. Sequence-specific insertion is exhibited to some degree by several elements and varies considerably in stringency [21]. Other elements exhibit regional preferences which are less obvious to determine [21]. IS elements frequently generate short target site duplication (TSD) flanking the IS upon insertion [21]–this feature was also observed for IS*629* in the four O157:H7 strains. IS*629* duplicated between 3 to 4 base pairs at the insertion site and was observed for 21 of the 47 IS*629* insertion sites with matching identical base pairs up- and down-stream of IS*629*. A comparison of 21 TSDs created by IS*629* in the four strains analyzed here did not reveal as many similarities as observed previously by Ooka et al (2009). The comparison of 25 bp up- and downstream of each insertion site did not show any similarities or patterns which would have suggested a target preference or "hot-spot" for IS*629* insertions. Hence, insertion site specificity for IS*629* remains unknown. However, IS*629* is frequently surrounded by other IS elements ('IS islands') and was found in the same gene (*gne*) inserted in different sites [4,13]. Although no specific "hot-spot" for IS629 insertions was observed, it seems highly possible that mobile elements like plasmids, phages or phage-like elements could have functioned as vectors for IS629 introduction into O157:H7 genomes. These observations suggest that an insertion might occur preferentially in a region of the chromosome however these events may not be sequence specific.

IS*629* insertion sites located on the backbone seem to be conserved in almost all of the strains studied here, whereby sites located on phages and phage-like areas appear to differ between all strains. These findings affirm the presence of regions of genomic stability and regions of genomic variability that exist within O157:H7 populations and closely related strains. It is noteworthy that sites associated with phages seem to be present predominantly in closely related strains. The majority of the phages present in the A6 CC strains appear to be unique to this complex. Since bacteriophages are known to contribute to the diversification of bacteria [22], they seem to be a major determinant in generating diversity among O55:H7, O157:H- and O157:H7 strains. The comparison of IS*629* prevalence in A5 and A6 CC as well as IS*629* insertion site prevalence in all strains allowed distinguishing strains from different complexes as it has been proposed in the evolution model for O157:H7 (Figure 1A) [11]. Adding the "same" strain from different collections, Sakai and EDL933 allowed confirmation of the stability of IS*629* sites. Minimal changes in IS*629* presence/absence were observed and could have occurred due to different storage conditions and passages. Despite these subtle changes, strains grouped tightly together on the parsimony tree. Therefore, this analysis can be used to further distinguish closely related O157:H7 strains. These findings are in agreement with a recently described IS*629* analysis in three O157 lineages [23]. Similarly to what was determined for A6 and A5 CC strains, Yokoyama et al (2011) determined that IS*629* distribution was biased in different O157 lineages, indicating the potential effectiveness of IS-printing for population genetics analysis of O157. Furthermore, Ooka et al. (2009) found that IS-printing can resolved about the same degree of diversity as PFGE. Since A1, A2 and A4 CC strains did not share IS*629* insertions, their population genetics analysis however, remains limited to closely related O157:H7 strains.

Comparison of IS*629*s found in O157:H7 and O55 pointed out extensive divergence between these elements. At least three different IS*629* types could be distinguished differing in 55 to 60 bp. The O157:H7 strains carry IS*629* elements subtype I and III whereby O55:H7 carries type II only. It is notable that only four nucleotide differences were observed among seven housekeeping genes comprising a current MLST scheme http://www.shigatox.net/ecmlst/cgi-bin/dcs between A1 CC strain DEC5A and A6 CC strain Sakai. These two strains, in particular, are taken to represent the most ancestral and most derived *E. coli*, respectively, in the stepwise evolutionary model for this pathogen. If the IS*629* type I and III observed in A6 CC strains resulted from divergent evolution of IS*629* type II, the amount of changes observed among these IS types should be similar to those observed for the MLST loci examined above. However, the number of nucleotide substitutions

Rump et al. BMC Microbiology 2011, 11:133
http://www.biomedcentral.com/1471-2180/11/133

between IS*629* type I and III in O157:H7 from type II in O55:H7 was 10-fold higher. Thus, the differences between IS*629* types are more significant than those observed for housekeeping genes. This indicates that IS*629*-type II was most likely lost and IS*629*-type I and III were acquired independently in distinct *E. coli* O157:H7 lineages. Further supporting this thesis was the fact that one of the IS*629* type II copies was found on the pO55 plasmid, which was subsequently lost during evolution towards O157:H7 strains. The other IS*629* copy in O55, with a unique internal deletion, is located in the chromosome and appears to be part of a mobile region [24] which is absent in O157:H7 strains.

