Kjersti Haugum

Studies of genetic characteristics in Shiga toxin-producing *Escherichia coli* (STEC) from patients with and without haemolytic uremic syndrome (HUS) in Norway

Thesis for the degree of Philosophiae Doctor

Trondheim, November 2014

Norwegian University of Science and Technology
Faculty of Medicine
Department of Laboratory Medicine,
Children's and Women's Health

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Medicine
Department of Laboratory Medicine,
Children's and Women's Health

© Kjersti Haugum

ISBN 978-82-326-0566-8 (printed ver.)
ISBN 978-82-326-0567-5 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2014:327

Printed by NTNU-trykk

Contents

Acknowledgement

The present work was performed at the Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine at the Norwegian University of Science and Technology (NTNU) during the years 2011-2014. The Ph.D. grant was financed by NTNU, while whole genome sequencing of Norwegian STEC strains was financed by funds from "Kontaktutvalget ved St. Olavs Hospital og Det medisinske fakultet, NTNU". An additional grant was received from "FUGE Midt-Norge" for bioinformatics analyses.

This work has been possible through the collaboration and support from many people. I would therefore like to give special thanks to my main supervisor Professor Jan Egil Afset and co-supervisors Professor Kåre Bergh and Senior Scientist Lin T. Brandal. To Jan Egil; thank you for employing me as a Ph.D. candidate and giving me the opportunity to work on virulence issues in Shiga toxin-producing *E. coli*. You have always been finding time for discussing the project with me, and in your friendly way you have been guiding and encouraged me through this project. Kåre; your experience and support have been invaluable. In addition, the whole genome sequencing of bacterial strains would not have been possible without you providing access to funding for this. And many thanks, Lin, for your encouraging involvement during the entire process.

The data analysis of the sequenced bacterial DNA would not have been possible without collaboration with the Bioinformatics core facility at NTNU. I would therefore like to give special thanks to Jostein Johansen for your excellent work and collaboration, and to Finn Drabløs for your excellent bioinformatics guidance. To David Ussery; thank you for sharing your knowledge and for giving us access to bioinformatics tools making this work possible. I would also like to thank Georg Kapperud and Bjørn-Arne Lindstedt for introducing me to research in the first place, as well as for important contribution in this project – and to Christina Gabrielsen, Inger Løbersli, and Astrid L. Wester; thank you for support and invaluable contribution.

I would like to thank the technical staff at the Department of Medical Microbiology at St. Olavs Hospital for help with information on STEC strains and patients at St. Olavs Hospital, and for providing equipment for DNA isolation. Then I would like to thank the staff at the Department of Foodborne Infections at the Norwegian Institute of Public Health in Oslo for skilful technical assistance on detection and characterization of STEC strains.

While a Ph.D. work is dependent on collaboration and support in a professional manner, it would not have been possible without support and cheering words from colleagues, friends and family. I would therefore like to thank present as well as previous colleagues for your support - with a special thanks to Hilde Lysvand, Kirsti Løseth and Anne Nor. And to all my friends and family; thank you for being there.

Last but not least, many thanks to my parents for your support and help in making this work possible – and to Jon and Sunniva; thank you for giving me the opportunity to live a "normal life" as well during this project!

Trondheim, November 2014.

Abbreviations

µm	Micrometre
A/E	Attaching and effacing
AEEC	Attaching and effacing *E. coli*
aEPEC	Atypical EPEC
AIEC	Adherent invasive *E. coli*
BFP	Bundle-forming pili
bp	Base pair
CDC	Centers for Disease Control and Prevention
CFA	Colonization factor antigen
DAEC	Diffusely adherent *E. coli*
E. coli	*Escherichia coli*
eae	Gene encoding intimin
EAEC	Enteroaggregative *E. coli*
EAF	EPEC adherence factor
ECDC	European Centre for Disease Prevention and Control
ECOR	*E. coli* reference collection
Efa-1	EHEC factor for adherence
EFSA	European Food Safety Authority
EHEC	Enterohaemorrhagic *E. coli*
EHEC-LST	EHEC that has lost the Shiga toxins
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
Esp	EHEC secreted protein
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extraintestinal *E. coli*
FDR	False Discovery Rate
Gb3	Globotriaocylceramide receptor
GO	Gene Ontology
HAS	HUS-associated serotype
HIV	Human immunodeficiency virus
HUS	Haemolytic uremic syndrome
IMS	Immunomagnetic separation
IS	Insertion sequence
kb	Kilobase
LA	Localized adherence
LEE	Locus of enterocyte effacement
Ler	LEE encoded regulator

LifaA	Lymphostatin inhibitory factor
Lpf (*lpf*)	Long polar fimbriae (encoding gene)
LT	Heat-labile enterotoxins
Mb	Megabase
MLST	Multi-locus sequence typing
MLVA	Multiple-locus variable number of tandem repeats analysis
MNEC	Meningitis-associated *E. coli*
MSIS	Meldingssystem for smittsomme sykdommer (Norwegian Surveillance System for Communicable Diseases)
Nle	Non-LEE
NSF	Non-sorbitol-fermenting
OI	O island
ORF	Open reading frame
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
Per	Plasmid encoded regulator
PFGE	Pulsed-field gel electrophoresis
PLS	Partial Least Squares Regression
RTX	Repeat in toxin
SF O157	Sorbitol-fermenting O157:H-
SMAC agar	Sorbitol-MacConkey agar
spp.	Species
ST	Heat-stable enterotoxins
STEC	Shiga toxin-producing *E. coli*
STEC-LST	STEC that has lost the Shiga toxins
Stx (*stx*)	Shiga toxin (encoding gene)
Stx1/Stx2 *(stx1/stx2)*	Shiga toxin 1 and 2 (encoding gene)
T3SS	Type III secretion system
Tccp	Tir cytoskeleton coupling protein
tEPEC	Typical EPEC
tir	Gene encoding Tir
Tir	Translocated intimin receptor
UPEC	Uropathogenic *E. coli*
VNTR	Variable number tandem repeat
VTEC	Vero cytotoxin-producing *E. coli*

List of papers

1. Haugum, Kjersti, Lindstedt, Bjørn-Arne, Løbersli, Inger, Kapperud, Georg, Brandal, Lin Thorstensen. Identification of the anti-terminator $q_{O111:H-}$ gene in Norwegian sorbitol-fermenting *Escherichia coli* O157:NM. FEMS Microbiol Lett. 2012 Apr;329(2):102-10. doi: 10.1111/j.1574-6968.2012.02505.x.

2. Haugum, K., Brandal, L.T., Lindstedt, B.-A., Wester, A.L., Bergh, K., Afset, J.E. PCR based detection of Shiga toxin-producing *Escherichia coli* (STEC) in a routine microbiology laboratory over 16 years: molecular characterization of strains. J Clin Microbiol. 2014 Jun 11. pii: JCM.00453-14. [Epub ahead of print]

3. Haugum, K., Johansen, J., Gabrielsen, C., Brandal, L.T., Bergh, K., Ussery, D.W., Drabløs, F., Afset, J.E. Comparative genomics to delineate pathogenic potential in non-O157 Shiga toxin-producing *Escherichia coli* (STEC) from patients with and without haemolytic uremic syndrome (HUS) in Norway. Manuscript accepted for publication in PLOS One with a tentative publication date October 31 2014.

Summary

Shiga toxin-producing *Escherichia coli* (STEC) is an important human pathogen that can cause symptoms ranging from asymptomatic carriage and mild disease to bloody diarrhoea and haemolytic uremic syndrome (HUS). Non-sorbitol-fermenting O157:H7 (NSF O157) was the first STEC serotype detected and is still the most common source of STEC disease and outbreaks worldwide. However non-O157 STEC serotypes are also known to cause severe disease and outbreaks, and in Norway non-O157 serotypes are more frequent than O157 among human STEC cases. There are several known virulence factors in STEC pathogenesis, among which the Shiga toxins (Stx) and the adherence factor intimin, encoded by the *eae* gene on the Locus of Enterocyte Effacement pathogenicity island (LEE), have long been known to be of key importance.

The main aim of the present study was to identify genetic characteristics which could distinguish between highly pathogenic STEC bacteria with the potential to cause HUS, and low-virulent STEC.

First, we investigated specific genes associated with regulation of Shiga toxin production in Norwegian sorbitol-fermenting O157:H- (SF O157) strains from the strain collection at the Norwegian Institute of Public Health, by sequencing of the genomic regions harbouring these particular genes. Secondly, we described severity of disease in patients with STEC infection and characterized a collection of STEC strains detected by polymerase chain reaction (PCR) in the years 1996-2011 at St. Olavs Hospital, Norway. Finally, we whole genome sequenced and compared 95 Norwegian non-O157 STEC strains from the strain collection at the Norwegian Institute of Public Health and St. Olavs Hospital isolated from persons with HUS, diarrhoea and asymptomatic carriage.

Sequencing of the anti-terminator *q* gene and genes upstream of the *stx2* genes in three selected SF O157 strains revealed that the DNA sequences were identical or similar to those observed in the STEC O111:H- strain AP010960, but different from the ones observed in the NSF O157 strain EDL933 (AE005174). In addition different DNA sequences were detected in the sequenced region among the three SF O157 strains. A PCR assay for detecting $q_{O111:H-}$ in SF O157 was developed, and analysis showed that all *stx2* positive SFO 157 strains harboured this *q* gene. Further investigations are needed to elucidate whether the $q_{O111:H-}$ gene observed in all Norwegian SF O157 contributes to the increased virulence seen in SF O157 compared to NSF O157.

In the period 1996 through 2011, STEC was isolated from 138 out of 12,651 (1.09%) patients tested at St. Olavs Hospital. Eleven of these patients (all ≤5 years old) suffered from HUS, but in neither case was the outcome lethal. Twenty (14.5%) of the strains were NSF O157, 78 (56.5%) belonged to other

common STEC serogroups frequently involved in disease and outbreaks (O26, O103, O111, O121, O145 and SF O157), and 40 (29.0%) belonged to other serogroups or were non-typeable. All HUS cases were infected with common STEC serogroups, except NSF O157. Twenty-four STEC strains were classified as HUS-associated. Low age (≤5 years old) and STEC containing *eae* and *stx2a* was significantly more frequent among HUS-associated cases ($p<0.05$ for each parameter), while STEC containing *stx1* was associated with non-HUS infection ($p<0.05$). Other putative virulence genes, apart from the gene encoding enterohaemolysin (*ehxA*), were significantly more frequent among HUS-associated than non-HUS strains ($p<0.05$ for each gene). However, these genes were also present in some of the non-HUS STEC strains and could therefore not reliably differentiate between HUS-associated and non-HUS STEC.

Whole genome sequencing and comparisons were performed across 95 non-O157 STEC strains. Twenty-three STEC of this collection were classified as HUS-associated, including strains from patients with HUS (n=19) and persons with an epidemiological link to a HUS-case (n=4). Genomic comparison and phylogenetic analysis revealed considerable heterogeneity in gene content across the 95 STEC strains. A clear difference in gene profile was observed between strains with and without the LEE pathogenicity island. All HUS-associated STEC strains were distributed in two distinct clusters within phylogroup B1. However, also non-HUS strains were found in these clusters. By genomic comparisons a number of genes were significantly overrepresented among HUS-associated STEC strains, but none of them were unique to these strains. The results from the present study indicate that STEC strains from different phylogenetic backgrounds have independently acquired virulence genes that determine pathogenic potential, and that the gene content among pathogenic and non-pathogenic STECs in part might be regarded as overlapping. Thus it was not possible to clearly distinguish between highly pathogenic STEC having caused HUS and low-virulent STEC having caused only mild disease or asymptomatic carriage.

In summary this study revealed that STEC causing human disease constitute a genetic and phylogenetic highly diverse nature. A clear difference in gene profile was observed between LEE positive and LEE negative STEC. Common characteristics among STEC classified as HUS-associated in this study were that all were LEE positive and contained the *stx2a* subtype, with the exception of strains that had lost the *stx* genes. In addition, all belonged to common STEC serogroups frequently involved in STEC disease and outbreaks. Genomic comparison revealed that the gene content in part was shared between HUS-associated and non-HUS STEC, and consequently it was not possible to reliably distinguish STEC with the potential to cause HUS and those without this potential.

Sammendrag

Analyse av genetiske egenskaper i shigatoksinproduserende *Escherichia coli* (STEC) fra norske pasienter med og uten hemolytisk-uremisk syndrom i Norge

Bakgrunn: Shigatoksinproduserende *Escherichia coli* (STEC) er bakterier som kan forårsake alvorlig sykdom som blodig diaré og hemolytisk-uremisk syndrom (HUS). Viktige sykdomsfremkallende egenskaper hos STEC er produksjon av shigatoksiner samt evne til å feste seg til tarmslimhinnen ved hjelp av adheranseproteinet intimin. I tillegg kommer også andre potensielle sykdomsfremkallende faktorer.

Mål: Målet med denne studien var å identifisere genetiske egenskaper som kan benyttes til å skille STEC-bakterier som gir alvorlig sykdom som HUS fra bakterier som gir mindre alvorlig sykdom.

Materialer og metoder: Vi undersøkte spesifikke gener assosiert med regulering av shigatoksinproduksjon i sorbitolfermenterende STEC O157:H- (SF O157) ved hjelp av sekvensering. Deretter beskrev vi forekomsten av human STEC diagnostisert ved PCR hos pasienter med diarésykdom fra 1996-2011 ved St. Olavs hospital. Til slutt ble 95 STEC-stammer assosiert med HUS, diaré og asymptomatisk bærerskap helgenomsekvensert og sammenlignet.

Resultater: Sekvensering av gener involvert i regulering av shigatoksinproduksjon i SF O157 viste at disse var forskjellig fra tilsvarende gener i den beslektede ikke-sorbitolfermenterende STEC O157:H7, men lik en fjernere beslektet STEC O111:H-. Praktisk betydning av dette funnet må undersøkes i videre studier. Fra 1996-2011 ble det funnet STEC i prøver fra 138 av 12 650 pasienter med diarésykdom ved St. Olavs hospital. Elleve pasienter, alle ≤5 år, ble diagnostisert med HUS. Alle HUS-pasientene bar STEC med Stx2a og intimin. Helgenomsekvensering av 95 STEC viste et stort utvalg ulike gener. Ved analyse av slektskap var det generelt sett store forskjeller mellom de ulike STEC-stammene, selv om stammer assosiert med HUS plasserte seg i to spesifikke grupper. Sammenligning av gener i HUS- og ikke-HUS STEC i de to siste studiene viste at enkelte gener hadde sterk assosiasjon med HUS. Imidlertid var disse genene også tilstede i enkelte stammer ikke assosiert med HUS, og kunne derfor ikke brukes til sikkert å skille mellom HUS og ikke-HUS STEC.

Konklusjon: Våre analyser viste at alle de undersøkte STEC-stammene i denne studien hadde store ulikheter med tanke på slektskap og geninnhold. Det var forskjeller mellom HUS- og ikke-HUS STEC, men siden geninnholdet i bakteriene til dels var overlappende var det ikke mulig sikkert å skille dem fra hverandre.

Navn kandidat: Kjersti Haugum
Institutt: Institutt for laboratoriemedisin, barne- og kvinnesykdommer
Veileder(e): Jan Egil Afset (hovedveileder), Kåre Bergh og Lin T. Brandal (biveiledere)

Finansieringskilde: NTNU

Ovennevnte avhandling er funnet verdig til å forsvares offentlig
for graden ph.d. i Molekylærmedisin.
Disputas finner sted i Auditoriet LA21, Laboratoriesenteret,
fredag 21. november 2014, kl. 12.15.

Errata

1. In List of papers, regarding Paper 3, the status has changes from “Manuscript submitted” to “Manuscript accepted for publication in PLOS One with a tentative publication date October 31 2014.”
2. The word “recognised” is changed to “recognized” in page 14 in the Introduction part.
3. Manuscript of Paper 2 is exchanged with the published version. Minor revisions regarding phrasing were done during the proofing process.

1 Introduction

1.1 *Escherichia coli*

Escherichia coli was first described as *Bacterium coli commune* in 1885 by Theodor Escherich during studies of the intestinal flora of infants. This name was used until 1919, when Castellani and Chalmers defined the genus *Escherichia* and the type species *E. coli* (1). The genus *Escherichia* consists of six species; *E. coli, E. hermannii, E. fergusonii, E. vulneris*, *E. blattae*, and *E. albertii*, which phenotypically can be differentiated by biochemical reactions (1, 2).

E. coli is a Gram-negative rod, 2.0-6.0 micrometre (µm) in length and 1.1-1.5 µm in width (Figure 1). *E. coli* is usually motile through peritrichous flagella that cover the cell surface but can also be non-motile. While *E. coli* is facultative anaerobic, and has both a respiratory and a fermentative metabolism, anaerobic biotypes do also occur (1).

The serotypes of *E. coli* are determined according to their O, K and H antigens. The H antigen is named from the German word “Hauch”, meaning film or veil, which refers to the veil-like spreading growth of the bacteria on solid medium. The “O” in the O antigen originates from “Ohne hauch”, meaning without film or veil and refers to bacterial growth without spread. The O antigens are heat-stable lipopolysaccharides which currently include 181 different types, while there are 53 antigens of the heat-labile flagellar H antigens, providing the basis for numerous combinations of the O and H antigens in different *E. coli* strains. The K antigen, which is named from the German word “Kapsule”, is defined by a group of capsular polysaccharides which covers the surface of *E. coli* cells (1, 3-5).

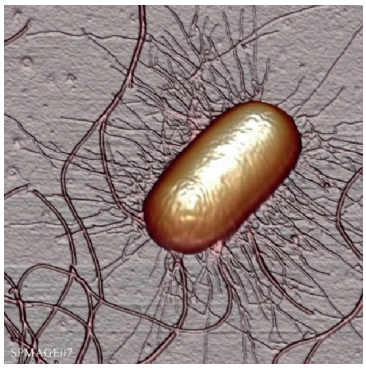


Figure 1. Scanning Probe Microscopy Image of *E. coli*. The size of the cell is about 1.9 µm long and 1.0 µm wide. The width of the pili is about 20 nm and of flagella about 30 nm (Photo: Mr. Ang Li, National University of Singapore, Singapore. SPMage Prize, http://www.icmm.csic.es/spmage/. Reprint with permission from publisher).

E. coli has its niche in the lower part of the intestines of mammals and birds where it is part of the normal flora (1). Attachment to the gastrointestinal surface occurs through type 1 somatic pili which adhere to the intestinal mucosa (6). *E. coli* is excreted with faeces and thus occurs in soil, sediments and water. However, it appears that there is no independent existence in the environment, and *E. coli* is therefore used as an indicator of faecal contamination (6).

1.2 Pathogenic *E. coli* – history and overview

The gastrointestinal tract of humans is colonized by *E. coli* shortly after birth, and the human host and the bacteria normally coexist with no detrimental effects in healthy individuals. However, in immunocompromised patients, and in individuals where the gastrointestinal barrier is broken, even commensal *E. coli* may cause disease (1, 7).

Some variants of *E. coli* have acquired specific virulence factors making them capable of causing a wide range of diseases in humans. Pathogenic strains of *E. coli* containing virulence factors associated with extra-intestinal infections are usually called ExPEC, including those causing meningitis which are termed meningitis-associated *E. coli* (MNEC) and those causing urinary tract disease, termed uropathogenic *E. coli* (UPEC) (Figure 2) (7). *E. coli* isolated from diarrhoeagenic disease is termed diarrhoeagenic *E. coli*. Based on pathogenic, phenotypic, clinical and epidemiological factors, there are currently seven recognized pathotypes of diarrhoeagenic *E. coli* (Figure 2); enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC), enteropathogenic *E. coli* (EPEC), Shiga toxin-producing *E. coli* (STEC), which may also be termed Vero cytotoxin-producing *E. coli* (VTEC) (7, 8), and adherent invasive *E. coli* (AIEC) which has recently been recognized as a new diarrhoeagenic *E. coli* pathotype (9).

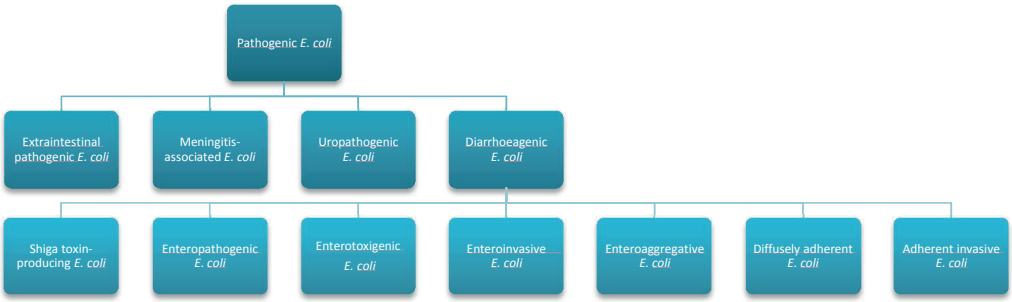


Figure 2. Presentation of pathogenic *E. coli,* displaying the seven pathotypes of diarrhoaegenic *E. coli*.

ETEC was first described as a cause of diarrhoea in piglets, but has later also been shown to cause diarrhoea in humans. ETEC is distributed worldwide, and contaminated food and water are the most common sources of ETEC infection. ETEC is a main cause of infant diarrhoea in less-developed

countries, as well as of neonatal diarrhoea in domesticated animals. Furthermore, ETEC is associated with diarrhoea in travellers to less-developed countries (10). Colonization of the intestinal mucosa and enterotoxin production by ETEC cause watery diarrhoea, which can range from mild to severe disease. Colonization of the intestinal mucosa is facilitated by surface fimbria, designated colonization factor antigen (CFA), and over the years there are identified numerous CFA's. Intestinal secretion is caused by either one or more of the heat-labile (LT) or heat-stable (ST) enterotoxins which are produced by ETEC (10).

EIEC is genetically, biochemically and pathogenically very similar to *Shigella* spp. Like *Shigella*, EIEC invades the colonic mucosa, where EIEC infection in most cases causes watery diarrhoea, similar to disease by ETEC. In the most severe cases, EIEC infection progresses to dysentery, characterized by bloody diarrhoea, lower abdominal cramps and fever. EIEC is mostly identified in food- and waterborne outbreaks, however transmission from person to person may also occur (10).

EAEC is characterized by a distinct aggregative adherence, described as "stacked brick-like", to the mucosal surface of the host intestines. The bacterium abundantly colonizes the mucosal surface, and induces significant mucosal damage (11). Colonization is followed by secretion of bacterial cytotoxins and enterotoxins. Furthermore, there are identified a large number of putative virulence factors that varies among different EAEC. EAEC is a cause of infant diarrhoea in less-developed as well as developed countries, traveller's diarrhoea and persistent diarrhoea in HIV patients. In addition, EAEC is a common cause of outbreaks of diarrhoea associated with contaminated food and/or water (11).

DAEC is recognized by a diffuse adherence pattern to HEp-2 and HeLa cells. This specific adherence pattern is mediated by both fimbrial and afimbrial adhesins, designated Afa-Dr adhesins. The bacterium has been associated with diarrhoea in children aged 1-5 years and in recurrent urinary tract infections in adults, however DAEC has also been identified in healthy persons. Identification and classification of DAEC has proved to be difficult, and methods for detecting diarrhoeagenic DAEC are still being developed. Further epidemiologic studies are therefore needed for the designation of DAEC as a diarrhoeagenic pathotype (9, 12).

Bacteria of the AIEC pathotype adhere to and invade epithelial cells, where the bacteria replicate. In addition, AIEC replicates in macrophages. These characteristics are utilized in AIEC detection. The AIEC pathotype has among other agents been implicated in Crohn's Disease, where it has been detected in more than 30% of affected patients in clinical studies (9).

EPEC was the first *E. coli* associated with diarrhoeal disease. Already in the 1940's, the organism was recognized as a cause of infant diarrhoea during large outbreaks in the United Kingdom (9). In

industrialized countries severe disease caused by EPEC has decreased significantly, however in less-developed countries the organism still is an important cause of infant diarrhoea (10, 12). A characteristic of EPEC is the ability to form attaching and effacing (A/E) lesions on intestinal epithelial cells, which has given this pathotype the designation attaching and effacing *E. coli* (AEEC). The bacteria attach intimately to the intestinal mucosa, efface the microvilli and induce cytoskeletal changes where actin from the host cells form pedestals at the attachment site. The attaching and effacing phenotype is encoded by a 35 kilobase (kb) pathogenicity island called the Locus of Enterocyte Effacement (LEE) (10, 12). The initial adherence of EPEC to intestinal cells is thought to be facilitated by bundle-forming pili (BFP). Bundle-forming pili are encoded by the *bfp* gene, located on the EPEC adherence factor (EAF) plasmid, and are responsible for EPEC adhering to intestinal cells in a distinct localized adherence (LA) pattern. In addition, the EAF plasmid harbours the plasmid encoded regulator (*per*) locus, encoding proteins that regulate the *bfp* operon and most of the LEE genes by the LEE encoded regulator, Ler. EPEC that harbours LEE and lack the EAF plasmid are traditionally called atypical EPEC (aEPEC), while typical EPEC (tEPEC) harbour both the LEE pathogenicity island and the EAF plasmid. aEPEC are more frequently isolated from diarrhoeal disease in industrialized countries than tEPEC, whereas in less-developed countries, outbreaks caused by tEPEC are more common (10).

1.3 Shiga toxin-producing *E. coli* (STEC)

1.3.1 Introduction

Shiga toxin-producing *E. coli* was first recognized as a cause of infection in two independent observations in 1982, reported by Karmali *et al.* (13) and Riley *et al.* (14). In the first report an association of haemolytic uremic syndrome (HUS) with a faecal cytotoxin and cytotoxin-producing *E. coli* in faecal samples from patients was suggested (13). In the second report non-sorbitol-fermenting *E. coli* O157:H7 (NSF O157) was identified as a cause of bloody diarrhoea (also called haemorrhagic colitis) due to contamination of undercooked hamburgers in a fast-food restaurant chain (14). In addition, Johnson *et al.* reported an outbreak of bloody diarrhoea in a Canadian institution for elderly patients in November 1982, where *E. coli* O157:H7 was isolated from the patients while no other recognized enteric pathogens were reported (15). Although STEC was not recognized as a pathogen before these observations in 1982, such bacteria had probably existed and caused serious disease for some time. It was already known that *Shigella dysenteria* 1 producing Shiga toxins (Stx) were associated with HUS. Then in the late 1960's it had been suggested that also particular strains of *E. coli* infected with a bacteriophage caused HUS. These suggestions were based on the fact that *E. coli* was identified in many HUS outbreaks, while pathogens known to cause HUS (e.g. *Shigella dysenteria* 1) were not detected (10). Until now, more than 400 *E. coli* serotypes producing Shiga toxins have

been isolated from humans with disease, and STEC is recognized as an important group of human pathogens (4). In addition, some EAEC strains contain the genes encoding Shiga toxins, where one particular strain of serotype O104:H4 was responsible for a large HUS outbreak in Germany in 2011 (9, 16).

The term STEC is used to describe any *E. coli* producing Shiga toxins or *E. coli* carrying the *stx* genes, whereas the term enterohaemorrhagic *E. coli* (EHEC) has different definitions; STEC responsible for bloody diarrhoea and HUS (4, 17), STEC carrying the LEE pathogenicity island (18), and STEC associated with human disease (4, 17). In the present study we will use the term STEC to describe both EHEC and all other STEC.

STEC of various serotypes contains a vast repertoire of genes encoding putative virulence factors with different functional characteristics. Table 1 lists both recognized and potential virulence factors found in STEC, the most important of which will be discussed in the following sections. It is also clear that infectious dose and host factors like the immune system, variations in expression of the Shiga toxin receptor and the intestinal environment might also affect STEC virulence, and therefore the clinical outcome of STEC disease (19). Young age of the patients is also known to be a risk factor for severe STEC disease (20, 21).

1.3.2 Pathogenesis (disease mechanisms)

Symptoms of STEC infection in humans range from asymptomatic carriage, mild diarrhoea, and bloody diarrhoea to HUS. After ingestion of STEC, there is a time interval of two to twelve days before onset of diarrhoea. In persons infected with NSF O157, bloody diarrhoea ensues after one to three days of diarrhoea in approximately 90% of the cases. Progression to HUS is more frequent in children and elderly, where HUS progression has been observed in about 15% of children diagnosed with NSF O157 (22). HUS is defined by acute renal failure, thrombocytopenia and microangiopathic haemolytic anaemia, and the severity of HUS varies from less severe to fulminant disease (22, 23). Although NSF O157 is implicated in most outbreaks causing STEC disease worldwide, other STEC serotypes have been detected both in sporadic cases of disease as well as in outbreaks. There is a risk of developing HUS from infection by sorbitol-fermenting O157:H- (SF O157) and various non-O157 STEC serotypes like O26:H11, O45:H2, O103:H2, O111:H8, O113:H21, O121:H19, and O145:H28 and their non-motile variants (17). In later years, progression to HUS has been reported in as many as 22-58% of cases in outbreaks with particular STEC serotypes, e.g. O103:H25 and O104:H4 (24, 25).

1.3.2.1 *The Shiga toxins*

The key virulence factors in STEC are the Shiga toxins (Table 1) (Figure 3) (4, 10). There are two types of Shiga toxins, i.e. Shiga toxin 1 and Shiga toxin 2 (Stx1 and Stx2), encoded by the *stx1* and *stx2* genes, respectively. Stx1 is similar to the Shiga toxin of *Shigella dysenteriae*, while the amino acid homology between Stx1 and Stx2 is about 55% (4).

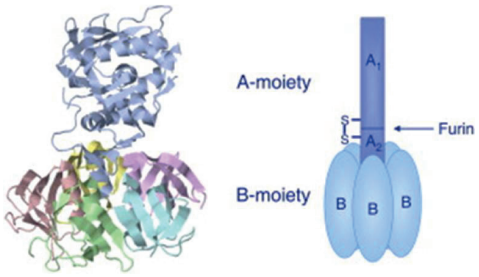


Figure 3. Architecture of the Shiga toxin (36). Image printed with permission from publisher.

The *stx* genes are encoded in lambdoid bacteriophages or in phage remnants (cryptic prophages) integrated into the bacterial host genome (26, 27). While phage remnants consist of phage sequences that are not able to synthesise intact functional phage particles, intact Shiga toxin bacteriophages can exist in either the lytic or lysogenic forms. In the lytic form, the phages infect its bacterial host and produce new phage particles which are released by lysis of the host cell, while in the lysogenic form the phages integrate into the host genome as prophages where they can remain in a quiescent state until induced. So far, there has not been described any Shiga toxin-specific secretion system, and it is therefore assumed that Shiga toxins are released into the gut lumen after bacterial lysis (19, 26). The Shiga toxin phages are morphologically highly diverse. Genomic comparison of available Shiga toxin phages has shown that they are of a highly mosaic nature, however they do display similar genomic structures and organization (26, 28, 29). In almost all Shiga toxin phages, the *stx* genes are located in the region of phage late genes, downstream of the phage late promoter $p_{R'}$ and the late terminator $t_{R'}$, which are expressed during the lytic life cycle. Expression of the *stx* genes is under control of the anti-terminator Q protein, which acts as a transcriptional activator at $p_{R'}$ by read-through of $t_{R'}$ (Figure 4) (28). Although the *stx* genes have a separate promoter, expression of both genes is dependent on the Q protein (30). Different variants of the *q* gene are divided into five main groups (groups I-V) based on ≥ 95% amino acid identity (31). Most studied are the two different *q* genes identified in NSF O157, q_{933} and q_{21}, of which the q_{933} gene has been associated with higher production of Shiga toxins (32-34).

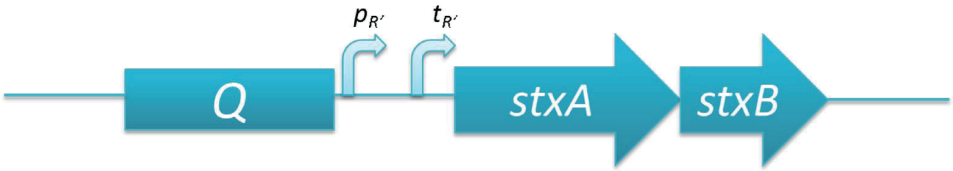


Figure 4. Schematic presentation of the *stx* genes and the upstream *q* gene. The anti-terminator Q protein acts as a transcriptional activator at $p_{R'}$ by read-through of $t_{R'}$. Adapted from (28).

Both Stx1 and Stx2 are A_1B_5 toxins in which the B subunit is a pentameric ring structure (Figure 3). The B subunits of Shiga toxins bind to specific glycosphingolipids, i.e. the globotriaosylceramide receptors Gb3 and Gb4 on the surface of host cells, preferentially microvascular endothelial cells of the renal glomeruli and the brain. All Stx1 subtypes bind both Gb3 and Gb4, while most Stx2 subtypes primarily bind Gb3. Stx2e is an exception, primarily binding Gb4 (35). It is not clear how Shiga toxins cross the intestinal epithelial barrier and gain access to the systemic circulation, although certain immune cells have been suggested as transporting vehicles. After binding to the host cell receptor, the toxin enters the host cell by retrograde endocytosis (35, 36). In the cell, the A subunit is cleaved by the protease furin into an enzymatically active A1-fragment and a carboxyl terminal A2-fragment, which remain linked by a disulphide bond. The A subunit is an N-glycosidase which inhibits protein synthesis in a similar manner as ricin, by disrupting the large eukaryotic ribosomal subunit (4, 22).

The amount and density of Gb3 on host endothelial cells is important for Shiga toxin cytotoxicity. In addition, the association of Gb3 with glycolipid-cholesterol enriched lipid rafts in the cell membrane is required. According to current knowledge, only Shiga toxins binding to Gb3 organized in lipid rafts will be transported backwards from the plasma membrane to the intracellular targets where it inhibits protein synthesis, while Shiga toxins binding Gb3 organized outside lipid rafts will be transported to lysosomes for degradation (35, 37). While Gb3 is expressed both in renal glomeruli and tubular cells, different Gb3 organization into lipid rafts is observed between glomeruli and tubular membranes, which might explain that Shiga toxins binding to Gb3 in glomeruli is important for HUS development. Current models suggest that different membrane organization of Gb3 also might explain the risk for age dependent development of HUS, as more Shiga toxins is binding to Gb3 in paediatric glomeruli than in adult glomeruli (37).

There are three subtypes of Stx1, designated Stx1a, Stx1c and Stx1d, while there are seven Stx2 subtypes, designated Stx2a through Stx2g (38). Stx2 is more highly associated with severe disease

than Stx1, although *eae* and *stx1* positive, *stx2* negative STEC has been isolated from patients with HUS (39, 40), However, there are to our knowledge no reports of outbreaks of severe disease with *eae* and *stx1* positive bacteria. Among Stx2 the subtypes Stx2a, Stx2c and Stx2d are more often associated with serious disease than other serotypes (41-43). In addition, purified Stx2a, Stx2d, and elastase-cleaved Stx2d are shown to be more potent than other Stx2 subtypes both *in vitro* and *in vivo* (44). Although the Shiga toxins are key virulence factors in STEC disease, the toxin genes are occasionally missing in STEC found in HUS patients and outbreaks. In such STEC, which are called EHEC-LST, here termed STEC-LST, it is believed that the bacteriophages harbouring the toxins genes are lost during the course of infection (45, 46).

1.3.2.2 *The LEE pathogenicity island*

STEC, like EPEC, has the ability to form attaching and effacing lesions on intestinal epithelial cells, with intimate attachment of the bacteria to the intestinal mucosa, effacement of the microvilli and formation of pedestals at the attachment site (Figure 5). The attaching and effacing phenotype is also in STEC encoded by the genes located in the LEE pathogenicity island (Table 1) (47, 48). LEE harbours genes encoding a type III secretion system (T3SS) and additional associated regulators, translocators, effector proteins and chaperones. The type III secretion system is a protein complex acting like a "molecular syringe" by injecting bacterial effector proteins into the host cytoplasm (12, 49). The approximately 41 genes encoded by LEE are distributed into five genetic units designated LEE1 to LEE5 and the pathogenicity island is integrated into different tRNA genes in different *E. coli*. LEE1 to LEE4 encode proteins involved in regulation and synthesis of the type III secretion system while LEE5 encodes the Translocated intimin receptor (Tir) and intimin. Whereas the genes encoding the structural components mostly are conserved among various LEE, other genes e.g. the genes encoding secreted effector proteins and intimin are more diverse (47-49). While the core LEE of about 35 kb is in general highly conserved, comparative analysis of LEE and its flanking regions in selected EPEC and STEC revealed a high degree of variation in the flanking regions due to integration of various genetic elements like IS elements, prophages, effector genes and virulence genes (49, 50).

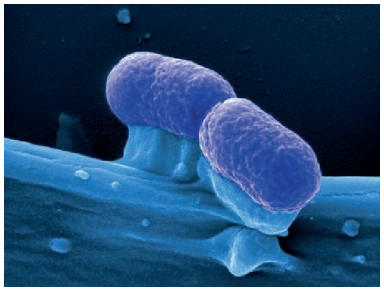


Figure 5. Electron microscope photo of STEC mediated pedestals. (Photo: Manfred Rohde, Helmholtz-Zentrum für Infektionsforschung). Printed with permission from the publisher.

A key colonization factor in STEC and EPEC is intimin, encoded by the LEE gene *eae*. Intimate attachment to intestinal epithelial cells is caused by intimin binding to Tir on host cells. Tir is first

translocated into the host cell cytoplasm by the type III secretion system, where it is integrated into the host cell wall and functions as a receptor for intimin (4, 12). The interaction of Tir and intimin subsequently causes Tir clustering, actin assembly and pedestal formation in the host cell. Actin assembly is induced by different signalling pathways in STEC and EPEC. In NSF O157, actin assembly is dependent on the non-LEE encoded effector Tir-cytoskeleton coupling protein (Tccp/$EspF_U$) signalling pathway, while SF O157 can use either the Tccp pathway and/or an alternative Tccp2 pathway. In some EPEC phylogenetically defined as EPEC 1 lineage, actin assembly is mediated by the phosphorylation of tyrosine residue 454, with subsequent binding to the host adaptor protein Nck. Other EPEC (EPEC phylogenetic lineage 2) and most non-O157 STEC can use both the Nck and TccP2 pathways, reviewed in (51, 52).

There are at least 29 known intimin variants, due to variation in the C-terminal amino acid part of the protein, and the different intimin variants are distributed among both EPEC and STEC (53, 54). There are also different variants of Tir, of which the alleles Tir-α, Tir-β, Tir-γ and Tir-γ2/θ have been identified in STEC. Various combinations of Tir and intimin are found in STEC, e.g. in serotype O157:H7 the combination *eae* γ1-*tir* γ1 is commonly observed whereas the combination *eae* β1-*tir* β1 is observed in O26:H11 (53, 55-58).

While most STEC causing human disease harbours the LEE pathogenicity island, there are a number of LEE negative STEC associated with human disease (59). One particular example is the STEC O104:H4 strain responsible for an outbreak causing HUS in 855 patients and 53 deaths in Germany in 2011 (24).

1.3.2.3 *Nle effectors*

In addition to the effectors encoded by the LEE pathogenicity island, there are several non-LEE encoded effectors (Nle) that are also translocated by the type III secretion system. Homology searches in the NSF O157 Sakai strain revealed more than 60 candidate effector genes, which at the protein level clustered into more than 20 families. Forty-nine of the 60 candidates were judged as potential effectors, of which 39 were confirmed to be translocated through proteomics and translocation assays (60). Nle effector sequences were found on specific pathogenicity islands called O islands (OI), in prophages and on the large EHEC plasmid (60, 61). The contribution of many of the Nle effectors to disease is not known. However, for some function has been experimentally verified: NleB, NleC, NleE and NleH have immune modulating activity, NleD and NleH inhibit apoptosis, Cif contributes to cell cycle arrest and NleG acts as an U-box ubiquitin ligase (Table 1), reviewed by (62).

In order to distinguish STEC causing severe disease from non-pathogenic STEC, the presence of *nle* genes in the genomic O island 122 (OI-122) encoding type III secreted virulence factors has been investigated. Karmali *et al.* (63) observed that a complete set of particular *nle* genes in OI-122 were found in the STEC serotypes O157:H7 and O157:H-. These are STEC serotypes that have a high frequency of association with disease, have common involvement in outbreaks and are associated with bloody diarrhoea and HUS (classified as seropathotype A) (63). However, in other seropathotypes (seropathotype B through E) which have been less strongly associated with disease and outbreaks, parts of the OI-122 genes were found less commonly: from 60% of seropathotype B isolates to 0% of seropathotype E isolates (63). Additional studies have supported that STEC harbouring an increased number of OI-122 *nle* genes are more associated with diarrhoea and HUS than strains with fewer of these genes (63-65). In addition to genes on the OI-122, genes on OI-71 have been shown to be associated with STEC having increased virulence (66, 67) (Table 1). Furthermore, the *nle* effector genes *ureD*, *espV*, *espK*, *espN*, *Z2098* and *espM1* have been more frequently identified in a subgroup of STEC more highly associated with disease and outbreaks (referred to as EHEC) than other STEC and EPEC, and could potentially be used as genetic markers for targeted detection of strains having the potential of causing severe disease (17).

1.3.2.4 *Adhesins in STEC*

Besides intimin, there are a number of other adhesins important for STEC colonization (Table 1). The long polar fimbriae (Lpf) in NSF O157 and homologues of the Lpf in non-O157 STEC increase STEC adhesion in vitro, indicating their role in adherence (68), reviewed in (59). The haemorrhagic coli pilus (HCP) mediates adherence to human and bovine epithelial cells and invasion of epithelial cells as well as biofilm formation, and it has furthermore been shown to play a role in the induction of proinflammatory cytokine secretion in intestinal epithelial cells (69, 70). EHEC factor for adherence (Efa-1), also known as lymphostatin (LifA) is a multi-functional protein found to be involved in STEC adhesion to human intestinal cells and has also been shown to mediate intestinal colonization in calves. Efa-1 has also been shown to be immune suppressive and to be involved in disruption of the intestinal epithelial barrier, in addition to be involved in regulation of protein expression of LEE encoded genes (71-73). Furthermore the IrgA homologue adhesin (Iha), the autotransporter Eha, ToxB and EspP mediate adhesion in STEC (74-77), while the STEC autoagglutinating adhesin (Saa) mediates STEC adhesion in LEE negative STEC (78, 79). For the Sfp fimbriae, encoded on the pSFO157 plasmid in SF O157, it has been shown that expression of Sfp fimbriae is correlated with increased adherence to human intestinal cells in vitro (80, 81).

1.3.2.5 *The pO157 plasmid*

The pO157 plasmid is present in almost all NSF O157 (82), for review see (83). There are approximately 100 open reading frames (ORFs) on the 92 kb plasmid, which encodes several potential virulence factors including enterohaemolysin (EhxA) (84), a zinc metalloprotease (StcE) (85), a catalase-peroxidase (KatP) (86), a type II secretion system apparatus (Etp) (87), a serine protease (EspP) (88) and a putative adhesin (ToxB) (75) (Table 1). The *ehxA* gene is commonly found among non-O157 STEC as well as O157 (4, 10). So far three enterohaemolysins have been detected in *E. coli*. Enterohaemolysin and alpha-enterhaemolysin (HlyA) are part of the pore-forming RTX (repeat in toxin) toxin family casing haemolysis of sheep erythrocytes, while the third enterohaemolysin, haemolysin E (HlyE) is not part of this toxin family (89). The precise role of enterohaemolysin in STEC disease is not clear, however patients with HUS develop antibodies against it, and EhxA from a STEC O128:H12 isolate has been observed to induce production of a proinflammatory cytokine (10, 90). The function of StcE is mucinase activity towards specific mucin types, which degrades and reduces the viscosity of the mucus layer. In addition it is believed to contribute to immune evasion (91, 92), while the catalase-peroxidase KatP scavenges hydrogen peroxide and therefore contributes to resistance towards hydrogen peroxide in O157:H7 (93). The 13 ORFs encoding the *etp* genes show similarities to genes encoding the type II secretion system (87). These genes are commonly found among human STEC serotype O157 strains, while they seem to be rare among strains from bovine sources (87). The serine protease EspP is a multifunctional protein involved in cleavage of pepsin A and human coagulation factor V, has cytopathic effect on Vero cells, and is involved in host colonization (74, 88, 94). ToxB, which is encoded by a homolog of the *efa-1/lifA* gene has been shown to be involved in EHEC adherence to epithelial cells, but does not seem to have lymphostatin activity (75, 95).

Table 1. Known or potential virulence factors found in STEC

Gene	Virulence factor	Function	Localization	Reference
astA	Enteroaggregative heat stable enterotoxin (EAST1)	Toxin	Chromosome, plasmid	(96)
cdtB	Cytolethal distending toxin	Damage to microvascular endothelial cells	Chromosome, plasmid, bacteriophage	(97, 98)
cif	Cycle inhibiting factor (Cif)	Toxin: Cell cycle arrest, modulates cellular processes	Prophage	(99-102)
eae	Intimin	Adhesin	LEE pathogenicity island	(103). For review, see (62)
efa-1/lifA	EHEC factor for adherence (Efa)	Adhesin, lymphostatin	OI-122	(71, 95, 104, 105).
ehaA, ehaB	Autotransporter EhaA, EhaB	Adhesin, autoaggregation, biofilm formation	Chromosome, OI-15	(77, 106)

Gene	Virulence factor	Function	Localization	Reference
ehxA	Enterohaemolysin	Production of proinflammatory cytokines	pO157	(84). For review see (4)
ent/espL2	EspL2	Induces actin microfilament aggregation	OI-122	(107-109)
epeA	EpeA	Serine protease	pO113	(94)
espB	*E. coli* secreted protein B (EspB)	Multifunctional effector protein	LEE pathogenicity island	(110). For review see (62, 111)
espF	EspF	Multifunctional effector protein	LEE pathogenicity island	(112-114). For review see (62, 111).
espG	EspG	Multifunctional effector protein	LEE pathogenicity island	(115). For review see (62, 111, 116)
espH	EspH	Modulating actin dynamics, cytoskeleton disruption	LEE pathogenicity island	(117, 118). For review see (62, 110, 111)
espJ	EspJ	Inhibits phagocytosis	CP-933U	(119-121)
espK	EspK	Unknown	CP-933N	(122)
espP	EspP	Serine protease: cleaves pepsin A and human coagulation factor V, cytopathic effect on Vero cells, host colonization	pO113	(74, 88, 94)
espZ	EspZ	Inhibits host cell apoptosis, regulate T3SS translocation	LEE pathogenicity island	(123-125). For review see (62, 111)
etpD	EtpD	Type II secretion system	Plasmid	(87)
hcpA	Haemorrhagic coli pilus (Hcp)	Adhesin	Chromosome	(69, 70)
iha	IrgA homologue adhesin	Adhesin	Tellurite resistance- and adherence-conferring island	(76)
katP	KatP	Catalase/peroxidase: hydrogen peroxide resistance	pO157	(86, 93, 126)
lpfA	Long polar fimbriae	Fimbrial adhesin	Chromosome	(68)
map	Mitochondrial-associated protein, Map	Multifunctional effector protein	LEE pathogenicity island	(110). For review see (62, 111)
nleA	NleA	TTSS serine protease: Inhibits cellular protein secretion, tight junction disruption	OI-71	(127-129)
nleB	NleB	Immune system modulation	OI-122	(63, 130-133)
nleC	NleC	TTSS metalloprotease: Immune system modulation	OI-36	(130, 134, 135)
nleD	NleD	TTSS metalloprotease: Immune system modulation, block apotosis	OI-36	(130, 135)
nleE	NleE	Immune system modulation	OI-122	(130-132, 134)
nleF	NleF	Caspase inhibitor	OI-71	(130, 136)
nleG	NleG	Ubiquitin ligase	Non-LEE	(130, 137)
nleH	NleH	Immune system modulation, block apoptosis	OI-71	(60, 138, 139)

Gene	Virulence factor	Function	Localization	Reference
paa	Porcine attaching and effacing associated factor (Paa)	Adhesin	OI-57	(140)
pagC	PagC	Bacterial survival within macrophages	OI-122	(63, 141)
saa	STEC autoagglutinating adhesion	Adhesin	Plasmid	(78)
sen	*Shigella flexnerii* enterotoxin (ShET)	Toxin	OI-122	(63, 142, 143)
sfp	Sfp	Adhesin	pSFO157	(80, 81, 144)
stcE	StcE	Metalloprotease: remodelling the mucosal lining during infection	pO157	(85)
stx1	Shiga toxin 1	Toxin	Bacteriophage	For review see (38, 145)
stx2	Shiga toxin 2	Toxin	Bacteriophage	For review see (38, 145)
subA	Subtilase cytotoxin	Toxin	pO113	(146)
tccP/espF$_u$	Tir cytoskeleton coupling protein (Tccp)	Actin pedestal formation	Prophage	(147-149). For review see (51)
tccP2	Tir cytoskeleton coupling protein 2	Actin pedestal formation	Prophage Sp4/CP-933 M	(147, 150)
tir	Translocated intimin receptor	Intimin receptor, actin pedestal formation, inhibits NF-κB function.	LEE pathogenicity island	(151). For review, see (62, 152)
toxB	ToxB	Adhesin	pO157	(75)

1.3.3 Epidemiology

STEC O157 is historically the most commonly reported serogroup associated with STEC disease worldwide (4, 10). In the years 2008-2011, STEC O157 was confirmed as the most common serogroup in Europe, with approximately half of the cases reported to the European Centre for Disease Control and Prevention. In addition, O157 was the serogroup most frequently associated with HUS, especially in children aged 0-4 years old (21, 153). In the years 2008-2011 the second most common STEC serogroup was O26, followed by O103 and O91. However, in 2011 as a single year, STEC O104:H4 was the second most common serotype with 20% of the cases due to the large HUS outbreak in Germany (21). In Europe there are variations among the different countries regarding which STEC serogroup that is most frequently detected. Whereas almost all reported STEC cases were of serogroup O157 in the United Kingdom and Ireland, many cases in e.g. Austria and Italy were caused by the O26 serogroup as well (154). In Norway, STEC of serogroup O103 was most frequent in 2012, followed by the serogroups O157, O145 and O26 (155).

The number of reported human cases of STEC infection in the European Union have increased since 2006, and in 2011 the overall notification rate was 2.54 cases per 100,000 population compared to 0.8 cases per 100,000 in 2006 and 1.00 in 2010. The higher number in 2011 was due to the German HUS outbreak (21, 154). Also in Norway the overall number of reported STEC disease has increased since 2006, although the number has varied in the last few years. The confirmed rate per 100,000 was 1.48 in 2012 compared to 2.25 in 2009, 1.07 in 2010, and 0.96 in 2011 (155).

Also in the U.S. was O157 the most frequently detected serogroup associated with human infection, based on reports from ten FoodNet sites during the years 2000–2010 (156). However, the rate of O157 STEC infection decreased from 2.17 per 100,000 in 2000 to 0.95 per 100,000 in 2010 in the U.S., while the rate of non-O157 STEC infections increased from 0.12 per 100,000 in 2000 to 0.95 per 100,000 in 2010 (156). Both in Europe and in the U.S., the high numbers of STEC O157 cases diagnosed compared to other STEC serogroups most likely at least in part are due to the use of laboratory methods designed to detect this specific serogroup. However, more focus on non-O157 STEC and change of laboratory practice now enables more frequent detection of non-O157 STEC than previously (153, 156).

The main reservoir of STEC is ruminants, but STEC has also been isolated from faeces from other animals such as pigs and wild boars, birds, cats and dogs (4, 10, 157, 158). In most of the world, cattle in particular are an important reservoir of STEC. However, sheep do also represent an important reservoir of STEC in many countries (4). The majority of STEC infections occur in the summer season, and the infectious dose for STEC disease is estimated to be low. Humans are infected by food or water contaminated with STEC, by direct transmission from animal to human, or from person to person (4). A number of sources are reported to be implicated in STEC disease, like beef and ground beef, municipal water, swimming water, unpasteurised milk and dairy products, fermented sausage, and sprouts (4, 22, 153).

1.3.4 Diagnosis

In the first years after NSF O157 was recognized as a cause of human disease, detection of this particular STEC relied on MacConkey agar containing sorbitol (SMAC agar). As NSF O157 is not able to ferment sorbitol these bacteria grow as distinct colour-less colonies on this medium. Later the SMAC agar was followed by media containing various supplements and now alternative plating media detecting both O157 and non-O157 STEC serotypes are available. While the use of selective media like SMAC agar is a sensitive method for detection of the NSF O157, a major drawback is that SF O157 and non-O157 STEC serotypes will usually not be detected by this strategy (10, 159-163). Therefore, for the diagnosis of all STEC it is recommended to use a combination of non-culture and

culture methods, where culturing is important for isolation of STEC bacteria, for confirmation of Shiga toxin production or the presence of the *stx* genes and other virulence associated genes, for serotyping and for genetic fingerprinting which is important in outbreak investigation (160, 164).

A widely used approach for many years in STEC isolation was to use DNA fragment probes targeting various genetic regions. Levine *et al.* reported the use of a 3.4 kb fragment DNA probe from the EHEC plasmid to detect STEC (165). This fragment was later shown to contain the *ehxA* gene encoding enterohaemolysin (10). Probes detecting the *stx1* and *stx2* genes as well as the *eae* gene encoding intimin have also been used for many years in STEC detection (10).

The Vero cell cytotoxicity assay is a highly specific and sensitive method for detecting Shiga toxins in human faecal samples, and has been regarded as gold standard for such detection. The method is based on the cytotoxic effect of Shiga toxins on Vero cell lines. Diluted faecal samples are centrifuged, and serial dilutions of the supernatant are incubated with Vero cell monolayers. The cytotoxic effect of the toxins on the Vero cells is confirmed by microscopic examination, with subsequent characterization of the toxins using specific neutralizing antisera against some of the toxin types (10, 159). Because the method is slow, labour intensive and specific laboratory facilities for cell culture is required, it is however not suited for use as a routine diagnostic method.

Methods like immunoassays for antigen detection or PCR are therefore more suited for use in a diagnostic laboratory. The Shiga toxins are detected by an immunoassay, while the *stx* genes are detected in a PCR assay. Both approaches enable detection of all STEC serotypes, regardless of specific biochemical characteristics of the bacterium, like their capability to ferment sorbitol. Several immunoassays are available for the detection of Shiga toxins. Some of these differentiate between Stx1 and Stx2, while others do not differentiate between the two Shiga toxin variants. Since faecal concentration of Shiga toxin may be low, it is recommended to enrich the sample in broth culture before testing by immunoassay (160). Immunoassay kits for screening of the O157 and H7 antigens in faecal samples are also available. One advantage of this test is that also SF O157 may be detected, however non-O157 STEC is not detected by this method (159, 160).

PCR is a sensitive and specific screening method, which is widely used in the detection of *stx* genes in STEC. The *stx1* and *stx2* genes may be detected using a primer pair detecting either type of *stx* gene or with primers that differentiate between *stx1* and *stx2*. In the latter case PCR may be performed for each gene alone or in multiplex, and both conventional and real-time PCR assays can be used for this purpose (10). In addition to detecting *stx* genes with PCR, other genes encoding known or potential virulence genes can be detected by PCR, like the *eae* gene encoding intimin, the *ehxA* gene encoding enterohaemolysin, and the *fliC* gene encoding the H7 antigen (10). In recent years STEC or EHEC

strains which easily lose their *stx* genes have been studied (45). To ensure that STEC that has lost the *stx* genes during infection not is missed in the laboratory, it is recommended to use detection methods that do not solely dependent on the presence of *stx* (45). One benefit of using PCR for STEC detection is that non-O157 serotypes are also detected by this method. However, the use of PCR as a primary test for the diagnosis of STEC also enables detection of strains with low potential for causing severe disease (10, 159, 160). As a consequence it becomes a challenge to assess the clinical and public health risk of STEC not commonly found in disease and outbreaks.

Immunomagnetic separation (IMS), using IMS beads labelled with antisera for O26, O103, O111, O145 and O157 is used for selective concentration of STEC in samples where the amount of bacteria is low. The method was initially used for concentration of STEC in food samples and bovine faeces. However, the method may also be used for human faeces in cases where the number of bacteria is suspected to be low, e.g. in HUS patients or during outbreak investigations (159).

Rapid and accurate identification of STEC is important to ensure proper treatment and improved patient outcome of infected patients. According to recommendations from Centers for Disease Control and Prevention (CDC) in the U.S.; "All stool specimens from patients with acute onset of community-acquired diarrhea and from patients with possible HUS should be tested for STEC" (164). While the recommendation of testing all faecal specimens from patients with acute onset of community-acquired diarrhoea and from patients with possible HUS for STEC is the ideal practice, this is not practical or possible in many laboratories. Therefore different practices for STEC testing and surveillance have been applied, and the methods for diagnosis of STEC may thus vary between laboratories and countries (153).

1.3.5 Typing of STEC

There is considerable genetic diversity within the *E. coli* species. Historically, different methods have been utilized to characterize the variation between different strains of *E. coli*. Serotyping was one of the first methods used to classify *E. coli*, and like all *E. coli*, STEC may be phenotypically serotyped according to their O, H and K antigens (1, 3), where some *E. coli* serogroups are more frequently found in relation to STEC disease and outbreaks than others, as discussed previously (4, 17, 61). Also mentioned above, Karmali *et al.* introduced the term seropathotype in order to assess the pathogenic potential of STEC (Table 2) (63). This classification of STEC into seropathotypes was recently reviewed by the European Food Safety Authority (EFSA), which concluded that the classification system did not define pathogenic STEC nor provide an exhaustive list of all pathogenic serotypes (166). The reviewing panel therefore suggested a modification of the seropathotype classification, in which all STEC with a serotype associated with severe disease like HUS "could be

categorised as seropathotype group “haemolytic uremic syndrome (HUS)-associated serotype(s)” or HAS. By this modified approach, in cases when full serotyping has been undertaken, all serotypes associated with severe disease are automatically categorised in the HAS group” (166).

While serotyping has been useful for phenotypic characterization of STEC, this method is not suited to characterize the genetic diversity between various strains. Multi-locus enzyme electrophoresis was used in early molecular studies of *E. coli* diversity, where electrophoretic analysis of 35 enzymes in the *E. coli* standard reference collection ECOR defined the six phylogenetic groups A, B1, B2, C, D and E (167). While most of the early phylogenetic studies focused on the ECOR collection, a couple of studies included STEC and other pathogenic *E. coli* (168, 169). When multi-locus enzyme electrophoresis was used to determine the genetic relationships of strains of O157:H7 to strains of other serotypes and pathotypes implicated in diarrhoeal disease, it was observed that multiple electrophoretic types were found among each serotype and pathotype (169). Also among the pathotypes ETEC, EIEC, EPEC and UPEC multiple electrophoretic types were observed, indicating that each pathotype had arisen by multiple parallel origins (168). In the study by Whittam *et al.*, it was suggested that the O157:H7 serotype clone had evolved from an O55:H7 clone (169). This theory was later supported by a stepwise evolutionary model where O157:H7 was suggested to have evolved from an O55:H7 progenitor (170).

When nucleotide sequencing became more available, multi-locus sequence typing (MLST) became the method of choice to discriminate between bacterial strains, and with MLST *E. coli* was separated into the phylogenetic groups A, B1, B2, D and E (C missing) (171). In one study on STEC evolution where seven housekeeping genes were sequenced, the phylogenetic analysis based on MLST data produced a phylogenetic tree which could be compared to trees generated by multi-locus enzyme data (172). The phylogenetic analysis suggested that gain and loss of mobile virulence elements had occurred several times in a parallel manner in different lineages of pathogenic *E. coli*, and the strains in the study were differentiated into the EHEC1, EHEC2, EPEC1 and EPEC2 groups (172). Three different MLST schemes have been developed for *E. coli* typing, and a number of studies on *E. coli* phylogeny have been published, reviewed in (173). In addition to MLST, multiplex PCR methods have been developed to distribute *E. coli* into the *E. coli* phylogenetic groups (174, 175).

Currently, whole genome sequencing of bacterial genomes has become a highly accessible and affordable analysis, and as more STEC have been whole genome sequenced the *E. coli* core and pan genomes have been estimated. The core genome represents the genes with high homology that are conserved and shared between *E. coli*, which encodes proteins required for survival and spread of the bacteria in the environment. The pan genome however represents the sum of all genes present

in the species (176-178). Both the *E. coli* core and pan genomes have been used for whole genome phylogenetic analyses, and have been shown to correspond well with the established phylogenetic groups (18, 173, 176).

For epidemiological purposes, various genetic fingerprinting methods have been developed that discriminate beyond serotype or species level to identify, trace and prevent dissemination of human pathogens like STEC. Pulsed-field gel electrophoresis (PFGE) is one of the methods which have been widely used for subtyping of STEC. This method is based on cutting of the bacterial chromosome with rare-cutting restriction enzymes, followed by separation of the DNA fragments by gel electrophoresis. PFGE has been shown to be well suited for outbreak investigation, and the method was regarded as the "gold standard" for STEC subtyping (4, 179). However, PFGE is a labour-intensive method, and due to the fact that it is an image-based method it is difficult to analyse and compare the results between laboratories.

Multiple-locus variable number tandem repeats analysis (MLVA) is another method which has been shown to be useful in subtyping of STEC. This method is based on polymorphic mini-satellites, referred to as variable number of tandem repeats (VNTRs). PCR is used to amplify the repeats regions, with subsequent (capillary) gel electrophoresis. The first MLVA method developed for STEC was an assay for NSF O157, and this method was shown to be useful in identifying outbreaks as well as discriminating between sporadic cases of infection (180-182). More recently, a generic *E. coli* MLVA assay has been developed for use in discrimination between pathogenic *E. coli* of non-O157 serotypes (183, 184). Also a third *E. coli* MLVA was developed to discriminate between epidemiologically unrelated SF O157 (185). Benefits of MLVA are that the method is fast, and has good discriminatory strength and typeability (4, 185).

Table 2. STEC seropathotypes[1]

Sero-pathotype	Relative incidence	Involvement in outbreaks	Association with HUS and HC[2]	Serotypes
A	High	Common	Yes	O157:H7, O157:H-
B	Moderate	Uncommon	Yes	O26:H11, 103:H2, 111:NM, O121:H19, O145:NM
C	Low	Rare	Yes	O91:H21, O104:H21, O113:H21, o121:NM, O165:H25 and others
D	Low	Rare	No	O7:H4, O69:H11, O103:H25, O113:H4, O117:H7, 119:H25, O132:NM, O146:H21, O171:H2, O172:NM, O174:H8 and others
E	Nonhuman only	Not implicated	Not implicated	O6:H34, O8:H19, O39:H49, O46:H38, O76:H7, O84:NM, O88:H25, O98:H25, O113:NM, O136:NM, O143:H31, O156:NM, O163:NM and others

[1]Adapted from (4, 63). [2]HUS = haemolytic uremic syndrome; HC = haemorrhagic colitis.

1.3.6 The STEC genome

Whole genome sequencing of bacterial genomes provides information on gene content and genetic organization, and gives an overview of how different organisms are related at the genetic level. The international prototype *E. coli* K-12, which is a commensal strain, was the first *E. coli* to be sequenced and was followed by the NSF O157 strains EDL933 and Sakai (186-188). Comparison of these *E. coli* genomes revealed that there was a high level of nucleotide sequence conservation in the 4.1 megabase (Mb) core genome shared between the strains. However, there was also considerable genetic heterogeneity between the genomes, as the core sequences in STEC O157 were interrupted by stretches of foreign DNA not present in K-12. About 1.5 Mb of the genome sequences in EDL933 and Sakai were strain specific and mostly made of horizontally transferrable genetic elements that originated from prophages and prophage-like elements, integrative elements and plasmids encoding potential virulence factors. These stretches of foreign DNA were referred to as “strain specific loops” in the Sakai genome and as “O islands” in the EDL933 genome (187, 188). Later, whole genome sequencing and genomic comparison of O157 and non-O157 LEE positive STEC with other *E. coli* and *Shigella* spp. showed that non-O157 STEC generally had larger genomes containing prophages, plasmids and integrative elements encoding potential virulence factors when compared to other *E. coli and Shigella* spp. (189). Although the prophages, plasmids and integrative elements in the different strains encoded the same or highly similar potential virulence factors, the genes showed

different evolutionary background. Furthermore, the gene content in the different STEC did not always correlate to the phylogenetic relationship between the strains (189). Additional whole genome comparison and phylogenetic analysis of 114 LEE positive *E. coli* strains revealed that strains referred to as EHEC (LEE positive, *stx* positive *E. coli*) could be subdivided into the two lineages EHEC1 and EHEC2, while tEPEC (LEE positive, *bfp* positive) were subdivided into the EPEC1, EPEC2 and EPEC4 lineages (18). Whereas a subset of aEPEC (LEE positive) strains were divided into the same lineages as tEPEC, many of the aEPEC strains were more similar to other *E. coli* pathovars and were not classified into any of the recognized EPEC lineages, demonstrating high diversity among LEE positive *E. coli* strains (18). Further analysis of these genomes revealed that although there were only few genomic features which distinguished the EHEC and EPEC lineages from each other, the number of virulence factors present in the different lineages varied. Genomic comparison of LEE islands showed that LEE with high similarity were found within the same EHEC/EPEC lineage, and that effectors encoded by LEE were detected in nearly all the investigated genomes (18). However, although most LEE islands followed the evolutionary history of the *E. coli* genomes, various LEE islands encoded different *eae* subtypes and the *eae* phylogeny did not always follow the phylogeny of the strains (18, 189). Comparison of different LEE positive *E. coli* genomes showed that the prevalence and level of similarity among T3SS effectors were associated with the *E. coli* phylogenetic lineage. In general, more *nle* genes were found in STEC than in EPEC (18, 189, 190). Furthermore, more *nle* genes were found among LEE positive *E. coli* that belong to the O157 EHEC1 lineage than among LEE positive *E. coli* of the EHEC2 lineage.

Also STEC bacteria that do not contain the LEE pathogenicity island have been shown to be genetically diverse (190). Genomic comparison of LEE negative STEC revealed that potential virulence factors frequently reported to be present in LEE negative STEC, were only partly present among the STEC investigated (189, 190). Further genomic comparisons between LEE negative and LEE positive STEC revealed that in addition to the absence of the LEE pathogenicity island, many other phage encoded effectors, including non-LEE encoded effectors, were absent in the LEE negative STEC. Phylogenetic analysis revealed that LEE negative STEC did not form a tight phylogenetic cluster, but was however distributed among LEE positive STEC in the *E. coli* phylogeny (190). As such there is clearly considerable genetic variety both within LEE positive and LEE negative STEC, as well as between these groups.

2 Aims of the studies

The main aim of the study was to identify genetic characteristics which could distinguish between highly pathogenic STEC with the potential to cause HUS, and STEC with a lower pathogenic potential.

Specific aims included

- ✓ To study genetic factors in the *stx2* encoding phage that might be associated with regulation and expression of *stx2* in Norwegian SF O157.
- ✓ To study potential differences between HUS and HUS-associated STEC compared to non-HUS STEC diagnosed by PCR from patients faecal specimens in the period 1996-2011 at St. Olavs Hospital, Norway, based on analysis of selected potential virulence genes.
- ✓ To distinguish between highly pathogenic STEC with the potential to cause HUS and less virulent STEC by performing whole genome comparison of 95 Norwegian non-O157 STEC.

3 Materials and Methods

3.1 Bacterial strains

Paper 1. Seventeen STEC strains harbouring the *stx2* gene were selected among 35 human clinical isolates of SF O157 recovered in Norway from 2005 through 2011. Because the aim of the study was to examine genetic characteristics in the phage encoding *stx2*, only strains that were *stx2* positive were included. In cases where multiple strains were present from a single patient, only one *stx2* positive strain from each patient was included. All isolates were from the strain collection at the Norwegian Institute of Public Health in Oslo, Norway.

Paper 2. All STEC isolates included in the study were isolated from patient stool specimens in the years 1996 through 2011 at the Department of Medical Microbiology, St. Olavs Hospital, Trondheim, Norway. Based on data from the laboratory information system, STEC was detected in 150 patients during the study period. Among these, 20 were excluded from the study because the laboratory did not succeed in obtaining STEC in pure culture, whereas for the remaining 130 patients STEC was identified in pure culture. Another eight *stx* negative (*eae* positive) *E. coli* isolates were included in the study because they were isolated from patients with HUS or bloody diarrhoea, or were epidemiologically linked to a HUS case and were of the same MLVA genotype as the STEC isolate from that case. In total 138 strains were included in the study.

Paper 3. In this study, 94 non-O157 STEC strains from the strain collection at the Norwegian Institute of Public Health, Oslo, Norway and three STEC (St. Olav104, St. Olav143 and St. Olav172) from the strain collection at St. Olavs Hospital, Trondheim, Norway were selected for whole genome sequencing. The strains included in the study were primarily selected to represent different MLVA genotypes (183, 184), a diversity of non-O157 STEC serotypes and patients with different severity of disease. All available non-O157 STEC strains isolated from patients with HUS (n=20) in Norway were included, apart from one strain (FHI10) which after whole genome sequencing was found to be contaminated. Thus, a total of 96 strains were included in the study. Some of the STEC strains from patients with HUS were from outbreaks and therefore had identical MLVA-genotypes or belonged to the same MLVA-genotype clusters. Four of the STEC strains included were furthermore classified as HUS-associated because they were epidemiologically linked and had identical MLVA-genotype as or belonged to the same MLVA-genotype cluster as a HUS case. Six strains of serotype O103:H25 did not have *stx* genes. The latter were included in the study because they were isolated from patients with HUS in an outbreak (five strains), or was isolated from fermented sausage linked to this outbreak (one strain). Five of the STEC strains were from non-human sources and were isolated during various

outbreak investigations related to STEC disease, of which one was designated as HUS-associated. Of the total 96 STEC strains included in the study, 95 strains were included for genomic comparison throughout the whole study whereas one strain (St. Olav104) was included for parts of the study only. In addition, 14 *E. coli* were included as reference strains for classification of the STEC strains into the *E. coli* phylogroups A, B1, B2, D, E, and F.

Both in Paper 2 and Paper 3, we included strains classified by us as STEC-LST. Although in general only *stx* positive strains are regarded as STEC, it is well known that STEC may lose its *stx* encoding prophage, either in the course of an infection or upon handling in the laboratory (40, 191, 192). The strains were classified as STEC-LST because they were isolated from patients with HUS or bloody diarrhoea, or were epidemiologically linked to a HUS case and were of the same MLVA genotype as the STEC isolate from that case.

3.2 Clinical data and ethical considerations

Patient information was collected from referral forms and from the Norwegian Surveillance System for Communicable Diseases (MSIS). We collected data on clinical symptoms (HUS, bloody diarrhoea, diarrhoea or no symptoms), age and sex, and correlated these data with laboratory results (Paper 2-3).

The study was approved by the Regional Committee for Medical and Health Research Ethics, REC South-East (REC number 2011/2314) (Paper 2 and Paper 3).

3.3 Identification of STEC

3.3.1 Detection and identification of STEC at St. Olavs Hospital

The primary detection of *stx1, stx2* and *eae* at St. Olavs Hospital was done by a two-step procedure where PCR for the *stx1*, *stx2* and *eae* genes first was done in mixed culture from a stool specimen. PCR was thereafter repeated on subcultures of discrete colonies from positive specimens with the aim to identify STEC in pure culture. During the period 1996-2004, screening for *stx1* and *stx2* was performed using primers and amplification conditions as described by Brian *et al.* (193). In 2004 conventional PCR for *stx1* and *stx2* was replaced by multiplex real-time PCR (for primers see Table 3). DNA isolation methods, amplification conditions, PCR reagents and PCR instruments varied during the study period (Paper 2 and Paper 3).

PCR for *eae* was done using the AE13 and AE14 primers, and amplification conditions were as described by Gannon *et al.* (194) from 2000-2004, and as described by Nielsen and Andersen (58)

from 2004-2008. Thereafter detection of *eae* was done by real-time PCR with primers described in Table 3 (Paper 2 and Paper 3).

STEC culturing was done by standard methods including SMAC agar, and identification of *E. coli* was done by standard biochemical tests (API 10S/20E, BioMerieux, Marcy l'Etoile, France) (Paper 2 and Paper 3).

To verify the primary PCR results we repeated PCR for *stx1*, *stx2* and *eae* for all strains included in the study, by real-time PCR using the primers described in Table 3 and the PerfeCTa Multiplex qPCR Supermix, UNG (Quanta Biosciences, Gaithersburg, USA) as described by the manufacturer. Real-time PCR was performed in a CFX instrument (BioRad, Hercules, USA), in a 20 µl volume with cycling conditions as follows for *stx1* and *stx2*; 95°C for 3 minutes, then 40 cycles with denaturation at 95°C for 10 seconds and annealing at 58°C for 10 seconds before elongation at 72°C for 10 seconds. PCR for *eae* was done using the following cycling conditions; 95°C for 3 minutes, then 40 cycles with denaturation at 95°C for 10 seconds and annealing at 50°C for 10 seconds before elongation at 72°C for 10 seconds (Paper 2).

Confirmation of *stx1*, *stx2* and *eae* was also done at the National Reference Laboratory for Enteropathogenic Bacteria at the Norwegian Institute of Public Health (195, 196) (Paper 1-3).

Table 3. PCR primers used for detection of STEC in Paper 2.

Primer	Sequence (5'-3' direction)	Target gene	Amplicon size (base pairs)	Reference
SLTI 1	AAA TCG CCA TTC GTT GAC TAC TTC T	*stx1*	368 bp	(193)
SLTI Rnew	CCA TTC TGG CAA CTC GCG ATG CA			Unpublished
SLTI TaqMan	FAM-AAC CTC ACT GAC GCA GTC TGT GGC AAG AGC-BHQ1			Unpublished
SLTIIFnew	CAG TCG TCA CTC ACT GGT TTC ATC	*stx2*	283 bp	Unpublished
SLTIIRnew	GGA TAT TCT CCC CAC TCT GAC AC			Unpublished
SLTII TaqMan	HEX-CTG TCA CGG CAG AAG CCT TAC GCT TCA GGC-BHQ1			Unpublished
eae-Fny	TTC ATT GAT CAG GAT TTT TCT GG	*eae*	105 bp	Unpublished
eae-Rny	GCT CAT GCG GAA ATA GCC			Unpublished
eae-P	FAM-ATA GTC TCG CCA GTA TTC GCC ACC AAT ACC-6-TAMRA			(58)
efa1-F	ATC AGA AGC CCG ACT ACG	*efa-1/lifA*	193 bp	Unpublished
efa1-R	AAC ATT TGC CAG ACC AAG G			Unpublished

3.3.2 Phenotypic characterization

Identification of STEC strains by culture and biochemical tests at St. Olavs Hospital was done according to standard microbial methods. Initial serogrouping was performed with O antisera using polyspecific Anti-coli I, II and III, and monospecific O-antisera for the O serogroups O26, O103, O111, O145 and O157, as described by the manufacturer (Sifin, Germany) (Paper 2 and Paper 3). Later more extensive serotyping was done at the National Reference Laboratory for Enteropathogenic Bacteria at the Norwegian Institute of Public Health using monospesific O:K and H antisera covering altogether 44 O-serogroups, including O26, O103, O111, O121, O145, O157; and 8 H-antigens (in-house antisera and antisera from Sifin and SSI, Denmark). Additional molecular serotyping by multiplex PCR was performed on selected STEC strains after 2009 (Lindstedt *et al*., unpublished) (Paper 1-3).

3.4 Statistical analyses

Fisher's exact test was used for statistical analyses (Paper 2 and Paper 3). In Paper 2, a p-value ≤0.05 was regarded as statistically significant. In Paper 3, Fisher's exact test was used to analyse if specific *stx* subtypes were differently distributed in LEE positive and LEE negative STEC. A p-value ≤0.05 was regarded as statistically significant. Fisher's exact test was also used to test if specific genes in the accessory genome were overrepresented, and for overrepresentation of gene variants in the core genome, in subgroups of the 95 STEC strains. Classification of the strains into subgroups was based on clinical and outbreak investigation information, phylogenetic analysis, and Principal Component Analysis and Partial Least Squares regression. For corrections of false discovery rate (FDR) in multiple testing the Benjamini-Hochberg method was used, with FDR≤0.01 regarded as statically significant. Whenever no significant association was identified after FDR correction, results for uncorrected analysis are given. The statistical analyses were performed using the R software package version 3.03 (R: A Language and Environment for Statistical Computing, http://www.R-project.org).

3.5 Genotypic identification and characterization of STEC

PCR

Paper 1. PCR was used for molecular detection of SF O157, as well as for detection of potential virulence genes. In addition, PCR was used to identify the anti-terminator *q* gene and to investigate the SF O157 strains with the stx8 primer set which is used to distinguish O157 lineage I from lineage II (197). PCR was used to amplify DNA for sequencing of the promoter region of the *stx2* genes, while inverse PCR was used to amplify DNA for sequencing of the *stx2* promoter and anti-terminator *q*

gene region in the SF-O157 strains. PCR was also used to amplify DNA for MLVA analysis of SF O157 and for subtyping of the *stx2* genes (see below).

Paper 2. As described in section 3.3.1, PCR was used for identification and verification of the *stx* and *eae* genes to identify STEC in mixed culture and for confirmation of STEC in pure culture. Furthermore, PCR was used for confirmation of *stx1, stx2* and *eae* at the Norwegian Institute of Public Health. PCR was also used to amplify bacterial DNA for MLVA analysis of all STEC serotypes and for *stx2* subtyping as described in Paper 1. In addition, PCR was used to identify potential virulence genes.

Paper 3. PCR was used for STEC detection at St. Olavs Hospital and for confirmation at the Norwegian Institute of Public Health as described in section 3.3.1. PCR was in addition used for amplification of DNA for MLVA analysis of the non-O157 STEC strains selected for the study as described above.

Subtyping of stx2 by molecular methods

The *stx2* subtype was determined at the Norwegian Institute of Public Health using modifications of previously published methods for PCR-restriction fragment length polymorphism (RFLP) and sequencing (43, 198, 199), and by PCR (38) (Paper 1 and Paper 2).

MLVA

MLVA was chosen for genotyping of STEC strains. Selected VNTR loci were amplified by PCR, and the fluorescently dyed DNA fragments were subsequently analysed by capillary electrophoresis. MLVA analysis of SF O157:H- and O157:H7 was performed with MLVA methods developed for the O157 serogroup (Paper 1 and Paper 2) (180, 181, 185). Generic MLVA methods developed for all *E. coli* serotypes were used for typing of all other STEC serotypes (183, 184) (Paper 2 and Paper 3).

DNA Sequencing with the Sanger method

PCR products were purified using the QIAquick PCR Purification Kit (Qiagen Gmbh, Hilden, Germany) and thereafter cycle sequencing was performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems by Life Technologies, Carlsbad, CA) as described by the manufacturer to determine *stx2* subtypes (Paper 1 and Paper 2), the *stx2* promoter region and the anti-terminator *q* gene in the SF O157 strains (Paper 1). After cycle sequencing, extension products were purified using the DyeEx 2.0 Spin Kit (Qiagen) before the DNA was separated by capillary electrophoresis.

Capillary electrophoresis

PCR-products. PCR-products labelled with a fluorochrome were run by capillary electrophoresis on an ABI-3100 or ABI-3130xl automated sequencer (Applied Biosystems by Life Technologies), where the PCR-products were separated according to size, charge, and colour (Paper 1).

MLVA. To determine the sizes of each VNTR locus, amplified VNTR loci labelled with a fluorochrome were run by capillary electrophoresis on an ABI-3100 or ABI-3130xl automated sequencer (Applied Biosystems by Life Technologies). Each locus was separated according to the size, charge, and colour of the loci (Paper 1-3).

DNA sequencing. Purified and amplified genomic DNA was run by capillary electrophoresis on an ABI-3100 or ABI-3130xl automated sequencer (Applied Biosystems by Life Technologies), which separates the amplified DNA with a resolution of one base pair. Specific fluorescence peaks for each nucleotide are analysed to determine the DNA sequence. The raw-data files were exported to the SEQMAN Pro sequencing analysis software (DNASTAR Lasergene 9 Core Suite, Madison, WI) for inspection and assembly (Paper 1 and 2).

PCR-restriction fragment length polymorphism (PCR-RFLP)

DNA from *stx2* variants amplified by PCR was digested with restriction enzymes, before the restriction fragments were separated by capillary electrophoresis on an Agilent BioAnalyzer 2100 (Agilent Technologies). The fragments were separated according to the lengths of the fragments. Specific combinations of fragment sizes corresponded with specific *stx2* subtypes (Paper 1 and 2).

3.6 Whole genome sequencing

STEC strains were whole genome sequenced with either Illumina or Pacific Biosciences (PacBio) Technology (Paper 3).

DNA isolation

Strains were grown overnight on MacConkey agar. Genomic DNA was isolated for each strain using the Qiagen MagAttract® DNA Mini M48 Kit and the Qiagen BioRobot M48 (Qiagen, Hilden Germany) as described by the manufacturer.

Sequencing

Ninety-six of the STEC strains were sequenced with the Illumina Technology, while one strain (St. Olav104) was sequenced with PacBio Technology.

For strains to be sequenced by Illumina technology a standard read library from the DNA was prepared, with an average fragment length of 370 bp. The DNA was sequenced by LGC Genomics (Berlin, Germany) on the Illumina HiSeq2000 platform (Illumina, San Diego, CA, USA) with 100 bp paired-end reads. Assembly of processed and error corrected paired-end reads was done using Velvet 1.2.04 (200).

Forty-eight of the 96 strains were selected for additional mate pair sequencing. For this purpose a 2 kb Illumina Mate Pair library was prepared and the DNA was sequenced by LGC Genomics (Berlin, Germany) on the Illumina HiSeq2000 platform with 100 bp paired-end reads. Assembly and scaffolding of processed and error corrected paired-end reads was performed using Allpaths-LG release 45553 (201). Gap closure of assembly scaffolds was done using SOAP GapCloser version 1.12 (202), while refinement of gap-closed scaffolds was done using SEQuel version 1.0.2 (203).

Genome sequencing on the PacBio platform was performed at the Norwegian Sequencing Centre (Oslo, Norway). A library was prepared using the Pacific Biosciences 10 kb library preparation protocol, and size selection of the final library was performed using Ampure beads. The library was sequenced on a Pacific Biosciences RS II instrument (Pacific Bioscience, Menlo Park, CA, USA) using P4-C2 chemistry and three SMRT cells. Processed reads were assembled using HGAP v2 (204).

Bioinformatics analyses

Gene annotation

Identification of open reading frames was performed using the Prodigal Microbial Gene Prediction Software (205). Functional gene annotation was done using myRAST (206).

Comparative analyses

The CMG-biotools (Comparative Microbial Genomics) package was used for genome comparison (207). Blastmatrix in CMG-biotools was used to identify proteins shared between genomes, while pancoreplot was used to identify the pan- and core- genome of the sequenced strains. In this context, genes were considered to be equal homologous having a minimum of 60% alignment length and 90% sequence identity. The accessory genome was defined by subtracting all core genes from the pan genome. Genome analysis and comparison was performed across 95 of the sequenced STEC.

A separate genomic comparison was done between the strain St. Olav104 and two HUS-strains (FHI58 and FHI63).

Core genome phylogeny

E. coli phylotypes were determined *in silico* based on a core gene tree. This was created as described by Kaas, et al. (176) using 1,861 core genes present in all the 95 STEC genomes and additional 14 *E. coli* reference genomes representing the *E. coli* phylotypes A, B1, B2, D, E and F (18, 171, 176, 208, 209).

Core gene analysis

Core gene nucleotide sequences (n=1,861) from the 95 STEC and the 14 reference *E. coli* were aligned separately and a consensus sequence was estimated for each gene using EMBOSS 6.3.1 (210). A python implementation of the edit distance method (211) was used to quantify the difference between the consensus sequence and the corresponding sequence of each core gene for all 109 strains included in the analysis. This resulted in various edit distances, representing different gene variants for each of the core genes. Edit distance values for all strains were normalized and transformed into a binary matrix for core gene comparisons.

To examine if any gene variant from the same core gene family showed different Pfam domains, we used pfam_scan.pl with the HMMER3 library of Pfam domains.

Principal component analysis and Partial least squares regression

For Principal component analysis (PCA) and Partial least squares (PLS) regression the Laydi software (http://www.laydi.org) (unpublished) was used. For PLS regression, dependent variables (for the Y-matrix) were the clinical diagnosis HUS or classification as HUS-associated, and the presence of *stx1* and/or *stx2*. HUS and HUS-associated STEC-LST were classified as *stx2* positive for these analyses.

Functional annotation and Gene Ontology enrichment analysis

Blast2GO was used for functional annotation based on gene ontology (GO) and for GO enrichment analysis. In Blast2GO, Fisher's exact test was used for GO enrichment analysis (212, 213).

Subtyping of Stx1 and Stx2

There are three known subtypes of Stx1; Stx1a, Stx1c and Stx1d, and seven known subtypes of Stx2; designated Stx2a through Stx2g. Reference protein sequences were downloaded for each Stx subtype and Stx type variant from GenBank (38). Amino acid sequences of the A and B subunits were

concatenated and aligned separately for Stx1 and Stx2 using Clustal O in Jalview (214, 215). For cluster analysis and tree calculations the Neighbour Joining algorithm in Jalview using % identity was used. Clustering of the Shiga toxin protein sequences of the sequenced strains with reference sequences was used to classify the former into Stx1 and Stx2 subtypes.

4 Results

4.1 Paper 1

In this study we investigated the nucleotide sequence of the anti-terminator *q* gene and the region upstream of the *stx2* gene involved in regulation of *stx2a* expression in 17 Norwegian SF O157 STEC strains. Sequencing of three selected SF O157 strains revealed that the anti-terminator *q* gene and genes upstream of *stx2a* were identical or similar to the ones observed in the STEC O111:H- strain AP010960, but different from the ones observed in the O157:H7 strain EDL933 (AE005174). The results suggested divergent *stx2a* encoding bacteriophages between NSF O157 and the three SF O157 strains (FR874039-41). Furthermore, stretches of different DNA sequences were detected in the *stx2* phages of the SF O157 strains, suggesting diversity among bacteriophages also within the SF O157 group. An assay for detecting $q_{O111:H-}$ was developed, and all *stx2* positive Norwegian SF O157 strains (n=17) harboured this specific *q* gene. Further investigations are needed to elucidate whether the $q_{O111:H-}$ gene observed in all our SF O157 contributes to the increased virulence seen in SF O157 compared to NSF O157.

4.2 Paper 2

In this work we studied STEC and STEC-LST strains which were detected by PCR for the *stx1*, *stx2* and *eae* genes in clinical faecal specimens in the years 1996 through 2011 at the Department of Medical Microbiology, St. Olavs Hospital, Norway. In this period *stx* genes were detected in faecal samples from 150 patients, while STEC was isolated in pure culture from 138 (1.09%) of 12,651 patients tested. The majority of patients with STEC infection were children <5 years old, and 11 of these suffered from HUS. None of the cases were fatal. Twenty (14.5%) of the strains were NSF O157, 78 (56.5%) belonged to common STEC serogroups frequently involved in disease and outbreaks (O26, O103, O111, O121, O145 and SF O157), and 40 (29.0%) belonged to other serogroups or were non-typeable. All HUS cases were infected with common STEC serogroups, except NSF O157. Twenty-four STEC strains were classified as HUS-associated. Age of the patient ≤5 years and the genes *eae* and *stx2a* were significantly associated with HUS-associated STEC ($p<0.05$ for each parameter), while *stx1* was associated with non-HUS STEC ($p<0.05$). Other putative virulence genes, apart from *ehxA,* were significantly more frequent among HUS-associated strains ($p<0.05$ for each gene). However, these genes were also present in some non-HUS STEC strains and could therefore not reliably differentiate between HUS-associated and non-HUS STEC.

4.3 Paper 3

In this study, we compared the genetic content across 95 Norwegian non-O157 STEC strains. The LEE pathogenicity island was present in 54 of the genomes, while 41 strains were LEE negative. *stx2a* was significantly more frequent among the LEE positive STEC, while *stx1c* and *stx2b* were significantly more frequent in LEE negative strains ($p<0.05$ for each gene). All the HUS-associated STEC included for analysis in this study (n=23) harboured the LEE. Apart from six STEC-LST strains of serotype O103:H25 from an outbreak all harboured the *stx2a* subtype. Furthermore, all HUS-associated STEC belonged to *E. coli* O serogroups known to be associated with STEC disease.

The phylogenetic analysis of the core genome distributed the 95 STEC strains into all the *E. coli* phylogroups A, B1, B2, D and E, except group F. In general, clusters of LEE negative STEC strains were distributed between clusters of LEE positive strains. Most of the strains belonged to the B1 phylogroup, and a majority of the LEE positive strains were also found within this group. All HUS-associated strains were found in phylogroup B1, in two clusters which we designated HUS-group 1 and HUS-group 2 (Figure 6), but also non-HUS strains were found in these clusters.

Genomic comparison revealed considerable variation in gene content across the 95 STEC strains. A total of 26,073 genes were identified, of which 1,954 were core genes. Thus the accessory genome consisted of approximately 24,000 genes. PCA of the accessory genome separated LEE positive (n=54) and LEE negative (n=41) STEC strains in two distinct groups. PCA and PLS regression of LEE positive strains (n=54) as well as of LEE positive/*stx2* positive strains (n=33) could not further separate these strains into subgroups. By comparison of all 54 LEE positive STEC strains, we identified eleven genes which were more frequent in the HUS-associated (n=23) than in non-HUS strains ($p\leq0.01$, FDR). None of these genes however, were present exclusively in HUS-associated strains. When STEC strains in HUS-group 1 (n=18) were compared with all other LEE positive strains (n=36), 357 genes were more frequent in HUS-group 1 strains ($p\leq0.01$, FDR), and one of these genes encoding a hypothetical protein was exclusive to HUS-group 1 in this analysis. When STEC strains in HUS-group 2 (n=23) were compared with all other LEE positive STEC (n=31), 576 genes were overrepresented in the former group ($p\leq0.01$, FDR). Four genes were present in all strains in HUS-group 2 while absent in the other strains.

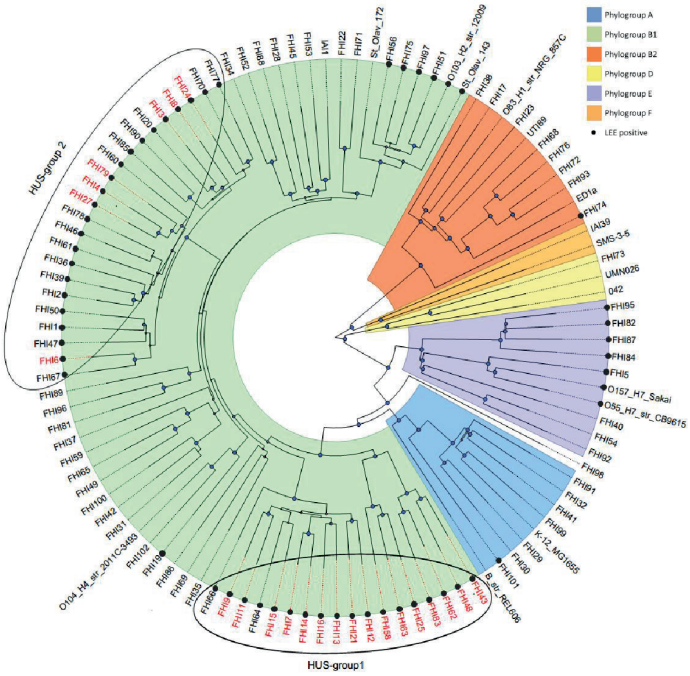
Analysis of the core genome identified in total approximately 13,000 gene variants. Different Pfam domains, and therefore different protein sequences, were observed in 13 of these gene variants. Comparison of core gene variants among LEE positive STEC (n=54) identified 281 gene variants that

were overrepresented in the HUS-associated (n=23) compared to non-HUS strains (n=31) ($p \leq 0.01$, FDR). None of these gene variants were found only among HUS-associated strains.

Comparative analysis on gene content was furthermore performed on specific HUS and non-HUS STEC that were epidemiologically linked and belonged to the same MLVA outbreak cluster. In HUS-strain FHI4 we identified 179 genes which were absent in the non-HUS strain FHI3 from the same outbreak. The majority of the genes were related to various mobile genetic elements integrated in the bacterial chromosome, Nle effectors or plasmid pO26_1 (AP010954). A closer search revealed additional plasmid genes on other contigs in the FHI4 draft genome, indicating the presence of a complete pO26_1 plasmid in this strain, while the corresponding genes were not found in strain FHI3. In HUS-strain FHI48 we identified 153 genes that were absent in the non-HUS strains FHI43 and FHI62 from the same outbreak cluster. Again, most of the genes were related to mobile genetic elements. In the two HUS strains FHI58 and FHI63 from another outbreak, we identified 54 genes exclusive to these two strains, while another 506 genes were present only in the non-HUS strain St. Olav104. The genes in the two HUS strains were related to various functions, while in the non-HUS strain, the majority of the genes were related to mobile genetic elements and several Nle effectors.

GO enrichment analysis of the genes significantly more present in the 23 HUS-associated STEC and the STEC in HUS-group 1 revealed that GO terms involved in degradation of L-idonate were enriched. Furthermore, in HUS-group 1 we identified enriched GO terms involved in protein secretion by the type II secretion system. Twenty-six GO terms which were involved in flagellar motility and siderophore biosynthesis were enriched in HUS-group 2.

In summary, we were not able to clearly distinguish between HUS-associated and non-HUS STEC by extensive genome comparisons in this study. Our results indicate that STECs from different phylogenetic backgrounds have independently acquired virulence genes that determine pathogenic potential, and that the content of such genes is overlapping between HUS-associated and non-HUS strains.



.

Figure 6. Core gene tree of 95 non-O157 STEC sequenced in this study, and 14 *E. coli* reference genomes. The *E. coli* phylogroups are marked with the colours blue (A), green (B1), orange (B2), yellow (D), ochre (F) and indigo (E). LEE positive STEC are marked with a •-sign, while all HUS and HUS-associated STEC included in the study are marked with red letters.