Interestingly, the ancestral IS*629*-deficient A2 O55:H7 strain 3256-97 is also lacking both IS*629* associated regions found in the O55:H7 strains. Our analysis of common IS*629* target sites demonstrated that strain 3256-97 seems to be more closely related to A4 and A5 CC strains than other A1 and A2 strains. Therefore, it is likely that IS*629* has been lost in strain 3256-97 as well as in the hypothetical A3 precursor. These results may indicate that strain 3256-97 or a similar strain lacking IS*629* might have given rise to IS*629*-deficient A4 CC strains.

E. coli O157:H7 strains carry multiple IS*629* copies while the non-pathogenic K-12 strain lacks IS*629* but carries other IS elements. Other pathogenic *E. coli* strains, amongst the top six non-O157 STEC O26:H11, O111:H- and O103:H2 [25], also harbor various copies of IS*629* elements in their genomes. Genome sequences for the other three most important pathogenic non-O157 STEC; O45, O145, and O121 are not available to date thus the presence of IS*629* elements is unknown. Interestingly, they also share the same reservoir with O157:H7 (*e.g.* cattle), shiga-toxins, haemolysin gene cluster, other virulence factors and several phages and phage-like elements [25]. Ooka et al (2009) postulated that IS-related genomic rearrangements may have significantly altered virulence and other phenotypes in O157 strains. These findings suggest that IS*629* might not only have a great impact in their genomic evolution but might increase the pathogenicity of those strains as well.

Conclusions

The genomic sequence analysis showed that IS*629* insertion sites exhibited a highly biased distribution. IS*629* was much more frequently located on phages or prophage-like elements than in the well-conserved backbone structure, which is consistent with the observations by Ooka et al (2009). IS*629* was found to be present in the A1 and one of two A2 CC strains examined as well as in all the O157:H7 strains of A5 and A6 CC, however it was totally absent in the 6 examined SFO157 strains of A4 CC. The A4 CC strains are related to but on a divergent evolution pathway from O157:H7. These results suggest that the absence of IS*629* in A4 strains probably occurred during the divergence, but it is uncertain if it contributed to the divergence. Overall, IS*629* had great impact on the genomic diversification of the *E. coli* O157:H7 lineage and might have contributed in the emergence of the highly pathogenic O157:H7.

Methods

Bacterial strains

The bacterial strains used in this study are listed in Table 2 and were chosen to represent typical EHEC and EPEC strains from the different clonal complexes from the evolution model for *E. coli* O157:H7 [11] with different serotypes (O157:H7, O157:H- and O55:H7) and different characteristics (e.g. β-glucuronidase activity (GUD), sorbitol fermentation (SOR).

"In silico" analysis

Various *E. coli* O157:H7 and non-O157 chromosomes and pO157 plasmids (Additional file 2, Table S1) deposited at the National Center for Biotechnology Information (NCBI) database were queried for IS*629* (accession number X51586) presence and insertion loci using BLAST analysis. Furthermore, approximately 400 bp up- and downstream of the flanking regions of each new localized IS*629* in the chromosome and the plasmids were compared with each other. We investigated whether an IS*629* was also present in the other strains or appears exclusively in either the chromosome or the plasmids.

Nucleic acid extraction and determination of IS*629* presence

DNA used as the template for PCR was prepared from overnight cultures grown in Luria-Bertani Broth (LB) and purified using the MASTER PURE™ DNA Purification kit (EpiCentre, Madison, WI). For determining IS*629* presence in the *E. coli* strains, we conducted a "touchdown" multiplex PCR using IS*629*-specific primers targeting conserved regions of the insertion element previously described by Ooka et al. (2009): IS*629*-insideF (5'- GAACGTCAGCGTCTGAAAGAGC-3') and IS*629*-insideR (5'- GTACTCCCTGTTGATGCCAG-3') and specific 16S rDNA primers: SRM86 (5'- AGAAGCACCGGCTAACTC -3') [7] and SRM87 (5'- CGCATTTCACCGCTACAC-3') [26]. The latter were used as internal amplification control. PCR amplifications were performed using 0.5 ng of template DNA and in a final volume of 30 μl. The PCR reaction mixture contained 2.5 U of HotStart Taq Polymerase (Qiagen, Valencia, CA), 1X Taq polymerase buffer, 2.0-3.5 mM $MgCl_2$, 400 μM each deoxynucleoside triphosphate (dNTP), 300 nM each IS*629* primer pair, and 300 nM each 16S rDNA primer pair. The "touchdown" PCR [27] conditions were: 1 cycle of 95°C for 15 min; 10 cycles of 95°C for 30 s, 69-59°C (-1°C/cycle)

for 15 s and 72°C for 1:30 min; followed by 35 cycles consisting of 95°C for 30 s, 58°C for 20 s, and 72°C for 1.5 min, and a final extension at 72°C for 4 min. Amplicons were visualized on a 1% agarose gel in Tris-Borate EDTA (TBE) buffer containing 0.3 μg/ml ethidium bromide.