5 Discussion

Despite extensive studies on STEC and HUS in the international research community, we are still unable to reliably differentiate between STEC that have the potential to cause severe infection with HUS from those that only cause mild or no disease. As long as one cannot differentiate between high and low risk STEC, in Norway, infection control measures are applied to most STEC cases, although very few of these will develop into HUS. Improved classification of STEC as HUS related or not HUS related would be of great importance for interpretation of results in diagnostic laboratories, for infection control measures, for assessment of prognosis in patients with STEC infection, and could save expenses if infection control measures could be avoided or minimized in low risk cases.

In this study we investigated STEC from persons in Norway with asymptomatic carriage or symptomatic infections ranging from mild to bloody diarrhoea and HUS, in order to identify genetic characteristics useful for differentiating between bacteria with the potential to cause HUS and those that do not have such potential. In summary this study revealed that STEC causing human disease constitute a genetic and phylogenetic highly diverse nature. A clear difference in gene profile was observed between LEE positive and LEE negative STEC. Common characteristics among STEC classified as HUS-associated in this study were that all were LEE positive and contained the *stx2a* subtype, apart from the STEC that had lost the *stx* genes. In addition, all belonged to common STEC serogroups frequently reported as cause of STEC disease and outbreaks elsewhere. Genomic comparison revealed that the gene content in part was shared between HUS-associated and non-HUS STEC, and consequently it was not possible to reliably distinguish STEC with the potential to cause HUS from those without this potential.

5.1 Detection of STEC using PCR

One major advantage of PCR in the diagnosis of STEC infection is that it improves the detection of both non-O157 as well as SF O157 STEC compared to culture based methods. However, it may also lead to detection of a range of STEC with a low potential for causing severe disease in humans. In our study (Paper 2) only 20 (14.5%) of the 138 STEC strains were NSF O157:H7, the leading STEC serotype detected worldwide, while 78 (56.5%) STEC strains belonged to common non-O157 STEC serogroups, and SF O157, frequently implicated in severe disease and outbreaks in both Norway and other countries (153, 155, 156). On the other hand, uncommon STEC serogroups or STEC not typeable with the available antisera were detected in 40 (28%) of patients. Consequently, use of PCR resulted in detection of a high number of O157 and non-157 STEC of common STEC serogroups that based on current knowledge may be viewed as high-risk strains (Croxen, Law et al. 2013), but also strains of non-STEC serogroups that most likely do not represent a high risk for HUS development. This

illustrates the importance of identifying microbiological characteristics that can be used to differentiate high risk from low risk STEC.

Although PCR is a sensitive and specific screening method in STEC detection, various PCR protocols might potentially have had an impact on which STEC strains that were detected in Paper 2. In the first part of the study period, only the *stx1* and *stx2* genes were analysed, whereas *eae* was included from year 2000 and onwards. These factors might have affected the sensitivity of STEC detection in the early phase of the study. In addition, different primers, reagents and equipment were used for PCR analysis through the study period, including a switch from conventional to real-time PCR, which might further have an impact on the sensitivity of STEC detection. Although different primers were used for detection of *stx1* and *stx2,* these primers were designed to detect all variants of the *stx* genes, except *stx2f*. However, despite the variation of PCR protocols used throughout the study period, the strain collection represented an unselected group of STEC diagnosed by PCR analysis of the *stx* and *eae* genes in a hospital laboratory through a period of 16 years.

5.2 STEC epidemiology in central Norway

In Paper 2 we characterized STEC diagnosed in our laboratory at St. Olavs Hospital during a 16 years period by use of PCR for *stx1*, *stx2* and *eae*. While *stx* genes were detected in faecal samples from 150 patients during this time period, STEC or STEC-LST strains were isolated in pure culture from 138 (1.09%) of the 12,651 patients tested. The age group <2 years old was most common among the patients isolated with STEC. This may partly be explained by the laboratory routine, which was to analyse all stool specimens from children <2 years old for STEC, whereas samples from older age groups only was examined for STEC if there was a clinical suspicion of HUS, bloody diarrhoea or if an STEC outbreak investigation was performed. However, a high number of STEC was detected also in children 2-4 years old. These findings support the notion that STEC infection in general is most common in low age groups (20, 21). All HUS cases at St. Olavs Hospital were from children ≤5 years old. In comparison, at the national level, this age group was found to account for 79% of all HUS-cases in the years 1999-2008, while 21% of the HUS patients were older than five years old (216).

As mentioned above, only 20 (14.5%) of the 138 STEC strains in the study at St. Olavs Hospital were NSF O157:H7 and none of these were from HUS cases (Paper 2). During the same time period NSF O157 was isolated from four cases of HUS in other parts of Norway (data from the Norwegian Surveillance System for Communicable Diseases, MSIS). At St. Olavs Hospital six (54.5%) of the 11 HUS cases were caused by non-O157 STEC, while five (45.5%) were caused by SF O157. Five of nine (55%) of the patients infected with SF O157:H- STEC developed HUS, and this serotype was the most common serotype isolated from HUS patients in the study (Paper 2). Besides, 11 of 16 children (68%)

infected with SF O157 in a national outbreak in Norway in 2009-2011 developed HUS (Paper 1). These results are in accordance with reports from other countries that SF O157 more often progresses to HUS compared to NSF O157 (217-221). In addition, STEC O103:H25 was another particularly aggressive STEC serotype causing HUS in this study (Paper 2 and Paper 3). In a national outbreak in 2006, ten out of 17 (58%) patients developed HUS (25). However, although all HUS cases in this study were caused by STEC serogroups known to be associated with severe STEC disease, the German HUS outbreak in 2011 (16) reminds us that new STEC serogroups might emerge and cause severe disease in a large number of people. Consequently, STEC detection should not be too dependent on identification of serotypes previously known to cause STEC disease and outbreaks.

While an ideal practice for diagnosis of STEC would be to test all stool specimens from patients with diarrhoea and possible HUS for this pathogen (164), this is not always practical or possible. Based on our findings that STEC was most frequently diagnosed in children <5 years old, prioritization of this age group for testing for STEC seems reasonable. In fact, the national recommendations for microbial diagnosis of STEC was updated in Norway in 2011 to test all stool specimens from children <5 years old for this pathogen. A higher number of STEC was detected in central Norway where St. Olavs Hospital is located compared to other regions of Norway in this period. Although less than a tenth of the population live in central Norway, more than one-fourth of STEC cases reported nationally were diagnosed in this region, by our laboratory (data from the Norwegian Surveillance System for Communicable Diseases, MSIS). Because our laboratory was one of the first in Norway to introduce PCR for the diagnosis of human STEC infection, the higher detection rate of STEC compared to other parts of Norway could to some extent be attributed to the diagnostic method used. However, the fact that during this same time period 11 out of 53 HUS cases (20.8%, data from Norwegian Surveillance System for Communicable Diseases, MSIS) in Norway also were from our region indicates that there may be epidemiological differences in risk of STEC infection and disease between different regions of Norway. In addition, there are differences in STEC epidemiology between European countries (21). Our findings might therefore not necessarily be representative for other regions of Norway and other countries.

The laboratory routine in the study presented in Paper 2 was to test samples for STEC dependent on age of the patients (Paper 2). However, this routine was not always strictly practised. Thirty-nine patients diagnosed with STEC in the study were older than 2 years of age and had diagnoses other than HUS and bloody diarrhoea and were not part of an outbreak investigation. Furthermore, information regarding bloody diarrhoea may have been incomplete, since bloody diarrhoea was recorded for only nine of the patients, while there was no information of bloody stools in 68 other

patients with diarrhoea. For these cases it might be that the information given by the referring physician was incomplete, or that such data was not always recorded or updated in the patient information database. These aspects might therefore represent limitations of the study.

5.3 Characterisation of STEC virulence factors

Stx2a and intimin are important virulence factors in STEC that have been associated with severe disease (9). Throughout this study all HUS-associated STEC strains carried *stx2a,* with the exception of STEC-LST strains, and all harboured *eae* (Paper 1-3). Analysis of the presence or absence of these virulence genes in Paper 2 revealed that *eae* and *stx2a* were significantly more frequent in HUS-associated than in non-HUS strains ($p<0.05$), whereas STEC containing *stx1* was exclusively associated with non-HUS infection ($p<0.05$) (Paper 2). In Paper 3, *stx2a* was more frequent among LEE positive strains ($p<0.05$) while *stx2b* and *stx1c* were more frequent among LEE negative strains ($p<0.05$) (Paper 3). Thus, our results are in line with results from previous studies where Stx2a has been shown to possess higher potency than Stx1 and certain other Stx2 subtypes (44), and that LEE positive and *stx2a* positive STEC strains are most often associated with severe disease (17, 38, 39, 43, 44).

While *eae* and *stx1* positive, *stx2* negative STEC has been isolated from patients with HUS (39, 222), there are to our knowledge no reports of outbreaks of severe disease with such bacteria. Furthermore, in line with reports from other regions in Norway (Brandal *et al*., manuscript in preparation) and other countries (156, 223-225), our results support the notion that infection with STEC that do not belong to common STEC serogroups, lack *eae* and/or *stx2a* even if *stx1* is present, represent a low risk for HUS development. As such it seems safe to suggest a classification of *stx1* positive, *stx2* negative STEC as low risk for HUS development.

Whereas *stx2a* and LEE were present among all the HUS-associated STEC, these characteristics were not unique for such STEC and thus not sufficient to clearly differentiate this group from non-HUS STEC. We therefore aimed to investigate other genetic factors in STEC in an attempt to differentiate between HUS-associated and non-HUS strains.

5.3.1 The anti-terminator *q* gene in SF O157

Although the *stx2* genes have their own promoter, the expression of Stx2 is dependent on the anti-terminator activity of the Q protein which regulates the bacteriophage late genes (26, 28). From previous studies in NSF O157 STEC it is known that two different *q* genes, q_{933} and q_{21}, are involved in the regulation of *stx2* expression, and that increased production of Shiga toxins and enhanced virulence are observed in strains harbouring the q_{933} gene compared to the ones carrying the q_{21}

gene (32-34). However, similar information has not been available for SF O157. By sequencing of the region harbouring the *q* and *stx* genes in SF O157 STEC strains (Paper 1), we surprisingly found that none of the three SF O157 strains investigated harboured any of the two *q* genes found in NSF O157 STEC. Instead the strains displayed identical sequences in the region including the *q and stx2* genes similar to that of the bacteriophage of O111:H− strain 11128 (AP010960) (189). Furthermore, PCR screening with primers designed for Paper 1 confirmed the presence of this specific *q* gene, $q_{O111:H-}$, among all *stx2* positive SF O157 strains (n=17) from the strain collection at the Norwegian Institute of Public Health.

Comparative genomics studies of *stx* encoding lambdoid bacteriophages have shown that although the bacteriophages possess an overall similarity in genetic organization, they constitute a heterogeneous group with high genetic and morphologic diversity due to various recombination events (26, 28, 29, 226). We did not sequence the complete bacteriophages in this work, and therefore do not know whether the rest of the bacteriophage DNA is similar to the phage sequence of the O111:H- strain. Nonetheless, our results demonstrate that SF O157 harbour a different *stx2a* bacteriophage compared to NSF O157 STEC. We also observed some degree of genetic diversity among *stx2a* encoding bacteriophages within the SF O157 group, as different DNA sequences were seen within the tRNA genes in one of the three sequenced strains.

It has been shown that the phylogenetic background of various *stx* bacteriophages mostly is concordant with their bacterial host (29, 226). Furthermore, recent bioinformatics investigations of *stx* and *q* genes have shown that different variants of the *q* gene exist, and that the *q* and *stx* genes combine into various *q-stx* genotypes (31). As phylogenetic analyses of SF O157 and NSF O157 strains show that these have diverged early in the evolution of *E. coli* O157 (221, 227), the finding of a *stx* phage with an alternative combination of the *q-stx2a* genes in SF O157 as compared to classic NSF O157 strains should not be surprising.

While higher levels of colonization, and thereby enhanced Stx2 exposure has been suggested to be partly responsible for the increased pathogenic potential of SF O157 STEC strains, there is no evidence of increased *stx2a* expression *in vitro* in SF O157 STEC as compared to NSF O157 STEC (220). However, *stx2* expression was not measured in our study and additional investigations are therefore needed to elucidate the activity of the $Q_{O111:H-}$ protein in the Norwegian SF O157 strains. In addition, there might be other, yet unknown factors in SF O157 that have an impact on virulence. Although the findings in this study showed that all *stx2a* positive Norwegian SF O157 harboured the same anti-terminator *q* gene, which was different compared to the ones seen in NSF O157 STEC, our strain collection constituted only the 17 Norwegian strains. Further investigations of these aspects in STEC

SF O157 strains from other countries would therefore gain knowledge on *stx2* regulation in this particular virulent STEC serotype.

5.3.2 Analysis of selected virulence genes in STEC

In Paper 2 we analysed the presence or absence of the virulence genes *ehxA, nleB, nleE, ent, efa-1/lifA, nleA, nleF, nleH1-2* and *espK* in the 138 STEC strains. The genes *nleB, nleE, ent* and *efa-1/lifA* encoded on the O islands OI-122, and *nleA, nleF* and *nleH1-2* encoded on OI-71, are all previously shown to be associated with STEC with increased virulence (63-67). In addition *ehxA* is frequently found both in O157 and non-O157 STEC (4), whereas *espK* has been reported to be strongly associated with severe STEC disease (17). We therefore selected this collection of genes to see if some of them could be useful for distinguishing between strains associated with severe disease and strains not associated with such disease in our collection of STEC strains. The results showed that all the potential virulence genes analysed, except *ehxA*, were significantly more frequent among HUS-associated than in non-HUS strains ($p<0.05$ for each gene). While the *ehxA* gene has been regarded as an important virulence marker in STEC and has been reported to be a marker of "typical EHEC" (67), it was the only potential virulence gene analysed that was not significantly more frequent in HUS-associated than non-HUS strains in this study. *ehxA* was also the sole potential virulence gene present among the subset of *eae* negative STEC strains. By cluster analysis of the selected potential virulence genes *eae* positive and *eae* negative STEC were separated in two main clusters, in which *eae* positive harboured more virulence genes than *eae* negative strains. In this respect, our results correspond with previous reports (60, 63, 64),

Although all the HUS and HUS-associated strains in the study in Paper 2 were of STEC serogroups that frequently are reported to be involved in severe STEC disease and harboured most of the virulence genes investigated, several of the non-HUS STEC strains showed identical serogroups and contained the same virulence gene profile as the HUS-associated strains (Paper 2). Especially non-HUS STEC strains of the serogroups O145, O103, O157 (NSF and SF), O26, O121, and O111 contained a high number of the potential virulence genes, in line with previous reports (228, 229). It was therefore not possible to reliably assess the risk profile of the STEC strains based on serotype and virulence gene profile, at least not from the gene set tested in this study. It may be that STEC harbouring such virulence genes actually are virulent and therefore should be interpreted as STEC with a high risk for HUS development. Alternatively, there may be other bacterial characteristics not analysed in these strains that might be of importance for the virulence potential of STEC.

The practice to analyse all stool specimens for STEC from children <2 years and from older children and adults only when clinical suspicion of STEC infection, might have led to identification of less

virulent STEC strains not related to clinical disease in the younger age group. However, comparison of the virulence profile of STEC strains isolated from children <2 years of age compared to those isolated from older children and adults did not support this idea (Paper 2). Strains from the younger age group contained at least as many virulence genes as those isolated from older children and adults.

Although we through the characterization of a limited set of virulence genes in the STEC strains in Paper 2 did not disclose new knowledge regarding STEC virulence, our results confirm what has been reported from previous studies (64, 66, 67, 230).

5.4 Whole genome comparison

Whole genome sequencing and comparison revealed that there was considerable heterogeneity in genetic content across the 95 non-O157 STEC strains included in the study (Paper 3). Analysis of the genomic content identified 26,073 genes in total. While the 1,954 core genes were shared by all the sequenced STEC genomes, the diversity was generated by the approximately 24,000 genes in the accessory genome. Much of the accessory genome contained various mobile genetic elements, which have also previously been shown to contribute to heterogeneity and pathogenic evolution in *E. coli* (178, 187-189, 231). By PCA, a clear difference in gene profile was observed between LEE positive and LEE negative STEC, which is in line with results from Paper 2 and several previous reports (17, 18, 60, 63, 64, 67, 190, 223, 230). However, although the accessory genome was not identical within the group of LEE positive STEC strains, further PCA analysis of LEE positive strains showed no distinct subgroups due to scattering of the strains but, indicating that the accessory genes were heterogeneously distributed within this group.

Core genome phylogeny revealed that the 95 non-O157 STEC strains were distributed in all the *E. coli* phylogroups apart from phylogroup F (Paper 3). This confirms that the strains included in this study were a heterogeneous collection. However, the majority of strains belonged to phylogroup B1 (Figure 6), and most of the LEE positive strains and all the HUS-associated strains in this study also clustered in this phylogroup, in accordance with previous studies (18, 39, 172, 176, 232). In addition, LEE negative STEC associated with HUS often belong to this phylogroup (176, 190), including the O104:H4 strain (FHI102) related to the 2011 German HUS outbreak, which did however not cluster with any of the HUS and HUS-associated STEC strains included in this study (Figure 6). LEE negative and LEE positive STEC did not form separate phylogenetic clusters, but were mixed in small clusters within several phylogroups as previously reported (190). This indicates that the LEE pathogenicity island has been independently taken up by different STEC lineages at different time points. Because HUS-associated O103, O121 and O145 strains were distributed in three related clusters in the

phylogenetic analysis, both in rooted and un-rooted trees, these STEC strains were classified as HUS-group 1, although they did not belong to one well-defined cluster. The other HUS-associated strains were located in another cluster which we termed HUS-group 2. This clustering of HUS-associated strains based on variation in core genes as observed in this study, indicates that the phylogenetic backgrounds of the bacteria at least to some extent determine the pathogenic potential of the organism.

Analysis of the accessory genome revealed that certain genes were overrepresented among HUS-associated STEC, suggesting that this gene profile may be associated with the virulence potential of these strains. Regardless, none of the genes were exclusive for the HUS-associated strains as a group, which suggest that the gene content in HUS-associated STEC at least in part is shared with non-HUS STEC strains. Furthermore, in an attempt to search for unique genes in HUS-group 1 and 2, we identified several hundred genes that were significantly overrepresented in each of these groups by analysing the accessory genome, suggesting that different sets of genes may contribute to the pathogenic potential in different phylogenetic STEC lineages. However, few of these genes were found to be unique to any of the groups which further suggest that the accessory genome is shared both between and within the different clusters defined by the core genome phylogeny. The majority of strains in HUS-group 2 were of serotype O26. The fact that HUS-associated O26 strains clustered with non-HUS strains of the same serogroup, suggests that accessory factors rather than core genes defines pathogenic potential within this group. It is also possible that other than bacterial factors, like host factors or infectious dose, may be at least in part be responsible for the difference in severity of disease between persons infected with STEC of the same sero- and genotype. Regardless, it was not possible to identify any genes in the accessory genome which distinguished HUS-associated from non-HUS strains of the same serogroup in HUS-group 2.

Approximately 13,000 different variants of the core genes were found among the strains. However, despite the high number of gene variants, differences in protein sequences were identified for only 13 of these variants. Comparison of the core gene variants revealed that although 281 gene variants were overrepresented in HUS-associated STEC, several of these were also present in strains not associated with HUS. The observation that none of the identified core gene variants were unique to the HUS-associated strains is supported by the fact that HUS-associated STEC clustered in more than one group in the core gene phylogeny.

By comparing the genomes of the STEC strains which were epidemiologically linked and belonged to the same MLVA outbreak cluster in Paper 3, we identified a number of genes that were different across HUS and non-HUS strains. The fact that different genes were present in strains from the same

outbreak might indicate that the infecting source consisted of a mixture of similar but not identical STEC strains which could have evolved from the same clone. This implicates that depending on the growth conditions and environment, STEC even within the time frame of an outbreak might gain and loose genetic elements which could further affect their pathogenic potential. Regardless, based on these results we could not identify any genes that clearly distinguished between the HUS and non-HUS associated strains of these outbreak clusters.

GO enrichment analyses are used to enable functional interpretation of genes and gene products. In this study (Paper 3), GO terms related to L-idonate degradation were found to be enriched both among all 23 HUS-associated STEC collectively and the STEC in HUS-group 1. *E. coli* is able to utilize L-idonate as a sole carbohydrate source through the Entner-Doudoroff metabolic pathway, which has been shown to be important for the ability of *E. coli* to colonize mammalian intestines (233). In addition, we identified enriched GO terms for protein secretion by the type II secretion system in HUS-group 1. The type II secretion system in Gram negative bacteria promotes protein transport across the outer membrane, and the majority of proteins exported by this system contribute to bacterial adaptation and colonization by generating nutrients available for uptake (234). Furthermore, certain exported lipoproteins have been shown to be involved in biofilm formation in EPEC (235). Genes responsible for the enriched GO terms may therefore contribute to enhanced bacterial colonization and adaptation, which might have an impact on bacterial virulence in these specific strains, however further investigations are needed to confirm this. Of the 26 GO terms that were enriched in HUS group 2, a few were related to flagellar motility, which in general are recognized as virulence factors in bacteria (236). In addition, enriched GO terms were related to siderophore biosynthesis. Siderophores, being iron chelating compounds, are important for iron acquisition in bacteria (237, 238). The specific siderophore identified among strains in HUS-group 2 was encoded on a high-pathogenicity island (HPI) found in distinct clonal lineages of STEC, including serogroup O26 (239, 240). These results indicate that both motility and iron acquisition might be important factors for bacterial virulence of STEC in HUS-group 2. However, it was not possible to further assess these aspects in the study, and the precise role of these genes therefore needs to be explored in further studies. In addition, although GO annotations were available for many of the genes analysed, this information was not available for all the genes examined and further information is therefore needed.

The strains included for whole genome sequencing in the study were primarily selected to represent different MLVA genotypes (183, 184), a diversity of non-O157 STEC serotypes and patients with different severity of disease, and all non-O157 STEC strains from HUS-cases (Paper 3). These inclusion criteria would potentially give a comprehensive overview of genetic content in a diverse collection of

non-O157 STEC. However, the strains collection contained only a limited number of STEC strains from each phylogenetic lineage or serotype. Furthermore, few of the HUS and non-HUS STEC strains in the study were epidemiologically linked, making genomic comparison difficult within each serotype or phylogenetic lineage with respect to disease severity. For future studies, if more STEC strains associated with HUS were included in the genomic comparisons this would give more strength both to phylogenetic analyses and statistical analyses on gene content.

5.5 Clinical classification

As previously mentioned, there are various risk factors for severe STEC disease, and different STEC strains do not have the same potential of causing severe disease. In Norway STEC infection is a notifiable disease, where clinical information on the patients and results from laboratory analyses of the bacterial strains are stored. In this study (Paper 2 and Paper 3) we collected data from the Norwegian Surveillance System for Communicable Diseases, MSIS on clinical symptoms (HUS, bloody diarrhoea, diarrhoea or no disease), age and sex of the patients, and correlated these data with laboratory results. Based on such information, all STEC strains from HUS patients and strains epidemiologically linked to a HUS case and with the same MLVA genotype as the STEC isolate from that case, were classified as HUS-associated. This was done to group strains with proven potential to cause severe disease, and compare these with non-HUS STEC. However, it might be difficult based on available information to classify correctly all STEC strains with respect to virulence potential. Some strains which have the potential to cause severe disease based on gene content, may not have caused severe disease in the persons infected due to host factors like high age, a low infectious dose or other unknown factors. Consequently such strains may not have been classified as HUS-associated in our study. This could explain the results seen especially in Paper 2, where non-HUS STEC strains of the classical STEC serogroups contained a high number of the potential virulence genes, without being associated with severe disease. It may be that such strains are actually virulent and should be interpreted as STEC with a high risk for HUS development. Thus, these aspects might potentially disturb the search for genes which could be associated with STEC virulence. In Paper 3 we therefore performed PCA of the sequenced strains as an alternative analysis without the use of clinical classification. In this analysis STEC strains would cluster based solely on gene content, and the analysis would therefore not be biased due to potential misclassification based on clinical data. Although the results from PCA of the accessory genome separated LEE positive from LEE negative strains, further PCA of LEE positive strains showed scattering of the strains without any distinct cluster which could represent highly virulent strains. Thus, the results indicate that the accessory genes of LEE positive STEC are heterogeneously distributed within this group, and it is likely that different sets of genes may contribute to the pathogenic potential in different STEC.

6 Main conclusions

- STEC infection was most common in children <5 years old and all HUS cases were among children ≤5 years old.
- All STEC strains from HUS patients were, apart from the STEC-LST strains, positive for the *stx2a* subtype as well as the LEE pathogenicity island. In addition they belonged to *E. coli* serogroups frequently associated with STEC disease.
- In SF O157, the anti-terminator *q* gene involved in regulation of *stx2* expression, as well as other genes upstream of *stx2a*, were different compared to the ones seen in NSF O157 STEC. These findings suggest divergent *stx2* encoding bacteriophages between SF and NSF O157 STEC.
- STEC causing human disease constituted a genetic and phylogenetic highly diverse nature.
- A clear difference in gene content was observed between LEE positive and LEE negative STEC.
- Genomic comparison revealed that the gene content in part was shared between HUS-associated and non-HUS STEC, and consequently it was not possible to reliably distinguish STEC with potential to cause HUS from those without such potential.

7 Future aspects

In this study the main aim was to identify genetic characteristics which could distinguish between highly pathogenic STEC with the potential to cause HUS, and STEC with a lower pathogenic potential. While the results showed that several potential virulence genes in LEE positive STEC were significantly associated with severe STEC disease, none of the genes were unique for the group of HUS-associated STEC. These genes could therefore not reliably differentiate between STEC with and without potential for causing severe disease. However, factors other than gene composition might cause different phenotype and virulence properties of bacteria. In an ongoing postdoctoral project in our group, the aim is to perform RNA sequencing to compare whether there are differences in transcribed genes in a selected material of HUS and non-HUS non-O157 STEC strains. Furthermore in this project, transcription factors are investigated. As a next step in an attempt to elucidate virulence factors in STEC, it is essential to characterize other factors which might be implicated in STEC gene expression and regulation, and which might have an impact on STEC pathogenicity, e.g. small regulatory RNAs. In addition, it will be of importance to study the differences between HUS and non-HUS STEC at the proteomic and phenotypic level. By implementing a combination of genomics, transcriptomics and proteomics analyses, it may be possible to get a comprehensive overview of factors that might differentiate between STEC with the potential to cause HUS, and STEC with a lower pathogenic potential.

References

1. **Topley WWC, Wilson GS, Balows A, Sussman M.** 1998. Topley & Wilson's microbiology and microbial infections, 9th ed, vol. 2. Arnold; Oxford University Press, London, New York.

2. **Huys G, Cnockaert M, Janda JM, Swings J.** 2003. Escherichia albertii sp. nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. Int. J. Syst. Evol. Micr. **53:**807-810.

3. **Scheutz F, Cheasty T, Woodward D, Smith HR.** 2004. Designation of O174 and O175 to temporary O groups OX3 and OX7, and six new E. coli O groups that include Verocytotoxin-producing E. coli (VTEC): O176, O177, O178, O179, O180 and O181. APMIS **112:**569-584.

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

5. **Gaastra W, de Graaf FK.** 1982. Host-specific fimbrial adhesins of noninvasive enterotoxigenic Escherichia coli strains. Microbiol. Rev. **46:**129-161.

6. **Hartl DL, Dykhuizen DE.** 1984. The population genetics of Escherichia coli. Annu. Rev. Genet. **18:**31-68.

7. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic Escherichia coli. Nat. Rev. Microbiol. **2:**123-140.

8. **Newell DG, Koopmans M, Verhoef L, Duizer E, Aidara-Kane A, Sprong H, Opsteegh M, Langelaar M, Threfall J, Scheutz F, van der Giessen J, Kruse H.** 2010. Food-borne diseases - the challenges of 20 years ago still persist while new ones continue to emerge. Int. J. Food Microbiol. **139 Suppl 1:**S3-15.

9. **Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB.** 2013. Recent Advances in Understanding Enteric Pathogenic Escherichia coli. Clin. Microbiol. Rev. **26:**822-880.

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

11. **Nataro JP.** 2005. Enteroaggregative Escherichia coli pathogenesis. Curr. Opin. Gastroen. **21:**4-8.

12. **Croxen MA, Finlay BB.** 2010. Molecular mechanisms of Escherichia coli pathogenicity. Nat. Rev. Microbiol. **8:**26-38.

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

14. **Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML.** 1983. Hemorrhagic colitis associated with a rare Escherichia coli serotype. N. Engl. J. Med. **308:**681-685.

15. **Johnson WM, Lior H, Bezanson GS.** 1983. Cytotoxic Escherichia coli O157:H7 associated with haemorrhagic colitis in Canada. Lancet **1:**76.

16. **Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, Bernard H, Fruth A, Prager R, Spode A, Wadl M, Zoufaly A, Jordan S, Kemper MJ, Follin P, Muller L, King LA, Rosner B, Buchholz U, Stark K, Krause G, Team HUSI.** 2011. Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. N. Engl. J. Med. **365:**1771-1780.

17. **Delannoy S, Beutin L, Fach P.** 2013. Discrimination of Enterohemorrhagic Escherichia coli (EHEC) from Non-EHEC Strains Based on Detection of Various Combinations of Type III Effector Genes. J. Clin. Microbiol. **51:**3257-3262.

18. **Hazen TH, Sahl JW, Fraser CM, Donnenberg MS, Scheutz F, Rasko DA.** 2013. Refining the pathovar paradigm via phylogenomics of the attaching and effacing Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. **110:**12810-12815.

19. **Schuller S.** 2011. Shiga toxin interaction with human intestinal epithelium. Toxins **3:**626-639.

20. **Centers for Disease Control and Prevention (CDC).** 2012. National Shiga toxin-producing Escherichia coli (STEC) Surveillance Annual Summary, 2009. Atlanta, Georgia: US Department of Health and Human Services, CDC, 2012. CDC.

21. **European Centre for Disease Prevention and Control.** 2013. Annual Epidemiological Report 2013 Q-AC-14-001-EN-C. European Centre for Disease Prevention and Control.

22. **Tarr PI, Gordon CA, Chandler WL.** 2005. Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. Lancet **365:**1073-1086.

23. **Karmali MA.** 2004. Infection by Shiga toxin-producing Escherichia coli: an overview. Mol. Biotechnol. **26:**117-122.

24. **Beutin L, Martin A.** 2012. Outbreak of Shiga toxin-producing Escherichia coli (STEC) O104:H4 infection in Germany causes a paradigm shift with regard to human pathogenicity of STEC strains. J. Food Prot. **75:**408-418.

25. **Schimmer B, Nygard K, Eriksen HM, Lassen J, Lindstedt BA, Brandal LT, Kapperud G, Aavitsland P.** 2008. Outbreak of haemolytic uraemic syndrome in Norway caused by stx2-positive Escherichia coli O103:H25 traced to cured mutton sausages. BMC Infect. Dis. **8:**41.

26. **Allison HE.** 2007. Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. Future microbiology **2:**165-174.

27. **Unkmeir A, Schmidt H.** 2000. Structural analysis of phage-borne stx genes and their flanking sequences in Shiga toxin-producing Escherichia coli and Shigella dysenteriae type 1 strains. Infect. Immun. **68:**4856-4864.

28. **Schmidt H.** 2001. Shiga-toxin-converting bacteriophages. Res. Microbiol. **152:**687-695.

29. **Smith DL, Rooks DJ, Fogg PC, Darby AC, Thomson NR, McCarthy AJ, Allison HE.** 2012. Comparative genomics of Shiga toxin encoding bacteriophages. BMC Genomics **13:**311.

30. **Wagner PL, Neely MN, Zhang X, Acheson DW, Waldor MK, Friedman DI.** 2001. Role for a phage promoter in Shiga toxin 2 expression from a pathogenic Escherichia coli strain. J. Bacteriol. **183:**2081-2085.

31. **Jacobsen A.** 2012. Variation in antiterminator Q in verocytotoxin phages. Master's Thesis. Technical University of Denmark.

32. **Koitabashi T, Vuddhakul V, Radu S, Morigaki T, Asai N, Nakaguchi Y, Nishibuchi M.** 2006. Genetic characterization of Escherichia coli O157: H7/- strains carrying the stx2 gene but not producing Shiga toxin 2. Microbiol. Immunol. **50:**135-148.

33. **Lejeune JT, Abedon ST, Takemura K, Christie NP, Sreevatsan S.** 2004. Human Escherichia coli O157:H7 genetic marker in isolates of bovine origin. Emerg. Infect. Dis. **10:**1482-1485.

34. **Matsumoto M, Suzuki M, Takahashi M, Hirose K, Minagawa H, Ohta M.** 2008. Identification and epidemiological description of enterohemorrhagic Escherichia coli O157 strains producing low amounts of Shiga toxin 2 in Aichi Prefecture, Japan. Jpn. J. Infec. Dis. **61:**442-445.

35. **Bauwens A, Betz J, Meisen I, Kemper B, Karch H, Muthing J.** 2013. Facing glycosphingolipid-Shiga toxin interaction: dire straits for endothelial cells of the human vasculature. Cell. Mol. Life Sci. **70:**425-457.

36. **Sandvig K, Bergan J, Dyve AB, Skotland T, Torgersen ML.** 2010. Endocytosis and retrograde transport of Shiga toxin. Toxicon **56:**1181-1185.

37. **Khan F, Proulx F, Lingwood CA.** 2009. Detergent-resistant globotriaosyl ceramide may define verotoxin/glomeruli-restricted hemolytic uremic syndrome pathology. Kidney Int. **75:**1209-1216.

38. **Scheutz F, Teel LD, Beutin L, Pierard D, Buvens G, Karch H, Mellmann A, Caprioli A, Tozzoli R, Morabito S, Strockbine NA, Melton-Celsa AR, Sanchez M, Persson S, O'Brien AD.** 2012. Multicenter evaluation of a sequence-based protocol for subtyping shiga toxins and standardizing stx nomenclature. J. Clin. Microbiol. **50:**2951-2963.

39. **Mellmann A, Bielaszewska M, Kock R, Friedrich AW, Fruth A, Middendorf B, Harmsen D, Schmidt MA, Karch H.** 2008. Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic Escherichia coli. Emerg. Infect. Dis. **14:**1287-1290.

40. **Mellmann A, Lu S, Karch H, Xu JG, Harmsen D, Schmidt MA, Bielaszewska M.** 2008. Recycling of Shiga toxin 2 genes in sorbitol-fermenting enterohemorrhagic Escherichia coli O157:NM. Appl. Environ. Microbiol. **74:**67-72.

41. **Bielaszewska M, Friedrich AW, Aldick T, Schurk-Bulgrin R, Karch H.** 2006. Shiga toxin activatable by intestinal mucus in Escherichia coli isolated from humans: predictor for a severe clinical outcome. Clin. Infect. Dis. **43:**1160-1167.

42. **Friedrich AW, Bielaszewska M, Zhang WL, Pulz M, Kuczius T, Ammon A, Karch H.** 2002. Escherichia coli harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. J. Infect. Dis. **185:**74-84.

43. **Persson S, Olsen KE, Ethelberg S, Scheutz F.** 2007. Subtyping method for Escherichia coli shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. J. Clin. Microbiol. **45:**2020-2024.

44. **Fuller CA, Pellino CA, Flagler MJ, Strasser JE, Weiss AA.** 2011. Shiga toxin subtypes display dramatic differences in potency. Infect. Immun. **79:**1329-1337.

45. **Bielaszewska M, Middendorf B, Kock R, Friedrich AW, Fruth A, Karch H, Schmidt MA, Mellmann A.** 2008. Shiga toxin-negative attaching and effacing Escherichia coli: distinct clinical associations with bacterial phylogeny and virulence traits and inferred in-host pathogen evolution. Clin. Infect. Dis. **47:**208-217.

46. **Mellmann A, Bielaszewska M, Karch H.** 2009. Intrahost genome alterations in enterohemorrhagic Escherichia coli. Gastroenterology **136:**1925-1938.

47. **Mcdaniel TK, Jarvis KG, Donnenberg MS, Kaper JB.** 1995. A Genetic-Locus of Enterocyte Effacement Conserved among Diverse Enterobacterial Pathogens. Proc. Natl. Acad. Sci. U. S. A. **92:**1664-1668.

48. **Garmendia J, Frankel G, Crepin VF.** 2005. Enteropathogenic and enterohemorrhagic Escherichia coli infections: translocation, translocation, translocation. Infect. Immun. **73:**2573-2585.

49. **Schmidt MA.** 2010. LEEways: tales of EPEC, ATEC and EHEC. Cellular microbiology **12:**1544-1552.

50. **Muller D, Benz I, Liebchen A, Gallitz I, Karch H, Schmidt MA.** 2009. Comparative analysis of the locus of enterocyte effacement and its flanking regions. Infect. Immun. **77:**3501-3513.

51. **Frankel G, Phillips AD.** 2008. Attaching effacing Escherichia coli and paradigms of Tir-triggered actin polymerization: getting off the pedestal. Cellular microbiology **10:**549-556.

52. **Lai Y, Rosenshine I, Leong JM, Frankel G.** 2013. Intimate host attachment: enteropathogenic and enterohaemorrhagic Escherichia coli. Cellular microbiology **15:**1796-1808.

53. **Mora A, Blanco M, Yamamoto D, Dahbi G, Blanco JE, Lopez C, Alonso MP, Vieira MA, Hernandes RT, Abe CM, Piazza RM, Lacher DW, Elias WP, Gomes TA, Blanco J.** 2009. HeLa-cell adherence patterns and actin aggregation of enteropathogenic Escherichia coli (EPEC) and Shiga-toxin-producing E. coli (STEC) strains carrying different eae and tir alleles. International microbiology: the official journal of the Spanish Society for Microbiology **12:**243-251.

54. **Adu-Bobie J, Frankel G, Bain C, Goncalves AG, Trabulsi LR, Douce G, Knutton S, Dougan G.** 1998. Detection of intimins alpha, beta, gamma, and delta, four intimin derivatives expressed by attaching and effacing microbial pathogens. J. Clin. Microbiol. **36:**662-668.

55. **China B, Goffaux F, Pirson V, Mainil J.** 1999. Comparison of eae, tir, espA and espB genes of bovine and human attaching and effacing Escherichia coli by multiplex polymerase chain reaction. FEMS Microbiol. Lett. **178:**177-182.

56. **Garrido P, Blanco M, Moreno-Paz M, Briones C, Dahbi G, Blanco J, Parro V.** 2006. STEC-EPEC oligonucleotide microarray: a new tool for typing genetic variants of the LEE pathogenicity island of human and animal Shiga toxin-producing Escherichia coli (STEC) and enteropathogenic E. coli (EPEC) strains. Clin. Chem. **52:**192-201.

57. **Afset JE, Anderssen E, Bruant G, Harel J, Wieler L, Bergh K.** 2008. Phylogenetic backgrounds and virulence profiles of atypical enteropathogenic Escherichia coli strains from a case-control study using multilocus sequence typing and DNA microarray analysis. J. Clin. Microbiol. **46:**2280-2290.

58. **Nielsen EM, Andersen MT.** 2003. Detection and characterization of verocytotoxin-producing Escherichia coli by automated 5' nuclease PCR assay. J. Clin. Microbiol. **41:**2884-2893.

59. **Farfan MJ, Torres AG.** 2012. Molecular mechanisms that mediate colonization of Shiga toxin-producing Escherichia coli strains. Infect. Immun. **80:**903-913.

60. **Tobe T, Beatson SA, Taniguchi H, Abe H, Bailey CM, Fivian A, Younis R, Matthews S, Marches O, Frankel G, Hayashi T, Pallen MJ.** 2006. An extensive repertoire of type III secretion effectors in Escherichia coli O157 and the role of lambdoid phages in their dissemination. Proc. Natl. Acad. Sci. U. S. A. **103:**14941-14946.

61. **Melton-Celsa A, Mohawk K, Teel L, O'Brien A.** 2012. Pathogenesis of Shiga-toxin producing escherichia coli. Curr. Top. Microbiol. Immunol. **357:**67-103.

62. **Wong AR, Pearson JS, Bright MD, Munera D, Robinson KS, Lee SF, Frankel G, Hartland EL.** 2011. Enteropathogenic and enterohaemorrhagic Escherichia coli: even more subversive elements. Mol. Microbiol. **80:**1420-1438.

63. **Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Reid-Smith R, Isaac-Renton J, Clark C, Rahn K, Kaper JB.** 2003. Association of genomic O island 122 of Escherichia coli EDL 933 with verocytotoxin-producing Escherichia coli seropathotypes that are linked to epidemic and/or serious disease. J. Clin. Microbiol. **41:**4930-4940.

64. **Coombes BK, Wickham ME, Mascarenhas M, Gruenheid S, Finlay BB, Karmali MA.** 2008. Molecular analysis as an aid to assess the public health risk of non-O157 Shiga toxin-producing Escherichia coli strains. Appl. Environ. Microbiol. **74:**2153-2160.

65. **Wickham ME, Lupp C, Mascarenhas M, Vazquez A, Coombes BK, Brown NF, Coburn BA, Deng W, Puente JL, Karmali MA, Finlay BB.** 2006. Bacterial genetic determinants of non-O157 STEC outbreaks and hemolytic-uremic syndrome after infection. J. Infect. Dis. **194:**819-827.

66. **Bugarel M, Beutin L, Fach P.** 2010. Low-density macroarray targeting non-locus of enterocyte effacement effectors (nle genes) and major virulence factors of Shiga toxin-producing Escherichia coli (STEC): a new approach for molecular risk assessment of STEC isolates. Appl. Environ. Microbiol. **76:**203-211.

67. **Bugarel M, Beutin L, Martin A, Gill A, Fach P.** 2010. Micro-array for the identification of Shiga toxin-producing Escherichia coli (STEC) seropathotypes associated with Hemorrhagic Colitis and Hemolytic Uremic Syndrome in humans. Int. J. Food Microbiol. **142:**318-329.

68. **Torres AG, Giron JA, Perna NT, Burland V, Blattner FR, Avelino-Flores F, Kaper JB.** 2002. Identification and characterization of lpfABCC'DE, a fimbrial operon of enterohemorrhagic Escherichia coli O157:H7. Infect. Immun. **70:**5416-5427.

69. **Ledesma MA, Ochoa SA, Cruz A, Rocha-Ramirez LM, Mas-Oliva J, Eslava CA, Giron JA, Xicohtencatl-Cortes J.** 2010. The hemorrhagic coli pilus (HCP) of Escherichia coli O157:H7 is an inducer of proinflammatory cytokine secretion in intestinal epithelial cells. PLoS ONE **5:**0012127.

70. **Xicohtencatl-Cortes J, Monteiro-Neto V, Ledesma MA, Jordan DM, Francetic O, Kaper JB, Puente JL, Giron JA.** 2007. Intestinal adherence associated with type IV pili of enterohemorrhagic Escherichia coli O157:H7. J. Clin. Invest. **117:**3519-3529.

71. **Badea L, Doughty S, Nicholls L, Sloan J, Robins-Browne RM, Hartland EL.** 2003. Contribution of Efa1/LifA to the adherence of enteropathogenic Escherichia coli to epithelial cells. Microb. Pathog. **34:**205-215.

72. **Deacon V, Dziva F, van Diemen PM, Frankel G, Stevens MP.** 2010. Efa-1/LifA mediates intestinal colonization of calves by enterohaemorrhagic Escherichia coli O26 : H- in a manner independent of glycosyltransferase and cysteine protease motifs or effects on type III secretion. Microbiology **156:**2527-2536.

73. **Klapproth JM.** 2010. The role of lymphostatin/EHEC factor for adherence-1 in the pathogenesis of gram negative infection. Toxins **2:**954-962.

74. **Weiss A, Brockmeyer J.** 2013. Prevalence, biogenesis, and functionality of the serine protease autotransporter EspP. Toxins **5:**25-48.

75. **Tatsuno I, Horie M, Abe H, Miki T, Makino K, Shinagawa H, Taguchi H, Kamiya S, Hayashi T, Sasakawa C.** 2001. toxB gene on pO157 of enterohemorrhagic Escherichia coli O157:H7 is required for full epithelial cell adherence phenotype. Infect. Immun. **69:**6660-6669.

76. **Tarr PI, Bilge SS, Vary JC, Jr., Jelacic S, Habeeb RL, Ward TR, Baylor MR, Besser TE.** 2000. Iha: a novel Escherichia coli O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure. Infect. Immun. **68:**1400-1407.

77. **Wells TJ, McNeilly TN, Totsika M, Mahajan A, Gally DL, Schembri MA.** 2009. The Escherichia coli O157:H7 EhaB autotransporter protein binds to laminin and collagen I and induces a serum IgA response in O157:H7 challenged cattle. Enviro. Microbiol. **11:**1803-1814.

78. **Paton AW, Srimanote P, Woodrow MC, Paton JC.** 2001. Characterization of Saa, a novel autoagglutinating adhesin produced by locus of enterocyte effacement-negative shiga-toxigenic Escherichia coli strains that are virulent for humans. Infect. Immun. **69:**6999-7009.

79. **Toma C, Nakasone N, Miliwebsky E, Higa N, Rivas M, Suzuki T.** 2008. Differential adherence of Shiga toxin-producing Escherichia coli harboring saa to epithelial cells. Int. J. Med. Microbiol. **298:**571-578.

80. **Brunder W, Khan AS, Hacker J, Karch H.** 2001. Novel type of fimbriae encoded by the large plasmid of sorbitol-fermenting enterohemorrhagic Escherichia coli O157:H(-). Infect. Immun. **69:**4447-4457.

81. **Musken A, Bielaszewska M, Greune L, Schweppe CH, Muthing J, Schmidt H, Schmidt MA, Karch H, Zhang W.** 2008. Anaerobic conditions promote expression of Sfp fimbriae and adherence of sorbitol-fermenting enterohemorrhagic Escherichia coli O157:NM to human intestinal epithelial cells. Appl. Environ. Microbiol. **74:**1087-1093.

82. **Karch H, Heesemann J, Laufs R, O'Brien AD, Tacket CO, Levine MM.** 1987. A plasmid of enterohemorrhagic Escherichia coli O157:H7 is required for expression of a new fimbrial antigen and for adhesion to epithelial cells. Infect. Immun. **55:**455-461.

83. **Lim JY, Yoon J, Hovde CJ.** 2010. A brief overview of Escherichia coli O157:H7 and its plasmid O157. J. Microbiol. Biotechnol. **20:**5-14.

84. **Bauer ME, Welch RA.** 1996. Characterization of an RTX toxin from enterohemorrhagic Escherichia coli O157:H7. Infect. Immun. **64:**167-175.

85. **Lathem WW, Grys TE, Witowski SE, Torres AG, Kaper JB, Tarr PI, Welch RA.** 2002. StcE, a metalloprotease secreted by Escherichia coli O157:H7, specifically cleaves C1 esterase inhibitor. Mol. Microbiol. **45:**277-288.

86. **Brunder W, Schmidt H, Karch H.** 1996. KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic Escherichia coli O157:H7. Microbiology **142:**3305-3315.

87. **Schmidt H, Henkel B, Karch H.** 1997. A gene cluster closely related to type II secretion pathway operons of gram-negative bacteria is located on the large plasmid of enterohemorrhagic Escherichia coli O157 strains. FEMS Microbiol. Lett. **148:**265-272.

88. **Brunder W, Schmidt H, Karch H.** 1997. EspP, a novel extracellular serine protease of enterohaemorrhagic Escherichia coli O157:H7 cleaves human coagulation factor V. Mol. Microbiol. **24:**767-778.

89. **Murase K, Ooka T, Iguchi A, Ogura Y, Nakayama K, Asadulghani M, Islam MR, Hiyoshi H, Kodama T, Beutin L, Hayashi T.** 2012. Haemolysin E- and enterohaemolysin-derived haemolytic activity of O55/O157 strains and other Escherichia coli lineages. Microbiology **158:**746-758.

90. **Taneike I, Zhang HM, Wakisaka-Saito N, Yamamoto T.** 2002. Enterohemolysin operon of Shiga toxin-producing Escherichia coli: a virulence function of inflammatory cytokine production from human monocytes. FEBS Lett. **524:**219-224.

91. **Grys TE, Siegel MB, Lathem WW, Welch RA.** 2005. The StcE protease contributes to intimate adherence of enterohemorrhagic Escherichia coli O157:H7 to host cells. Infect. Immun. **73:**1295-1303.

92. **Yu AC, Worrall LJ, Strynadka NC.** 2012. Structural insight into the bacterial mucinase StcE essential to adhesion and immune evasion during enterohemorrhagic E. coli infection. Structure **20:**707-717.

93. **Uhlich GA.** 2009. KatP contributes to OxyR-regulated hydrogen peroxide resistance in Escherichia coli serotype O157 : H7. Microbiology **155:**3589-3598.

94. **Leyton DL, Sloan J, Hill RE, Doughty S, Hartland EL.** 2003. Transfer region of pO113 from enterohemorrhagic Escherichia coli: similarity with R64 and identification of a novel plasmid-encoded autotransporter, EpeA. Infect. Immun. **71:**6307-6319.

95. **Abu-Median AB, van Diemen PM, Dziva F, Vlisidou I, Wallis TS, Stevens MP.** 2006. Functional analysis of lymphostatin homologues in enterohaemorrhagic Escherichia coli. FEMS Microbiol. Lett. **258:**43-49.

96. **Savarino SJ, McVeigh A, Watson J, Cravioto A, Molina J, Echeverria P, Bhan MK, Levine MM, Fasano A.** 1996. Enteroaggregative Escherichia coli heat-stable enterotoxin is not restricted to enteroaggregative E. coli. J. Infect. Dis. **173:**1019-1022.

97. **Janka A, Bielaszewska M, Dobrindt U, Greune L, Schmidt MA, Karch H.** 2003. Cytolethal distending toxin gene cluster in enterohemorrhagic Escherichia coli O157:H- and O157:H7: characterization and evolutionary considerations. Infect. Immun. **71:**3634-3638.

98. **Bielaszewska M, Stoewe F, Fruth A, Zhang W, Prager R, Brockmeyer J, Mellmann A, Karch H, Friedrich AW.** 2009. Shiga toxin, cytolethal distending toxin, and hemolysin repertoires in clinical Escherichia coli O91 isolates. J. Clin. Microbiol. **47:**2061-2066.

99. **Cui J, Yao Q, Li S, Ding X, Lu Q, Mao H, Liu L, Zheng N, Chen S, Shao F.** 2010. Glutamine deamidation and dysfunction of ubiquitin/NEDD8 induced by a bacterial effector family. Science **329:**1215-1218.

100. **Jubelin G, Taieb F, Duda DM, Hsu Y, Samba-Louaka A, Nobe R, Penary M, Watrin C, Nougayrede JP, Schulman BA, Stebbins CE, Oswald E.** 2010. Pathogenic bacteria target NEDD8-conjugated cullins to hijack host-cell signaling pathways. PLoS pathogens **6:**1001128.

101. **Marches O, Ledger TN, Boury M, Ohara M, Tu X, Goffaux F, Mainil J, Rosenshine I, Sugai M, De Rycke J, Oswald E.** 2003. Enteropathogenic and enterohaemorrhagic Escherichia coli deliver a novel effector called Cif, which blocks cell cycle G2/M transition. Mol. Microbiol. **50:**1553-1567.

102. **Samba-Louaka A, Nougayrede JP, Watrin C, Oswald E, Taieb F.** 2009. The enteropathogenic Escherichia coli effector Cif induces delayed apoptosis in epithelial cells. Infect. Immun. **77:**5471-5477.

103. **Jerse AE, Yu J, Tall BD, Kaper JB.** 1990. A genetic locus of enteropathogenic Escherichia coli necessary for the production of attaching and effacing lesions on tissue culture cells. Proc. Natl. Acad. Sci. U. S. A. **87:**7839-7843.

104. **Nicholls L, Grant TH, Robins-Browne RM.** 2000. Identification of a novel genetic locus that is required for in vitro adhesion of a clinical isolate of enterohaemorrhagic Escherichia coli to epithelial cells. Mol. Microbiol. **35:**275-288.

105. **Knutton S, Rosenshine I, Pallen MJ, Nisan I, Neves BC, Bain C, Wolff C, Dougan G, Frankel G.** 1998. A novel EspA-associated surface organelle of enteropathogenic Escherichia coli involved in protein translocation into epithelial cells. EMBO J. **17:**2166-2176.

106. **Wells TJ, Sherlock O, Rivas L, Mahajan A, Beatson SA, Torpdahl M, Webb RI, Allsopp LP, Gobius KS, Gally DL, Schembri MA.** 2008. EhaA is a novel autotransporter protein of enterohemorrhagic Escherichia coli O157:H7 that contributes to adhesion and biofilm formation. Environ. Microbiol. **10:**589-604.

107. **Miyahara A, Nakanishi N, Ooka T, Hayashi T, Sugimoto N, Tobe T.** 2009. Enterohemorrhagic Escherichia coli effector EspL2 induces actin microfilament aggregation through annexin 2 activation. Cellular microbiology **11:**337-350.

108. **Tobe T.** 2010. Cytoskeleton-modulating effectors of enteropathogenic and enterohemorrhagic Escherichia coli: role of EspL2 in adherence and an alternative pathway for modulating cytoskeleton through Annexin A2 function. Febs J **277:**2403-2408.

109. **Ogura Y, Ooka T, Asadulghani, Terajima J, Nougayrede JP, Kurokawa K, Tashiro K, Tobe T, Nakayama K, Kuhara S, Oswald E, Watanabe H, Hayashi T.** 2007. Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic Escherichia coli strains of O157 and non-O157 serotypes. Genome Biol. **8:**R138.

110. **Kenny B, Ellis S, Leard AD, Warawa J, Mellor H, Jepson MA.** 2002. Co-ordinate regulation of distinct host cell signalling pathways by multifunctional enteropathogenic Escherichia coli effector molecules. Mol. Microbiol. **44:**1095-1107.

111. **Dean P, Kenny B.** 2009. The effector repertoire of enteropathogenic E. coli: ganging up on the host cell. Curr. Opin. Microbiol. **12:**101-109.

112. **McNamara BP, Donnenberg MS.** 1998. A novel proline-rich protein, EspF, is secreted from enteropathogenic Escherichia coli via the type III export pathway. FEMS Microbiol. Lett. **166:**71-78.

113. **McNamara BP, Koutsouris A, O'Connell CB, Nougayrede JP, Donnenberg MS, Hecht G.** 2001. Translocated EspF protein from enteropathogenic Escherichia coli disrupts host intestinal barrier function. J. Clin. Invest. **107:**621-629.

114. **Zhao S, Zhou Y, Wang C, Yang Y, Wu X, Wei Y, Zhu L, Zhao W, Zhang Q, Wan C.** 2013. The N-terminal domain of EspF induces host cell apoptosis after infection with enterohaemorrhagic Escherichia coli O157:H7. PLoS One **8:**e55164.

115. **Elliott SJ, Krejany EO, Mellies JL, Robins-Browne RM, Sasakawa C, Kaper JB.** 2001. EspG, a novel type III system-secreted protein from enteropathogenic Escherichia coli with similarities to VirA of Shigella flexneri. Infect. Immun. **69:**4027-4033.

116. **Glotfelty LG, Hecht GA.** 2012. Enteropathogenic E. coli effectors EspG1/G2 disrupt tight junctions: new roles and mechanisms. Ann. N. Y. Acad. Sci. **1258:**149-158.

117. **Wong AR, Raymond B, Collins JW, Crepin VF, Frankel G.** 2012. The enteropathogenic E. coli effector EspH promotes actin pedestal formation and elongation via WASP-interacting protein (WIP). Cellular microbiology **14:**1051-1070.

118. **Tu X, Nisan I, Yona C, Hanski E, Rosenshine I.** 2003. EspH, a new cytoskeleton-modulating effector of enterohaemorrhagic and enteropathogenic Escherichia coli. Mol. Microbiol. **47:**595-606.

119. **Dahan S, Wiles S, La Ragione RM, Best A, Woodward MJ, Stevens MP, Shaw RK, Chong Y, Knutton S, Phillips A, Frankel G.** 2005. EspJ is a prophage-carried type III effector protein of attaching and effacing pathogens that modulates infection dynamics. Infect. Immun. **73:**679-686.

120. **Kurushima J, Nagai T, Nagamatsu K, Abe A.** 2010. EspJ effector in enterohemorrhagic E. coli translocates into host mitochondria via an atypical mitochondrial targeting signal. Microbiol. Immunol. **54:**371-379.

121. **Marches O, Covarelli V, Dahan S, Cougoule C, Bhatta P, Frankel G, Caron E.** 2008. EspJ of enteropathogenic and enterohaemorrhagic Escherichia coli inhibits opsono-phagocytosis. Cellular microbiology **10:**1104-1115.

122. **Vlisidou I, Marches O, Dziva F, Mundy R, Frankel G, Stevens MP.** 2006. Identification and characterization of EspK, a type III secreted effector protein of enterohaemorrhagic Escherichia coli O157:H7. FEMS Microbiol. Lett. **263:**32-40.

123. **Roxas JL, Wilbur JS, Zhang X, Martinez G, Vedantam G, Viswanathan VK.** 2012. The enteropathogenic Escherichia coli-secreted protein EspZ inhibits host cell apoptosis. Infect. Immun. **80:**3850-3857.

124. **Berger CN, Crepin VF, Baruch K, Mousnier A, Rosenshine I, Frankel G.** 2012. EspZ of enteropathogenic and enterohemorrhagic Escherichia coli regulates type III secretion system protein translocation. mBio **3**(5): e00317-12.

125. **Kanack KJ, Crawford JA, Tatsuno I, Karmali MA, Kaper JB.** 2005. SepZ/EspZ is secreted and translocated into HeLa cells by the enteropathogenic Escherichia coli type III secretion system. Infect. Immun. **73:**4327-4337.

126. **Uhlich GA, Chen CY, Cottrell BJ, Irwin PL, Phillips JG.** 2012. Peroxide resistance in Escherichia coli serotype O157 : H7 biofilms is regulated by both RpoS-dependent and -independent mechanisms. Microbiology **158:**2225-2234.

127. **Gruenheid S, Sekirov I, Thomas NA, Deng W, O'Donnell P, Goode D, Li Y, Frey EA, Brown NF, Metalnikov P, Pawson T, Ashman K, Finlay BB.** 2004. Identification and characterization of NleA, a non-LEE-encoded type III translocated virulence factor of enterohaemorrhagic Escherichia coli O157:H7. Mol. Microbiol. **51:**1233-1249.

128. **Kim J, Thanabalasuriar A, Chaworth-Musters T, Fromme JC, Frey EA, Lario PI, Metalnikov P, Rizg K, Thomas NA, Lee SF, Hartland EL, Hardwidge PR, Pawson T, Strynadka NC, Finlay BB, Schekman R, Gruenheid S.** 2007. The bacterial virulence factor NleA inhibits cellular protein secretion by disrupting mammalian COPII function. Cell host & microbe **2:**160-171.

129. **Thanabalasuriar A, Koutsouris A, Weflen A, Mimee M, Hecht G, Gruenheid S.** 2010. The bacterial virulence factor NleA is required for the disruption of intestinal tight junctions by enteropathogenic Escherichia coli. Cellular microbiology **12:**31-41.

130. **Deng W, Puente JL, Gruenheid S, Li Y, Vallance BA, Vazquez A, Barba J, Ibarra JA, O'Donnell P, Metalnikov P, Ashman K, Lee S, Goode D, Pawson T, Finlay BB.** 2004. Dissecting virulence: systematic and functional analyses of a pathogenicity island. Proc. Natl. Acad. Sci. U. S. A. **101:**3597-3602.

131. **Nadler C, Baruch K, Kobi S, Mills E, Haviv G, Farago M, Alkalay I, Bartfeld S, Meyer TF, Ben-Neriah Y, Rosenshine I.** 2010. The type III secretion effector NleE inhibits NF-kappaB activation. PLoS pathogens **6:**e1000743.

132. **Newton HJ, Pearson JS, Badea L, Kelly M, Lucas M, Holloway G, Wagstaff KM, Dunstone MA, Sloan J, Whisstock JC, Kaper JB, Robins-Browne RM, Jans DA, Frankel G, Phillips AD, Coulson BS, Hartland EL.** 2010. The type III effectors NleE and NleB from enteropathogenic E. coli and OspZ from Shigella block nuclear translocation of NF-kappaB p65. PLoS pathogens **6:**e1000898.

133. **Gao X, Wang X, Pham TH, Feuerbacher LA, Lubos ML, Huang M, Olsen R, Mushegian A, Slawson C, Hardwidge PR.** 2013. NleB, a bacterial effector with glycosyltransferase activity, targets GAPDH function to inhibit NF-kappaB activation. Cell host & microbe **13:**87-99.

134. **Yen H, Ooka T, Iguchi A, Hayashi T, Sugimoto N, Tobe T.** 2010. NleC, a type III secretion protease, compromises NF-kappaB activation by targeting p65/RelA. PLoS pathogens **6:**e1001231.

135. **Baruch K, Gur-Arie L, Nadler C, Koby S, Yerushalmi G, Ben-Neriah Y, Yogev O, Shaulian E, Guttman C, Zarivach R, Rosenshine I.** 2011. Metalloprotease type III effectors that specifically cleave JNK and NF-kappaB. EMBO J. **30:**221-231.

136. **Blasche S, Mortl M, Steuber H, Siszler G, Nisa S, Schwarz F, Lavrik I, Gronewold TM, Maskos K, Donnenberg MS, Ullmann D, Uetz P, Kogl M.** 2013. The E. coli Effector Protein NleF Is a Caspase Inhibitor. PLoS ONE **8:**e58937.

137. **Wu B, Skarina T, Yee A, Jobin MC, Dileo R, Semesi A, Fares C, Lemak A, Coombes BK, Arrowsmith CH, Singer AU, Savchenko A.** 2010. NleG Type 3 effectors from enterohaemorrhagic Escherichia coli are U-Box E3 ubiquitin ligases. PLoS pathogens **6:**e1000960.

138. **Gao X, Wan F, Mateo K, Callegari E, Wang D, Deng W, Puente J, Li F, Chaussee MS, Finlay BB, Lenardo MJ, Hardwidge PR.** 2009. Bacterial effector binding to ribosomal protein s3 subverts NF-kappaB function. PLoS pathogens **5:**e1000708.

139. **Hemrajani C, Marches O, Wiles S, Girard F, Dennis A, Dziva F, Best A, Phillips AD, Berger CN, Mousnier A, Crepin VF, Kruidenier L, Woodward MJ, Stevens MP, La Ragione RM, MacDonald TT, Frankel G.** 2008. Role of NleH, a type III secreted effector from attaching and effacing pathogens, in colonization of the bovine, ovine, and murine gut. Infect. Immun. **76:**4804-4813.

140. **Batisson I, Guimond MP, Girard F, An H, Zhu C, Oswald E, Fairbrother JM, Jacques M, Harel J.** 2003. Characterization of the novel factor paa involved in the early steps of the adhesion mechanism of attaching and effacing Escherichia coli. Infect. Immun. **71:**4516-4525.

141. **Pulkkinen WS, Miller SI.** 1991. A Salmonella typhimurium virulence protein is similar to a Yersinia enterocolitica invasion protein and a bacteriophage lambda outer membrane protein. J. Bacteriol. **173:**86-93.

142. **Nataro JP, Seriwatana J, Fasano A, Maneval DR, Guers LD, Noriega F, Dubovsky F, Levine MM, Morris JG, Jr.** 1995. Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive Escherichia coli and Shigella strains. Infect. Immun. **63:**4721-4728.

143. **Farfan MJ, Toro CS, Barry EM, Nataro JP.** 2011. Shigella enterotoxin-2 is a type III effector that participates in Shigella-induced interleukin 8 secretion by epithelial cells. FEMS Immunol. Med. Microbiol. **61:**332-339.

144. **Brunder W, Karch H, Schmidt H.** 2006. 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.

145. **Scheutz F, Strockbine NA.** 2005. Genus I. Escherichia in Volume 2, Part B The Gammaproteobacteria, p. 607. *In* Garrity GM (ed.), Bergey's Manual of Systematic Bacteriology, vol. 2. Springer.

146. **Paton AW, Srimanote P, Talbot UM, Wang H, Paton JC.** 2004. A new family of potent AB(5) cytotoxins produced by Shiga toxigenic Escherichia coli. J. Exp. Med. **200:**35-46.

147. **Garmendia J, Phillips AD, Carlier MF, Chong Y, Schuller S, Marches O, Dahan S, Oswald E, Shaw RK, Knutton S, Frankel G.** 2004. TccP is an enterohaemorrhagic Escherichia coli O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton. Cellular microbiology **6:**1167-1183.

148. **Vingadassalom D, Kazlauskas A, Skehan B, Cheng HC, Magoun L, Robbins D, Rosen MK, Saksela K, Leong JM.** 2009. Insulin receptor tyrosine kinase substrate links the E. coli O157:H7 actin assembly effectors Tir and EspF(U) during pedestal formation. Proc. Natl. Acad. Sci. U. S. A. **106:**6754-6759.

149. **Campellone KG, Robbins D, Leong JM.** 2004. EspFU is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. Dev. Cell **7:**217-228.

150. **Ogura Y, Ooka T, Whale A, Garmendia J, Beutin L, Tennant S, Krause G, Morabito S, Chinen I, Tobe T, Abe H, Tozzoli R, Caprioli A, Rivas M, Robins-Browne R, Hayashi T, Frankel G.** 2007. TccP2 of O157:H7 and non-O157 enterohemorrhagic Escherichia coli (EHEC): challenging the dogma of EHEC-induced actin polymerization. Infect. Immun. **75:**604-612.

151. **Kenny B, DeVinney R, Stein M, Reinscheid DJ, Frey EA, Finlay BB.** 1997. Enteropathogenic E. coli (EPEC) transfers its receptor for intimate adherence into mammalian cells. Cell **91:**511-520.

152. **Ruchaud-Sparagano MH, Muhlen S, Dean P, Kenny B.** 2011. The enteropathogenic E. coli (EPEC) Tir effector inhibits NF-kappaB activity by targeting TNFalpha receptor-associated factors. PLoS pathogens **7:**e1002414.

153. **European Centre for Disease Prevention and Control and European Food Safety Authority.** 2011. Shiga toxin/verotoxin-producing Escherichia coli in humans, food and animals in the EU/EEA, with special reference to the German outbreak strain STEC O104. Report ISBN: 978-92-9193-298-6. ECDC.

154. **European Centre for Disease Prevention and Control.** 2013. Annual Epidemiological Report 2012. Reporting on 2010 surveillance data and 2011 epidemic intelligence data. Report 978-92-9193-443-0/1830-6160. European Centre for Disease Prevention and Control.

155. **Norwegian Institute of Public Health.** 2013. Årsrapport. Næringsmiddelbårne infeksjoner i 2012. Norwegian Institute of Public Health.

156. **Gould LH, Mody RK, Ong KL, Clogher P, Cronquist AB, Garman KN, Lathrop S, Medus C, Spina NL, Webb TH, White PL, Wymore K, Gierke RE, Mahon BE, Griffin For The Emerging Infections Program Foodnet Working Group PM.** 2013. Increased Recognition of Non-O157 Shiga Toxin-Producing Escherichia coli Infections in the United States During 2000-2010: Epidemiologic Features and Comparison with E. coli O157 Infections. Foodborne pathogens and disease **10:**453-460.

157. **Miko A, Pries K, Haby S, Steege K, Albrecht N, Krause G, Beutin L.** 2009. Assessment of Shiga toxin-producing Escherichia coli isolates from wildlife meat as potential pathogens for humans. Appl. Environ. Microbiol. **75:**6462-6470.

158. **Wallace JS, Cheasty T, Jones K.** 1997. Isolation of vero cytotoxin-producing Escherichia coli O157 from wild birds. J. Appl. Microbiol. **82:**399-404.

159. **Karch H, Bielaszewska M, Bitzan M, Schmidt H.** 1999. Epidemiology and diagnosis of Shiga toxin-producing Escherichia coli infections. Diagn. Microbiol. Infect. Dis. **34:**229-243.

160. **Gould LH, Bopp C, Strockbine N, Atkinson R, Baselski V, Body B, Carey R, Crandall C, Hurd S, Kaplan R, Neill M, Shea S, Somsel P, Tobin-D'Angelo M, Griffin PM, Gerner-Smidt P.** 2009. Recommendations for diagnosis of shiga toxin--producing Escherichia coli infections by clinical laboratories. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control **58:**1-14.

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

162. **Harris AA, Kaplan RL, Goodman LJ, Doyle M, Landau W, Segreti J, Mayer K, Levin S.** 1985. Results of a screening method used in a 12-month stool survey for Escherichia coli O157:H7. J. Infect. Dis. **152:**775-777.

163. **Hirvonen JJ, Siitonen A, Kaukoranta SS.** 2012. Usability and performance of CHROMagar STEC medium in detection of Shiga toxin-producing Escherichia coli strains. J. Clin. Microbiol. **50:**3586-3590.

164. **Gould LH.** 2012. Update: Recommendations for Diagnosis of Shiga Toxin-Producing Escherichia coli Infections by Clinical Laboratories. Clinical Microbiology Newsletter **34:**75-83.

165. **Levine MM, Xu JG, Kaper JB, Lior H, Prado V, Tall B, Nataro J, Karch H, Wachsmuth K.** 1987. 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.

166. **EFSA Panel on Biological Hazards (BIOHAZ).** 2013. Scientific Opinion on VTEC-seropathotype and scientific criteria regarding pathogenicity assessment.

167. **Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:**873-884.

168. **Pupo GM, Karaolis DK, Lan R, Reeves PR.** 1997. Evolutionary relationships among pathogenic and nonpathogenic Escherichia coli strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. Infect. Immun. **65:**2685-2692.

169. **Whittam TS, Wolfe ML, Wachsmuth IK, Orskov F, Orskov I, Wilson RA.** 1993. Clonal relationships among Escherichia coli strains that cause hemorrhagic colitis and infantile diarrhea. Infect. Immun. **61:**1619-1629.

170. **Feng P, Lampel KA, Karch H, Whittam TS.** 1998. Genotypic and phenotypic changes in the emergence of Escherichia coli O157:H7. J. Infect. Dis. **177:**1750-1753.

171. **Tenaillon O, Skurnik D, Picard B, Denamur E.** 2010. The population genetics of commensal Escherichia coli. Nat. Rev. Microbiol. **8:**207-217.

172. **Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS.** 2000. Parallel evolution of virulence in pathogenic Escherichia coli. Nature **406:**64-67.

173. **Chaudhuri RR, Henderson IR.** 2012. The evolution of the Escherichia coli phylogeny. Infection, genetics and evolution: Journal of molecular epidemiology and evolutionary genetics in infectious diseases **12:**214-226.

174. **Clermont O, Bonacorsi S, Bingen E.** 2000. Rapid and simple determination of the Escherichia coli phylogenetic group. Appl. Environ. Microbiol. **66:**4555-4558.

175. **Clermont O, Christenson JK, Denamur E, Gordon DM.** 2013. The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ. Microbiol. reports **5:**58-65.

176. **Kaas RS, Friis C, Ussery DW, Aarestrup FM.** 2012. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. BMC Genomics **13:**577.

177. **Lukjancenko O, Wassenaar TM, Ussery DW.** 2010. Comparison of 61 sequenced Escherichia coli genomes. Microb. Ecol. **60:**708-720.

178. **Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J.** 2008. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J. Bacteriol. **190:**6881-6893.

179. **Lindstedt BA.** 2005. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. Electrophoresis **26:**2567-2582.

180. **Lindstedt BA, Heir E, Gjernes E, Vardund T, Kapperud G.** 2003. DNA fingerprinting of Shiga-toxin producing Escherichia coli O157 based on Multiple-Locus Variable-Number Tandem-Repeats Analysis (MLVA). Annals of clinical microbiology and antimicrobials **2:**12.

181. **Lindstedt BA, Vardund T, Kapperud G.** 2004. Multiple-Locus Variable-Number Tandem-Repeats Analysis of Escherichia coli O157 using PCR multiplexing and multi-colored capillary electrophoresis. J. Microbiol. Methods **58:**213-222.

182. **Noller AC, McEllistrem MC, Pacheco AG, Boxrud DJ, Harrison LH.** 2003. Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic Escherichia coli O157:H7 isolates. J. Clin. Microbiol. **41:**5389-5397.

183. **Lindstedt BA, Brandal LT, Aas L, Vardund T, Kapperud G.** 2007. Study of polymorphic variable-number of tandem repeats loci in the ECOR collection and in a set of pathogenic Escherichia coli and Shigella isolates for use in a genotyping assay. J. Microbiol. Methods **69:**197-205.

184. **Lobersli I, Haugum K, Lindstedt BA.** 2012. Rapid and high resolution genotyping of all Escherichia coli serotypes using 10 genomic repeat-containing loci. J. Microbiol. Methods **88:**134-139.

185. **Lindstedt BA.** 2011. Genotyping of selected bacterial enteropathogens in Norway. Int. J. Med. Microbiol. **301:**648-653.

186. **Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y.** 1997. The complete genome sequence of Escherichia coli K-12. Science **277:**1453-1462.

187. **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.** 2001. Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11-22.

188. **Perna NT, Plunkett G, 3rd, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Posfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR.** 2001. Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature **409:**529-533.

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

190. **Steyert SR, Sahl JW, Fraser CM, Teel LD, Scheutz F, Rasko DA.** 2012. Comparative genomics and stx phage characterization of LEE-negative Shiga toxin-producing Escherichia coli. Frontiers in cellular and infection microbiology **2:**133.

191. **Bielaszewska M, Kock R, Friedrich AW, von Eiff C, Zimmerhackl LB, Karch H, Mellmann A.** 2007. Shiga toxin-mediated hemolytic uremic syndrome: time to change the diagnostic paradigm? PLoS ONE **2:**e1024.

192. **Friedrich AW, Zhang W, Bielaszewska M, Mellmann A, Kock R, Fruth A, Tschape H, Karch H.** 2007. Prevalence, virulence profiles, and clinical significance of Shiga toxin-negative variants of enterohemorrhagic Escherichia coli O157 infection in humans. Clin. Infect. Dis. **45:**39-45.

193. **Brian MJ, Frosolono M, Murray BE, Miranda A, Lopez EL, Gomez HF, Cleary TG.** 1992. Polymerase chain reaction for diagnosis of enterohemorrhagic Escherichia coli infection and hemolytic-uremic syndrome. J. Clin. Microbiol. **30:**1801-1806.

194. **Gannon VP, Rashed M, King RK, Thomas EJ.** 1993. Detection and characterization of the eae gene of Shiga-like toxin-producing Escherichia coli using polymerase chain reaction. J. Clin. Microbiol. **31:**1268-1274.

195. **Brandal LT, Lindstedt BA, Aas L, Stavnes TL, Lassen J, Kapperud G.** 2007. Octaplex PCR and fluorescence-based capillary electrophoresis for identification of human diarrheagenic Escherichia coli and Shigella spp. J. Microbiol. Methods **68:**331-341.

196. **Brandal LT, Sekse C, Lindstedt BA, Sunde M, Lobersli I, Urdahl AM, Kapperud G.** 2012. Norwegian sheep are an important reservoir for human-pathogenic Escherichia coli O26:H11. Appl. Environ. Microbiol. **78:**4083-4091.

197. **Dowd SE, Williams JB.** 2008. Comparison of Shiga-like toxin II expression between two genetically diverse lineages of Escherichia coli O157:H7 5. J. Food Prot. **71:**1673-1678.

198. **Jelacic JK, Damrow T, Chen GS, Jelacic S, Bielaszewska M, Ciol M, Carvalho HM, Melton-Celsa AR, O'Brien AD, Tarr PI.** 2003. Shiga toxin-producing Escherichia coli in Montana: bacterial genotypes and clinical profiles. J. Infect. Dis. **188:**719-729.

199. **Russmann H, Schmidt H, Heesemann J, Caprioli A, Karch H.** 1994. Variants of Shiga-like toxin II constitute a major toxin component in Escherichia coli O157 strains from patients with haemolytic uraemic syndrome. J. Med. Microbiol. **40:**338-343.

200. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821-829.

201. **Ribeiro FJ, Przybylski D, Yin S, Sharpe T, Gnerre S, Abouelleil A, Berlin AM, Montmayeur A, Shea TP, Walker BJ, Young SK, Russ C, Nusbaum C, MacCallum I, Jaffe DB.** 2012. Finished bacterial genomes from shotgun sequence data. Genome Res. **22:**2270-2277.

202. **Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J.** 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience **1:**18.

203. **Ronen R, Boucher C, Chitsaz H, Pevzner P.** 2012. SEQuel: improving the accuracy of genome assemblies. Bioinformatics **28:**i188-196.

204. **Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J.** 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Meth. **10:**563-569.

205. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. Bmc Bioinformatics **11:**119.

206. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9:**75.

207. **Vesth T, Lagesen K, Acar O, Ussery D.** 2013. CMG-biotools, a free workbench for basic comparative microbial genomics. PLoS One **8:**e60120.

208. **Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbonnelle E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S.** 2008. Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. BMC Genomics **9:**560.

209. **Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguenec C, Lescat M, Mangenot S, Martinez-Jehanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Medigue C, Rocha EP, Denamur E.** 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS genetics **5:**e1000344.

210. **Rice P, Longden I, Bleasby A.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. **16:**276-277.

211. **Levenshtein VI.** 1966. Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics Doklady **10:**707-710.

212. **Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M.** 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21:**3674-3676.

213. **Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A.** 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. **36:**3420-3435.

214. **Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ.** 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics **25:**1189-1191.

215. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology **7:**539.

216. **Jenssen GR, Hovland E, Bjerre A, Bangstad HJ, Nygard K, Vold L.** 2014. Incidence and etiology of hemolytic-uremic syndrome in children in Norway, 1999-2008 - a retrospective study of hospital records to assess the sensitivity of surveillance. BMC Infect. Dis. **14:**265.

217. **Alpers K, Werber D, Frank C, Koch J, Friedrich AW, Karch H, An Der Heiden M, Prager R, Fruth A, Bielaszewska M, Morlock G, Heissenhuber A, Diedler A, Gerber A, Ammon A.** 2009. Sorbitol-fermenting enterohaemorrhagic Escherichia coli O157:H(-) causes another outbreak of haemolytic uraemic syndrome in children. Epidemiol. Infect. **137:**389-395.

218. **Ammon A, Petersen LR, Karch H.** 1999. 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.

219. **Norwegian Institute of Public Health.** 2010. Årsrapport 2010: Matbårne infeksjoner og utbrudd i 2009. Norwegian Institute of Public Health.

220. **Rosser T, Dransfield T, Allison L, Hanson M, Holden N, Evans J, Naylor S, La Ragione R, Low JC, Gally DL.** 2008. Pathogenic potential of emergent sorbitol-fermenting Escherichia coli O157:NM. Infect. Immun. **76:**5598-5607.

221. **Karch H, Bielaszewska M.** 2001. Sorbitol-fermenting Shiga toxin-producing Escherichia coli O157:H(-) strains: epidemiology, phenotypic and molecular characteristics, and microbiological diagnosis. J. Clin. Microbiol. **39:**2043-2049.

222. **Karch H, Friedrich AW, Gerber A, Zimmerhackl LB, Schmidt MA, Bielaszewska M.** 2006. New aspects in the pathogenesis of enteropathic hemolytic uremic syndrome. Semin. Thromb. Hemost. **32:**105-112.

223. **Buvens G, Pierard D.** 2012. Virulence profiling and disease association of verocytotoxin-producing Escherichia coli O157 and non-O157 isolates in Belgium. Foodborne pathogens and disease **9:**530-535.

224. **Kawano K, Ono H, Iwashita O, Kurogi M, Haga T, Maeda K, Goto Y.** 2012. stx genotype and molecular epidemiological analyses of Shiga toxin-producing Escherichia coli O157:H7/H- in human and cattle isolates. Eur. J. Clin. Microbiol. Infect. Dis. **31:**119-127.

225. **Pradel N, Bertin Y, Martin C, Livrelli V.** 2008. Molecular analysis of shiga toxin-producing Escherichia coli strains isolated from hemolytic-uremic syndrome patients and dairy samples in France. Appl. Environ. Microbiol. **74:**2118-2128.

226. **Laing CR, Zhang Y, Gilmour MW, Allen V, Johnson R, Thomas JE, Gannon VP.** 2012. A comparison of Shiga-toxin 2 bacteriophage from classical enterohemorrhagic Escherichia coli serotypes and the German E. coli O104:H4 outbreak strain. PLoS ONE **7:**e37362.

227. **Feng PC, Monday SR, Lacher DW, Allison L, Siitonen A, Keys C, Eklund M, Nagano H, Karch H, Keen J, Whittam TS.** 2007. Genetic diversity among clonal lineages within Escherichia coli O157:H7 stepwise evolutionary model 1. Emerg. Infect. Dis. **13:**1701-1706.

228. **Delannoy S, Beutin L, Fach P.** 2013. Towards a molecular definition of enterohemorrhagic Escherichia coli (EHEC): detection of genes located on O island 57 as markers to distinguish EHEC from closely related enteropathogenic E. coli strains. J. Clin. Microbiol. **51:**1083-1088.

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

230. **Bugarel M, Martin A, Fach P, Beutin L.** 2011. Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) Escherichia coli strains: a basis for molecular risk assessment of typical and atypical EPEC strains. Bmc Microbiol **11:**142.

231. **Jackson RW, Vinatzer B, Arnold DL, Dorus S, Murillo J.** 2011. The influence of the accessory genome on bacterial pathogen evolution. Mobile genetic elements **1:**55-65.

232. **Escobar-Paramo P, Clermont O, Blanc-Potard AB, Bui H, Le Bouguenec C, Denamur E.** 2004. A specific genetic background is required for acquisition and expression of virulence factors in Escherichia coli. Mol. Biol. Evol. **21:**1085-1094.

233. **Peekhaus N, Conway T.** 1998. What's for dinner?: Entner-Doudoroff metabolism in Escherichia coli. J. Bacteriol. **180:**3495-3502.

234. **Nivaskumar M, Francetic O.** 2014. Type II secretion system: A magic beanstalk or a protein escalator. Biochim. Biophys. Acta 2014 Aug;**1843**(8):1568-77.

235. **Baldi DL, Higginson EE, Hocking DM, Praszkier J, Cavaliere R, James CE, Bennett-Wood V, Azzopardi KI, Turnbull L, Lithgow T, Robins-Browne RM, Whitchurch CB, Tauschek M.** 2012. The type II secretion system and its ubiquitous lipoprotein substrate, SslE, are required for biofilm formation and virulence of enteropathogenic Escherichia coli. Infect. Immun. **80:**2042-2052.

236. **Duan Q, Zhou M, Zhu L, Zhu G.** 2013. Flagella and bacterial pathogenicity. J. Basic Microbiol. **53:**1-8.

237. **Ratledge C, Dover LG.** 2000. Iron metabolism in pathogenic bacteria. Annu. Rev. Microbiol. **54:**881-941.

238. **Saha R, Saha N, Donofrio RS, Bestervelt LL.** 2013. Microbial siderophores: a mini review. J. Basic Microbiol. **53:**303-317.

239. **Karch H, Schubert S, Zhang D, Zhang W, Schmidt H, Olschlager T, Hacker J.** 1999. A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing Escherichia coli clonal lineages. Infect. Immun. **67:**5994-6001.

240. **Bielaszewska M, Zhang W, Mellmann A, Karch H.** 2007. Enterohaemorrhagic Escherichia coli O26:H11/H-: a human pathogen in emergence. Berl. Munch. Tierarztl. Wochenschr. **120:**279-287.

Paper I

FEMS MICROBIOLOGY LETTERS

RESEARCH LETTER

Identification of the anti-terminator $q_{O111:H-}$ gene in Norwegian sorbitol-fermenting *Escherichia coli* O157:NM

Kjersti Haugum[1,2], Bjørn-Arne Lindstedt[1], Inger Løbersli[1], Georg Kapperud[1] & Lin Thorstensen Brandal[1]

[1]Department of Foodborne Infections, The Norwegian Institute of Public Health, Oslo, Norway; and [2]Department of Laboratory Medicine, Children's and Women's Health, The Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence: Lin Thorstensen Brandal, The Norwegian Institute of Public Health, Division of Infectious Disease Control, Department of Foodborne Infections, PO Box 4404 Nydalen, N-0403 Oslo, Norway. Tel.: +47 21076652; fax: +47 21076518; e-mail: lin.thorstensen.brandal@fhi.no

Received 30 November 2011; revised 10 January 2012; accepted 13 January 2012. Final version published online 8 February 2012.

DOI: 10.1111/j.1574-6968.2012.02505.x

Editor: Rob Delahay

Keywords
sorbitol-fermenting O157:NM; *stx2*-encoding bacteriophages; anti-terminator *q* gene; tRNA.

Abstract

Sorbitol-fermenting *Escherichia coli* O157:NM (SF O157) is an emerging pathogen suggested to be more virulent than nonsorbitol-fermenting *Escherichia coli* O157:H7 (NSF O157). Important virulence factors are the Shiga toxins (stx), encoded by *stx1* and/or *stx2* located within prophages integrated in the bacterial genome. The *stx* genes are expressed from $p_{R'}$ as a late protein, and anti-terminator activity from the Q protein is necessary for read through of the late terminator $t_{R'}$ and activation of $p_{R'}$. We investigated the regulation of $stx2_{EDL933}$ expression at the genomic level in 17 Norwegian SF O157. Sequencing of three selected SF O157 strains revealed that the anti-terminator *q* gene and genes upstream of $stx2_{EDL933}$ were identical or similar to the ones observed in the *E. coli* O111:H− strain AP010960, but different from the ones observed in the NSF O157 strain EDL933 (AE005174). This suggested divergent $stx2_{EDL933}$-encoding bacteriophages between NSF O157 and the SF O157 strains (FR874039-41). Furthermore, different DNA structures were detected in the SF O157 strains, suggesting diversity among bacteriophages also within the SF O157 group. Further investigations are needed to elucidate whether the $q_{O111:H-}$ gene observed in all our SF O157 contributes to the increased virulence seen in SF O157 compared to NSF O157. An assay for detecting $q_{O111:H-}$ was developed.

Introduction

Sorbitol-fermenting *Escherichia coli* O157:NM (SF O157) was first identified in an outbreak in Bavaria in Germany in 1988 (Karch & Bielaszewska, 2001). Since then, these highly pathogenic bacteria have been isolated in many European countries (Allerberger *et al.*, 2000; Karch & Bielaszewska, 2001; Allison, 2002; Editorial Team, 2006; Eklund *et al.*, 2006; Jakubczak *et al.*, 2008; Alpers *et al.*, 2009; Buvens *et al.*, 2009), including Norway (Norwegian Institute of Public Health, 2010). The first isolate of SF O157 in Norway was recovered from a patient in 2005, and until 2009, only eight sporadic cases of SF O157 infection were detected. In 2009, we had an outbreak with SF O157 affecting 13 children, of whom nine developed haemolytic uraemic syndrome (HUS) and one died. The source of infection was not found (Norwegian Institute of Public Health, 2010). The same outbreak strain was also isolated from a cluster of three children with HUS in 2010 (Norwegian Institute of Public Health, 2011), and in May 2011, another child, without HUS, was diagnosed with this specific strain (The Norwegian Surveillance System for Communicable Diseases (MSIS)). Outside Europe, SF O157 has been isolated in Australia and Brazil (Bettelheim *et al.*, 2002; Moreira *et al.*, 2003).

There are reports suggesting that SF O157 more often progresses to HUS compared to nonsorbitol-fermenting *Escherichia coli* O157:H7 (NSF O157), and epidemiological and phenotypical characteristics as well as the presence of specific virulence genes differ between SF O157 and NSF O157 (Karch & Bielaszewska, 2001; Rosser *et al.*, 2008). Additionally, phylogenetic analyses show that SF

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

O157 and NSF O157 most probably have diverged early in the evolution of *E. coli* O157 and belong to different clones (Karch & Bielaszewska, 2001; Feng *et al.*, 2007).

Important virulence factors in enterohaemorrhagic *E. coli* (EHEC) are the Shiga toxins (stx1 and stx2), encoded by the *stx1* and *stx2* genes, both of which may be divided into subtypes. In NSF O157, *stx1*, $stx2_{EDL933}$ and *stx2c* have been detected, either alone or in combinations, whereas $stx2_{EDL933}$ is the only *stx* subtype found so far in SF O157 (Karch & Bielaszewska, 2001). In some SF O157, two identical copies of the $stx2_{EDL933}$ gene have been reported, resulting in increased production of stx. However, an association between increased stx production and enhanced virulence as compared to strains with only one $stx2_{EDL933}$ copy was not observed (Bielaszewska *et al.*, 2006). Furthermore, loss of the *stx2* phage in SF O157 followed by regain of the phage in the same SF O157 strain has been reported, thus giving SF O157 the ability to recycle *stx2* (Mellmann *et al.*, 2008).

The *stx* genes are encoded in the late region of lambdoid prophages, where they are located downstream of the late promoter $p_{R'}$ and late terminator $t_{R'}$. The *stx* genes are expressed from $p_{R'}$ as a late protein, and the anti-terminator activity from the Q protein is necessary for read through of the late terminator, $t_{R'}$ and activation of $p_{R'}$ (Schmidt, 2001). Although the *stx* genes have their own functional promoters (Calderwood *et al.*, 1987; Schmidt, 2001), induction of the prophage and transcription from $p_{R'}$ is important for the expression of the *stx* genes as well as for the release of stx from the bacteria (Wagner *et al.*, 2001). Two different *q* genes, q_{933} and q_{21}, have been identified in NSF O157, giving evidence of higher production of stx2 in strains positive for the q_{933} gene (LeJeune *et al.*, 2004; Koitabashi *et al.*, 2006; Matsumoto *et al.*, 2008). Additionally, mutations in the *stx2* promoter region have been observed in strains containing the q_{21} gene, which further affects the expression of *stx2* negatively (Matsumoto *et al.*, 2008).

Dowd and Williams compared expression of *stx2* between two genetically diverse lineages of *E. coli* O157:H7 and observed that lineage I produced significantly more toxin than lineage II (Dowd & Williams, 2008). Furthermore, when using the stx_8 primer set, all the lineage I strains were positive, whereas all lineage II strains were negative (Dowd & Williams, 2008). They, therefore, predicted that the stx_8 primers were useful to differentiate lineage I and lineage II (Dowd & Williams, 2008).

Draft genome sequences of two SF O157 strains are published (Rump *et al.*, 2011), but to our knowledge, little is known about the genomic regulation of $stx2_{EDL933}$ expression in such strains. Thus, in the present study, we aimed to examine factors at the genomic level that might influence the expression of *stx2* in SF O157.

Materials and methods

Bacterial strains

Among the 35 human clinical isolates of SF O157 recovered in Norway, 17 harboured the *stx2* gene and were included in the present study (Table 1). Only one *stx2* positive strain from each patient belonging to the 2009–2011 outbreak cluster was included. All isolates were from the strain collection at the Norwegian Institute of Public Health and were recovered from 2005 through 2011.

Extraction of DNA

For PCR analyses and sequencing, suspensions of bacterial cells were boiled for 15 min, and the supernatant was used directly in the PCR after centrifugation at 15 871 ***g***. for one minute. For inverse PCR, DNA was isolated using the BioRobot EZ1 as described by the manufacturer (Qiagen Gmbh, Hilden, Germany).

Genetic characterization of SF O157, screening of 10 virulence genes, *stx2* subtyping, and MLVA genotyping

Molecular identification of SF O157 was carried out by a multiplex-PCR (M-PCR) detecting the genes rbf_{O157} (Maurer *et al.*, 1999), $fliC_{H7}$ (Lindstedt *et al.*, unpublished), *terE* (Taylor *et al.*, 2002) and the *Shigella* resistance locus (SRL) (Janka *et al.*, 2005). The *dinB* gene (Lindstedt *et al.*, unpublished) was used as an internal amplification control. All strains were screened for the *stx1*, *stx2* and *eae* genes (modification of Brandal *et al.*, 2007) as well as the *ehxA*, *nleB*, *stcE*, $stcE_{O103}$, *cdt*, *subA* and *saa* genes (Brandal *et al.*, manuscript in preparation). The *stx2* subtype was determined using PCR-restriction fragment length polymorphism (RFLP) and sequencing (modifications of Jelacic *et al.*, 2003; Russmann *et al.*, 1994; and Persson *et al.*, 2007a). All strains were genotyped with MLVA for SF O157 (Lindstedt, 2011).

Sequencing of PCR products

PCR products for sequencing were purified using the QIAquick PCR Purification Kit (Qiagen). All sequencing were performed with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems by Life Technologies, Carlsbad, CA) as described by the manufacturer. The extension products were purified using the DyeEx 2.0 Spin Kit (Qiagen). The samples were run on an ABI-3100 or ABI-3130*xl* automated sequencer (Applied Biosystems), and the raw-data files were exported to the SeqMan Pro sequencing analysis software (DNASTAR

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Table 1. Genetic characterization, genotyping, virulence gene status and *q* gene status of the SF O157 strains in relation to HUS status

Isolate number*	rfb_{O157}	$fliC_{H7}$	SRL	*dinB*	MLVA profile	*stx*	*eae*	*stx2* subtype	*ehxA*	*nleB*	*stcE*	*cdt*	stx_8	$q_{O111:H-}$	HUS
1105-3298	+	+	+	+	A	+	+	$stx2_{EDL933}$	+	+	+	+	+	+	+
1106-4002	+	+	+	+	B	+	+	$stx2_{EDL933}$	+	+	+	−	−	+	+
1108-2781	+	+	+	+	C	+	+	$stx2_{EDL933}$	+	+	+	−	+	+	−
1109-0113	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	+
1109-0435	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	+
1109-0500	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	−
1109-0621	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	+
1109-0634	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	+
1109-0823-5	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	−	−	+	+
1109-0923	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	+
1109-1280	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	−
1109-2176	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	+
1109-3001	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	+
1110-2506	+	+	+	+	D	+	+	$stx2_{EDL933}$	−	+	+	+	−	+	+
1110-2531	+	+	+	+	D	+	+	$stx2_{EDL933}$	−	+	+	+	−	+	+
1110-2621	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	+
1111-0887	+	+	+	+	D	+	+	$stx2_{EDL933}$	+	+	+	+	−	+	−

HUS, haemolytic ureamic syndrome; +, HUS; −, no HUS.
*All strains were negative for the *terE*, $stcE_{O103}$, *subA*, *saa*, q_{933} and q_{21} genes.