Determination of IS*629* specific location and IS*629* insertion sites

For the analysis of the IS*629* insertion sites, primers were designed to target the different IS*629* flanking regions in each strain and the plasmids. The presence/absence of amplicons would determine the presence/absence of the specific insertion sites and the sizes of each amplicons would indicate the presence/absence of IS*629* at those loci. Potential primers were analyzed for their ability to produce stable base pairing with the template using the NetPrimer software (PREMIER Biosoft International http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html). The size of the PCR products were between 1,500 - 2,500 bp in the case of IS*629* presence in a strain or between 200 - 800 bp in the case that the specific flanking region existed in the chromosome but did not contain an IS*629* element. Each multiplex PCR contained a set of 16S rDNA primers as PCR internal control (either set SRM86/SRM87 or VMP5 (5'-AGAAGCACCGGCTAACTC-3') and VMP6 (5'-CGCATTTCACCGCTACAC-3') [28]), and IS*629* insertion site specific primers. The list of the 40 primer combinations for each IS*629* site and PCR conditions can be found in Additional file 5, Table S4.

IS*629* presence/absence parsimony tree analysis

IS*629* PCR fragments sizes indicating IS*629* presence/absence and IS*629* target site presence/absence identified by PCR using primers specific for each IS*629* observed in 4 *E. coli* O157:H7 genomes were entered as binary characters (+ or -) into BioNumerics version 6.0 (Applied Maths, Saint-Martens-Latem, Belgium). IS*629* presence/absence and IS*629* target site presence/absence were used to create a phylogenetic parsimony tree rooted to A5 CC strains for A5/A6 CC strains analysis (Figure 1B) and statistical support of the nodes was assessed by 1000 bootstrap re-sampling. IS*629* target site presence/absence were used to create a phylogenetic parsimony tree rooted to A1/A2 CC strains for strains of the entire model (A1 - A6) (Figure 1C) and statistical support of the nodes was assessed by 1000 bootstrap re-sampling.

IS*629* phylogenetic analysis

Minimum evolution tree for IS*629* sequences present in 4 *E. coli* O157:H7 genomes, two IS*629* in O55:H7 genome, IS*629* sequences from *Shigella*, two other IS*629* isoforms (IS*1203* and IS*3411*), and IS*Psy21* (a member of the IS*3* family and sharing only 68% homology with IS*629*) as out-group (*Pseudomonas syringae* pv. savastanoi TK2009-5) was constructed using Mega version 4.0 [29]. The evolutionary distances were computed using the Kimura 2-parameter method [30] and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 299 positions in the final dataset. The statistical support of the nodes in the ME tree was assessed by 1000 bootstrap re-sampling.

Additional material

Additional file 1: "Figure S1". Schematic representation of the strategy used for primer design. Primer pairs: A: presence/absence of IS*629* at specific loci, B: IS*629* internal primer. A) Amplification product for locations where the IS629 element is present; B) Amplification product for locations where the IS629 element is absent, although the up-and downstream flanking region is present in the genome but not carrying an insertion.

Additional file 2: "Table S1". Genomes and plasmids investigated by "in silico" analysis.

Additional file 3: "Table S2". IS*629* insertion sites in O157:H7 strains with complete genomes available in Genbank (Additional Table 1). In bold are the locations shared by the four O157:H7 strains. The direct repeats (duplication are in red). IS*629* sites were numbered from 1 - 47 starting with all sites in Sakai, followed by all additional, unshared sites from EDL933, EC4115, the sites found in the plasmids and unshared sites of strain TW1435. The newly found IS*629* insertion in O *rough*:H7 strain MA6 was numbered IS.39.

Additional file 4: "Table S3". IS*629* target site presence/absence in CC strains from the O157:H7 stepwise evolutionary model.