Lasergene 9 Core Suite, Madison, WI) for inspection and assembly.

Screening with the stx_8 primer set, identification of the anti-terminator *q* gene and sequencing of the promoter region of the *stx2* genes

A M-PCR including the stx_8 primer set (Dowd & Williams, 2008), the q933 forward primer, the q21 forward primer and the 595 reverse primer (Unkmeir & Schmidt, 2000; LeJeune *et al.*, 2004) was designed (Supporting Information, Table S1). Each forward primer was labelled with a fluorochrome, and PCR was performed using the Qiagen Multiplex PCR kit (Qiagen), as described by the manufacturer. The PCR was run in a GeneAmp 9700 machine (Applied Biosystems) with the temperature profile as described for the Qiagen Multiplex PCR kit (Qiagen) and an annealing temperature of 60 °C. The PCR products were identified by capillary electrophoresis on an ABI PRISM 3130*xl* Genetic analyzer (Applied Biosystems) as follows: 0.5 μL PCR product was mixed with 0.5 μL GeneScan 1200 LIZ size standard (Applied Biosystems) and 9 μL HiDi formamide (Applied Biosystems). After denaturation, the capillary electrophoresis was run for 2 h at 60 °C, using POP7 polymer (Applied Biosystems) with an injection voltage of 1.8 kV for 11 s and running voltage of 6.5 kV. For data analysis, GENEMAPPER Software 4.0 (Applied Biosystems) was used. The primers slt2s-2 (Matsumoto *et al.*, 2008) and 595 (Unkmeir & Schmidt, 2000; Table S1) were used for PCR amplification and sequencing of the promoter region of the *stx2* genes. The PCR was run as described earlier, with an annealing temperature of 60 °C. The PCR products were sequenced using the primers slt2s-2 and 595 (Table S1).

Inverse PCR

Three SF O157 strains from different years, with different MLVA profiles, different outbreak and clinical status, as well as different results from the stx_8 screening, were selected for inverse PCR (Table 2). Strain EDL933 (FHBa 667) was included as positive control for NSF O157. DNA digestion was performed as described earlier (Zhang *et al.*, 2010) and checked on a BioAnalyzer (Agilent Technologies, Santa Clara, CA) using the Agilent DNA 7500 Kit (Agilent Technologies) as recommended by the manufacturer. Digested DNA was purified with the QIAquick PCR Purification Kit (Qiagen) and ligated as described by Zhang (Zhang *et al.*, 2010). Ligated DNA was purified with the QIAquick PCR Purification Kit (Qiagen) and used as template for inverse PCR. The primers PS7-rev and PS8-rev [reverse complement of PS7 and PS8 (Persson *et al.*, 2007b; Table S1)] and the Advantage 2 PCR Kit (Clontech, Mountain View, CA) were used for PCR amplification as described by the manufacturer. The PCR was run as described earlier (Zhang *et al.*, 2010). Positive amplification was checked on a BioAnalyzer using the Agilent DNA 7500 Kit (Agilent Technologies), and the PCR products were sequenced as described earlier, using primers listed in Table S1. The primers designed in this study for sequencing of the inverse PCR

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Table 2. Sorbitol-fermenting O157 strains used for sequencing of the *q* gene and downstream of the *stx2* gene

Strain number	Year isolated	MLVA profile	Outbreak	Clinical outcome	PCR stx_8
1106-4002	2006	B	Single case	HUS	Negative
1108-2781	2008	C	Strain 1108-2781 (*stx2*+) was isolated from an asymptomatic brother of a HUS patient (strain *stx2*−) during a family outbreak	None	Positive
1109-0113	2009	D	Index patient of the 2009 outbreak with 13 children, 9 HUS	HUS, patient died	Negative

product were designed by the Primer Walk function in SeqMan Pro sequencing analysis software (DNASTAR Lasergene 9 Core Suite).

Confirmation of the anti-terminator *q* gene and promoter region of SF O157 with PCR and sequencing

Inspection and assembly of the sequences were performed using the the SeqMan Pro sequencing analysis software (DNASTAR Lasergene 9 Core Suite). BLAST search of the sequences revealed that the *q* gene, the promoter region of *stx2* and the *stx2* gene of the SF O157 strains 1106-4002 and 1109-0113 were identical to the sequence of the GenBank accession number AP010960 (*E. coli* O111:H−, strain 11128), whereas strain 1108-2781 nearly was identical to this specific sequence (AP010960). Therefore, the sequence of AP010960 was used as template for primer design for the confirmation of the anti-terminator *q* gene and *stx2* promoter region of the three strains. For strain 1108-2781, GenBank accession number AE005174 (*E. coli* O157:H7 EDL933) was used as template for primer design downstream of the *stx2* gene, whereas AP010960 was used for the other two strains. The primers were designed using PrimerSelect (DNASTAR Lasergene 9 Core Suite; Table S1). The PCR was run as described earlier with annealing temperatures of 55 °C for primer sets SF2 and SF7-SF10, 58 °C for primer sets SF1, SF5, SF6, SF11, and 60 °C for primer sets nySF3, nySF4, stx_8 and SF11-2. Sequencing was performed as described earlier, with primers listed in Table S1.

Screening of the $q_{O111:H-}$ gene in all SF O157

All 17 SF O157 were screened for the $q_{O111:H-}$ gene by using the SF1-F and SF1-R primer set (Table S1). The PCR was run as described earlier with an annealing temperature of 58 °C. Strain EDL933 (FH-Ba 667) was included as a negative control. All PCR products were run on a BioAnalyzer, using the Agilent DNA 1000 Kit (Agilent Technologies) as described by the manufacturer to check for positive amplification.

Results

Genetic characterization of SF O157, screening of 10 virulence genes, *stx2* subtyping, and MLVA genotyping

All 17 SF O157 isolates were positive for rfb_{O157}, $fliC_{H7}$, SRL and *dinB*, as well as s*tx2* and *eae*. $Stx2_{EDL933}$ was the only *stx2* subtype detected. Additionally, the strains harboured *nleB* and *stcE*. Fifteen of 17 isolates (88%) were positive for the *ehxA* gene, and 14/17 strains (82%) carried *cdt* (Table 1). *terE*, $stcE_{O103}$, *saa* and *subA* were not present in any of the strains examined (Table 1). The SF O157 isolates recovered from Norwegian patients before 2009 showed distinct MLVA profiles, indicating that the cases concerned did not belong to common-source outbreaks. However, all isolates obtained from 2009 through May 2011 grouped into one MLVA genotype (Table 1).

Screening with the stx_8 primer set, identification of the anti-terminator *q* gene and sequencing of the promoter region of the *stx2* gene in SF O157

Screening with the stx_8 primer set showed that only two SF O157 were positive for these primers, whereas 15 were negative (Table 1). All isolates failed to amplify the *q933* and *q21* genes. PCR and sequencing of the *stx2* promoter region with the primers slt2s-2 and 595 showed that 15 of the SF O157 (all stx_8 negative) were identical and differed from the NSF O157 strain EDL933 sequence (AE005174) by five nucleotides (Fig 1). The sequence differences were seen between the tRNA genes *argN* and *argO* located proximately to the *stx2* promoter region, and in the *argO* gene, between the −35 and the −10 region within the *stx2* promoter. However, these isolates showed identical sequence to the O111:H− strain 11128 (AP010960) in this region (Fig 1). The two last SF O157 (both stx_8 positive) differed from the EDL933 sequence (AE005174) in one nucleotide only, located in the tRNA gene *argN* (Fig 1).

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

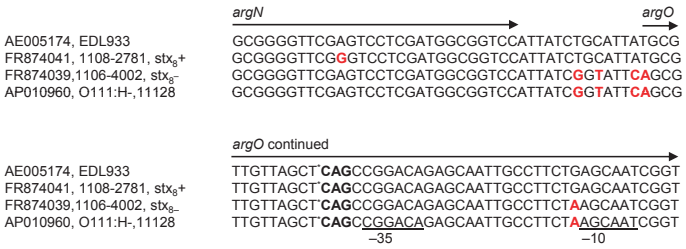


Fig. 1. Alignment of parts of the *argN* and the *argO* genes in AE005174, FR874041, FR874039, and AP010960. Bases differing from the sequence of AE005174 are shown in red. The start of the *stx2* promoter (*CAG) are given in bold with an asterisk. The −35 and −10 regions of the *stx2* promoter are underlined.

PCR and sequencing of the anti-terminator *q* gene, the *stx2* gene and downstream of the *stx2* gene in SF O157

We determined the nucleotide sequences of the *q* gene, the region between the *q* gene and the *stx2* gene, the *stx2* gene and the 500-bp region downstream of the *stx2* gene in the SF O157 strains 1106-4002 (EMBL/GenBank accession number FR874039), 1109-0113 (outbreak strain from 2009, EMBL/GenBank accession number FR874040) and 1108-2781 (EMBL/GenBank accession number FR874041). Strains 1106-4002 and 1109-0113 showed identical sequences in reversed complement to the *E. coli* O111:H− strain 11128 (AP010960) in all the examined regions, except for a single nucleotide polymorphism (SNP) in position 371 in the *stx2A* subunit (Fig 2). Strain 1108-2781 had an identical *q* gene to the O111:H− strain (AP010960), but differed from this strain in 14 nucleotides in the region between the *q* gene and the *stx2* gene as well as in the *stx2* gene (Fig 2). Additionally, downstream of the *stx2* gene, 1108-2781 was different from the O111:H− strain and showed identical sequence to the NSF O157 strain EDL933 (AE005174) (Fig 2).

Screening of the $q_{O111:H-}$ gene in SF O157

The $q_{O111:H-}$ gene was detected in all SF O157 included in the present study (Table 1).

Discussion

Infection with SF O157 has been reported to be more aggressive than infection with NSF O157, as SF O157 more often progresses to HUS compared to NSF O157 (Karch & Bielaszewska, 2001; Rosser *et al.*, 2008). It is well known that stx2 play a key role in the development of HUS (Gyles, 2007). In NSF O157, two different *q* genes, q_{933} and q_{21}, have been identified, giving evidence of higher production of stx2 in strains positive for q_{933} (LeJeune *et al.*, 2004; Koitabashi *et al.*, 2006; Matsumoto

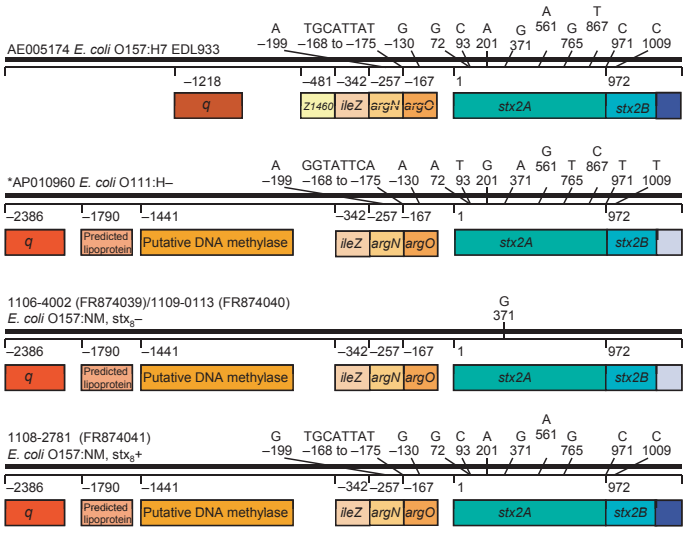


Fig. 2. Comparison of the structure of the *stx* genes and upstream genes in the *E. coli* strains O157:H7 EDL933 (AE005174), O111:H–11128 (AP010960), O157:NM strains 1106-4002 (FR874039), 1109-0113 (FR874040), and 1108-2781 (FR874041). Additionally, approximately 500 bp downstream of the *stx* genes are shown. The numericals are indicating base positions or gene start positions (translational start for *stx2A*; +1). *Reversed complement of AP010960.

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

et al., 2008). Additionally, mutations in the *stx2* promoter region have been observed in strains carrying the q_{21} gene, which probably also contribute to the reduced expression of *stx2* (Matsumoto *et al.*, 2008). However, the knowledge about the genomic regulation of *stx2* expression in SF O157 is sparse.

In the present study, the sequence upstream (including the *q* gene) and approximately 500-bp downstream of the *stx2* gene in three Norwegian SF O157 isolates were sequenced, and a distinct *q* gene and different genes upstream of the $stx2_{EDL933}$ gene, as compared to the NSF O157:H7 strain EDL933 (AE005174), were detected. The *q* gene and the genes upstream of $stx2_{EDL933}$ in SF O157 had identical or similar sequence to the O111:H− strain 11128 (AP010960), a strain isolated from a patient with bloody diarrhoea in Japan in 2001 (Ogura *et al.*, 2007). *stx*-encoding lamboid bacteriophages show similarities in DNA sequences, yet they might be heterogeneous as evidenced by divergent gene organization and chromosomal location, as well as harbouring high degree of mosaic DNA structures (Unkmeir & Schmidt, 2000; Allison, 2007; Ogura *et al.*, 2009). Based on these observations, our results indicate that the sequenced SF O157 isolates harboured different $stx2_{EDL933}$-encoding phages than the NSF O157 strain EDL933 (Allison, 2007; Ogura *et al.*, 2009). Furthermore, mosaic DNA structure was seen within the bacteriophage of strain 1108-2781 (FR874041), but not within the other two sequenced SF O157 strains, demonstrating that considerable diversity also exists among $stx2_{EDL933}$-encoding bacteriophages within the group of SF O157.

Two of 17 SF O157 strains were positive for the stx_8 primer set. Strain 1108-2781 (stx_8+) had identical sequence with the NSF O157:H7 strain EDL933 in this region, whereas strains 1106-4002 (FR874039) and 1109-0113 (FR874040) (both stx_8−) showed identical sequence to the O111:H− strain 11128, thus explaining the PCR results. The stx_8 primer set was suggested to differentiate NSF O157 into lineage I and II, where lineage I strains, positive for stx_8, were shown to express more stx proteins and to have a higher pathogenic potential than the lineage II strains (stx_8 negative) (Dowd & Williams, 2008). We did not investigate the expression of stx. However, one of the two stx_8+ SF O157 isolates was obtained from a HUS patient, whereas as many as 80% (12/15) of the patients with SF O157 negative for stx_8 developed HUS. This implies that stx_8 screening was not an appropriate tool to differentiate SF O157 into lineage I and II, and probably a separate lineage exists for the SF O157, different from both lineages I and II, as suggested by Laing *et al.* (2009).

It has been suggested that higher levels of colonization and thereby enhanced stx2 exposure might be responsible, at least partly, for the increased pathogenic potential seen in SF O157 compared to NSF O157 (Rosser *et al.*, 2008). Although no evidence of *in vitro* increased $stx2_{EDL933}$ expression in SF O157 compared to NSF O157 has been observed, so far only two SF O157 have been examined (Rosser *et al.*, 2008), and to our knowledge, the *in vivo* level has never been investigated. It cannot be excluded that the $q_{O111:H-}$ gene identified in all Norwegian SF O157 isolates, as opposed to *q933* and *q21* previously found in NSF O157 (LeJeune *et al.*, 2004; Koitabashi *et al.*, 2006; Matsumoto *et al.*, 2008), may contribute to the increased virulence observed in SF O157, by virtue of increased stx2 level and/or enhanced resistance of the bacteria concerned (Ferens & Hovde, 2011). Additional investigations are needed to elucidate the activity of the $Q_{O111:H-}$ protein in SF O157 strains.

Lambdoid phages are introducing tRNA genes into the bacteria, which may be required for efficient expression of foreign genes encoded by the phages, as for instance the *stx* genes (Plunkett *et al.*, 1999; Hayashi *et al.*, 2001; Schmidt, 2001). The tRNA genes *ileZ-argN-argO* located close to the *stx* genes, in both SF O157 and NSF O157 as well as in other EHEC, might thus serve as a supplement to the host tRNA pool and lead to a more efficient translation of foreign genes (Plunkett *et al.*, 1999). In strains 1106-4002 (FR874039) and 1109-0113 (FR874040), the tRNA genes *ileZ-argN-argO* showed identical sequences to the ones in the O111:H− strain 11128 (AP010960). However, strain 1108-2781 (FR874041) exhibited an *ileZ-argN-argO* sequence identical to that found in NSF O157 EDL933 (AE005174), except for a single nucleotide polymorphism in the *argN* gene observed neither in the O111:H− strain 11128, the O157:H7 strain EDL933 nor the SF O157 strains 1106-4002 and 1109-0113. These observations might suggest different origin of the bacteriophages and/or rearrangements within the phage DNA in SF O157 (Allison, 2007). Whether the observed base substitutions seen within the tRNA *ileZ-argN-argO* sequence contribute to enhanced virulence in the SF O157, compared to NSF O157, needs to be further explored.

Phenotypic characteristics as well as the presence of specific virulence genes are in part different between SF O157 and NSF O157 (Karch & Bielaszewska, 2001; Rosser *et al.*, 2008). Genetic characterization of SF O157 showed that our results were in concordance with previous reports (they all carried rfb_{O157}, $fliC_{H7}$, SRL and *dinB*) (Karch & Bielaszewska, 2001; Taylor *et al.*, 2002; Janka *et al.*, 2005; Orth *et al.*, 2007). Additionally, all SF O157 harboured the *eae* and $stx2_{EDL933}$ genes. Eighty-eight per cent of the strains carried *ehxA*, and *cdt* was present in 82% of our SF O157 isolates, which is in agreement with earlier studies (Karch & Bielaszewska, 2001; Janka *et al.*,

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

2003). The *nleB* gene, which was seen in all our SF O157, has recently been reported to be highly associated with virulent EHEC and EPEC seropathotypes (Bugarel *et al.*, 2011). Additionally, all SF O157 carried the *stcE* gene encoding a metalloprotease shown to promote the intimate adherence and inhibit the inflammatory system (Szabady *et al.*, 2009). However, both *nleB* and *stcE* are also common in NSF O157 (Szabady *et al.*, 2009; Bugarel *et al.*, 2011).

MLVA genotyping and data from MSIS indicate that the cases of SF O157 infection recorded in Norway before 2009 were sporadic. However, in the period 2009 through May 2011, SF O157 was isolated from 16 children, of whom 11 developed HUS. The isolates fell into one distinct MLVA cluster (cluster D), indicating a common source of infection. However, like unsuccessful attempts to identify the source and the reservoir of SF O157 in other countries (Allerberger *et al.*, 2000; Karch & Bielaszewska, 2001; Allison, 2002; Editorial Team, 2006; Eklund *et al.*, 2006; Jakubczak *et al.*, 2008; Buvens *et al.*, 2009), the source in the Norwegian cases could not be determined despite a considerable amount of work invested (Norwegian Institute of Public Health, 2010).

In conclusion, all the Norwegian SF O157 showed a distinct *q* gene, as well as different genes upstream of the $stx2_{EDL933}$ gene compared to the NSF O157:H7 strain EDL933 (AE005174). This indicated that Norwegian SF O157 harboured divergent $stx2_{EDL933}$-encoding bacteriophages compared to the NSF O157 strain EDL933 (AE005174). The SF O157 carried a *q* gene identical to the *q* gene in O111:H− strain 11128 (AP010960). Interestingly, different DNA sequences were observed within the region sequenced in the three SF O157 strains (FR874039-41), suggesting that considerable diversity exists among $stx2_{EDL933}$-encoding bacteriophages also within the SF O157 group. It is possible that the $q_{O111:H-}$ gene identified in SF O157 contributes to the increased virulence seen in SF O157 compared to NSF O157. However, further investigations are needed to elucidate the activity of the $Q_{O111:H-}$ protein in SF O157.

We have developed an assay for detecting the $q_{O111:H-}$ gene in SF O157, and this might be a useful supplement to differentiate SF O157 from NSF O157.

References

Allerberger F, Dierich MP, Gruber-Moesenbacher U, Liesegang A, Prager R, Hartmann G, Rabsch W, Tschape H & Karch H (2000) Nontoxigenic sorbitol-fermenting *Escherichia coli* O157:H- associated with a family outbreak of diarrhoea. *Wien Klin Wochenschr* **112**: 846–850.

Allison L (2002) HUS due to a sorbitol-fermenting verotoxigenic *E. coli* O157 in Scotland. *Euro Surveill* **6**: pii=1955. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=1955

Allison HE (2007) Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. *Future Microbiol* **2**: 165–174.

Alpers K, Werber D, Frank C *et al.* (2009) Sorbitol-fermenting enterohaemorrhagic *Escherichia coli* O157:H- causes another outbreak of haemolytic uraemic syndrome in children. *Epidemiol Infect* **137**: 389–395.

Bettelheim KA, Whipp M, Djordjevic SP & Ramachandran V (2002) First isolation outside Europe of sorbitol-fermenting verocytotoxigenic *Escherichia coli* (VTEC) belonging to O group O157. *J Med Microbiol* **51**: 713–714.

Bielaszewska M, Prager R, Zhang W, Friedrich AW, Mellmann A, Tschape H & Karch H (2006) Chromosomal dynamism in progeny of outbreak-related sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM. *Appl Environ Microbiol* **72**: 1900–1909.

Brandal LT, Lindstedt BA, Aas L, Stavnes TL, Lassen J & Kapperud G (2007) Octaplex PCR and fluorescence-based capillary electrophoresis for identification of human diarrheagenic *Escherichia coli* and *Shigella* spp. *J Microbiol Methods* **68**: 331–341.

Bugarel M, Martin A, Fach P & Beutin L (2011) Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) *Escherichia coli* strains: a basis for molecular risk assessment of typical and atypical EPEC strains. *BMC Microbiol* **11**: 142.

Buvens G, Pierard D, Hachimi-Idrissi S & Lauwers S (2009) First sorbitol-fermenting Verocytotoxin-producing *Escherichia coli* O157: H- isolated in Belgium. *Acta Clin Belg* **64**: 59–64.

Calderwood SB, Auclair F, Donohue-Rolfe A, Keusch GT & Mekalanos JJ (1987) Nucleotide sequence of the Shiga-like toxin genes of *Escherichia coli*. *P Natl Acad Sci USA* **84**: 4364–4368.

Dowd SE & Williams JB (2008) Comparison of Shiga-like toxin II expression between two genetically diverse lineages of *Escherichia coli* O157:H7. *J Food Prot* **71**: 1673–1678.

Editorial Team (2006) *E. coli* O157 infections in the UK. *Euro Surveill* **11**: pii=2964. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2964

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

Feng PC, Monday SR, Lacher DW *et al.* (2007) Genetic diversity among clonal lineages within *Escherichia coli* O157: H7 stepwise evolutionary model. *Emerg Infect Dis* **13**: 1701–1706.

Ferens WA & Hovde CJ (2011) *Escherichia coli* O157:H7: animal reservoir and sources of human infection. *Foodborne Pathog Dis* **8**: 465–487.

Gyles CL (2007) Shiga toxin-producing *Escherichia coli*: an overview. *J Anim Sci* **85**: E45–E62.

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Hayashi T, Makino K, Ohnishi M *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.

Jakubczak A, Szych J & Januszkiewicz K (2008) Characterization of first sorbitol-fermenting shiga toxin-producing *Escherichia coli* O157:H- strain isolated in Poland. *Med Dosw Mikrobiol* **60**: 173–181.

Janka A, Bielaszewska M, Dobrindt U, Greune L, Schmidt MA & Karch H (2003) Cytolethal distending toxin gene cluster in enterohemorrhagic *Escherichia coli* O157:H- and O157: H7: characterization and evolutionary considerations. *Infect Immun* **71**: 3634–3638.

Janka A, Becker G, Sonntag AK, Bielaszewska M, Dobrindt U & Karch H (2005) Presence and characterization of a mosaic genomic island which distinguishes sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H- from *E. coli* O157:H7. *Appl Environ Microbiol* **71**: 4875–4878.

Jelacic JK, Damrow T, Chen GS, Jelacic S, Bielaszewska M, Ciol M, Carvalho HM, Melton-Celsa AR, O'Brien AD & Tarr PI (2003) Shiga toxin-producing *Escherichia coli* in Montana: bacterial genotypes and clinical profiles. *J Infect Dis* **188**: 719–729.

Karch H & Bielaszewska M (2001) Sorbitol-fermenting Shiga toxin-producing *Escherichia coli* O157:H(-) strains: epidemiology, phenotypic and molecular characteristics, and microbiological diagnosis. *J Clin Microbiol* **39**: 2043–2049.

Koitabashi T, Vuddhakul V, Radu S, Morigaki T, Asai N, Nakaguchi Y & Nishibuchi M (2006) Genetic characterization of *Escherichia coli* O157: H7/- strains carrying the stx2 gene but not producing Shiga toxin 2. *Microbiol Immunol* **50**: 135–148.

Laing CR, Buchanan C, Taboada EN, Zhang Y, Karmali MA, Thomas JE & Gannon VP (2009) In silico genomic analyses reveal three distinct lineages of *Escherichia coli* O157:H7, one of which is associated with hyper-virulence. *BMC Genomics* **10**: 287.

LeJeune JT, Abedon ST, Takemura K, Christie NP & Sreevatsan S (2004) Human *Escherichia coli* O157:H7 genetic marker in isolates of bovine origin. *Emerg Infect Dis* **10**: 1482–1485.

Lindstedt BA (2011) Genotyping of selected bacterial enteropathogens in Norway. *Int J Med Microbiol* **301**: 648–653.

Matsumoto M, Suzuki M, Takahashi M, Hirose K, Minagawa H & Ohta M (2008) Identification and epidemiological description of enterohemorrhagic *Escherichia coli* O157 strains producing low amounts of Shiga toxin 2 in Aichi Prefecture, Japan. *Jpn J Infect Dis* **61**: 442–445.

Maurer JJ, Schmidt D, Petrosko P, Sanchez S, Bolton L & Lee MD (1999) Development of primers to O-antigen biosynthesis genes for specific detection of *Escherichia coli* O157 by PCR. *Appl Environ Microbiol* **65**: 2954–2960.

Mellmann A, Lu S, Karch H, Xu JG, Harmsen D, Schmidt MA & Bielaszewska M (2008) Recycling of Shiga toxin 2 genes in sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM. *Appl Environ Microbiol* **74**: 67–72.

Moreira CN, Pereira MA, Brod CS, Rodrigues DP, Carvalhal JB & Aleixo JA (2003) Shiga toxin-producing *Escherichia coli* (STEC) isolated from healthy dairy cattle in southern Brazil. *Vet Microbiol* **93**: 179–183.

Norwegian Institute of Public Health (2010) Årsrapport Matbårne infeksjoner og utbrudd i 2009. [Annual report Foodborne infections and outbreaks 2009] [in Norwegian] [http://www.fhi.no/eway/default.aspx?pid=233&trg=MainLeft_5583&MainArea_5661=5583:0:15,1134:1:0:0:::0:0&MainLeft_5583=5603:84108::1:5699:35:::0:0].

Norwegian Institute of Public Health (2011) Årsrapport Matbårne infeksjoner og utbrudd i 2010. [Annual report Foodborne infections and outbreaks 2010] [in Norwegian] [http://www.fhi.no/eway/default.aspx?pid=233&trg=MainLeft_5583&MainArea_5661=5583:0:15,1359:1:0:0:::0:0&MainLeft_5583=5603:90234::1:5693:11:::0:0].

Ogura Y, Ooka T, Asadulghani M *et al.* (2007) Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes. *Genome Biol* **8**: R138.

Ogura Y, Ooka T, Iguchi A *et al.* (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *P Natl Acad Sci USA* **106**: 17939–17944.

Orth D, Grif K, Dierich MP & Wurzner R (2007) Variability in tellurite resistance and the ter gene cluster among Shiga toxin-producing *Escherichia coli* isolated from humans, animals and food. *Res Microbiol* **158**: 105–111.

Persson S, Olsen KE, Ethelberg S & Scheutz F (2007a) Subtyping method for *Escherichia coli* shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *J Clin Microbiol* **45**: 2020–2024.

Persson S, Olsen KE, Scheutz F, Krogfelt KA & Gerner-Smidt P (2007b) A method for fast and simple detection of major diarrhoeagenic *Escherichia coli* in the routine diagnostic laboratory. *Clin Microbiol Infect* **13**: 516–524.

Plunkett G, Rose DJ, Durfee TJ & Blattner FR (1999) 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.

Rosser T, Dransfield T, Allison L, Hanson M, Holden N, Evans J, Naylor S, La RR, Low JC & Gally DL (2008) Pathogenic potential of emergent sorbitol-fermenting *Escherichia coli* O157:NM. *Infect Immun* **76**: 5598–5607.

Rump LV, Strain EA, Cao G, Allard MW, Fischer M, Brown EW & Gonzalez-Escalona N (2011) Draft genome sequences of six *Escherichia coli* isolates from the stepwise model of emergence of *Escherichia coli* O157:H7. *J Bacteriol* **193**: 2058–2059.

Russmann H, Schmidt H, Heesemann J, Caprioli A & Karch H (1994) Variants of Shiga-like toxin II constitute a major toxin component in *Escherichia coli* O157 strains from patients with haemolytic uraemic syndrome. *J Med Microbiol* **40**: 338–343.

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Schmidt H (2001) Shiga-toxin-converting bacteriophages. *Res Microbiol* **152**: 687–695.

Szabady RL, Lokuta MA, Walters KB, Huttenlocher A & Welch RA (2009) Modulation of neutrophil function by a secreted mucinase of *Escherichia coli* O157:H7. *PLoS Pathog* **5**: e1000320.

Taylor DE, Rooker M, Keelan M, Ng LK, Martin I, Perna NT, Burland NT & Blattner FR (2002) Genomic variability of O islands encoding tellurite resistance in enterohemorrhagic *Escherichia coli* O157:H7 isolates. *J Bacteriol* **184**: 4690–4698.

Unkmeir A & Schmidt H (2000) Structural analysis of phage-borne stx genes and their flanking sequences in shiga toxin-producing *Escherichia coli* and Shigella dysenteriae type 1 strains. *Infect Immun* **68**: 4856–4864.

Wagner PL, Neely MN, Zhang X, Acheson DW, Waldor MK & Friedman DI (2001) Role for a phage promoter in Shiga toxin 2 expression from a pathogenic *Escherichia coli* strain. *J Bacteriol* **183**: 2081–2085.

Zhang Y, Laing C, Zhang Z *et al.* (2010) Lineage and host source are both correlated with levels of Shiga toxin 2 production by *Escherichia coli* O157:H7 strains. *Appl Environ Microbiol* **76**: 474–482.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Primers used for PCR and sequencing.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

S1. Primers used for PCR and sequencing.

Primer name	Sequence	Application	Reference
q933	VIC-CGGAGGGGATTGTTGAAGGC	PCR	Unkmeir et al., 2000
q21	NED-GAAATCCTCAATGCCTCGTTG	PCR	LeJeune et al., 2004
slt2s-2	GTTATTATCCTACGCTCGGCCCTT	PCR and sequencing	Matsumoto et al., 2008
595	CCGAAGAAAAACCCAGTAACAG	PCR and sequencing (of inverse PCR product)	Unkmeir et al., 2000
PS8-rev	GAGTGACGACTGATTTGCATTCC	Inverse PCR and sequencing	Persson et al., 2007
PS7-rev	TGTCAGATAACTGGCGACAGGC	Inverse PCR and sequencing	Persson et al., 2007
slt2s-2rev	AAGGGCCGAGCGTAGGATAATAAC	Sequencing of inverse PCR product	Matsumoto et al., 2008
q933-rev	GCCTTCAACAATCCCCTCCG	Sequencing of inverse PCR product EDL933	Unkmeir et al., 2000
qSF-1	AAGCAATCAGCGGAGCCATCAC	Sequencing of inverse PCR product SF O157	This study[1]
qSF-2	ATGGGCCTGAGGTTTGCTTGTG	Sequencing of inverse PCR product SF O157	This study[1]
qSF-3	TACCCCCAGCACAGACAAACATCC	Sequencing of inverse PCR product SF O157	This study[1]
stx_8-F	AGCGCTGGAACGTTCCGGAATGCAAATCAGTCGT	PCR and sequencing	Dowd and Williams, 2008
stx_8-R	CGGCCTGGTACCACCGAACCCGTGTCGATAATGAAT	PCR and sequencing	Dowd and Williams, 2008
SF1-F	ATACCGTGGCATTTGAAGAGAAGT	PCR and sequencing	This study[2]
SF1-R	TTTTTAGCAGCCAGTCGTCCA	PCR and sequencing	This study[2]

	TC		
SF2-F	TGCGAATATCAAAAAGACAGA	PCR and sequencing	This study[2]
SF2-R	CCAGACCCGCCATAACA	PCR and sequencing	This study[2]
nySF3-F	CACGTTAACCGGAAAAGGGAAAAA	PCR and sequencing	This study[2]
nySF3-R	GCTCTGTGGCGGGGAAATACG	PCR and sequencing	This study[2]
nySF3-sekvF	GCAAAAGCGGTATGAGC	Sequencing	This study[2]
nySF3-sekvR	ATTCCAGGCAGTCGGCGTTGATT	Sequencing	This study[2]
nySF4-F	CCAGTGGACGGGTGATGAGGATT	PCR and sequencing	This study[2]
nySF4-R	CGGTTTTTCGCACGGATGTTTC	PCR and sequencing	This study[2]
nySF4-sekvF	CGAAAGATGCCGGGTATGAGG	Sequencing	This study[2]
nySF4-sekvR	GGCATCCACTATCTGTTTT	Sequencing	This study[2]
SF5-F	CGTATTTCCCCGCCACAGAG	PCR and sequencing	This study[2]
SF5-R	CACCGGTTTATGCGTCCACAC	PCR and sequencing	This study[2]
SF6-F	CCCGGGTGGCAGAAGAG	PCR and sequencing	This study[2]
SF6-R	CGATGAGCTAATGGCGGTATGTGA	PCR and sequencing	This study[2]
SF7-F	GAGCAGACGGTCAGGGAAGT	PCR and sequencing	This study[2]
SF7-R	TTGTTGAGTCGAAAAGTCTAT	PCR and sequencing	This study[2]
SF8-F	TCAGCCAAAAGGAACACC	PCR and sequencing	This study[2]
SF8-R	ATTAACGCCAGATATGATGAAA	PCR and sequencing	This study[2]

SF9-F	TTAATACGGCAACAAATACTT	PCR and sequencing	This study[2]
SF9-R	ACGCGCCCCCTGATGATGG	PCR and sequencing	This study[2]
SF10-F	CTTACGCTTCAGGCAGATAC	PCR and sequencing	This study[2]
SF10-R	GTTACCCACATACCACGAA	PCR and sequencing	This study[2]
SF11-F	AGTGAAGGTTGACGGGAAAG AATA	PCR and sequencing	This study[2]
SF11-R	ACACCAACCCGGACAGGCGT AATA	PCR and sequencing	This study[2]
SF11-2F	AGGCTCCGGATTTGCTGAAGT GC	PCR and sequencing	This study[3]
SF11-2R	CCGGCCGGGATAATATTGTGA GTA	PCR and sequencing	This study[3]
[1]Primers were designed by primer walking			
[2]Primer sequences were based on NCBI accession number AP010960 O111:H- (Ogura et al., 2009)			
[3]Primer sequenced were based on NCBI accession number AE005174 O157:H7 (Perna et al., 2001)			

Paper II

Journal of Clinical Microbiology

PCR-Based Detection and Molecular Characterization of Shiga Toxin-Producing Escherichia coli Strains in a Routine Microbiology Laboratory over 16 years

K. Haugum, L. T. Brandal, B.-A. Lindstedt, A. L. Wester, K. Bergh and J. E. Afset

J. Clin. Microbiol. **2014, 52(9):3156. DOI: 10.1128/JCM.00453-14.**
Published Ahead of Print 11 June 2014.

Updated information and services can be found at:
http://jcm.asm.org/content/52/9/3156

These include:

SUPPLEMENTAL MATERIAL Supplemental material

REFERENCES This article cites 47 articles, 23 of which can be accessed free at: http://jcm.asm.org/content/52/9/3156#ref-list-1

CONTENT ALERTS Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: **http://journals.asm.org/site/misc/reprints.xhtml**
To subscribe to to another ASM Journal go to: **http://journals.asm.org/site/subscriptions/**

Journals.ASM.org

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

JCM
Journals.ASM.org

PCR-Based Detection and Molecular Characterization of Shiga Toxin-Producing *Escherichia coli* Strains in a Routine Microbiology Laboratory over 16 years

K. Haugum,[a] L. T. Brandal,[b] B.-A. Lindstedt,[c] A. L. Wester,[b] K. Bergh,[a,d] J. E. Afset[a,d]

Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway[a]; Department of Foodborne Infections, Norwegian Institute of Public Health, Oslo, Norway[b]; Gene Technology Section, Akershus University Hospital, Lørenskog, Norway[c]; Department of Medical Microbiology, St. Olavs University Hospital, Trondheim, Norway[d]

Shiga toxin-producing *Escherichia coli* (STEC) is a heterogeneous group of bacteria causing disease ranging from asymptomatic carriage and mild infection to hemolytic uremic syndrome (HUS). Here, we describe patients with STEC infection and characterize the STEC strains detected in our laboratory by use of PCR for stx_1, stx_2, and *eae* from 1996 through 2011. Patient information was collected from referral forms and from the Norwegian Surveillance System for Communicable Diseases. STEC isolates were characterized with respect to serogroup or serotype, selected potential virulence genes, and multilocus variable-number tandem-repeat analysis (MLVA) genotype. STEC strains were isolated from 138 (1.09%) of 12,651 patients tested. STEC strains of serogroups O26, O103, O121, O145, and O157 were the most frequent. These serogroups, except non-sorbitol-fermenting O157, were also the most frequent among the 11 patients (all ≤5 years old) who developed HUS. Twenty-four STEC strains were classified as being HUS associated based on an epidemiological link to a HUS case, including an MLVA genotype identical to that of the STEC strain. The age of the patient (≤5 years) and the genes *eae* and stx_{2a} were significantly associated with HUS-associated STEC ($P < 0.05$ for each parameter), while stx_1 was associated with non-HUS-associated STEC ($P < 0.05$). All of the potential virulence genes analyzed, except *ehxA*, were significantly more frequent among HUS-associated than non-HUS-associated strains ($P < 0.05$ for each gene). However, these genes were also present in some non-HUS-associated STEC strains and could therefore not reliably differentiate between HUS-associated and non-HUS-associated STEC strains.

Shiga toxin-producing *Escherichia coli* (STEC) was recognized as a cause of bloody diarrhea and hemolytic uremic syndrome (HUS) for the first time in two independent studies in 1982 (1, 2). Later this pathogen was found to be the main cause of diarrhea-associated HUS with a high number of cases worldwide. Non-sorbitol-fermenting STEC (NSF) O157:H7 was the first STEC serotype that was isolated in association with HUS and has been the most frequently reported cause of diarrhea-associated HUS (3). However, STEC strains of other serogroups like O26, O103, O111, O121, and O145 have also been recognized to cause severe disease and outbreaks (4, 5).

Shiga toxins 1 and 2 (Stx1 and Stx2) are essential virulence factors of STEC. The term STEC is used to describe any *E. coli*-producing Shiga toxin, whereas the term enterohemorrhagic *Escherichia coli* (EHEC) is often used to describe the subset of STEC strains responsible for causing hemorrhagic colitis and HUS (3). Shiga toxins are encoded by the stx_1 and stx_2 genes located in lambdoid bacteriophages integrated into the bacterial host genome. The toxins exist as various subtypes, in which the Stx2 subtypes Stx2a, Stx2c, and Stx2d are more often associated with HUS than are the other Stx subtypes (6–9). In addition to the Shiga toxins, most STEC strains possess the locus of enterocyte effacement (LEE) pathogenicity island in which the virulence gene *eae* encoding the adherence factor intimin is located (3, 10).

In addition to stx_1, stx_2, and *eae*, the presence and absence of various other genes have been investigated as potential virulence markers for HUS and outbreaks. Karmali et al. reported that several genes located on the genomic O island OI-122 were present in 60 to 100% of STEC strains of serotypes highly associated with severe disease and outbreaks, while the same genes were detected in only 0 to 15% of strains of serotypes not associated with severe disease or outbreaks (11). In addition, the presence of non-LEE-encoded (*nle*) genes from the genomic O islands OI-71 and OI-57 has been associated with STEC virulence (12–14). Pathogenic STEC also harbors a large virulence plasmid (frequently termed pO157) encoding factors involved in STEC virulence (3, 15).

Since STEC was first recognized as a cause of diarrhea-associated HUS, the microbial detection of this pathogen has in most laboratories until recently relied on culturing on sorbitol-MacConkey agar (SMAC), which is a selective medium for NSF O157:H7 (16), with subsequent agglutination with anti-O157 antiserum. While a selective medium for NSF O157:H7 STEC is a sensitive method for detection of this specific STEC serotype, a major drawback is that other STEC serotypes and sorbitol-fermenting (SF) O157:H- not will be detected. Therefore, a more suitable strategy for the diagnosis of STEC infection is to combine culture and nonculture methods detecting both O157 and non-O157 STEC serotypes (17, 18). In recent years, many diagnostic

Received 14 February 2014 Returned for modification 2 April 2014
Accepted 6 June 2014
Published ahead of print 11 June 2014
Editor: B. A. Forbes
Address correspondence to J. E. Afset, jan.afset@ntnu.no.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.00453-14.
Copyright © 2014, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JCM.00453-14

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

TABLE 1 Characteristics of *stx*-negative strains, isolated at St. Olavs Hospital, Trondheim, Norway, during the 1996-2011 period, classified as STEC-lost Shiga toxin in this study

Strain	stx_2 subtype	*eae*	Serotype	Reason the strain was included in the study
St. Olav49	STEC-LST[a]	+	O103:H25	HUS[b]
St. Olav59	STEC-LST	+	O103:H25	HUS
St. Olav75	2a[c]	+	SF O157:H-	Outbreak investigation
St. Olav77	STEC-LST	+	SF O157:H-	HUS
St. Olav84	STEC-LST	+	SF O157:H-	Outbreak investigation
St. Olav97	STEC-LST	+	O145:H?[d]	Outbreak investigation
St. Olav131	STEC-LST	+	O103:H25	BD[e]
St. Olav154	2b	−	0[f]	Previously positive for stx_{2b}
St. Olav156	STEC-LST	+	0[f]	BD
St. Olav165	STEC-LST	+	O145:H28	BD

[a] *eae*-positive *E. coli* isolate classified as STEC that may have lost its *stx* genes (STEC-lost Shiga toxin [LST]).
[b] HUS, hemolytic uremic syndrome.
[c] Strain previously positive for stx_2, which after storage was found to be stx_2 negative.
[d] H?, motile but unknown H-type.
[e] BD, bloody diarrhea.
[f] Serotype 0, the strains did not belong to any of the serogroups tested for in the study.

laboratories have switched to PCR for detection of stx_1 and stx_2 and other putative virulence genes (e.g., *eae* and *ehxA*) in STEC strains. Although the use of PCR improves the detection of non-O157 STEC serotypes which may be the cause of HUS, it may also lead to detection of a range of STEC strains with a probable low potential for causing severe disease in humans (10, 17, 19). In many cases, we therefore still do not know how to differentiate between high- and low-risk STEC strains, and it is a challenge to make a reliable assessment of the clinical and public health risk related to the diagnosis of non-O157 STEC infections.

The aim of this study was to present the results of PCR-based diagnosis of STEC infection from patient stool specimens during the period 1996-2011 and to search for differences between HUS and HUS-associated compared to non-HUS-associated STEC isolates based on serotypes and analysis of selected potential virulence genes.

MATERIALS AND METHODS

Bacterial strains and clinical information. All STEC isolates included in the present study were retrieved from patient stool specimens in the years 1996 through 2011 at the Department of Medical Microbiology, St. Olavs Hospital, Trondheim, Norway. From 1996 to 2011 the laboratory routine was to analyze stool specimens from children <2 years old for stx_1 and stx_2 (and *eae* from 2000) irrespective of clinical information by PCR and to analyze stool specimens from patients in age groups >2 years old if there was information on HUS or bloody diarrhea. In addition, specimens from persons epidemiologically associated with a HUS case or a STEC outbreak were analyzed for STEC. Based on data from the laboratory information system, stx_1 and/or stx_2 was detected in mixed stool cultures from 150 patients during the study period. Among these, 20 patients were excluded from the study because the laboratory did not succeed in obtaining STEC isolates in pure cultures, whereas for the remaining 130 patients, STEC isolates were identified in pure cultures. Another eight *stx*-negative (*eae*-positive) *E. coli* isolates were included in the study because they were isolated from patients with HUS or bloody diarrhea or were epidemiologically linked to a HUS case and were of the same MLVA genotype as the STEC isolate from that case (Table 1). STEC strains that have lost *stx* genes are often termed EHEC/STEC-lost Shiga toxin (LST) (20). In total, 138 strains were included in the study.

In Norway, STEC infection is notifiable to the Norwegian Surveillance System for Communicable Diseases (MSIS) at the Norwegian Institute of Public Health (NIPH), where clinical information on the patients and results from laboratory analyses of the bacterial strains are stored. We collected data from the MSIS on clinical symptoms (HUS, bloody diarrhea, diarrhea, or no symptoms), age, and sex and correlated these data with laboratory results.

The study was approved by the Regional Committee for Medical and Health Research Ethics, REC South-East (REC number 2011/2314).

Primary detection and identification of STEC. stx_1, stx_2, and *eae* (from the year 2000) were detected by a two-step procedure where PCRs for the stx_1, stx_2, and *eae* genes first were done in mixed cultures from a stool specimen and thereafter repeated on subcultures of discrete colonies from positive specimens with the aim of identifying STEC strains in pure cultures. STEC isolate culturing was done by standard methods, including SMAC agar, and *E. coli* was identified by standard biochemical tests (API 10S/20E; bioMérieux, Marcy l'Etoile, France). During the period 1996-2004, screening for stx_1 and stx_2 was performed using primers and amplification conditions as described by Brian et al. (21). In 2004, conventional PCRs for stx_1 and stx_2 were replaced by multiplex real-time PCRs (for primers, see Table S1 in the supplemental material). DNA isolation methods, amplification conditions, PCR reagents, and PCR instruments varied in the years 2004-2011.

PCR for *eae* was done using the AE13 and AE14 primers, and amplification conditions were as described by Gannon et al. (22) from 2000 to 2004 and as described by Nielsen and Andersen (23) from 2004 to 2008. Thereafter detection of *eae* was done by real-time PCR with the primers described in Table S1 in the supplemental material. Confirmation of stx_1, stx_2, and *eae* was done at the National Reference Laboratory for Enteropathogenic Bacteria (NRL) at the NIPH (24, 25).

Serotyping. Initial serogrouping was performed with O antisera using polyspecific anti-coli I, II, and III and monospecific O antisera for the O serogroups O26, O103, O111, O145, and O157, as described by the manufacturer (Sifin, Germany). Later, more extensive serotyping was done at the NRL, NIPH, using monospecific O:K and H antisera covering altogether 44 O serogroups, including O26, O103, O111, O121, O145, and O157 and 8 H antigens (in-house antisera and antisera from Sifin and SSI, Denmark).

***stx*$_2$ subtyping and MLVA genotyping.** The stx_2 subtype was determined at the NRL, NIPH, using modifications of previously published methods for PCR-restriction fragment length polymorphism (RFLP) and sequencing (8, 26, 27) and by PCR (9).

Two MLVA protocols, one for the O157 serogroup and one for all *E. coli* isolates, were used for MLVA typing (28–30) at the NRL, NIPH.

Verification of stx_1, stx_2, and *eae* and detection of potential virulence genes. To verify the primary PCR results, we repeated PCRs for stx_1, stx_2, and *eae* for all strains included in the study. For PCR analyses, bacterial strains were grown overnight on MacConkey agar. One colony of bacterial cells was suspended in 100 μl lysis buffer (50 mM KCl, 10 mM Tris-HCl

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

TABLE 2 Age distribution and other characteristics of patients with STEC infections diagnosed at St. Olavs Hospital during the 1996-2011 period

Age group (yr)	No. (%) of patients:		Positive rate	Sex (no.)		Clinical presentation in patients with STEC (no.):				
	With STEC	Tested		Female	Male	HUS[a]	BD[b]	Diarrhea	Asymptomatic[c]	ND[d]
<2	50 (36.3)	6,860 (54.2)	0.73	25	25	6	5	25	5	9
2–4	32 (23.2)	1,301 (10.3)	2.46	17	15	4	0	15	6	7
5–9	9 (6.5)	344 (2.7)	2.62	2	7	1	1	1	2	4
≥10	47 (34.0)	4,146 (32.8)	1.13	25	22	0	3	27	7	10
Total	138 (100.0)	12,651 (100.0)	1.09	69	69	11	9	68	20	30

[a] HUS, hemolytic uremic syndrome.
[b] BD, bloody diarrhea.
[c] Some of the strains were from persons tested in outbreak investigations.
[d] ND, no clinical information. Some of these strains were from outbreak investigations.

[pH 8.3], 2.5 mM $MgCl_2$, 0.1 mg/ml gelatin, 0.45% NP-40, and 0.45% Tween 20) and 100 μl Tris-EDTA (TE) buffer solution (pH 7.4) and boiled for 15 min at 95°C. After centrifugation at 13,000 rpm for 1 min, the supernatant was used directly for PCR analysis.

Real-time PCRs for stx_1, stx_2, and *eae* were done using the primers described in Table S1 in the supplemental material and PerfeCTa Multiplex qPCR Supermix, UNG (Quanta Biosciences, Gaithersburg, MD, USA) as described by the manufacturer. Real-time PCR was performed in a CFX instrument (Bio-Rad, Hercules, CA, USA), in a 20-μl volume with cycling conditions for stx_1 and stx_2 as follows: 95°C for 3 min and then 40 cycles with denaturation at 95°C for 10 s and annealing at 58°C for 10 s before elongation at 72°C for 10 s. The PCR for *eae* was done using the following cycling conditions: 95°C for 3 min and then 40 cycles with denaturation at 95°C for 10 s and annealing at 50°C for 10 s before elongation at 72°C for 10 s.

Analyses for the *ehxA*, *ent*/*espL2*, *nleB*, *nleE*, *nleF*, *nleH1-2*, and *nleA* genes were done using the primers described by Bugarel et al. (31). In addition, PCR for the *espK* gene was done using the primers described by Bugarel et al. (12), and the *efa-1*/*lifA* gene was analyzed with the 88AT and 88TN primers from Nicholls et al. (32). A subset of strains was also tested for *efa-1*/*lifA* using an alternative *efa-1* primer pair (see Table S1 in the supplemental material). The PCRs were done in singleplex in a 20-μl volume using the SsoFast EvaGreen Supermix (Bio-Rad), as described by the manufacturer. The cycling conditions for the *nleB* and *nleE* genes were 98°C for 2 min and then 40 cycles with 95°C for 5 s and 60°C for 5 s, and the cycling conditions for *ehxA*, *ent*/*espL2*, *nleF*, *nleH1-2*, *nleA*, *espK*, and *efa-1*/*lifA* were 95°C for 3 min and then 40 cycles with 95°C for 10 s and 57°C for 30 s.

HUS-associated strains. A STEC strain was classified as HUS associated if it either was isolated from a patient with HUS or was epidemiologically linked to a HUS case and was of the same MLVA genotype as the STEC isolate from that case.

Statistical methods. Fisher's exact test was used for statistical calculations. A P value of ≤0.05 was considered statistical significant.

Cluster analysis of virulence genes with construction of dendrograms was performed by BioNumerics version 6.6 (Applied Maths NV, Sint-Martens-Latem, Belgium) using the Dice correlation and the unweighted-pair group method using average linkages (UPGMA).

RESULTS

The present study included 138 patients among a total of 12,651 patients tested in whom STEC infection had been diagnosed by PCR, and STEC strains had successfully been isolated in pure cultures (positive rate, 1.09%) in the period 1996-2011 at the Department of Medical Microbiology, St. Olavs Hospital, Trondheim, Norway. Stool specimens from children <2 years old were most frequently tested for STEC, and this age group was also the most common among the 138 patients diagnosed with STEC infection (36.3%) (Table 2). Sixty-nine of the patients with STEC infections were female and 69 were male. Eleven patients had HUS, while bloody diarrhea was recorded for 9 patients, and nonbloody diarrhea was recorded for 68 patients. All HUS patients were ≤5 years old ($P = 0.007$) (Table 2). Six of the HUS patients were female and 5 were male (Table 3).

The STEC strains included in the present study had the following distribution of serogroups: O157, $n = 29$ (21.0%); other common STEC serogroups, $n = 69$ (50.0%), including O145 ($n = 28$), O103 ($n = 21$), and O26 ($n = 11$); and less common STEC serogroups, $n = 17$ (12.3%). Twenty-three (16.7%) strains, one of which was Orough, did not belong to any of the serogroups tested (Table 4).

A total of 128 strains contained the stx_1 and/or stx_2 gene, and 108 strains contained *eae* (Table 5). A combination of stx_1 and stx_2 was found in 21 (15.2%) strains, while stx_1 only was found in 57 (41.3%) strains and stx_2 only in 50 (36.2%) strains. The stx_2 subtypes most frequently detected were stx_{2a} ($n = 36$) and stx_{2c} ($n = 18$) (Table 5). Two strains which had previously been confirmed to be stx_2 positive were negative for *stx* after frozen storage (Table 1).

PCRs for *ehxA*, *nleB*, *nleE*, *ent*, *efa-1*/*lifA*, *nleA*, *nleF*, *nleH1-2*,

TABLE 3 Characteristics of STEC strains from patients with hemolytic uremic syndrome and information from the patients with STEC infections diagnosed at St. Olavs Hospital during the 1996-2011 period

Strain	yr	Sex[a]	Age (yr)	Serotype	stx_2 subtype
St. Olav26	2002	F	4	O26:H-	2a
St. Olav49	2006	F	1	O103:H25	—[b]
St. Olav56	2006	M	2	SF O157:H-	2a
St. Olav59	2006	M	1	O103:H25	—
St. Olav77	2008	F	1	SF O157:H-	—
St. Olav80	2009	F	3	SF O157:H-	2a
St. Olav81	2009	M	5	SF O157:H-	2a
St. Olav91	2009	F	1	O121:H19	2a
St. Olav100	2009	M	1	O145:H?[c]	2a
St. Olav164	2011	F	2	O145:H?	2a
St. Olav166	2011	M	0	SF O157:H-	2a

[a] F, female; M, male.
[b] Strains were negative for stx_{2a}. The O103:H25 strains (St. Olav49 and St. Olav59) were part of a national outbreak in 2006 where stx_{2a} was identified in some of the other strains included in the same MLVA genotype cluster (50). St. Olav77 was part of a small family outbreak with St. Olav75 (see Table 1).
[c] H?; motile but unknown H-type.

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

TABLE 4 Serogroups of 138 HUS-associated and non-HUS-associated STEC strains isolated by PCR and culture at St. Olavs Hospital, Trondheim, Norway, during the 1996-2011 period

Serogroup	No. (%) of STEC strains		
	HUS associated	Non-HUS associated	Total
O145	7 (29.1)	21 (18.4)	28 (20.3)
O103	2 (8.4)	19 (16.6)	21 (15.3)
O157	0 (0)	20 (17.5)	20 (14.5)
O26	1 (4.2)	10 (8.8)	11 (7.9)
SF O157	9 (37.5)	0 (0)	9 (6.5)
O121	5 (20.8)	4 (3.5)	9 (6.5)
Other[a]	0 (0)	40 (35.1)	40 (28.9)
Total	24 (100)	114 (100)	138 (100)

[a] Other serogroups: O2 ($n = 1$), O76 ($n = 1$), O91 ($n = 1$), O104 ($n = 3$), O111 ($n = 1$), O113 ($n = 3$), O117 ($n = 1$), O118 ($n = 1$), O119 ($n = 1$), O128 ($n = 3$), O177 ($n = 1$), and unknown O serogroups ($n = 23$), of which one strain was Orough.

and *espK* revealed that *ehxA* was the most frequent ($n = 123$, 89.1%), whereas *nleA* was the least frequent ($n = 87$, 63.0%) in the 138 STEC strains examined (Table 5). All of the potential virulence genes were found among strains of serogroups O145, O103, O157, O26, O121, SF O157, and O111 (see Table S2 in the supplemental material), whereas 11 strains with other serogroups contained none of the genes analyzed in the study.

The STEC strains isolated from patients with HUS belonged to the serotypes SF O157:H-, O145:H? (H unknown), O103:H25, O26:H-, and O121:H19 (Table 3). Another 13 STEC strains were epidemiologically linked to an HUS case and of the same MLVA genotype as the STEC strain isolated from that case. Altogether 24 strains were therefore classified as HUS associated. Among these strains SF O157:H- and O145:H? were the most common (Table 4). The age of the patient (≤5 years old) was significantly associated with HUS-associated STEC infection, as 20 of the 24 HUS-associated strains were isolated from children within this age group ($P = 0.035$). The *eae* gene was present in all of the HUS-associated strains compared with its presence in 84/114 (73.7%) non-HUS-associated strains ($P = 0.002$) (Table 5). stx_2 was significantly more frequent among HUS-associated than non-HUS-associated strains ($P = 0.013$) (Table 5), especially the stx_{2a} subtype, which was present in 18/24 HUS-associated strains compared with 22/114 non-HUS-associated strains ($P < 0.0001$) (Table 5). In contrast, stx_1 was not detected in any of the HUS-associated strains, but was present in 78 of the non-HUS-associated strains ($P < 0.0001$) (Table 5). The genes *ehxA*, *nleB*, *nleE*, *ent*, *efa-1/lifA*, *nleA*, *nleF*, and *nleH1-2* were present in all HUS-associated strains, whereas *espK* was present in all but one of these strains. All of the potential virulence genes analyzed, except *ehxA*, were significantly more frequent among HUS-associated than non-HUS-associated strains ($P < 0.05$ for each gene) (Table 5).

By cluster analysis of potential virulence genes, *eae*-positive and *eae*-negative STEC strains were separated into two main clusters (Fig. 1; see also Fig. S1 in the supplemental material). One exception was an

TABLE 5 Virulence genes identified by PCR analysis and stx_2 subtypes of 138 HUS-associated and non-HUS-associated STEC strains isolated at St. Olavs Hospital, Trondheim, Norway, during the 1996-2011 period

Virulence gene	No. ([%]) of genes in:			
	HUS-associated STEC ($n = 24$)	Non-HUS associated STEC ($n = 114$)	Total	*P*
stx_1	0 (0)	57 (50)	57 (41.3)	<0.0001
$stx_1 + stx_2$	0 (0)	21 (18.4)	21 (15.2)	0.02
stx_2	18 (75)[a]	32 (28)[b]	50 (36.2)	<0.0001
eae	24 (100)	84 (74)	108 (78.3)	0.002
ehxA	24 (100)	99 (87)	123 (89.1)	0.073
nleB	24 (100)	82 (72)	106 (76.8)	0.001
nleE	24 (100)	81 (71)	105 (76.1)	0.001
ent	24 (100)	82 (72)	106 (76.8)	0.001
efa-1/lifA	24 (100)	73 (64)	97 (70.3)	0.0001
nleA	24 (100)	63 (55)	87 (63)	<0.0001
nleF	24 (100)	84 (74)	108 (78.3)	0.002
nleH1-2	24 (100)	85 (75)	109 (79)	0.004
espK	23 (96)	83 (73)	106 (76.8)	0.002
Total	24 (100)	114 (100)	138 (100)	
stx_2 subtype				
stx_{2a}	18 (75)[a]	18 (15.8)	36 (50.7)	$P < 0.0001$
$stx_{2a} + stx_{2c}$	0 (0)	4 (3.5)	4 (5.6)	ND[c]
stx_{2b}	0 (0)	9 (7.9)[b]	9 (12.7)	ND
stx_{2c}	0 (0)	18 (15.8)	18 (25.4)	ND
stx_{2d}	0 (0)	2 (1.8)	2 (2.8)	ND
stx_{2g}	0 (0)	1 (0.9)	1 (1.4)	ND
ND	0 (0)	1 (0.9)[d]	1 (1.4)	ND
Total	18 (75)	53 (47)	71 100	

[a] This number does not include one SF O157:H- strain (St. Olav75; see Table 1) previously found to be positive for stx_{2a}.
[b] This number does not include one strain of unknown serotype (St. Olav154; see Table 1) previously found to be positive for stx_{2b}.
[c] ND, not determined.
[d] The stx_2 phage was lost in the STEC strain at arrival at the Norwegian Institute of Public Health and therefore was not stx_2 subtyped.

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

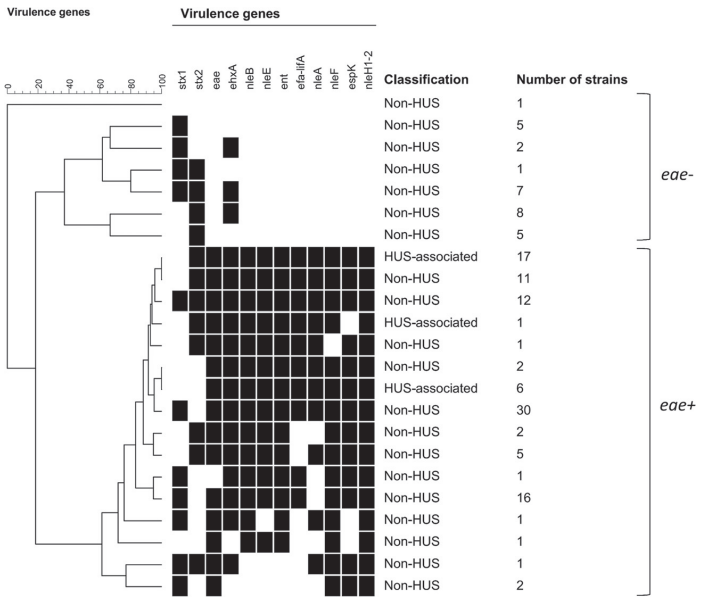


FIG 1 Cluster analysis of potential virulence genes in STEC strains. *eae*-negative and *eae*-positive STEC strains were mainly separated into two clusters. One exception was one eae-negative strain that clustered among the *eae*-positive strains due to the presence of some of the potential virulence genes. All HUS-associated strains clustered among the *eae*-positive strains and harbored all of the potential virulence genes investigated in the study. For further details (serotype, stx_2 subtype, etc.), see Fig. S1 in the supplemental material.

eae-negative, stx_1-positive strain (St. Olav12) that clustered with the group of *eae*-positive strains due to the presence of some of the potential virulence genes in this strain. In the cluster of 29 *eae*-negative STEC strains, *ehxA* was the only potential virulence gene present (17 strains) (Fig. 1; see also Fig. S1). All *eae*-positive strains harbored more than three of the potential virulence genes investigated (Fig. 1; see also Fig. S1). Although all the HUS-associated strains clustered among strains which were *eae* and stx_2 positive but stx_1 negative, non-HUS-associated strains were also found in the same cluster (Fig. 1; see also Fig. S1).

The 109 non-O157 STEC strains were distributed in 48 distinct MLVA genotypes (see Fig. S2 in the supplemental material). Thirty-one of these MLVA genotypes were represented by only one strain each, including one O26:H- strain from an HUS patient. The other 17 MLVA genotypes, with 78 non-O157 strains, were found in 2 to 15 strains, respectively. Some of these MLVA genotypes included STEC strains from local or national outbreaks. Among the 29 O157 and SF O157 STEC strains, 17 distinct MLVA genotypes were observed (see Fig. S3 in the supplemental material). Twelve of these MLVA genotypes were present in single strains, including two SF O157:H- strains from HUS patients, while five genotypes included two to five strains. For the two MLVA methods, some of the HUS-associated strains showed the same MLVA genotype as non-HUS-associated strains.

DISCUSSION

In this study, we present the results of STEC infection diagnosis in our laboratory based on PCRs for the stx_1, stx_2, and *eae* genes in cultures from stool samples in the years 1996-2011. During this period *stx* genes were detected in stool samples from 150 patients, and STEC or STEC-LST strains were isolated in pure cultures from 138 patients (Table 2). Similar to what has been reported elsewhere, the highest number of STEC infections was diagnosed in children <5 years old (4, 33). In our study, this may partly be explained by the routine of the laboratory to analyze all stool specimens from children <2 years old for STEC, while samples from older age groups were analyzed only if there was a clinical suspicion of HUS or bloody diarrhea or the samples were part of a STEC outbreak investigation. However, the fact that a high number of STEC infections were also detected in children 2 to 4 years old (Table 2) in whom tests for STEC were done only because of a specific suspicion, supports the notion that STEC infection is most common in young age groups.

As shown in Table 4, the STEC serogroups isolated most often in this study, including strains associated with HUS, belonged to the STEC serogroups frequently implicated in severe disease and outbreaks described elsewhere (5, 34). However, only 20 (14.5%) of the 138 STEC strains belonged to serotype NSF O157:H7, the only STEC serotype that is selected for by SMAC agar. Seventy-eight (56.5%) STEC strains belonged to other common STEC serogroups, including SF O157, that would easily have been missed on SMAC agar since they could not be differentiated from the majority of commensal *E. coli*, and 40 strains belonged to serogroups not common for STEC or unknown serogroups. Consequently, use of PCRs resulted in detection of a high number of non-O157 STEC strains, both of STEC serogroups that based on current knowledge may be viewed as high-risk strains (35), and of non-O157 STEC serogroups that most likely do not represent a high risk for HUS development.

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

More than half (55%) of the patients infected with SF O157:H- STEC developed HUS, and this serotype was the most common serotype isolated from HUS patients in this study (Table 3). These results are in line with previous reports suggesting that there is a high risk for development of HUS with SF O157 infection (36, 37). Six (54.5%) of the 11 HUS cases were caused by non-O157 STEC serotypes, while NSF O157, which is the most common STEC serotype causing HUS worldwide (3–5), was not the cause of any of the HUS cases in this study. During the same time period, this serotype was isolated from four cases of HUS in other parts of Norway (MSIS).

In this study, HUS-associated STEC strains contained the following characteristics: (i) all of them, except STEC-LST strains, carried stx_2 (stx_{2a}), (ii) all harbored *eae*, (iii) all but one contained the other nine potential virulence genes tested, and (iv) all belonged to STEC serogroups frequently associated with severe disease, many of them non-O157. Stx2a and intimin are important virulence factors in STEC strains that have been associated with severe disease (3, 38). Analysis of the presence or absence of virulence genes in this study revealed that *eae* and stx_{2a} were significantly more frequent in HUS-associated than non-HUS-associated strains, whereas STEC strains containing stx_1 were exclusively associated with non-HUS infection (Table 5). While the *ehxA* gene has been regarded as an important virulence marker in STEC infection and has been reported to be a marker of "typical EHEC" (14), it was the only potential virulence gene analyzed that was not significantly more frequent in HUS-associated than non-HUS-associated strains in this study. *ehxA* was also the sole potential virulence gene present among the subset of *eae*-negative STEC strains (Fig. 1; see also Fig. S1 in the supplemental material). However, although most of the potential virulence genes were significantly more present among HUS-associated than non-HUS-associated strains, several of the non-HUS-associated STEC strains contained a virulence gene profile similar to that for the group of HUS-associated strains. In particular, non-HUS-associated STEC strains of the serogroups O145, O103, O157 (NSF and SF), O26, O121, and O111 contained a high number of the potential virulence genes, which is in line with previous reports (39, 40). This made reliable differentiation between HUS-associated and non-HUS-associated STEC strains based on serotype and potential virulence genes impossible in this strain collection.

Two of the STEC strains were negative for *stx* after frozen storage. Although in general only *stx*-positive strains are regarded as STEC, it is well known that STEC may lose its *stx* encoding prophage, either in the course of an infection or upon handling in the laboratory (20, 41, 42).

Although *eae*- and stx_1-positive, stx_2-negative STEC strains have been isolated from patients with HUS (43, 44), there are to our knowledge no reports of outbreaks of severe disease with such bacteria. Furthermore, in line with reports from other regions in Norway (L. T. Brandal, A. L. Wester, H. Lange, I. Løbersli, B.-A. Lindstedt, G. Kapperud, and L. Vold, unpublished data) and other countries (5, 45–47), our results support the notion that infections with STEC strains that do not belong to common STEC serogroups and lack *eae* and/or stx_{2a} even if stx_1 is present represent a low risk for HUS development. Therefore, based on the results from the present study, it seems safe to suggest a classification of stx_1-positive, stx_2-negative STEC strains with a low risk for HUS development.

In the current study, it was difficult to assess the risk profile of STEC strains with a serotype and virulence profile similar to that of HUS-associated strains. It may be that such strains are actually virulent and should be interpreted as STEC with a high risk for HUS development. Alternatively, there may be other bacterial characteristics not analyzed in these strains that may be of importance for the virulence potential of STEC. In addition to bacterial virulence, young age (48) of the infected person is a risk factor for HUS development. This is also evident in this study where all HUS cases were in children ≤5 years old. Most likely, development of HUS may be influenced by other host factors as well.

As expected MLVA genotyping revealed that STEC isolates of some MLVA genotypes clustered with isolates from other parts of the country, in relation to local or national outbreaks. The largest local outbreak occurred in a kindergarten in 2009 where 15 isolates of STEC serotype O145:H28 were of the same MLVA genotype. Although the index child presented with bloody diarrhea, none of the children affected in that outbreak developed HUS (49).

Our laboratory was one of the first in Norway to introduce PCR for detection of human STEC infection. During the period from 1996 through 2011, a higher number of STEC infections were detected in our laboratory than in other regions of Norway, as more than one-fourth of the cases of STEC infections on the national level were detected here (MSIS), although less than one-tenth of the population live in this region. The high detection rate of STEC infections in our laboratory compared to rates in other parts of Norway most likely was, at least to some extent, due to early introduction of PCR in our laboratory. However, the fact that during this period, a higher proportion than expected of HUS cases (11 of 53 [20.8%], data from MSIS) also were from central Norway where our hospital is located may indicate that there may be epidemiological differences in the risks of STEC infection and disease between different regions in Norway.

Starting with PCR for STEC detection in 1996, the routine in our laboratory was to analyze all stool specimens from children <2 years for STEC irrespective of the clinical diagnosis and only those from older children and adults on clinical suspicion. One might expect that this practice could have led to identification of less virulent STEC strains not related to clinical disease in the younger age group. However, a comparison of the virulence profile of STEC strains isolated from children <2 years of age compared with that of STEC strains isolated from older children and adults does not support this idea. Strains from the younger age group contained at least as many virulence genes as those isolated from older children and adults (see Table S3 in the supplemental material).

A limitation of this study was that the laboratory routine to use different indications for testing of samples for STEC dependent on the age of the patient was not always followed. This was reflected in that the 39 patients diagnosed with STEC infections in the present study who were >2 years of age had diagnoses other than HUS and bloody diarrhea and were not part of an outbreak investigation. Furthermore, information regarding bloody diarrhea may have been incomplete since bloody diarrhea was recorded for only 9 of the patients, while there was no information on bloody stools in 68 other patients with diarrhea. For these cases, the information given by the referring physician might have been incomplete, or such data were not always recorded or updated in the patient information database. Another weakness was that in the first part of the study period, only the stx_1 and stx_2 genes were analyzed,

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

whereas *eae* was included from year 2000 onward. In addition, different primers, reagents, and equipment were used for PCR analysis through the study period, including a switch from conventional to real-time PCR. Although different primers were used for detection of stx_1 and *stx2*, the primers used were designed to detect all variants of the *stx* genes, except stx_{2f}. The use of different PCR methods for STEC detection might potentially have had an impact on which STEC strains were detected. However, despite the variations in the PCR protocols used throughout the study period, the strain collection represents an unselected group of STEC infections diagnosed by PCR analysis of the *stx* and *eae* genes in a hospital laboratory throughout a period of 16 years. In this study, we tested the STEC strains for a limited number of virulence genes. Although through this virulence characterization of the STEC strains, we did not disclose new knowledge, we were able to confirm results from previous studies regarding STEC virulence in a unselected collection of STEC isolates from hospital routine diagnosis based on PCR (12–14, 31).

In summary, STEC infection was diagnosed by PCR and STEC strains were isolated from stool samples from 138 (1.09%) of 12,651 patients tested at St. Olavs Hospital, Norway, during the period 1996-2011. More than half of the patients diagnosed with STEC infections were <5 years old. Eleven patients (all ≤5 years old) had HUS, but no one died. All HUS patients were infected with STEC strains of serogroups frequently involved in severe disease and outbreaks elsewhere. Six of the 11 HUS patients were infected with non-O157 serogroups. Twenty-four STEC strains were classified as HUS associated. Young age (≤5 years old) and STEC strains containing *eae* and stx_{2a} were significantly associated with HUS ($P < 0.05$ for each parameter), while STEC strains containing stx_1 were associated with non-HUS-associated STEC infections ($P < 0.05$). Also, the other potential virulence genes analyzed, except for *ehxA*, were significantly associated with HUS ($P < 0.05$ for each gene). However, as they were also present in some of the non-HUS-associated STEC strains, these genes could not reliably differentiate between HUS-associated and non-HUS-associated STEC strains.

ACKNOWLEDGMENTS

We thank the staff of the Department of Medical Microbiology at St. Olavs Hospital and the Department of Foodborne Infections at the Norwegian Institute of Public Health for skillful technical assistance. We also thank Frode Width Gran and Hilde Fossum of the Department of Medical Microbiology at St. Olavs Hospital for help with information on STEC strains and patients included in the study, Anne Nor for design of the primers SLTI Rnew, SLTI TaqMan, SLTIIFnew, SLTIIRnew, and SLTII TaqMan, and Tonje Haukeberg for design of the primers eae-Fny, eae-Rny, and eae-P.

The funding for this project was part of Ph.D. grant 81723800 to K.H. from the Norwegian University of Science and Technology.

REFERENCES

1. **Karmali MA, Steele BT, Petric M, Lim C.** 1983. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. Lancet **i:**619–620.
2. **Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N. Engl. J. Med. **308:**681–685. http://dx.doi.org/10.1056/NEJM198303243081203.
3. **Gyles CL.** 2007. Shiga toxin-producing *Escherichia coli*: an overview. J. Anim. Sci. **85**(13 Suppl)**:**E45–E62. http://dx.doi.org/10.2527/jas.2006-508.
4. **European Centre for Disease Prevention and Control.** 2013. Annual Epidemiological Report 2012. Reporting on 2010 surveillance data and 2011 epidemic intelligence data. European Centre for Disease Prevention and Control, Stockholm, Sweden.
5. **Gould LH, Mody RK, Ong KL, Clogher P, Cronquist AB, Garman KN, Lathrop S, Medus C, Spina NL, Webb TH, White PL, Wymore K, Gierke RE, Mahon BE, Griffin RM, Emerging Infections Program Foodnet Working Group.** 2013. Increased recognition of non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States during 2000-2010: epidemiologic features and comparison with *E. coli* O157 infections. Foodborne Pathog. Dis. **10:**453–460. http://dx.doi.org/10.1089/fpd.2012.1401.
6. **Bielaszewska M, Friedrich AW, Aldick T, Schurk-Bulgrin R, Karch H.** 2006. Shiga toxin activatable by intestinal mucus in *Escherichia coli* isolated from humans: predictor for a severe clinical outcome. Clin. Infect. Dis. **43:**1160–1167. http://dx.doi.org/10.1086/508195.
7. **Friedrich AW, Bielaszewska M, Zhang WL, Pulz M, Kuczius T, Ammon A, Karch H.** 2002. *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. J. Infect. Dis. **185:**74–84. http://dx.doi.org/10.1086/338115.
8. **Persson S, Olsen KE, Ethelberg S, Scheutz F.** 2007. Subtyping method for *Escherichia coli* Shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. J. Clin. Microbiol. **45:**2020–2024. http://dx.doi.org/10.1128/JCM.02591-06.
9. **Scheutz F, Teel LD, Beutin L, Pierard D, Buvens G, Karch H, Mellmann A, Caprioli A, Tozzoli R, Morabito S, Strockbine NA, Melton-Celsa AR, Sanchez M, Persson S, O'Brien AD.** 2012. Multicenter evaluation of a sequence-based protocol for subtyping shiga toxins and standardizing stx nomenclature. J. Clin. Microbiol. **50:**2951–2963. http://dx.doi.org/10.1128/JCM.00860-12.
10. **Nataro JP, Kaper JB.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
11. **Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Reid-Smith R, Isaac-Renton J, Clark C, Rahn K, Kaper JB.** 2003. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J. Clin. Microbiol. **41:**4930–4940. http://dx.doi.org/10.1128/JCM.41.11.4930-4940.2003.
12. **Bugarel M, Martin A, Fach P, Beutin L.** 2011. Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) *Escherichia coli* strains: a basis for molecular risk assessment of typical and atypical EPEC strains. BMC Microbiol. **11:**142. http://dx.doi.org/10.1186/1471-2180-11-142.
13. **Coombes BK, Wickham ME, Mascarenhas M, Gruenheid S, Finlay BB, Karmali MA.** 2008. Molecular analysis as an aid to assess the public health risk of non-O157 Shiga toxin-producing *Escherichia coli* strains. Appl. Environ. Microbiol. **74:**2153–2160. http://dx.doi.org/10.1128/AEM.02566-07.
14. **Bugarel M, Beutin L, Martin A, Gill A, Fach P.** 2010. Micro-array for the identification of Shiga toxin-producing *Escherichia coli* (STEC) seropathotypes associated with hemorrhagic colitis and hemolytic uremic syndrome in humans. Int. J. Food Microbiol. **142:**318–329. http://dx.doi.org/10.1016/j.ijfoodmicro.2010.07.010.
15. **Johnson TJ, Nolan LK.** 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. Microbiol. Mol. Biol. Rev. **73:**750–774. http://dx.doi.org/10.1128/MMBR.00015-09.
16. **March SB, Ratnam S.** 1986. Sorbitol-MacConkey medium for detection of *Escherichia coli* O157:H7 associated with hemorrhagic colitis. J. Clin. Microbiol. **23:**869–872.
17. **Gould LH, Bopp C, Strockbine N, Atkinson R, Baselski V, Body B, Carey R, Crandall C, Hurd S, Kaplan R, Neill M, Shea S, Somsel P, Tobin-D'Angelo M, Griffin PM, Gerner-Smidt P.** 2009. Recommendations for diagnosis of Shiga toxin-producing *Escherichia coli* infections by clinical laboratories. MMWR Recomm. Rep. **58**(RR-12)**:**1–14.
18. **Gould LH.** 2012. Update: recommendations for diagnosis of Shiga toxin-producing *Escherichia coli* infections by clinical laboratories. Clin. Microbiol. Newsl. **34:**75–83. http://dx.doi.org/10.1016/j.clinmicnews.2012.04.004.
19. **Karch H, Bielaszewska M, Bitzan M, Schmidt H.** 1999. Epidemiology and diagnosis of Shiga toxin-producing *Escherichia coli* infections. Diagn. Microbiol. Infect. Dis. **34:**229–243. http://dx.doi.org/10.1016/S0732-8893(99)00031-0.
20. **Bielaszewska M, Kock R, Friedrich AW, von Eiff C, Zimmerhackl LB, Karch H, Mellmann A.** 2007. Shiga toxin-mediated hemolytic uremic syndrome: time to change the diagnostic paradigm? PLoS One **2:**e1024. http://dx.doi.org/10.1371/journal.pone.0001024.

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

21. **Brian MJ, Frosolono M, Murray BE, Miranda A, Lopez EL, Gomez HF, Cleary TG.** 1992. Polymerase chain reaction for diagnosis of enterohemorrhagic *Escherichia coli* infection and hemolytic-uremic syndrome. J. Clin. Microbiol. **30:**1801–1806.
22. **Gannon VP, Rashed M, King RK, Thomas EJ.** 1993. Detection and characterization of the *eae* gene of Shiga-like toxin-producing *Escherichia coli* using polymerase chain reaction. J. Clin. Microbiol. **31:**1268–1274.
23. **Nielsen EM, Andersen MT.** 2003. Detection and characterization of verocytotoxin-producing *Escherichia coli* by automated 5′ nuclease PCR assay. J. Clin. Microbiol. **41:**2884–2893. http://dx.doi.org/10.1128/JCM.41.7.2884-2893.2003.
24. **Brandal LT, Lindstedt BA, Aas L, Stavnes TL, Lassen J, Kapperud G.** 2007. Octaplex PCR and fluorescence-based capillary electrophoresis for identification of human diarrheagenic *Escherichia coli* and *Shigella* spp. J. Microbiol. Methods **68:**331–341. http://dx.doi.org/10.1016/j.mimet.2006.09.013.
25. **Brandal LT, Sekse C, Lindstedt BA, Sunde M, Lobersli I, Urdahl AM, Kapperud G.** 2012. Norwegian sheep are an important reservoir for human-pathogenic *Escherichia coli* O26:H11. Appl. Environ. Microbiol. **78:**4083–4091. http://dx.doi.org/10.1128/AEM.00186-12.
26. **Jelacic JK, Damrow T, Chen GS, Jelacic S, Bielaszewska M, Ciol M, Carvalho HM, Melton-Celsa AR, O'Brien AD, Tarr PI.** 2003. Shiga toxin-producing *Escherichia coli* in Montana: bacterial genotypes and clinical profiles. J. Infect. Dis. **188:**719–729. http://dx.doi.org/10.1086/376999.
27. **Rüssmann H, Schmidt H, Heesemann J, Caprioli A, Karch H.** 1994. Variants of Shiga-like toxin II constitute a major toxin component in *Escherichia coli* O157 strains from patients with haemolytic uraemic syndrome. J. Med. Microbiol. **40:**338–343. http://dx.doi.org/10.1099/00222615-40-5-338.
28. **Lindstedt BA, Vardund T, Kapperud G.** 2004. Multiple-Locus Variable-Number Tandem-Repeats Analysis of *Escherichia coli* O157 using PCR multiplexing and multi-colored capillary electrophoresis. J. Microbiol. Methods **58:**213–222. http://dx.doi.org/10.1016/j.mimet.2004.03.016.
29. **Lindstedt BA, Brandal LT, Aas L, Vardund T, Kapperud G.** 2007. Study of polymorphic variable-number of tandem repeats loci in the ECOR collection and in a set of pathogenic *Escherichia coli* and *Shigella* isolates for use in a genotyping assay. J. Microbiol. Methods **69:**197–205. http://dx.doi.org/10.1016/j.mimet.2007.01.001.
30. **Lindstedt BA.** 2011. Genotyping of selected bacterial enteropathogens in Norway. Int. J. Med. Microbiol. **301:**648–653. http://dx.doi.org/10.1016/j.ijmm.2011.09.005.
31. **Bugarel M, Beutin L, Fach P.** 2010. Low-density macroarray targeting non-locus of enterocyte effacement effectors (*nle* genes) and major virulence factors of Shiga toxin-producing *Escherichia coli* (STEC): a new approach for molecular risk assessment of STEC isolates. Appl. Environ. Microbiol. **76:**203–211. http://dx.doi.org/10.1128/AEM.01921-09.
32. **Nicholls L, Grant TH, Robins-Browne RM.** 2000. Identification of a novel genetic locus that is required for in vitro adhesion of a clinical isolate of enterohaemorrhagic *Escherichia coli* to epithelial cells. Mol. Microbiol. **35:**275–288. http://dx.doi.org/10.1046/j.1365-2958.2000.01690.x.
33. **Centers for Disease Control and Prevention.** 2012. National Shiga toxin-producing *Escherichia coli* (STEC) surveillance annual summary, 2009. CDC, Atlanta, GA.
34. **European Centre for Disease Prevention and Control, European Food Safety Authority.** 2011. Shiga toxin/verotoxin-producing *Escherichia coli* in humans, food and animals in the EU/EEA, with special reference to the German outbreak strain STEC O104. European Centre for Disease Prevention and Control, Stockholm, Sweden.
35. **Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB.** 2013. Recent advances in understanding enteric pathogenic *escherichia coli*. Clin. Microbiol. Rev. **26:**822–880. http://dx.doi.org/10.1128/CMR.00022-13.
36. **Alpers K, Werber D, Frank C, Koch J, Friedrich AW, Karch H, An Der Heiden M, Prager R, Fruth A, Bielaszewska M, Morlock G, Heissenhuber A, Diedler A, Gerber A, Ammon A.** 2009. Sorbitol-fermenting enterohaemorrhagic *Escherichia coli* O157:H$^-$ causes another outbreak of haemolytic uraemic syndrome in children. Epidemiol. Infect. **137:**389–395. http://dx.doi.org/10.1017/S0950268808001465.
37. **Rosser T, Dransfield T, Allison L, Hanson M, Holden N, Evans J, Naylor S, La Ragione R, Low JC, Gally DL.** 2008. Pathogenic potential of emergent sorbitol-fermenting *Escherichia coli* O157:NM. Infect. Immun. **76:**5598–5607. http://dx.doi.org/10.1128/IAI.01180-08.
38. **Melton-Celsa A, Mohawk K, Teel L, O'Brien A.** 2012. Pathogenesis of Shiga-toxin producing *Escherichia coli*. Curr. Top. Microbiol. Immunol. **357:**67–103. http://dx.doi.org/10.1007/82_2011_176.
39. **Delannoy S, Beutin L, Fach P.** 2013. Towards a molecular definition of enterohemorrhagic *Escherichia coli* (EHEC): detection of genes located on O island 57 as markers to distinguish EHEC from closely related enteropathogenic *E. coli* strains. J. Clin. Microbiol. **51:**1083–1088. http://dx.doi.org/10.1128/JCM.02864-12.
40. **Johnson KE, Thorpe CM, Sears CL.** 2006. The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. Clin. Infect. Dis. **43:**1587–1595. http://dx.doi.org/10.1086/509573.
41. **Friedrich AW, Zhang W, Bielaszewska M, Mellmann A, Kock R, Fruth A, Tschape H, Karch H.** 2007. Prevalence, virulence profiles, and clinical significance of Shiga toxin-negative variants of enterohemorrhagic *Escherichia coli* O157 infection in humans. Clin. Infect. Dis. **45:**39–45. http://dx.doi.org/10.1086/518573.
42. **Mellmann A, Lu S, Karch H, Xu JG, Harmsen D, Schmidt MA, Bielaszewska M.** 2008. Recycling of Shiga toxin 2 genes in sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM. Appl. Environ. Microbiol. **74:**67–72. http://dx.doi.org/10.1128/AEM.01906-07.
43. **Mellmann A, Bielaszewska M, Kock R, Friedrich AW, Fruth A, Middendorf B, Harmsen D, Schmidt MA, Karch H.** 2008. Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. Emerg. Infect. Dis. **14:**1287–1290. http://dx.doi.org/10.3201/eid1408.071082.
44. **Karch H, Friedrich AW, Gerber A, Zimmerhackl LB, Schmidt MA, Bielaszewska M.** 2006. New aspects in the pathogenesis of enteropathic hemolytic uremic syndrome. Semin. Thromb. Hemost. **32:**105–112. http://dx.doi.org/10.1055/s-2006-939766.
45. **Buvens G, Pierard D.** 2012. Virulence profiling and disease association of verocytotoxin-producing *Escherichia coli* O157 and non-O157 isolates in Belgium. Foodborne Pathog. Dis. **9:**530–535. http://dx.doi.org/10.1089/fpd.2011.1073.
46. **Kawano K, Ono H, Iwashita O, Kurogi M, Haga T, Maeda K, Goto Y.** 2012. *stx* genotype and molecular epidemiological analyses of Shiga toxin-producing *Escherichia coli* O157:H7/H− in human and cattle isolates. Eur. J. Clin. Microbiol. Infect. Dis. **31:**119–127. http://dx.doi.org/10.1007/s10096-011-1283-1.
47. **Pradel N, Bertin Y, Martin C, Livrelli V.** 2008. Molecular analysis of Shiga toxin-producing *Escherichia coli* strains isolated from hemolytic-uremic syndrome patients and dairy samples in France. Appl. Environ. Microbiol. **74:**2118–2128. http://dx.doi.org/10.1128/AEM.02688-07.
48. **Khan F, Proulx F, Lingwood CA.** 2009. Detergent-resistant globotriaosyl ceramide may define verotoxin/glomeruli-restricted hemolytic uremic syndrome pathology. Kidney Int. **75:**1209–1216. http://dx.doi.org/10.1038/ki.2009.7.
49. **Wahl E, Vold L, Lindstedt BA, Bruheim T, Afset JE.** 2011. Investigation of an *Escherichia coli* O145 outbreak in a child day-care centre—extensive sampling and characterization of *eae*- and stx_1-positive *E. coli* yields epidemiological and socioeconomic insight. BMC Infect. Dis. **11:**238. http://dx.doi.org/10.1186/1471-2334-11-238.
50. **Schimmer B, Nygard K, Eriksen HM, Lassen J, Lindstedt BA, Brandal LT, Kapperud G, Aavitsland P.** 2008. Outbreak of haemolytic uraemic syndrome in Norway caused by stx_2-positive *Escherichia coli* O103:H25 traced to cured mutton sausages. BMC Infect. Dis. **8:**41. http://dx.doi.org/10.1186/1471-2334-8-41.

Downloaded from http://jcm.asm.org/ on August 22, 2014 by UNIVERSITETSBIBLIOTEKET I TRONDHEIM

Haugum et al., Supplemental material

Table S1. PCR primers used in the study.

Primer	Sequence (5'-3' direction)	Target gene	Amplicon size	Reference
SLTI 1	AAA TCG CCA TTC GTT GAC TAC TTC T	*stx1*	368 bp	Brian et al. 1992
SLTI Rnew	CCA TTC TGG CAA CTC GCG ATG CA			This study
SLTI TaqMan	FAM-AAC CTC ACT GAC GCA GTC TGT GGC AAG AGC-BHQ1			This study
SLTIIFnew	CAG TCG TCA CTC ACT GGT TTC ATC	*stx2*	283 bp	This study
SLTIIRnew	GGA TAT TCT CCC CAC TCT GAC AC			This study
SLTII TaqMan	HEX-CTG TCA CGG CAG AAG CCT TAC GCT TCA GGC-BHQ1			This study
eae-Fny	TTC ATT GAT CAG GAT TTT TCT GG	*eae*	105 bp	This study
eae-Rny	GCT CAT GCG GAA ATA GCC			This study
eae-P	FAM-ATA GTC TCG CCA GTA TTC GCC ACC AAT ACC-6-TAMRA			Nielsen et al. 2003
efa1-F	ATC AGA AGC CCG ACT ACG	*efa-1/lifA*	193 bp	This study
efa1-R	AAC ATT TGC CAG ACC AAG G			This study

Table S2. Number of virulence genes identified in various STEC serotypes by PCR analysis of 138 HUS-associated and non-HUS STEC strains diagnosed at St. Olavs Hospital, Trondheim, Norway during the period 1996-2011.

Serogroup	*stx1*	*stx2*	*stx2* subtype	*eae*	*ehxA*	*nleB*	*nleE*	*ent*	*efa-1/lifA*	*nleA*	*nleF*	*nleH1-2*	*espK*	HUS-associated	non-HUS	Total
O145	18	8	2a	28	28	28	28	28	26	26	28	28	28	7	21	28
O103	18	0	0	20	21	21	21	21	21	4	21	21	21	2	19	21
O157	9	20	2a (n=2), 2a+2c (n=1), 2c (n=17)	20	20	20	20	20	15	20	19	20	20	0	20	20
O26	8	3	2a	11	11	11	11	11	11	11	11	11	11	1	10	11
O121	0	9	2a	9	9	9	9	9	9	9	9	9	8	5	4	9
SF O157	0	6	2a	9	9	9	9	9	9	9	9	9	9	9	0	9
O111	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1
Others	24	25	2a (n=8), 2a+2c (n=3), 2b (n=10), 2c (n=1), 2d (n=2), 2g (n=1)	10	24	7	6	7	5	7	10	10	8	0	39	39
Total	78	71	-	108	123	106	105	106	97	87	108	109	106	24	114	138

Table S3. Comparison of potential virulence genes in 138 STEC strains diagnosed in patients where STEC was analyzed irrespective of symptoms (patients <2 years old) and on specific suspicion of STEC disease (patients ≥2 years old) at St. Olavs Hospital, Trondheim, Norway during the period 1996-2011.

Gene	Age <2 years n=50		Age ≥2 years n=88		P
	No. of strains	(%)	No. of strains	(%)	
stx2	20	(40)	51[1]	(58)	0.05
stx2a	9	(18)	28	(31.8)	0.1
stx1	27	(54)	51	(58)	0.7
eae	44	(88)	64	(72.7)	0.05
ehxA	47	(94)	76	(86.4)	0.3
nleB	44	(88)	62	(70.5)	0.02
nleE	43	(86)	62	(70.5)	0.06
ent	44	(88)	62	(70.5)	0.02
efa-1/lifA	39	(78)	58	(65.9)	0.2
nleA	34	(68)	53	(60.2)	0.5
nleF	45	(90)	63	(71.6)	0.02
nleH1-2	45	(90)	64	(72.7)	0.02
espK	42	(84)	64	(72.7)	0.1

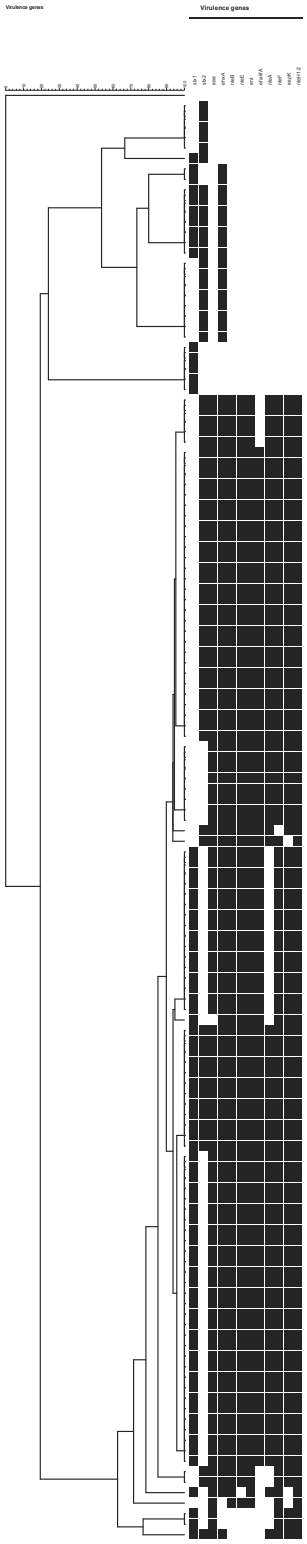
[1]This number does not include one SF O157:H- strain (St. Olav75, see Table 1) previously found to be positive for *stx2a* and one strain of unknown serotype (St. Olav154, see Table 1) previously found to be positive for *stx2b*.