Additional file 5: "Table S4". Primer sequences for the amplification of each flanking IS*629* regions on the four *E. coli* genomes available (see Additional Table 2). If IS absent size equal to 0 bp means that the primer pair was designed with one target region inside IS*629* therefore the IS*629* target site could not be observed.

Acknowledgements and Funding

The authors thank Eric W. Brown for his helpful comments. This project was supported by an appointment to LVR through the Research Fellowship Program for the Center for Food Safety and Applied Nutrition administered by the Oak Ridge Associated Universities through a contract with the FDA.

Author details

[1]Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, 5100 Paint Branch Parkway, College Park, MD 20740, USA. [2]Institute of Food Chemistry, University of Hamburg, Hamburg, Germany.

Authors' contributions

LVR conceived the study, participated in the experimental design, performed all the experiments, and participated in the production of the draft of the manuscript. MF participated in the experimental design, and production of the draft of the manuscript. NGE participated in the experimental design and coordination, performed most of the sequence analysis and phylogeny, and participated in production of the draft of the manuscript. All authors have read and approved the final manuscript.

Received: 17 December 2010 Accepted: 14 June 2011
Published: 14 June 2011

References

1. Feng P: ***Escherichia coli*** **serotype O157:H7: novel vehicles of infection and emergence of phenotypic variants.** *Emerg Infect Dis* 1995, **1**:47-52.
2. Griffin PM, Tauxe RV: **The epidemiology of infections caused by *Escherichia coli* O157:H7, other enterohemorrhagic *E. coli*, and the associated hemolytic uremic syndrome.** *Epidemiol Rev* 1991, **13**:60-98.
3. Monday SR, Minnich SA, Feng PC: **A 12-base-pair deletion in the flagellar master control gene flhC causes nonmotility of the pathogenic German sorbitol-fermenting *Escherichia coli* O157:H- strains.** *J Bacteriol* 2004, **186**:2319-2327.
4. Rump LV, Feng PC, Fischer M, Monday SR: **Genetic analysis for the lack of expression of the O157 antigen in an O Rough:H7 *Escherichia coli* strain.** *Appl Environ Microbiol* 2010, **76**:945-947.
5. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM: **Foodborne illness acquired in the United States–major pathogens.** *Emerg Infect Dis* 2011, **17**:7-15.
6. Feng P, Sandlin RC, Park CH, Wilson RA, Nishibuchi M: **Identification of a rough strain of *Escherichia coli* O157:H7 that produces no detectable O157 antigen.** *J Clin Microbiol* 1998, **36**:2339-2341.
7. Ooka T, Ogura Y, Asadulghani M, Ohnishi M, Nakayama K, Terajima J, Watanabe H, Hayashi T: **Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes.** *Genome Res* 2009, **19**:1809-1816.
8. Arbeit RD: **Laboratory procedures for the epidemiologic analysis of microorganisms.** In *Manual of clinical microbiology*.. 6 edition. Edited by: Murray PJ, Baron EJ, Pfaller MA, Tenover FC, Yolken RH. Washington, D.C.: ASM Press; 1995:190-208.
9. Whittam TS, Wolfe ML, Wachsmuth IK, Orskov F, Orskov I, Wilson RA: **Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea.** *Infect Immun* 1993, **61**:1619-1629.
10. Feng P, Lampel KA, Karch H, Whittam TS: **Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7.** *J Infect Dis* 1998, **177**:1750-1753.
11. Feng PC, Monday SR, Lacher DW, Allison L, Siitonen A, Keys C, Eklund M, Nagano H, Karch H, Keen J, Whittam TS: **Genetic diversity among clonal lineages within *Escherichia coli* O157:H7 stepwise evolutionary model.** *Emerg Infect Dis* 2007, **13**:1701-1706.
12. Wick LM, Qi W, Lacher DW, Whittam TS: **Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7.** *J Bacteriol* 2005, **187**:1783-1791.
13. Rump LV, Beutin L, Fischer M, Feng PC: **Characterization of a gne::IS629 O rough:H7 *Escherichia coli* strain from a hemorrhagic colitis patient.** *Appl Environ Microbiol* 2010, **76**:5290-5291.
14. Zhou Z, Li X, Liu B, Beutin L, Xu J, Ren Y, Feng L, Lan R, Reeves PR, Wang L: **Derivation of *Escherichia coli* O157:H7 from its O55:H7 precursor.** *PLoS One* 2010, **5**:e8700.
15. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H: **Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12.** *DNA Res* 2001, **8**:11-22.
16. Ooka T, Terajima J, Kusumoto M, Iguchi A, Kurokawa K, Ogura Y, Asadulghani M, Nakayama K, Murase K, Ohnishi M, Iyoda S, Watanabe H, Hayashi T: **Development of a multiplex PCR-based rapid typing method for enterohemorrhagic *Escherichia coli* O157 strains.** *J Clin Microbiol* 2009, **47**:2888-2894.
17. Rump LV, Strain EA, Cao G, Allard MW, Fischer M, Brown EW, Gonzalez-Escalona N: **Draft Genome Sequences of Six *Escherichia coli* Isolates from the Stepwise Model of Emergence of *Escherichia coli* O157:H7.** *J Bacteriol* 2011, **193**:2058-2059.
18. Paton AW, Paton JC: **Characterization of IS1203, an insertion sequence in *Escherichia coli* O111:H-.** *Gene* 1994, **150**:67-70.
19. Matsutani S, Ohtsubo E: **Complete sequence of IS*629*.** *Nucleic Acids Res* 1990, **18**:1899.
20. Zhang W, Mellmann A, Sonntag AK, Wieler L, Bielaszewska M, Tschape H, Karch H, Friedrich AW: **Structural and functional differences between disease-associated genes of enterohaemorrhagic *Escherichia coli* O111.** *Int J Med Microbiol* 2007, **297**:17-26.
21. Mahillon J, Chandler M: **Insertion sequences.** *Microbiol Mol Biol Rev* 1998, **62**:725-774.
22. Ohnishi M, Kurokawa K, Hayashi T: **Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors?** *Trends Microbiol* 2001, **9**:481-485.
23. Yokoyama E, Hashimoto R, Etoh Y, Ichihara S, Horikawa K, Uchimura M: **Biased distribution of IS629 among strains in different lineages of enterohemorrhagic *Escherichia coli* serovar O157.** *Infect Genet Evol* 2010.
24. Brunder W, Karch H, Schmidt H: **Complete sequence of the large virulence plasmid pSFO157 of the sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H- strain 3072/96.** *Int J Med Microbiol* 2006, **296**:467-474.
25. Fey PD, Wickert RS, Rupp ME, Safranek TJ, Hinrichs SH: **Prevalence of non-O157:H7 shiga toxin-producing *Escherichia coli* in diarrheal stool samples from Nebraska.** *Emerg Infect Dis* 2000, **6**:530-533.
26. Monday SR, Minnich SA, Feng PC: **A 12-base-pair deletion in the flagellar master control gene flhC causes nonmotility of the pathogenic German sorbitol-fermenting *Escherichia coli* O157:H- strains.** *J Bacteriol* 2004, **186**:2319-2327.
27. Sambrook J, Russell RG: **Molecular Cloning. A Laboratory Manual.** Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2001.
28. Monday SR, Beisaw A, Feng PC: **Identification of Shiga toxigenic *Escherichia coli* seropathotypes A and B by multiplex PCR.** *Mol Cell Probes* 2007, **21**:308-311.
29. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596-1599.
30. Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.** *J Mol Evol* 1980, **16**:111-120.
31. Rzhetsky A, Nei M: **Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference.** *J Mol Evol* 1992, **35**:367-375.
32. Nagano H, Hirochi T, Fujita K, Wakamori Y, Takeshi K, Yano S: **Phenotypic and genotypic characterization of beta-D-glucuronidase-positive Shiga toxin-producing *Escherichia coli* O157:H7 isolates from deer.** *J Med Microbiol* 2004, **53**:1037-1043.
33. Nagano H, Okui T, Fujiwara O, Uchiyama Y, Tamate N, Kumada H, Morimoto Y, Yano S: **Clonal structure of Shiga toxin (Stx)-producing and beta-D-glucuronidase-positive *Escherichia coli* O157:H7 strains isolated from outbreaks and sporadic cases in Hokkaido, Japan.** *J Med Microbiol* 2002, **51**:405-416.
34. Eklund M, Bielaszewska M, Nakari UM, Karch H, Siitonen A: **Molecular and phenotypic profiling of sorbitol-fermenting *Escherichia coli* O157:H- human isolates from Finland.** *Clin Microbiol Infect* 2006, **12**:634-641.

doi:10.1186/1471-2180-11-133
Cite this article as: Rump *et al.*: Prevalence, distribution and evolutionary significance of the IS*629* insertion element in the stepwise emergence of *Escherichia coli* O157:H7. *BMC Microbiology* 2011 **11**:133.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central