Figure legends

Figure S1. Cluster analysis of potential virulence genes in STEC. *eae* negative and *eae* positive STEC were mainly separated in two clusters. One exception was one *eae* negative strain that clustered among the *eae* positive strains due to the presence of some of the potential virulence genes. All HUS-associated strains clustered among *eae* positive strains and harbored all the potential virulence genes investigated in the study. Strains with serotype 0 did not belong to any of the serotypes tested for in the study.

Figure S2. MLVA dendrograms of the 109 non-O157 STEC strains, using seven repeat loci. The strains were distributed in 48 distinct MLVA genotypes. The virulence gene profile of each strain is displayed.

Figure S3. MLVA dendrograms of O157 and SF O157 with the virulence gene profile displayed. The strains were distributed in 17 distinct MLVA genotypes. The virulence gene profile of each strain is displayed.

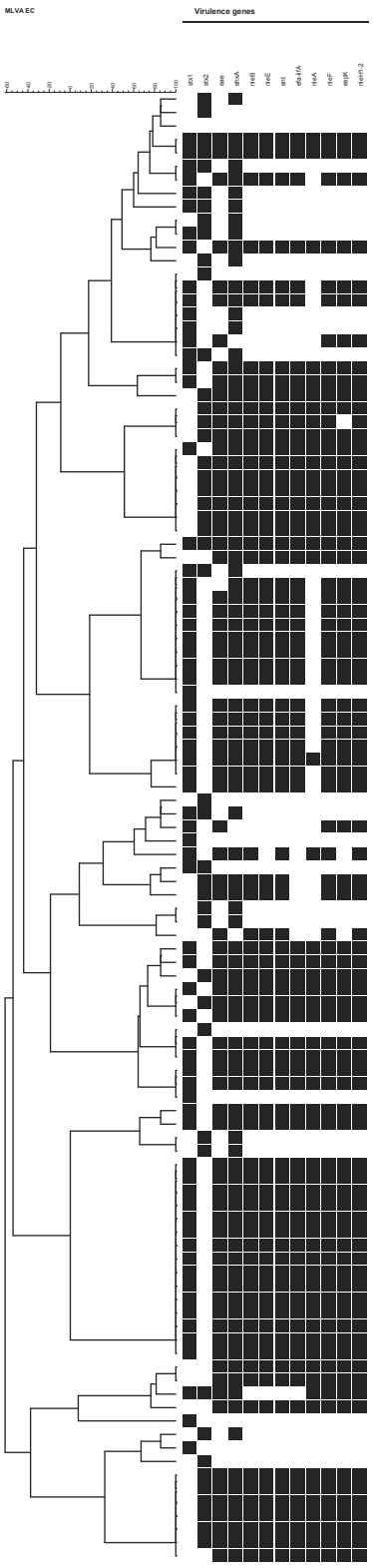


Key	Year	Sex	Clinic	Serotype	stx2 subtype	Age	Classification
St. Olav154	2011	M	ND	0	2b	58	Non-HUS
St. Olav14	2001	M	Gastroenteritis	0	2b	13	Non-HUS
St. Olav153	2011	F	Gastroenteritis	O2	2a	84	Non-HUS
St. Olav155	2011	F	Gastroenteritis	0	2a	90	Non-HUS
St. Olav159	2011	M	ND	0	2b	18	Non-HUS
St. Olav62	2006	M	ND	0	2b	73	Non-HUS
St. Olav1	1996	M	Gastroenteritis	O113:H21	2d	4	Non-HUS
St. Olav78	2009	M	BD	O104		0	Non-HUS
St. Olav79	2009	F	Asymptomatic	O104:H-		22	Non-HUS
St. Olav10	2000	M	Gastroenteritis	O7:H8	2b	19	Non-HUS
St. Olav11	2000	F	ND	0	2a+2c	59	Non-HUS
St. Olav168	2010	M	BD	O7:H28	2b	72	Non-HUS
St. Olav18	2001	M	ND	O81:H21/H8	2a+2c	0	Non-HUS
St. Olav4	1997	M	Gastroenteritis	O128	ND	13	Non-HUS
St. Olav48	2005	F	Gastroenteritis	0	2d	43	Non-HUS
St. Olav8	1999	F	Gastroenteritis	0	2a	3	Non-HUS
St. Olav130	2010	F	ND	O128:[H2]	2b	0	Non-HUS
St. Olav133	2010	F	ND	O128:H2	2b	29	Non-HUS
St. Olav146	2010	M	Gastroenteritis	0	2a	10	Non-HUS
St. Olav3	1997	F	ND	O113:H4	2b	49	Non-HUS
St. Olav38	2004	F	Gastroenteritis	0	2a	3	Non-HUS
St. Olav41	2004	F	Gastroenteritis	0	2c	1	Non-HUS
St. Olav5	1998	M	Gastroenteritis	O113:H4	2b	75	Non-HUS
St. Olav9	2000	F	ND	0	2g	3	Non-HUS
St. Olav132	2010	F	Annet	O117:H7		22	Non-HUS
St. Olav27	2002	M	Gastroenteritis	0		4	Non-HUS
St. Olav66	2007	F	Gastroenteritis	O76		2	Non-HUS
St. Olav74	2008	M	Gastroenteritis	O118		1	Non-HUS
St. Olav94	2009	F	Gastroenteritis	O104		11	Non-HUS
St. Olav44	2005	F	Gastroenteritis	O157:H7	2c	1	Non-HUS
St. Olav88	2009	M	Gastroenteritis	O157:H7	2c	1	Non-HUS
St. Olav89	2009	M	Gastroenteritis	O157:H-	2c	3	Non-HUS
St. Olav90	2009	F	Asymptomatic	O157:H-	2c	0	Non-HUS
St. Olav93	2009	F	Asymptomatic	O157:H-	2c	25	Non-HUS
St. Olav100	2009	M	HUS	O145	2a	1	HUS-associated
St. Olav101	2009	M	Asymptomatic	O145	2a	2	HUS-associated
St. Olav102	2009	M	Asymptomatic	O145	2a	28	HUS-associated
St. Olav104	2009	F	Asymptomatic	O145	2a	1	HUS-associated
St. Olav107	2009	F	Gastroenteritis	O145	2a	2	HUS-associated
St. Olav151	2011	F	BD	O157:H7	2c	1	Non-HUS
St. Olav157	2011	M	Gastroenteritis	O121:[H19]	2a	1	Non-HUS
St. Olav164	2011	F	HUS	O145	2a	2	HUS-associated
St. Olav166	2011	M	HUS	SF O157:H-	2a	0	HUS-associated
St. Olav17	2001	M	Gastroenteritis	O121:H19	2a	1	Non-HUS
St. Olav25	2002	F	ND	O26:H-	2a	31	Non-HUS
St. Olav26	2002	F	HUS	O26:H-	2a	4	HUS-associated
St. Olav51	2006	F	Asymptomatic	O121	2a	5	Non-HUS
St. Olav53	2006	F	Gastroenteritis	O26	2a	36	Non-HUS
St. Olav56	2006	M	HUS	SF O157:H-	2a	2	HUS-associated
St. Olav6	1999	M	Gastroenteritis	O157:H7	2a	27	Non-HUS
St. Olav63	2006	F	Gastroenteritis	O121	2a	2	Non-HUS
St. Olav64	2007	F	Gastroenteritis	O157:H7	2c	1	Non-HUS
St. Olav7	1999	M	Gastroenteritis	O157:H7	2a	15	Non-HUS
St. Olav72	2007	M	ND	O157:H-	2c	0	Non-HUS
St. Olav80	2009	F	HUS	SF O157:H-	2a	3	HUS-associated
St. Olav81	2009	M	HUS	SF O157:H-	2a	5	HUS-associated
St. Olav82	2009	F	Asymptomatic	SF O157:H-	2a	35	HUS-associated
St. Olav83	2009	F	ND	SF O157:H-	2a	66	HUS-associated
St. Olav85	2009	M	Gastroenteritis	O121:H19	2a	1	HUS-associated
St. Olav87	2009	F	Asymptomatic	O121:H19	2a	31	HUS-associated
St. Olav91	2009	F	HUS	O121:H19	2a	1	HUS-associated
St. Olav99	2009	F	Gastroenteritis	O121:H19	2a	2	HUS-associated
St. Olav131	2010	M	BD	O103:[H25]		1	Non-HUS
St. Olav165	2011	M	BD	O145:H28		6	Non-HUS
St. Olav49	2006	F	HUS	O103:H25		1	HUS-associated
St. Olav59	2006	M	HUS	O103:H25		1	HUS-associated
St. Olav75	2008	M	Gastroenteritis	SF O157:H-	2a	5	HUS-associated
St. Olav77	2008	F	HUS	SF O157:H-		1	HUS-associated
St. Olav84	2009	M	Gastroenteritis	SF O157:H-		1	HUS-associated
St. Olav97	2009	F	Gastroenteritis	O145		1	HUS-associated
St. Olav65	2007	M	Gastroenteritis	O157:H7	2a+2c	19	Non-HUS
St. Olav86	2009	M	Asymptomatic	O121:H19	2a	1	HUS-associated
St. Olav125	2009	F	Gastroenteritis	O103:[H2]		1	Non-HUS
St. Olav126	2009	F	Asymptomatic	O103:H2		24	Non-HUS
St. Olav127	2010	M	Gastroenteritis	O103:H2		1	Non-HUS
St. Olav128	2010	M	Asymptomatic	O103:[H2]		26	Non-HUS
St. Olav136	2010	M	ND	O103		6	Non
St. Olav139	2010	F	Gastroenteritis	O103:H2		1	Non
St. Olav140	2010	F	Gastroenteritis	O103:[H2]		41	Non
St. Olav143	2010	F	BD	O103:[H2]		0	Non
St. Olav144	2010	M	Gastroenteritis	O103:[H2]		3	Non
St. Olav152	2011	F	BD	O103:H2		1	Non
St. Olav19	2001	F	ND	O103		29	Non
St. Olav45	2005	F	Gastroenteritis	O103		76	Non-HUS
St. Olav47	2005	M	Gastroenteritis	O103		1	Non-HUS
St. Olav50	2006	F	Gastroenteritis	O103		1	Non-HUS
St. Olav73	2008	F	Gastroenteritis	O103:H2		60	Non-HUS
St. Olav96	2009	M	Gastroenteritis	O103:H2		1	Non-HUS
St. Olav12	2000	F	ND	O103		1	Non-HUS
St. Olav142	2010	F	BD	0	2a+2c	69	Non-HUS
St. Olav20	2001	M	Gastroenteritis	O157:H-	2c	31	Non-HUS
St. Olav23	2002	M	Gastroenteritis	O157:H-	2c	13	Non-HUS
St. Olav24	2002	F	Gastroenteritis	O157:H-	2c	30	Non-HUS
St. Olav28	2002	F	Gastroenteritis	O157:H-	2c	47	Non-HUS
St. Olav29	2003	F	Gastroenteritis	O157:H-	2c	1	Non-HUS
St. Olav30	2003	F	ND	O157:H-	2c	1	Non-HUS
St. Olav31	2003	M	ND	O157:H-	2c	5	Non-HUS
St. Olav32	2003	M	ND	O157:H-	2c	8	Non-HUS
St. Olav35	2003	F	Gastroenteritis	O157:H-	2c	42	Non-HUS
St. Olav39	2004	F	Gastroenteritis	0	2a	1	Non-HUS
St. Olav40	2004	F	ND	0	2a	4	Non-HUS
St. Olav103	2009	M	Gastroenteritis	O103:H2		3	Non-HUS
St. Olav108	2009	M	Gastroenteritis	O145:H28		2	Non-HUS
St. Olav111	2009	M	Gastroenteritis	O145:H28		3	Non-HUS
St. Olav112	2009	F	ND	O145:H28		3	Non-HUS
St. Olav113	2009	M	Gastroenteritis	O145:H28		2	Non-HUS
St. Olav114	2009	F	ND	O145:H28		2	Non-HUS
St. Olav115	2009	F	ND	O145:[H28]		1	Non-HUS
St. Olav116	2009	F	Asymptomatic	O145:H28		4	Non-HUS
St. Olav117	2009	F	Asymptomatic	O145:H28		3	Non-HUS
St. Olav118	2009	M	Asymptomatic	O145:H28		1	Non-HUS
St. Olav119	2009	M	Asymptomatic	O145:H28		3	Non-HUS
St. Olav120	2009	M	Asymptomatic	O145:H28		1	Non-HUS
St. Olav121	2009	M	ND	O145:H28		3	Non-HUS
St. Olav122	2009	F	Asymptomatic	O145:H28		2	Non-HUS
St. Olav123	2009	M	Asymptomatic	O145:H28		4	Non-HUS
St. Olav129	2010	M	Annet	O26:[H11]		0	Non-HUS
St. Olav13	2000	M	ND	O26		2	Non-HUS
St. Olav147	2010	F	ND	0		1	Non-HUS
St. Olav15	2001	M	Gastroenteritis	0		1	Non-HUS
St. Olav16	2001	M	Gastroenteritis	O26		50	Non-HUS
St. Olav2	1997	M	Gastroenteritis	O26:H11		32	Non-HUS
St. Olav21	2001	M	Gastroenteritis	O145		0	Non-HUS
St. Olav33	2003	F	Gastroenteritis	O26		2	Non-HUS
St. Olav34	2003	F	Gastroenteritis	O26		0	Non-HUS
St. Olav37	2004	F	Gastroenteritis	O145		20	Non-HUS
St. Olav43	2005	M	Gastroenteritis	O111:H-		23	Non-HUS
St. Olav54	2006	F	Gastroenteritis	O145		0	Non-HUS
St. Olav57	2006	F	Gastroenteritis	O26		1	Non-HUS
St. Olav60	2006	M	Gastroenteritis	O26		12	Non-HUS
St. Olav69	2007	M	Gastroenteritis	O145		16	Non-HUS
St. Olav141	2010	M	BD	O145:[H28]	2a	21	Non-HUS
St. Olav61	2006	F	ND	O145	2a	7	Non-HUS
St. Olav42	2005	M	Gastroenteritis	O119:H-		1	Non-HUS
St. Olav156	2011	M	BD	0		1	Non-HUS
St. Olav135	2010	M	ND	O177:H-		3	Non-HUS
St. Olav36	2003	M	Gastroenteritis	0		0	Non-HUS
St. Olav124	2009	M	Asymptomatic	0	2a	5	Non-HUS

eae-

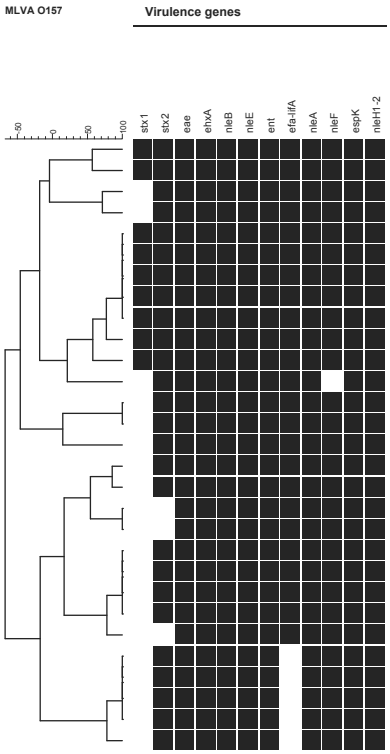
eae+

Figure S1.



St. Olav146	2010	M	Gastroenteritis	0	2a	10	Non-HUS
St. Olav62	2006	M	ND	0	2b	73	Non-HUS
St. Olav154	2011	M	ND	0	2b	58	Non-HUS
St. Olav39	2004	F	Gastroenteritis	0	2a	1	Non-HUS
St. Olav40	2004	F	ND	0	2a	4	Non-HUS
St. Olav8	1999	F	Gastroenteritis	0	2a	3	Non-HUS
St. Olav152	2011	F	BD	O103:H2		1	Non-HUS
St. Olav168	2010	M	BD	O?:H28	2b	72	Non-HUS
St. Olav48	2005	F	Gastroenteritis	0	2d	43	Non-HUS
St. Olav9	2000	F	ND	0	2g	3	Non-HUS
St. Olav18	2001	M	ND	O91:H21/H8	2a+2c	0	Non-HUS
St. Olav21	2001	M	Gastroenteritis	O145		0	Non-HUS
St. Olav41	2004	F	Gastroenteritis	0	2c	1	Non-HUS
St. Olav14	2001	M	Gastroenteritis	0	2b	13	Non-HUS
St. Olav125	2009	F	Gastroenteritis	O103:[H2]		1	Non-HUS
St. Olav126	2009	F	Asymptomatic	O103:H2		24	Non-HUS
St. Olav78	2009	M	BD	O104		0	Non-HUS
St. Olav79	2009	F	Asymptomatic	O104:H-		22	Non-HUS
St. Olav135	2010	M	ND	O177:H-		3	Non-HUS
St. Olav10	2000	M	Gastroenteritis	O?:H8	2b	19	Non-HUS
St. Olav57	2006	F	Gastroenteritis	O26		1	Non-HUS
St. Olav60	2006	M	Gastroenteritis	O26		12	Non-HUS
St. Olav26	2002	F	HUS	O26:H-	2a	4	HUS-associated
St. Olav85	2009	M	Gastroenteritis	O121:H19	2a	1	HUS-associated
St. Olav86	2009	M	Asymptomatic	O121:H19	2a	1	HUS-associated
St. Olav87	2009	F	Asymptomatic	O121:H19	2a	31	HUS-associated
St. Olav43	2005	M	Gastroenteritis	O111:H-		23	Non-HUS
St. Olav51	2006	F	Asymptomatic	O121	2a	5	Non-HUS
St. Olav63	2006	F	Gastroenteritis	O121	2a	2	Non-HUS
St. Olav157	2011	M	Gastroenteritis	O121:[H19]	2a	1	Non-HUS
St. Olav17	2001	M	Gastroenteritis	O121:H19	2a	1	Non-HUS
St. Olav91	2009	F	HUS	O121:H19	2a	1	HUS-associated
St. Olav99	2009	F	Gastroenteritis	O121:H19	2a	2	HUS-associated
St. Olav142	2010	F	BD	0	2a+2c	69	Non-HUS
St. Olav165	2011	M	BD	O145:H28		6	Non-HUS
St. Olav11	2000	F	ND	0	2a+2c	59	Non-HUS
St. Olav12	2000	F	ND	O103		1	Non-HUS
St. Olav19	2001	F	ND	O103		29	Non-HUS
St. Olav45	2005	F	Gastroenteritis	O103		76	Non-HUS
St. Olav47	2005	M	Gastroenteritis	O103		1	Non-HUS
St. Olav50	2006	F	Gastroenteritis	O103		1	Non-HUS
St. Olav128	2010	M	Asymptomatic	O103:[H2]		26	Non-HUS
St. Olav127	2010	M	Gastroenteritis	O103:H2		1	Non-HUS
St. Olav73	2008	F	Gastroenteritis	O103:H2		60	Non-HUS
St. Olav94	2009	F	Gastroenteritis	O104		11	Non-HUS
St. Olav140	2010	F	Gastroenteritis	O103:[H2]		41	Non-HUS
St. Olav96	2009	M	Gastroenteritis	O103:H2		1	Non-HUS
St. Olav143	2010	F	BD	O103:[H2]		0	Non-HUS
St. Olav144	2010	M	Gastroenteritis	O103:[H2]		3	Non-HUS
St. Olav103	2009	M	Gastroenteritis	O103:H2		3	Non-HUS
St. Olav139	2010	F	Gastroenteritis	O103:H2		1	Non-HUS
St. Olav136	2010	M	ND	O103		6	Non-HUS
St. Olav155	2011	F	Gastroenteritis	0	2a	90	Non-HUS
St. Olav4	1997	M	Gastroenteritis	O128	ND	13	Non-HUS
St. Olav36	2003	M	Gastroenteritis	0		0	Non-HUS
St. Olav132	2010	F	Annet	O117:H7		22	Non-HUS
St. Olav42	2005	M	Gastroenteritis	O119:H-		1	Non-HUS
St. Olav1	1996	M	Gastroenteritis	O113:H21	2d	4	Non-HUS
St. Olav61	2006	F	ND	O145	2a	7	Non-HUS
St. Olav141	2010	M	BD	O145:[H28]	2a	21	Non-HUS
St. Olav3	1997	F	ND	O113:H4	2b	49	Non-HUS
St. Olav5	1998	M	Gastroenteritis	O113:H4	2b	75	Non-HUS
St. Olav156	2011	M	BD	0		1	Non-HUS
St. Olav147	2010	F	ND	0		1	Non-HUS
St. Olav129	2010	M	Annet	O26:[H11]		0	Non-HUS
St. Olav53	2006	F	Gastroenteritis	O26	2a	36	Non-HUS
St. Olav13	2000	M	ND	O26		2	Non-HUS
St. Olav25	2002	F	ND	O26:H-	2a	31	Non-HUS
St. Olav2	1997	M	Gastroenteritis	O26:H11		32	Non-HUS
St. Olav159	2011	M	ND	0	2b	18	Non-HUS
St. Olav33	2003	F	Gastroenteritis	O26		2	Non-HUS
St. Olav34	2003	F	Gastroenteritis	O26		0	Non-HUS
St. Olav15	2001	M	Gastroenteritis	0		1	Non-HUS
St. Olav16	2001	M	Gastroenteritis	O26		50	Non-HUS
St. Olav66	2007	F	Gastroenteritis	O76		2	Non-HUS
St. Olav54	2006	F	Gastroenteritis	O145		0	Non-HUS
St. Olav69	2007	M	Gastroenteritis	O145		16	Non-HUS
St. Olav130	2010	F	ND	O128:[H2]	2b	0	Non-HUS
St. Olav133	2010	F	ND	O128:H2	2b	29	Non-HUS
St. Olav37	2004	F	Gastroenteritis	O145		20	Non-HUS
St. Olav115	2009	F	ND	O145:[H28]		1	Non-HUS
St. Olav108	2009	M	Gastroenteritis	O145:H28		2	Non-HUS
St. Olav111	2009	M	Gastroenteritis	O145:H28		3	Non-HUS
St. Olav112	2009	F	ND	O145:H28		3	Non-HUS
St. Olav113	2009	M	Gastroenteritis	O145:H28		2	Non-HUS
St. Olav114	2009	F	ND	O145:H28		2	Non-HUS
St. Olav116	2009	F	Asymptomatic	O145:H28		4	Non-HUS
St. Olav117	2009	F	Asymptomatic	O145:H28		3	Non-HUS
St. Olav118	2009	M	Asymptomatic	O145:H28		1	Non-HUS
St. Olav119	2009	M	Asymptomatic	O145:H28		3	Non-HUS
St. Olav120	2009	M	Asymptomatic	O145:H28		1	Non-HUS
St. Olav121	2009	M	ND	O145:H28		3	Non-HUS
St. Olav122	2009	F	Asymptomatic	O145:H28		2	Non-HUS
St. Olav123	2009	M	Asymptomatic	O145:H28		4	Non-HUS
St. Olav49	2006	F	HUS	O103:H25		1	HUS-associated
St. Olav59	2006	M	HUS	O103:H25		1	HUS-associated
St. Olav124	2009	M	Asymptomatic	0	2a	5	Non-HUS
St. Olav131	2010	M	BD	O103:[H25]		1	Non-HUS
St. Olav27	2002	M	Gastroenteritis	0		4	Non-HUS
St. Olav38	2004	F	Gastroenteritis	0	2a	3	Non-HUS
St. Olav74	2008	M	Gastroenteritis	O118		1	Non-HUS
St. Olav153	2011	F	Gastroenteritis	O2	2a	84	Non-HUS
St. Olav100	2009	M	HUS	O145	2a	1	HUS-associated
St. Olav101	2009	M	Asymptomatic	O145	2a	2	HUS-associated
St. Olav102	2009	M	Asymptomatic	O145	2a	28	HUS-associated
St. Olav104	2009	F	Asymptomatic	O145	2a	1	HUS-associated
St. Olav107	2009	F	Gastroenteritis	O145	2a	2	HUS-associated
St. Olav164	2011	F	HUS	O145	2a	2	HUS-associated
St. Olav97	2009	F	Gastroenteritis	O145		1	HUS-associated

Figure S2.



Key	Year	Sex	Clinic	Serotype	stx2 subtype	Age	Classification
St. Olav24	2002	F	Gastroenteritis	O157:H-	2c	30	Non-HUS
St. Olav35	2003	F	Gastroenteritis	O157:H-	2c	42	Non-HUS
St. Olav64	2007	F	Gastroenteritis	O157:H7	2c	1	Non-HUS
St. Olav72	2007	M	ND	O157:H-	2c	0	Non-HUS
St. Olav23	2002	M	Gastroenteritis	O157:H-	2c	13	Non-HUS
St. Olav28	2002	F	Gastroenteritis	O157:H-	2c	47	Non-HUS
St. Olav29	2003	F	Gastroenteritis	O157:H-	2c	1	Non-HUS
St. Olav30	2003	F	ND	O157:H-	2c	1	Non-HUS
St. Olav32	2003	M	ND	O157:H-	2c	8	Non-HUS
St. Olav31	2003	M	ND	O157:H-	2c	5	Non-HUS
St. Olav20	2001	M	Gastroenteritis	O157:H-	2c	31	Non-HUS
St. Olav65	2007	M	Gastroenteritis	O157:H7	2a+2c	19	Non-HUS
St. Olav6	1999	M	Gastroenteritis	O157:H7	2a	27	Non-HUS
St. Olav7	1999	M	Gastroenteritis	O157:H7	2a	15	Non-HUS
St. Olav151	2011	F	BD	O157:H7	2c	1	Non-HUS
St. Olav166	2011	M	HUS	SF O157:H-	2a	0	HUS-associated
St. Olav56	2006	M	HUS	SF O157:H-	2a	2	HUS-associated
St. Olav75	2008	M	Gastroenteritis	SF O157:H-	2a	5	HUS-associated
St. Olav77	2008	F	HUS	SF O157:H-		1	HUS-associated
St. Olav80	2009	F	HUS	SF O157:H-	2a	3	HUS-associated
St. Olav81	2009	M	HUS	SF O157:H-	2a	5	HUS-associated
St. Olav82	2009	F	Asymptomatic	SF O157:H-	2a	35	HUS-associated
St. Olav83	2009	F	ND	SF O157:H-	2a	66	HUS-associated
St. Olav84	2009	M	Gastroenteritis	SF O157:H-		1	HUS-associated
St. Olav88	2009	M	Gastroenteritis	O157:H7	2c	1	Non-HUS
St. Olav89	2009	M	Gastroenteritis	O157:H-	2c	3	Non-HUS
St. Olav90	2009	F	Asymptomatic	O157:H-	2c	0	Non-HUS
St. Olav93	2009	F	Asymptomatic	O157:H-	2c	25	Non-HUS
St. Olav44	2005	F	Gastroenteritis	O157:H7	2c	1	Non-HUS

Figure S3.

Paper III

Comparative Genomics to Delineate Pathogenic Potential in Non-O157 Shiga Toxin-producing *Escherichia coli* (STEC) from Patients with and without Haemolytic Uremic Syndrome (HUS) in Norway

Haugum, K.[1]#, Johansen, J.[2], Gabrielsen, C.[1], Brandal, L.T[3]., Bergh, K.1[1, 5], Ussery, D.W.[4], Drabløs, F.[2], Afset, J.E.[1, 5]

[1]Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway

[3]Department of Foodborne Infections, Norwegian Institute of Public Health, Oslo, Norway

[4]Biosciences Division, Oak Ridge National Labs, Oak Ridge, Tennessee, USA

[5]Department of Medical Microbiology, St. Olavs University Hospital, Trondheim, Norway

 # Corresponding Author: Kjersti Haugum, Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway. E-mail; kjersti.haugum@stolav.no

Running title: Genomic Comparison of Non-O157 Shiga Toxin-producing *Escherichia coli*

Abstract

Shiga toxin-producing *Escherichia coli* (STEC) cause infections in humans ranging from asymptomatic carriage to bloody diarrhoea and haemolytic uremic syndrome (HUS). Here we present whole genome comparison of Norwegian non-O157 STEC strains with the aim to distinguish between strains with the potential to cause HUS and less virulent strains.

Whole genome sequencing and comparisons were performed across 95 non-O157 STEC strains. Twenty-three of these were classified as HUS-associated, including strains from patients with HUS (n=19) and persons with an epidemiological link to a HUS-case (n=4). Genomic comparison revealed considerable heterogeneity in gene content across the 95 STEC strains. A clear difference in gene profile was observed between strains with and without the Locus of Enterocyte Effacement (LEE) pathogenicity island. Phylogenetic analysis of the core genome showed high degree of diversity among the STEC strains, but all HUS-associated STEC strains were distributed in two distinct clusters within phylogroup B1. However, also non-HUS strains were found in these clusters. A number of accessory genes were found to be significantly overrepresented among HUS-associated STEC, but none of them were unique to this group of strains, suggesting that different sets of genes may contribute to the pathogenic potential in different phylogenetic STEC lineages.

In this study we were not able to clearly distinguish between HUS-associated and non-HUS non-O157 STEC by extensive genome comparisons. Our results indicate that STECs from different phylogenetic backgrounds have independently acquired virulence genes that determine pathogenic potential, and that the content of such genes is overlapping between HUS-associated and non-HUS strains.

Author summary

Shiga toxin-producing *Escherichia coli* (STEC) are recognized as important disease-causing pathogens worldwide. Non-O157 serotypes dominate among human STEC cases in Norway, where such serotypes have caused several cases of haemolytic uremic syndrome (HUS). We have whole genome sequenced and compared 95 Norwegian non-O157 STEC strains, from HUS-patients and non-HUS patients, to search for potential genetic differences suitable for distinguishing between highly pathogenic STEC with the potential to cause HUS and low-virulent STEC without this potential. Comparative genomics revealed that the genomic content in our strain collection was highly heterogeneous. Our analyses indicate that STEC from different phylogenetic backgrounds have independently acquired virulence genes that determine pathogenic potential, and that different sets of genes may contribute to the pathogenic potential in different phylogenetic lineages. Furthermore, it appears that both highly pathogenic and low-virulent STEC share a lot of this genetic material, and we were therefore not able to identify any unique genes which could clearly distinguish between HUS-associated and non-HUS STEC in this study. Despite this, our results provide insight into the genetic background of an important group of human pathogens.

Introduction

Shiga toxin producing *E. coli* (STEC) are important human pathogens known to cause infections ranging from diarrhoea and haemorrhagic colitis to haemorrhagic uremic syndrome (HUS) [1]. Since the first reports of disease caused by O157:H7 in 1982 [2,3], this serotype has been the most frequently reported cause of severe STEC disease and outbreaks worldwide [1]. However, several non-O157 STEC serogroups (*e.g.* O26, O45, O103, O111, O121 and O145) have also been recognized to be responsible for severe disease and outbreaks [4,5].

The STEC pathotype is defined by the presence of Shiga toxins Stx1 and Stx2 encoded by the *stx1* and *stx2* genes, which are acquired through horizontal gene transfer of a heterogeneous group of lambdoid bacteriophages [6-9]. There are several subtypes of Shiga toxins, of which the Stx2 subtypes Stx2a, Stx2c and Stx2d are more often associated with HUS than other Stx subtypes [10-13]. In addition, the adherence factor intimin, encoded by the *eae* gene located in the Locus of Enterocyte Effacement (LEE) pathogenicity island, is important for STEC pathogenicity. STEC causing severe disease and outbreaks do usually harbour LEE [1,14,15], although also LEE negative STEC are sometimes found in patients with severe disease [15-21]. LEE encodes several genes responsible for the attaching and effacing nature of STEC, a feature these bacteria share with the closely related enteropathogenic *E. coli* (EPEC). In addition, the LEE encodes additional associated regulators, translocators, effector proteins and chaperones [22,23].

Whole genome sequencing of bacterial genomes has become an accessible and affordable analysis. Comparison of whole genome sequences provides information on gene content and organization, and gives an overview of how organisms are related. Whole genome sequences available of STEC and other *E. coli* have demonstrated high diversity among different *E. coli* genomes, due to horizontal gene transfer, gene loss and other genomic alterations [20,24-34]. Genomic comparisons of O157 and non-O157 LEE positive STEC genomes with other *E. coli* and *Shigella* have also revealed that LEE

positive STEC in general have larger genomes, mostly due to horizontally transferred DNA such as prophage DNA, plasmids and integrative elements encoding potential virulence factors [24-26].

In Norway non-O157 STEC are more frequently isolated from patients with STEC disease than O157, and are also more common than O157 STEC in patients suffering from HUS [35]. Although whole genome sequence comparisons of O157 and non-O157 STEC are available [20,25,28], it is still unclear whether it is possible to differentiate between STEC strains based on their potential to cause HUS. In this study our main aim was to compare whole genome sequences from 95 non-O157 human STEC strains to investigate potential genetic differences suitable for distinguishing between highly pathogenic STEC having caused HUS and low-virulent STEC having caused only mild disease or asymptomatic carriage. We were not able to clearly distinguish between HUS-associated and non-HUS non-O157 STEC by extensive genome comparisons in this study. Our results indicate that STEC from different phylogenetic backgrounds have independently acquired virulence genes that determine pathogenic potential, and that the content of such genes is overlapping between HUS-associated and non-HUS strains. Despite this, to our knowledge this is the largest collection of non-O157 STEC strains that has been sequenced to date, thus providing valuable data on the less characterized STEC serotypes.

Results

Sequencing and whole genome comparison of the 95 non-O157 STEC strains included in this study revealed high degree of variation in gene content as well as diversity in whole genome phylogeny. A total of 1,954 genes represented the core genome among the 95 strains, while 26,073 genes represented the pan genome. The LEE pathogenicity island was identified in 54 (57%) of the genomes, whereas 41 (43%) of the sequenced STEC strains were LEE negative (Table 1). Stx genes were detected in 84 (88%) genomes; *stx1* in 35 (37%), *stx2* in 37 (39%) and a combination of *stx1* and *stx2* in 12 (13%) of the genomes (Table 1). Eleven (12%) of the sequenced genomes which did not harbour stx genes, were classified as STEC-LST (Table 1). The *stx2* subtypes were differently distributed: *stx2a* was significantly more frequent among LEE positive STEC, while *stx2b* was more frequent among LEE negative strains ($p<0.05$ for both analyses) (Table 1). Of the *stx1* subtypes, *stx1c* was significantly associated with LEE negative STEC.

In the present study, all 19 STEC strains from patients with HUS were from children <5 years old (Table S1). An additional four strains were linked to a HUS-case, and consequently 23 strains were classified as HUS-associated (Table S1). All the HUS-associated STEC strains harboured the LEE pathogenicity island, and except for six STEC-LST of serotype O103:H25 from the same outbreak, all contained the *stx2a* subtype. Only one STEC from a HUS patient (FHI6) harboured *stx1* (subtype *stx1a*), in addition to *stx2a* (Table S1).

Phylogenetic analysis of the core genome

A core gene tree was constructed from alignment of the 1,861 core genes present in all the 95 STEC and 14 *E. coli* reference genomes representing the *E. coli* phylogroups (109 genomes in total). In this phylogenetic tree, the 95 strains were distributed in the *E. coli* phylogroups A, B1, B2, D and E (Figure 1). In general, clusters of LEE negative STEC strains were distributed between clusters of LEE positive strains. Most of the strains belonged to the B1 phylogroup, and a majority of the LEE positive STEC strains were also found within this group. All HUS-associated strains were found in phylogroup B1, in

two clusters which we designated HUS-group 1 and 2 (Figure 1). The STEC strains in HUS-group 1 were distributed in three related clusters, consisting of mainly HUS and HUS-associated strains of serotypes O103:H25, O145:H[unknown], and O121:H- (n=18) (Figure 1, Figure S1). HUS-group 2 consisted of one distinct cluster of strains (n=23), in which all HUS and HUS-associated STEC strains of serogroups O26, O86 and O111 were located. Sixteen of the strains in HUS-group 2, of which 13 strains were of serogroup O26, were not associated with HUS (Figure 1, Figure S1). This group therefore appeared to be more heterogeneous than HUS-group 1 with respect to pathogenicity.

Comparative analysis of the accessory genome

The accessory genome consisted of approximately 24,000 genes. PCA of the accessory genome separated LEE positive (n=54, S1) and LEE negative (n=41) STEC strains in two distinct groups (data not shown). PCA and PLS regression of LEE positive strains (n=54) as well as of LEE positive/*stx2* positive strains (n=33, Table S1) could not further separate the strains into subgroups.

By comparison of all 54 LEE positive STEC strains, we identified eleven genes which were more frequent in the HUS-associated strains (n=23) and four genes which were more frequent in non-HUS STEC strains (n=31) ($p<0.01$, FDR) (Table 2, Table S2). None of these genes were however present exclusively in one of these two groups. Among the 33 LEE positive STEC strains containing *stx2* (Table S1), we identified 69 genes which were overrepresented in HUS-associated strains (n=23) and 44 genes which were more frequent in non-HUS STEC strains (n=10) ($p<0.01$, data uncorrected) (Table 2, Table S3).

When STEC strains in HUS-group 1 (n=18, Figure 1) were compared with all other LEE positive strains (n=36), 357 genes were more frequent in HUS-group 1 strains ($p<0.01$, FDR) (Table 2, Table S4). One gene encoding a hypothetical protein (Table S4) was present in all strains in HUS-group 1 but absent in other LEE positive STEC. This gene was however present in one LEE negative strain. When STEC strains in HUS-group 2 (n=23) were compared with all other LEE positive STEC strains (n=31), 576 genes were overrepresented in the former group ($p<0.01$, FDR) (Table 2, Table S5). Four genes were

present in all strains in HUS-group 2 while absent in the other strains (Table 2, Table S5). Seventeen genes were overrepresented in HUS-associated serogroup O26 strains in HUS-group 2 (n=5), whereas 13 genes were more frequent in the non-HUS strains of the same serogroup (n=13) ($p<0.01$, uncorrected) (Table 2, Table S6).

Comparative analysis of the core genome: Analysis of the core genome identified in total approximately 13,000 gene variants. Different Pfam domains, and therefore different protein sequences, were observed in 13 of these gene variants (Table S7). Comparison of core gene variants among LEE positive STEC (n=54, Table S1) identified 281 gene variants that were overrepresented in the HUS-associated (n=23) compared to non-HUS strains (n=31) ($p<0.01$, FDR) (Table 2, Table S8). None of the gene variants were however found only among HUS-associated strains.

PLS regression of the core gene variants in serogroup O26 strains in HUS-group 2 discriminated *stx2* positive O26 (n=8) from *stx1* positive O26 strains (n=10) (data not shown). Eighty-seven gene variants were more frequent in *stx2* positive O26 strains compared to *stx1* positive strains of the same serogroup ($p<0.01$, data uncorrected) (Table 2, Table S9), but none of these gene variants were exclusive for the group of *stx2* positive strains. Eighty-four gene variants were more common in the O26 strains which were HUS or HUS-associated (n=5) compared to the O26 non-HUS strains (n=13) ($p<0.1$, data uncorrected) (Table 2, Table S10), but also in this case none of the variants were exclusive to either group.

Comparison of HUS and non-HUS STEC from specific outbreaks

Comparative analysis on gene content was furthermore performed on specific HUS and non-HUS STEC that were epidemiologically linked and belonged to the same MLVA outbreak cluster (Table S1). In HUS-strain FHI4 (Table S11) we identified 179 genes (Table S12) which were absent in the non-HUS strain FHI3 (Table S11) from the same outbreak. The majority of the genes were related to various mobile genetic elements integrated in the bacterial chromosome, Nle effectors (Table S12) or plasmid pO26_1 (AP010954) (Table 12). A closer search revealed additional plasmid genes on other

contigs in the FHI4 draft genome, indicating the presence of a complete pO26_1 plasmid in this strain, while the corresponding genes were not found in strain FHI3. In HUS-strain FHI48 we identified 153 genes (Table S13) that were absent in the non-HUS strains FHI43 and FHI62 from the same outbreak cluster. Again, most of the genes were related to mobile genetic elements (Table S13). In the two HUS strains FHI58 and FHI63 from another outbreak, we identified 54 genes exclusive to these two strains (Table S14), while another 506 genes were present only in the non-HUS strain St. Olav104. The genes in the two HUS strains were related to various functions, while in the non-HUS strain, the majority of the genes were related to mobile genetic elements and several Nle effectors (Table S14).

Gene ontology enrichment analysis

GO analysis of genes significantly overrepresented in the 23 HUS-associated STEC strains (Table S2) revealed that nine GO terms in biological processes were enriched in these strains. The enriched terms specified metabolic and catabolic processes related to degradation of L-idonate (GO:0046183) (Table S15).

Among the 357 genes more frequent in HUS-group 1 (Table S4), we identified six enriched GO terms, in biological processes (n=4), molecular functions (n=1) and cellular components (n=1) (Table S16). Also in this case we identified enrichment in GO terms related to degradation of L-idonate (GO:0046183, GO:0003939, GO0019523). In biological processes and cellular components, we furthermore identified enriched GO terms related to protein secretion by the type II secretion system (GO:0015628, GO:0015627) (Table S16).

Twenty-six GO terms were enriched in HUS-group 2 (Table S5); in biological processes (n=11), molecular function (n=4) and cellular components (n=11) (Table S17). In biological processes, enriched terms were for siderophore biosynthetic process (GO:0019290) and ciliary or bacterial-type flagellar motility (GO:0001539) (Table S17). For molecular function we identified enrichment in terms for motor activity (GO:0003774), isochorismate synthase activity (GO:0008909) and oxo-acid-lyase

198 activity (GO:0016833), while for cellular components we found enrichment in terms for bacterial-
199 type flagellum basal body and rod (GO:0030694) (Table S17). None of the enriched GO terms were
200 however unique for the HUS-associated STEC strains.

201

Discussion

In this study we have performed comparative genomic analyses on what is to our knowledge is the largest collection so far of genome-sequenced non-O157 STEC strains, in order to investigate if there were genetic differences suitable for distinguishing between highly pathogenic STEC having caused HUS and low-virulent STEC having caused only mild disease or asymptomatic carriage. Whole genome sequencing and comparison revealed that there was considerable heterogeneity in genetic content across the 95 non-O157 STEC strains included in the study. The approximately 24,000 genes constituting the accessory genome contribute to this heterogeneity, while 1,954 core genes were shared by all the sequenced strains. Much of the accessory genome contained various mobile genetic elements, which have also previously been shown to contribute to heterogeneity and pathogenic evolution in *E. coli* [24-26,30,36]. The results from principal component analysis of the accessory genome where LEE positive strains were separated from LEE negative strains, is in line with several previous reports [14,20,28,37-42]. Although the accessory genome was not identical within LEE positive STEC strains, further PCA analysis of LEE positive strains showed scattering of the strains without any distinct subgroups, indicating that variable accessory gene content was heterogeneously distributed within this group.

The various *stx1* and *stx2* subtypes were differentially distributed between LEE positive and LEE negative STEC strains, i.e. *stx2b* and *stx1c* were more frequent among LEE negative strains ($p<0.05$) while *stx2a* was more frequent among LEE positive strains ($p<0.05$) (Table 1). All HUS associated STEC in this study were LEE positive and contained *stx2a*, except for the STEC-LST strains. Thus our results are in line with previous studies where Stx2a has been shown to possess higher potency than Stx1 and other Stx2 subtypes [43], and that LEE positive and *stx2a* positive STEC strains are more often associated with severe disease [12-15,43]. Furthermore, all the HUS-associated STEC belonged to *E. coli* O serogroups known to be associated with STEC disease (Table S1) [1]. However, although *stx2a* and LEE were typical for the HUS-associated STEC, these characteristics are not unique for such STEC, and thus not sufficient to clearly distinguish HUS-associated from non-HUS STEC. We therefore aimed

to compare the accessory genomes of *stx2a/*LEE positive STEC in an attempt to differentiate between HUS-associated and non-HUS strains. This analysis revealed that certain genes were overrepresented among HUS-associated STEC (Table 2, Table S2, S3), suggesting that strains with this gene profile have a high pathogenic potential. Regardless, none of the genes were exclusive for these strains, which further suggest that the gene content in HUS-associated STEC at least in part is shared with non-HUS STEC strains, or that different HUS strains have different gene content.

Core genome phylogeny revealed that the 95 non-O157 STEC strains were distributed in all the *E. coli* phylogroups except phylogroup F, confirming that the strains included in this study were heterogeneous. The majority of strains belonged to phylogroup B1 (Figure 1). Most of the LEE positive and all the HUS-associated strains in this study also clustered in this phylogroup, similarly to what has been reported in previous studies [15,28,31,33,44]. In addition, LEE negative STEC associated with HUS often belong to this phylogroup [20,33], including the 104:H4 strain (FHI102) related to the 2011 German outbreak, which did however not cluster with any of the HUS and HUS-associated STEC strains included in this study (Figure 1). LEE negative and LEE positive STEC did not form separate phylogenetic clusters, but were mixed in small clusters within several phylogroups as previously reported [20]. This indicates that the LEE pathogenicity island has been independently taken up by different STEC lineages at different time points. Because HUS-associated O103, O121 and O145 strains were distributed in three related clusters in the phylogenetic analysis, these STEC strains were classified as HUS-group 1, although they did not belong to one defined cluster. The remaining HUS-associated strains were located in one cluster which we termed HUS-group 2. This clustering of HUS-associated strains based on variation in core genes as observed in this study indicates that the phylogenetic backgrounds of the bacteria at least to some extent determine the pathogenic potential of the organism. In an attempt to search for unique genes in these groups, we analysed the accessory genome and identified several hundred genes that were significantly overrepresented in both groups, suggesting that different sets of genes may contribute to the pathogenic potential in different phylogenetic STEC lineages. However, few of these genes were

found to be unique to any of the groups (Table 2, Table S4, S5), which further suggest that the accessory genome is shared both between and within the different clusters defined by the core genome phylogeny.

The majority of strains in HUS-group 2 were of serotype O26, of which HUS-associated strains clustered with non-HUS strains, suggesting that accessory factors rather than core genes defines pathogenic potential within this group. Regardless, it was not possible to identify any genes in the accessory genome which could reliably distinguish HUS-associated from non-HUS strains of the same serogroup in HUS-group 2 (Table 2).

In the core genome of the 95 STEC strain included in this study, we identified approximately 13,000 different gene variants by edit distance analysis. However, despite the high number of gene variants, differences in protein sequences were identified for only 13 of these variants. Comparison of the core gene variants revealed that although 281 gene variants were overrepresented in HUS-associated STEC, several of these were also present in strains not associated with HUS (Table 2, Table S8). The observation that none of the identified gene variants were unique to HUS-associated STEC is supported by the fact that HUS-associated strains clustered in more than one group in the core gene phylogeny. Also for the O26 strains, although PLS regression revealed that serogroup O26 strains in HUS group 2 containing *stx2* were separated from strains that contained *stx1*, no core gene variants were found to be significantly over- or underrepresented in either of these two groups (Table 2, Table S9).

By comparing the genomes of the STEC strains which were epidemiologically linked and belonged to the same MLVA outbreak cluster (Table S1), we identified a number of genes that were different across HUS and non-HUS strains (Table S11). The fact that different genes were present in strains from the same outbreak might indicate that the infecting source consisted of a mixture of similar but not identical STEC which could have evolved from the same clone. Regardless, we could not identify

any genes that clearly distinguished between the HUS and non-HUS associated strains within each of these outbreak clusters.

GO terms related to L-idonate degradation were found to be enriched both among all 23 HUS-associated STEC strains collectively and the strains in HUS-group 1. *E. coli* is able to utilize L-idonate as a sole carbohydrate source through the Entner-Doudoroff metabolic pathway, which has been shown to be important for the ability of *E. coli* to colonize mammalian intestines [45]. In addition, we identified enriched GO terms for protein secretion by the type II secretion system in HUS-group 1. The type II secretion system in Gram negative bacteria promotes protein transport across the outer membrane, and the majority of proteins exported by this system contribute to bacterial adaptation and colonization by generating nutrients available for uptake [46]. Furthermore, certain exported lipoproteins have been shown to be involved in biofilm formation in EPEC [47]. Genes responsible for the enriched GO terms could therefore contribute to enhanced bacterial colonization and adaptation, which might have an impact on bacterial virulence in these specific strains. This however needs to be confirmed in further investigations. Of the 26 GO terms that were enriched in HUS group 2, a few were related to flagellar motility, which in general are recognized as virulence factors in bacteria [48]. In addition, enriched GO terms were related to siderophore biosynthesis. Siderophores, being iron chelating compounds, are important for iron acquisition in bacteria [49,50]. The specific siderophore identified among strains in HUS-group 2 was encoded on a high-pathogenicity island (HPI) found in a distinct clonal lineages of STEC, which includes serogroup O26 [51,52]. These results indicate that both motility and iron acquisition might be important factors for bacterial virulence of STEC in HUS-group 2. However, the precise role of these genes for STEC pathogenesis needs to be explored in further studies.

In addition to bacterial factors, it is clear that infection dose and host factors like the immune system, expression of the Shiga toxin receptor and intestinal environment might also affect STEC virulence, and thus severity of STEC disease [53]. In this study, all patients with HUS were <5 years old, which is

known to be a risk factor for severe STEC disease [54,55]. Therefore, it is possible that host factors play an important part in explaining why highly similar strains lead to such different clinical outcomes in different patients.

In our study we included all non-O157 STEC strains from HUS-patients in Norway. However, these represent only a limited number of STEC strains from each phylogenetic lineage or serotype. Furthermore, few epidemiologically linked HUS and non-HUS STEC strains were included in the study. For future studies, if more STEC strains associated with HUS were included in the genomic comparisons this would give more strength both to phylogenetic and to statistical analyses. In addition, even if highly virulent STEC strains share overlapping genetic content with less pathogenic strains, further investigations regarding factors regulating transcription and translation as well as transcriptomics and proteomics analyses could shed further light into STEC virulence and pathogenicity.

Conclusion

In this study whole genome sequencing and comparisons of 95 non-O157 STEC strains revealed that there were considerable genetic and phylogenetic heterogeneity between the strains. Although all HUS-associated STEC strains belonged to the B1 phylogroup, all non-O157 STEC from HUS patients did not cluster together, but were found in two separate clusters within this phylogenetic group. A clear difference in gene profile was observed between LEE positive and LEE negative STEC. A number of accessory genes were found to be significantly overrepresented among HUS-associated STEC, but none of them were unique to this group of strains. Our results indicate that STEC from different phylogenetic backgrounds independently have acquired virulence genes that determine pathogenic potential, and that specific genes overrepresented among HUS strains are not necessarily shared by all such strains, but that different sets of genes may contribute to the pathogenic potential in different phylogenetic STEC lineages.

Materials and Methods

Bacterial strains and clinical information

We selected 94 non-O157 STEC strains from the strain collection at the Norwegian Institute of Public Health (Oslo, Norway) isolated in 2000-2012 for sequencing in this study (Table S1). In addition, three STEC strains (St. Olav104, St. Olav143 and St. Olav172, Table S1) were selected from the strain collection at St. Olavs Hospital (Trondheim, Norway). The strains included in the study were primarily selected to represent different MLVA genotypes [56,57], a diversity of non-O157 STEC serotypes and patients with different severity of disease (Table S1). All available non-O157 STEC strains isolated from patients with HUS (n = 20) in Norway were included, except one strain (FHI10) which after whole genome sequencing was shown to be contaminated (Table S1). Thus, a total of 96 strains were included in the study (Table S1).

Some of the STEC strains from patients with HUS were from outbreaks and therefore had identical MLVA-genotypes or belonged to the same MLVA-genotype clusters (Table S1). Four of the STEC strains included were furthermore classified as HUS-associated because they had identical MLVA-genotype as or belonged to the same MLVA-genotype cluster as a HUS case (Table S1). Five of the STEC strains were from non-human sources and were isolated during various outbreak investigations related to STEC disease (Table S1). One of these was designated as HUS-associated (FHI16, Table S1). Of the total 96 STEC strains included in the study, 95 strains were included for genomic comparison throughout the whole study whereas one strain (St. Olav104) was included for parts of the study only. In addition, 14 *E. coli* were included as reference strains for classification of the STEC strains into the *E. coli* phylogroups A, B1, B2, D, E, and F (Table S1).

Primary characterization of STEC at the Norwegian Institute of Public Health and St. Olavs Hospital had been based on PCR for the *stx1*, *stx2* and *eae* genes [58-60]. Ninety-one strains then had contained the *stx1* and/or *stx2* genes, while six strains of serotype O103:H25 did not have *stx* genes. The latter were included in the study because they were isolated from patients with HUS in an

outbreak (five strains), or was isolated from fermented sausage linked to this outbreak (one strain) [61] (Table S1). In this particular outbreak, *stx2a* was detected in only two of the isolated strains, and the strains without *stx* genes were regarded as STEC that had lost their *stx* genes, often termed EHEC/STEC-LST [62]. Fifty-five of the STEC strains were positive for the LEE pathogenicity island, as detected by the presence of the *eae* gene.

Serotyping was performed at the National Reference Laboratory for Enteropathogenic Bacteria at the Norwegian Institute of Public Health, using monospesific O:K and H antisera by a combination of in-house antisera before 2002, and by antisera from Sifin (Germany) and SSI (Denmark) after 2002, covering altogether 44 O-serogroups including O26, O103, O111, O121, O145, O157; and 8 H-antigens. Twenty-four of the strains included in the study did not belong to any of the serogroups tested for (Table S1).

Ethics Statement

This experimental study was approved by the Regional Committee for Medical and Health Research Ethics, REC South-East (REC number 2011/2314). Clinical data (including age and gender) required for classification of patients into the groups HUS, bloody diarrhoea, diarrhoea and no disease were obtained from Norwegian Surveillance System for Communicable Diseases (MSIS) at the Norwegian Institute of Public Health (Table S1). Dispensation from professional secrecy requirements was given by the REC. As data were analysed anonymously informed consent was not obtained.

DNA isolation

Strains were grown overnight on MacConkey agar. Genomic DNA was isolated for each strain using the Qiagen MagAttract® DNA Mini M48 Kit and the Qiagen BioRobot M48 (Qiagen, Hilden Germany) as described by the manufacturer.

Whole genome sequencing

Ninety-six of the STEC strains were sequenced with the Illumina Technology, while one strain (St. Olav104) was sequenced with Pacific Biosciences (PacBio) Technology (Table S1).

For the strains to be sequenced by Illumina technology a standard read library of bacterial genomic DNA was prepared, with an average fragment length of 370 base pairs (bp). The DNA was sequenced by LGC Genomics (Berlin, Germany) on the Illumina HiSeq2000 platform (Illumina, San Diego, CA, USA) with 100 bp paired-end reads. Assembly and scaffolding of processed and error corrected paired-end reads was done using Velvet 1.2.04 [63]. Information on the resulting draft genome for each strain is given in Table S1.

Fourty-eight of the 96 strains were selected for additional mate pair sequencing (Table S1). For this purpose a 2 kb Illumina Mate Pair library was prepared and the DNA was sequenced by LGC Genomics (Berlin, Germany) on the Illumina HiSeq2000 platform with 100 bp paired-end reads. Assembly and scaffolding of processed and error corrected paired-end reads was performed using Allpaths-LG release 45553 [64]. Gap closure of assembly scaffolds was done using SOAP GapCloser version 1.12 [65], while refinement of gap-closed scaffolds was done using SEQuel version 1.0.2 [66]. Information on the resulting draft genome for each strain is given in Table S1.

Genome sequencing on the PacBio platform was performed at the Norwegian Sequencing Centre (Oslo, Norway). A library was prepared using the Pacific Biosciences 10 kb library preparation protocol, and size selection of the final library was performed using Ampure beads. The library was sequenced on a Pacific Biosciences RS II instrument (Pacific Bioscience, Menlo Park, CA, USA) using P4-C2 chemistry and three SMRT cells. Processed reads were assembled using HGAP v2 [67]. Information on the resulting draft genome is given in Table S1.

400 **Gene annotation**

401 Identification of open reading frames was performed using the Prodigal Microbial Gene Prediction
402 Software [68], and functional gene annotation was done using myRAST [69].

403 **Comparative analyses**

404 The CMG-biotools (Comparative Microbial Genomics) package was used for genome comparison
405 [70]. Blastmatrix in CMG-biotools was used to identify proteins shared between genomes, while
406 pancoreplot was used to identify the pan- and core- genome of the sequenced strains. In this
407 context, genes were considered to be equal homologous having a minimum of 60% alignment length
408 and 90% sequence identity. The accessory genome was defined by subtracting all core genes from
409 the pan genome. Genome analysis and comparison was performed across all sequenced STECs (Table
410 S1) except strain St. Olav104 which was only used for comparison with two HUS-strains (FHI58 and
411 FHI63).

412 **Core genome phylogeny**

413 *E. coli* phylotypes were determined *in silico* based on a core gene tree. This was created as described
414 by Kaas et al. [33] using 1,861 core genes present in all the 95 STEC genomes and additional 14 *E. coli*
415 reference genomes (109 genomes in total) representing the *E. coli* phylotypes A, B1, B2, D, E and F
416 (Table S1) [28,33,71-73].

417 **Core gene analysis**

418 Core gene nucleotide sequences (n=1,861) from the 95 STEC and 14 reference *E. coli* (Table S1) were
419 aligned separately and a consensus sequence was estimated for each of the 1,861 genes using
420 EMBOSS 6.3.1 [74]. A python implementation of the edit distance method [75] was used to quantify
421 the difference between the consensus sequence and the corresponding sequence of each core gene
422 for all 109 strains included in the analysis. This resulted in various edit distances, representing

423 different gene variants for each of the core genes. Edit distance values for all strains were normalized
424 and transformed into a binary matrix for core gene comparisons.

425 To examine if any gene variant from the same core gene family showed different Pfam domains, we
426 used pfam_scan.pl with the HMMER3 library of Pfam domains.

427 **Principal component analysis and Partial least squares regression**

428 For Principal component analysis (PCA) and Partial least squares (PLS) regression the Laydi software
429 (http://www.laydi.org) (unpublished) was used. For PLS regression, dependent variables (for the Y-
430 matrix) were the clinical diagnosis HUS or classification as HUS-associated, and the presence of *stx1*
431 and/or *stx2*. HUS and HUS-associated STEC-LST were classified as *stx2* positive for these analyses.

432 **Functional annotation and Gene Ontology enrichment analysis**

433 Blast2GO was used for functional annotation based on gene ontology (GO) and for GO enrichment
434 analysis [76,77].

435 **Subtyping of Stx1 and Stx2**

436 There are three known subtypes of Stx1 and seven known subtypes of Stx2, designated Stx1a, Stx1c
437 and Stx1d, and Stx2a through Stx2g, respectively. Reference protein sequences were downloaded for
438 each Stx subtype and Stx type variant from GenBank [13]. Amino acid sequences of the A and B
439 subunits were concatenated and aligned separately for Stx1 and Stx2 using Clustal O in Jalview
440 [78,79]. For cluster analysis and tree calculations the Neighbour Joining algorithm in Jalview using %
441 identity was used. Clustering of the Stx protein sequences of the sequenced strains with reference
442 sequences was used to classify the former into Stx1 and Stx2 subtypes.

443 **Statistical analysis**

444 Fisher's exact test was used to analyse if specific *stx* subtypes were differently distributed in LEE
445 positive and LEE negative STEC, with a p-value ≤0.05 regarded as statistically significant. Fisher's

446 exact test was also used to test if specific genes in the accessory genome were overrepresented, and
447 for overrepresentation of gene variants in the core genome, in subgroups of the 95 STEC strains.
448 Classification of the strains into subgroups was based on clinical and outbreak investigation
449 information, phylogenetic analysis, and PCA and PLS regression (Table 2). For corrections of false
450 discovery rate (FDR) in multiple testing the Benjamini-Hochberg method was used, with FDR≤0.01
451 regarded as statically significant. Whenever no significant association was identified after FDR
452 correction, results for uncorrected analysis are given. The statistical analyses were performed using
453 the R software package version 3.03 (R: A Language and Environment for Statistical Computing,
454 http://www.R-project.org). In addition, in Blast2GO, Fisher's exact test was used for GO enrichment
455 analysis.

Acknowledgement

The authors acknowledge the microbiological laboratories in Norway for sending in their isolates to the Norwegian Institute of Public Health (NIPH). They thank the staff at The Department of Food-borne Infections, NIPH, for their skilful assistance.

References

1. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, et al. (2013) Recent Advances in Understanding Enteric Pathogenic Escherichia coli. Clin Microbiol Rev 26: 822-880.

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

3. Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, et al. (1983) Hemorrhagic colitis associated with a rare Escherichia coli serotype. New Eeng J Med 308: 681-685.

4. Gould LH, Mody RK, Ong KL, Clogher P, Cronquist AB, et al. (2013) Increased Recognition of Non-O157 Shiga Toxin-Producing Escherichia coli Infections in the United States During 2000-2010: Epidemiologic Features and Comparison with E. coli O157 Infections. Foodborne Pathog Dis 10: 453-460.

5. European Centre for Disease Prevention and Control and European Food Safety Authority (2011) Shiga toxin/verotoxin-producing Escherichia coli in humans, food and animals in the EU/EEA, with special reference to the German outbreak strain STEC O104. ECDC. ISBN: 978-92-9193-298-6 ISBN: 978-92-9193-298-6.

6. Allison HE (2007) Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. Future Microbiol 2: 165-174.

7. Laing CR, Zhang Y, Gilmour MW, Allen V, Johnson R, et al. (2012) A comparison of Shiga-toxin 2 bacteriophage from classical enterohemorrhagic Escherichia coli serotypes and the German E. coli O104:H4 outbreak strain. PLoS ONE 7: e37362.

8. O'Brien AD, Newland JW, Miller SF, Holmes RK, Smith HW, et al. (1984) Shiga-like toxin-converting phages from Escherichia coli strains that cause hemorrhagic colitis or infantile diarrhea. Science 226: 694-696.

9. Smith DL, Rooks DJ, Fogg PC, Darby AC, Thomson NR, et al. (2012) Comparative genomics of Shiga toxin encoding bacteriophages. BMC Genomics 13: 311.

10. Bielaszewska M, Friedrich AW, Aldick T, Schurk-Bulgrin R, Karch H (2006) Shiga toxin activatable by intestinal mucus in Escherichia coli isolated from humans: predictor for a severe clinical outcome. Clin Infect Dis 43: 1160-1167.

11. Friedrich AW, Bielaszewska M, Zhang WL, Pulz M, Kuczius T, et al. (2002) Escherichia coli harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. J Infect Dis 185: 74-84.

12. Persson S, Olsen KE, Ethelberg S, Scheutz F (2007) Subtyping method for Escherichia coli shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. J Clin Microbiol 45: 2020-2024.

13. Scheutz F, Teel LD, Beutin L, Pierard D, Buvens G, et al. (2012) Multicenter evaluation of a sequence-based protocol for subtyping shiga toxins and standardizing stx nomenclature. J Clin Microbiol 50: 2951-2963.

14. Delannoy S, Beutin L, Fach P (2013) Discrimination of Enterohemorrhagic Escherichia coli (EHEC) from Non-EHEC Strains Based on Detection of Various Combinations of Type III Effector Genes. J Clin Microbiol 51: 3257-3262.

15. Mellmann A, Bielaszewska M, Kock R, Friedrich AW, Fruth A, et al. (2008) Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic Escherichia coli. Emerg Infect Dis 14: 1287-1290.

16. Bonnet R, Souweine B, Gauthier G, Rich C, Livrelli V, et al. (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.

17. Elliott EJ, Robins-Browne RM, O'Loughlin EV, Bennett-Wood V, Bourke J, et al. (2001) Nationwide study of haemolytic uraemic syndrome: clinical, microbiological, and epidemiological features. Arch Dis Child 85: 125-131.

18. Kappeli U, Hachler H, Giezendanner N, Beutin L, Stephan R (2011) Human infections with non-O157 Shiga toxin-producing Escherichia coli, Switzerland, 2000-2009. Emer Infect Dis 17: 180-185.

19. Buvens G, De Rauw K, Roisin S, Vanfraechem G, Denis O, et al. (2013) Verocytotoxin-producing Escherichia coli O128ab:H2 bacteremia in a 27-year-old male with hemolytic-uremic syndrome. J Clin Microbiol 51: 1633-1635.

20. Steyert SR, Sahl JW, Fraser CM, Teel LD, Scheutz F, et al. (2012) Comparative genomics and stx phage characterization of LEE-negative Shiga toxin-producing Escherichia coli. Front Cell Infect Microbiol 2: 133.

21. Frank C, Werber D, Cramer JP, Askar M, Faber M, et al. (2011) Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. N Engl J Med 365: 1771-1780.

22. Garmendia J, Frankel G, Crepin VF (2005) Enteropathogenic and enterohemorrhagic Escherichia coli infections: translocation, translocation, translocation. Infect Immun 73: 2573-2585.

23. Mcdaniel TK, Jarvis KG, Donnenberg MS, Kaper JB (1995) A Genetic-Locus of Enterocyte Effacement Conserved among Diverse Enterobacterial Pathogens. P Natl Acad Sci USA 92: 1664-1668.

24. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, et al. (2001) Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. DNA Research 8: 11-22.

25. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, et al. (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic Escherichia coli. P Natl Acad Sci USA 106: 17939-17944.

26. Perna NT, Plunkett G, 3rd, Burland V, Mau B, Glasner JD, et al. (2001) Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature 409: 529-533.

27. Rohde H, Qin J, Cui Y, Li D, Loman NJ, et al. (2011) Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. N Engl J Med 365: 718-724.

28. Hazen TH, Sahl JW, Fraser CM, Donnenberg MS, Scheutz F, et al. (2013) Refining the pathovar paradigm via phylogenomics of the attaching and effacing Escherichia coli. P Natl Acad Sci USA 110: 12810-12815.

29. Lawrence JG, Ochman H (1998) Molecular archaeology of the Escherichia coli genome. P Natl Acad Sci USA 95: 9413-9417.

30. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, et al. (2008) The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol 190: 6881-6893.

31. Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS (2000) Parallel evolution of virulence in pathogenic Escherichia coli. Nature 406: 64-67.

32. Welch RA, Burland V, Plunkett G, 3rd, Redford P, Roesch P, et al. (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. P Natl Acad Sci USA 99: 17020-17024.

33. Kaas RS, Friis C, Ussery DW, Aarestrup FM (2012) Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. BMC Genomics 13: 577.

34. Lukjancenko O, Wassenaar TM, Ussery DW (2010) Comparison of 61 sequenced Escherichia coli genomes. Microbial Ecol 60: 708-720.

35. Jenssen GR, Hovland E, Bjerre A, Bangstad HJ, Nygard K, et al. (2014) Incidence and etiology of hemolytic-uremic syndrome in children in Norway, 1999-2008 - a retrospective study of hospital records to assess the sensitivity of surveillance. BMC Infect Dis 14: 265.

36. Jackson RW, Vinatzer B, Arnold DL, Dorus S, Murillo J (2011) The influence of the accessory genome on bacterial pathogen evolution. Mob Genet Elements 1: 55-65.

37. Bugarel M, Martin A, Fach P, Beutin L (2011) Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) Escherichia coli strains: a basis for molecular risk assessment of typical and atypical EPEC strains. BMC Microbiology 11: 142.

38. Coombes BK, Wickham ME, Mascarenhas M, Gruenheid S, Finlay BB, et al. (2008) Molecular analysis as an aid to assess the public health risk of non-O157 Shiga toxin-producing Escherichia coli strains. Appl Environ Microb 74: 2153-2160.

39. Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, et al. (2003) Association of genomic O island 122 of Escherichia coli EDL 933 with verocytotoxin-producing Escherichia coli seropathotypes that are linked to epidemic and/or serious disease. J Clin Microbiol 41: 4930-4940.

570 40. Bugarel M, Beutin L, Martin A, Gill A, Fach P (2010) Micro-array for the identification of Shiga
571 toxin-producing Escherichia coli (STEC) seropathotypes associated with Hemorrhagic Colitis
572 and Hemolytic Uremic Syndrome in humans. Int J Food Microbiol 142: 318-329.

573 41. Buvens G, Pierard D (2012) Virulence profiling and disease association of verocytotoxin-producing
574 Escherichia coli O157 and non-O157 isolates in Belgium. Foodborne Pathog Dis 9: 530-535.

575 42. Tobe T, Beatson SA, Taniguchi H, Abe H, Bailey CM, et al. (2006) An extensive repertoire of type III
576 secretion effectors in Escherichia coli O157 and the role of lambdoid phages in their
577 dissemination. P Natl Acad Sci USA 103: 14941-14946.

578 43. Fuller CA, Pellino CA, Flagler MJ, Strasser JE, Weiss AA (2011) Shiga toxin subtypes display
579 dramatic differences in potency. Infect Immun 79: 1329-1337.

580 44. Escobar-Paramo P, Clermont O, Blanc-Potard AB, Bui H, Le Bouguenec C, et al. (2004) A specific
581 genetic background is required for acquisition and expression of virulence factors in
582 Escherichia coli. Mol Biol Evol 21: 1085-1094.

583 45. Peekhaus N, Conway T (1998) What's for dinner?: Entner-Doudoroff metabolism in Escherichia
584 coli. J Bacteriol 180: 3495-3502.

585 46. Nivaskumar M, Francetic O (2014) Type II secretion system: A magic beanstalk or a protein
586 escalator. Biochim Biophys Acta.

587 47. Baldi DL, Higginson EE, Hocking DM, Praszkier J, Cavaliere R, et al. (2012) The type II secretion
588 system and its ubiquitous lipoprotein substrate, SslE, are required for biofilm formation and
589 virulence of enteropathogenic Escherichia coli. Infect Immun 80: 2042-2052.

590 48. Duan Q, Zhou M, Zhu L, Zhu G (2013) Flagella and bacterial pathogenicity. J Basic Microbiol 53: 1-
591 8.

592 49. Ratledge C, Dover LG (2000) Iron metabolism in pathogenic bacteria. Annu Rev Microbiol 54: 881-
593 941.

594 50. Saha R, Saha N, Donofrio RS, Bestervelt LL (2013) Microbial siderophores: a mini review. J Basic
595 Microbiol 53: 303-317.

596 51. Karch H, Schubert S, Zhang D, Zhang W, Schmidt H, et al. (1999) A genomic island, termed high-
597 pathogenicity island, is present in certain non-O157 Shiga toxin-producing Escherichia coli
598 clonal lineages. Infect Immun 67: 5994-6001.

599 52. Bielaszewska M, Zhang W, Mellmann A, Karch H (2007) Enterohaemorrhagic Escherichia coli
600 O26:H11/H-: a human pathogen in emergence. Berl Munch Tierarztl Wochenschr 120: 279-
601 287.

602 53. Schuller S (2011) Shiga toxin interaction with human intestinal epithelium. Toxins (Basel) 3: 626-
603 639.

604 54. Centers for Disease Control and Prevention (CDC) (2012) National Shiga toxin-producing
605 Escherichia coli (STEC) Surveillance Annual Summary, 2009. Atlanta, Georgia: US Department
606 of Health and Human Services, CDC, 2012. CDC.

607 55. European Centre for Disease Prevention and Control (2013) Annual Epidemiological Report 2013.
608 Stockholm: ECDC; 2013: European Centre for Disease Prevention and Control. Q-AC-14-001-
609 EN-C Q-AC-14-001-EN-C.

610 56. Lindstedt BA, Brandal LT, Aas L, Vardund T, Kapperud G (2007) Study of polymorphic variable-
611 number of tandem repeats loci in the ECOR collection and in a set of pathogenic Escherichia
612 coli and Shigella isolates for use in a genotyping assay. J Microbiol Meth 69: 197-205.

613 57. Lobersli I, Haugum K, Lindstedt BA (2012) Rapid and high resolution genotyping of all Escherichia
614 coli serotypes using 10 genomic repeat-containing loci. J Microbiol Meth 88: 134-139.

615 58. Brandal LT, Lindstedt BA, Aas L, Stavnes TL, Lassen J, et al. (2007) Octaplex PCR and fluorescence-
616 based capillary electrophoresis for identification of human diarrheagenic Escherichia coli and
617 Shigella spp. J Microbiol Meth 68: 331-341.

618 59. Brandal LT, Sekse C, Lindstedt BA, Sunde M, Lobersli I, et al. (2012) Norwegian sheep are an
619 important reservoir for human-pathogenic Escherichia coli O26:H11. Appl Environ Microbiol
620 78: 4083-4091.

621 60. Haugum K, Brandal LT, Lindstedt BA, Wester AL, Bergh K, et al. (2014) PCR based detection of
622 Shiga toxin-producing Escherichia coli (STEC) in a routine microbiology laboratory over 16
623 years: molecular characterization of strains. J Clin Microbiol.

624 61. Schimmer B, Nygard K, Eriksen HM, Lassen J, Lindstedt BA, et al. (2008) Outbreak of haemolytic
625 uraemic syndrome in Norway caused by stx2-positive Escherichia coli O103:H25 traced to
626 cured mutton sausages. BMC Infect Dis 8: 41.

627 62. Bielaszewska M, Kock R, Friedrich AW, von Eiff C, Zimmerhackl LB, et al. (2007) Shiga toxin-
628 mediated hemolytic uremic syndrome: time to change the diagnostic paradigm? PLoS ONE 2:
629 e1024.

630 63. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn
631 graphs. Genome Res 18: 821-829.

632 64. Ribeiro FJ, Przybylski D, Yin S, Sharpe T, Gnerre S, et al. (2012) Finished bacterial genomes from
633 shotgun sequence data. Genome Res 22: 2270-2277.

634 65. Luo R, Liu B, Xie Y, Li Z, Huang W, et al. (2012) SOAPdenovo2: an empirically improved memory-
635 efficient short-read de novo assembler. Gigascience 1: 18.

636 66. Ronen R, Boucher C, Chitsaz H, Pevzner P (2012) SEQuel: improving the accuracy of genome
637 assemblies. Bioinformatics 28: i188-196.

638 67. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, et al. (2013) Nonhybrid, finished microbial
639 genome assemblies from long-read SMRT sequencing data. Nat Meth 10: 563-569.

640 68. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, et al. (2010) Prodigal: prokaryotic gene
641 recognition and translation initiation site identification. BMC Bioinformatics 11: 119.

642 69. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. (2008) The RAST Server: rapid annotations
643 using subsystems technology. BMC Genomics 9: 75.

644 70. Vesth T, Lagesen K, Acar O, Ussery D (2013) CMG-biotools, a free workbench for basic
645 comparative microbial genomics. PLoS One 8: e60120.

646 71. Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, et al. (2008) Phylogenetic and genomic
647 diversity of human bacteremic Escherichia coli strains. BMC Genomics 9: 560.

648 72. Tenaillon O, Skurnik D, Picard B, Denamur E (2010) The population genetics of commensal
649 Escherichia coli. Nat Rev Microbiol 8: 207-217.

650 73. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, et al. (2009) Organised genome dynamics
651 in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet 5:
652 e1000344.

653 74. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software
654 Suite. Trends Genet 16: 276-277.

655 75. Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions and reversals.
656 Soviet Physics Doklady 10: 707-710.

657 76. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al. (2005) Blast2GO: a universal tool for
658 annotation, visualization and analysis in functional genomics research. Bioinformatics 21:
659 3674-3676.

660 77. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, et al. (2008) High-throughput
661 functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36: 3420-
662 3435.

663 78. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple
664 sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191.

665 79. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-
666 quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7: 539.

667

668

Figure legends

Figure 1. Core gene phylogeny of the 95 sequenced non-O157 STEC and 14 *E. coli* reference genomes. The tree was rooted in Figtree (http://tree.bio.ed.ac.uk/software/figtree/) by midpoint rooting. For an unrooted version of the phylogenetic tree, see Figure S1. The *E. coli* phylogroups are marked with the colours blue (A), green (B1), orange (B2), yellow (D), ochre (F) and indigo (E). Bootstrap values were scaled from 0-1, and blue circles indicate a bootstrap value of ≥0.8. LEE positive STEC were marked with •, while all HUS and HUS-associated STEC included in the study were indicated with red letters.

Figure S1: Unrooted core gene phylogeny. Unrooted core gene tree of the 95 sequenced non-O157 STEC and 14 *E. coli* reference genomes. The *E. coli* phylogroups are marked with the colours blue (A), green (B1), orange (B2), yellow (D), ochre (F) and indigo (E). LEE positive STEC were marked with a #-sign, while all HUS and HUS-associated STEC included in the study were coloured with red letters. HUS-group 1 consisted of 18 STEC strains in three related clusters, mainly strains of serotypes O103:H25, O145:H[unknown], and O121:H-. HUS-group 2 consisted of 23 STEC strains in one cluster, mainly strains of serogroups O26, O86 and O111.

Tables

Table 1. Distribution of *stx1*, *stx2* and their subtypes in 95 Norwegian non-O157 LEE positive and LEE negative STEC strains.

	LEE positive n=54		**LEE negative n=41**		**Total n=95**		
Gene	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**p-value**
stx1	21	38.9	14	34.1	35	36.8	ND
stx2	22	40.7	15	36.6	37	38.9	ND
stx1+stx2	4	7.4	8	19.5	12	12.6	ND
STEC-LST[1]	7[2]	13.0	4[3]	9.8	11	11.6	ND
stx1 subtype							
stx1a	24	44.4	10	24.4	34	35.8	>0.05
stx1c	1	1.9	11	26.8	12	12.6	0.0003
stx1d	0	0.0	1	2.4	1	1.1	>0.05
stx2 subtype							
stx2a	24	44.4	5	12.2	29	30.5	0.00073
stx2b	0	0.0	14	34.1	14	14.7	0.0000017
stx2c	2	3.7	1	2.4	3	3.2	>0.05
stx2d	0	0.0	2	4.9	2	2.1	>0.05
stx2e	0	0.0	1	2.4	1	1.1	>0.05

[1]STEC-LST: STEC that has lost Shiga toxin.

[2]Six of these strains, which were stx2 negative *E. coli* when initially tested at the Norwegian Public Health Institute, had been isolated from a patient with HUS or had a MLVA profile identical to an outbreak STEC strain and was epidemiologically related to that HUS case. The last strain had been stx2 positive when initially tested, but had lost the stx-gene at a later stage.

[3]When initially tested, three of these strains contained stx1 and one strain contained stx2.

694 Table 2. Overview of the different subgroups of STEC that were compared in this study. In the upper
695 half of the table, different groups were compared with respect to gene content in the accessory
696 genome. In the lower half, the groups were compared with respect to gene variants in the core
697 genome.

Gene source	Groups defined by	Groups of strains that were compared	Number of genes overrepresented in group	False discovery rate (FDR)	Number of genes or gene variants unique to group
Accessory genome	Clinical and epidemiological information	LEE+/*stx2+* HUS[1] n=23	11	≤0.01	0
		Other LEE+ n=31	4	≤0.01	0
		LEE+/*stx2+* HUS[1] n=23	69	≥0.01	0
		LEE+/*stx2+* non-HUS n=10	44	≥0.01	0
	Core gene phylogeny	HUS-group 1 (LEE+) n=18	357	≤0.01	1[2]
		Other LEE+ n=36	365	≤0.01	0
		HUS-group 2 (LEE+) n=23	576	≤0.01	4
		Other LEE+ n=31	218	≤0.01	0
		LEE+ O26 HUS[1] n=5	17	≥0.01	0
		O26 non-HUS n=13	13	≥0.01	0
Core genome	Clinical and epidemiological information	LEE+/*stx2+* HUS[1] n=23	281	≤0.01	0
		Other LEE+ n=31	0	≤0.01	0
	Core gene phylogeny	O26 *stx2+* n=8	87	≥0.01	0
		O26 *stx1+* n=10	83	≥0.01	1
		O26 HUS[1] n=5	84	≥0.01	0
		O26 non-HUS n=13	78	≥0.01	0

[1]HUS: HUS-associated STEC.

[2]The gene was not exclusive to this group as it was also found in one LEE negative STEC.

Figures

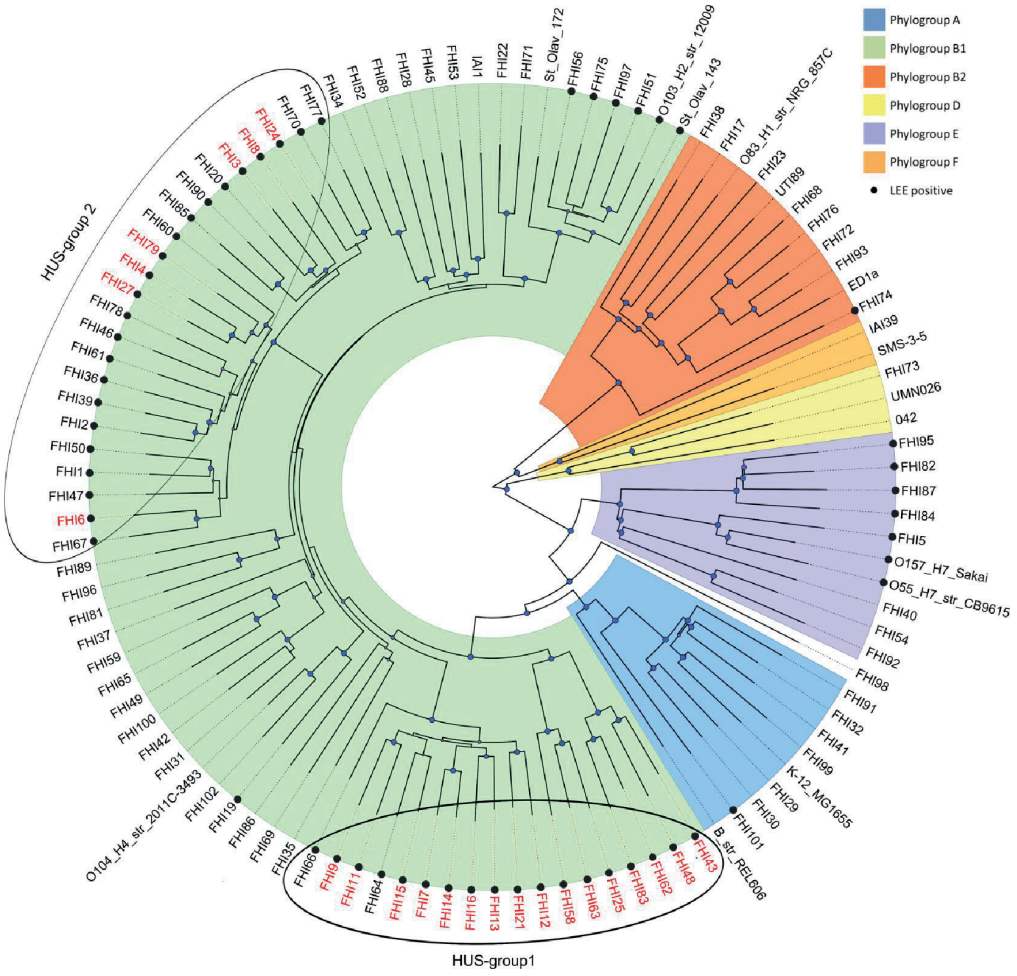
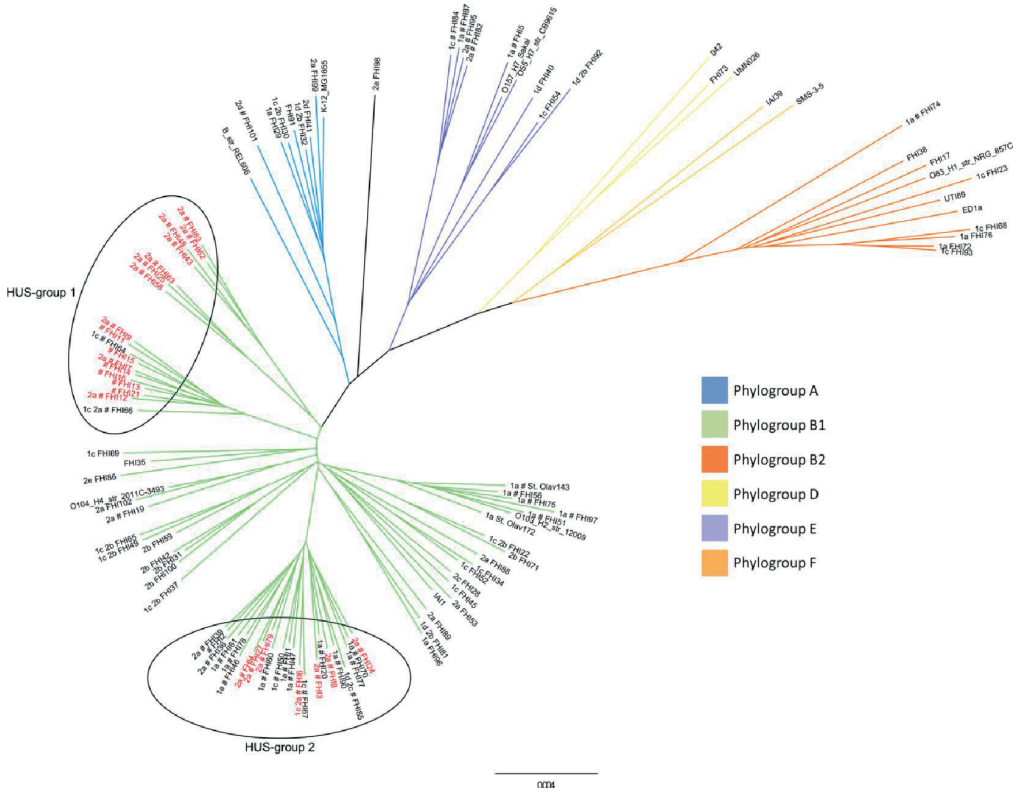


Figure 1.



705

706 Figure S1.

Table S1. Information on 95 Norwegian non-O157 STEC genomes sequenced and analyzed in this study.

		Strain ID	O-type	Lab_stx1[6]	Lab_stx2[6]	Lab_eae[6]	Pred_stx1[7]	Pred_stx2[7]	Pred_eae[7]	stx1 subtype	stx2 subtype	MLVA Allele[8]	Source	Clinic	HUS-associated	Outbreak	Sex	Age	Accession No.	Sequencing method	Contigs
LEE positive STEC n=54	LEE positive and stx2 positive STEC n=33[1] (35)	HUS and HUS-associated STEC n=23[1] (25)	FHI3	26	0	1	1	0	1	1		stx2a	6-1-0-8-3-5-1-6-11-15	Human faeces	Unknown	1	1	F	31	ERS480135	Illumina PE[9]	130
			FHI4	26	0	1	1	0	1	1		stx2a	6-1-0-8-3-5-1-6-11-15	Human faeces	HUS	1	1	F	4	ERS480136	Illumina PE, MP[10]	35
			FHI24	26	0	1	1	0	1	1		stx2a	6-1-0-8-3-4-1	Human faeces	HUS	1	0	F	2	ERS480154	Illumina PE, MP	42
			FHI27	26	1	1	1	0	1	1		stx2a	6-1-0-8-3-4-1	Human faeces	HUS	1	0	M	1	ERS480156	Illumina PE, MP	27
			FHI79	26	0	1	1	0	1	1		stx2a	6-0-0-8-3-4-1-6-13-12	Human faeces	HUS	1	0	M	0	ERS480204	Illumina PE, MP	28
			FHI8	86	0	1	1	0	1	1		stx2a	6-3-0-8-3-7-1	Human	HUS	1	0	F	1	ERS480140	Illumina PE, MP	33
			FHI7	103	0	1	1	0	1	1		stx2a	7-3-0-5-0-8-1-16-9-11	Human faeces	HUS	1	0	F	1	ERS480139	Illumina PE, MP	30
			FHI9	103	0	1	1	0	1	1		stx2a	7-3-0-5-0-3-1-16-9-11	Human faeces	HUS	1	0	F	0	ERS480141	Illumina PE, MP	34
			FHI10[2,3]	103	0	0	1	0	0	1			7-3-0-5-0-7-1-16-9-12	Human faeces	HUS	1	2	F	4	-	Illumina PE	ND
			FHI11[3]	103	0	0	1	0	0	1			7-3-0-5-0-7-1-16-9-12	Human faeces	HUS	1	2	F	2	ERS480142	Illumina PE, MP	25
			FHI12	103	0	1	1	0	1	1		stx2a	7-3-0-5-0-7-1-16-9-12	Human faeces	HUS	1	2	F	4	ERS480143	Illumina PE, MP	31
			FHI13[3]	103	0	0	1	0	0	1			7-3-0-5-0-7-1	Human faeces	HUS	1	2	F	1	ERS480144	Illumina PE	120
			FHI14[3]	103	0	1	1	0	0	1			7-3-0-5-0-7-1	Human faeces	HUS	1	2	F	2	ERS480145	Illumina PE, MP	29
			FHI15[3]	103	0	0	1	0	0	1			7-3-0-5-0-7-1	Human faeces	HUS	1	2	F	1	ERS480146	Illumina PE, MP	42
			FHI16[3]	103	0	0	1	0	0	1			7-3-0-5-0-7-1-16-9-11	Fermented sausage	NA	1	2	NA	NA	ERS480147	Illumina PE	129
			FHI21[3]	103	0	0	1	0	0	1			7-3-0-5-0-5-1	Human faeces	HUS	1	2	M	1	ERS480151	Illumina PE, MP	32
			FHI6	111	1	1	1	1	1	1	stx1a	stx2a	6-3-0-5-3-10-1	Human faeces	HUS	1	0	F	1	ERS480138	Illumina PE, MP	33
			FHI43	121	0	1	1	0	1	1		stx2a	6-3-0-5-3-6-1-6-11-13	Human faeces	Gastroenteritis	1	3	M	1	ERS480171	Illumina PE, MP	17
			FHI48	121	0	1	1	0	1	1		stx2a	6-3-0-5-3-6-1-6-11-16	Human faeces	HUS	1	3	F	1	ERS480175	Illumina PE, MP	34
			FHI62	121	0	1	1	0	1	1		stx2a	6-3-0-5-3-5-1-6-11-15	Human faeces	Gastroenteritis	1	3	F	2	ERS480187	Illumina PE, MP	23
			FHI83	121	0	1	1	0	1	1		stx2a	8-0-0-6-3-6-1-6-11-0	Human faeces	HUS	1	0	F	1	ERS480207	Illumina PE, MP	27
			FHI25	145	0	1	1	0	1	1		stx2a	5-3-0-8-4-1-1-16-8-12	Human faeces	HUS	1	0	M	2	ERS480155	Illumina PE, MP	27
			FHI58	145	0	1	1	0	1	1		stx2a	5-3-0-8-4-1-1-16-9	Human faeces	HUS	1	4	M	1	ERS480183	Illumina PE, MP	28
			FHI63	145	0	1	1	0	1	1		stx2a	5-3-0-8-4-1-1-16-9	Human faeces	HUS	1	4	M	1	ERS480188	Illumina PE, MP	43
			St. Olav104[4]	145	0	1	1	0	1	1		stx2a	5-3-0-8-4-1-1-16-9	Human faeces	Asymptomatic	1	4	F	1	ERS480228	Pacific Biosciences	78
			FHI5	0	1	0	0	1	1	1	stx1a	stx2c	6-10-3-5-4-7-2-6-9	Human faeces	Gastroenteritis	0	0	M	4	ERS480137	Illumina PE	89
			FHI66	0	1	1	1	1	1	1	stx1c	stx2a	7-3-0-5-4-7-1-16-11	Human faeces	Asymptomatic	0	0	M	5	ERS480191	Illumina PE, MP	32
			FHI85	0	1	1	1	1	1	1	stx1a	stx2c	5-3-0-8-3-4-1-6-20-0	Human faeces	ND	0	0	F	69	ERS480209	Illumina PE, MP	33
			FHI101	0	0	1	1	0	1	1		stx2d	5-1-0-8-3-9-1-64-0-13	Human faeces	Gastroenteritis	0	0	F	1	ERS480224	Illumina PE	152
			FHI36	26	0	1	1	0	1	1		stx2a	6-1-0-8-3-5-1-6-16	Human faeces	Bloody diarrhoea	0	0	F	1	ERS480164	Illumina PE	137
			FHI39	26	0	1	1	0	1	1		stx2a	6-1-0-8-3-5-1-6-15	Sheep faeces	NA	0	0	NA	NA	ERS480167	Illumina PE	133
			FHI2	26	0	1	1	0	0	1			6-0-0-8-3-4-1-6-15-10	Human faeces	ND	0	0	F	36	ERS480134	Illumina PE	157
			FHI19	104	0	1	0	0	1	1		stx2a	6-3-10-8-3-6-1-6-6-0	Human faeces	Gastroenteritis	0	0	M	74	ERS480149	Illumina PE	283
			FHI82	145	0	1	1	0	1	1		stx2a	7-0-0-8-3-2-1-35-0-0	Human faeces	Bloody diarrhoea	0	0	M	21	ERS480206	Illumina PE, MP	29
			FHI95	145	0	1	1	0	1	1		stx2a	7-3-0-8-3-2-1-35-0-0	Human faeces	ND	0	0	F	2	ERS480218	Illumina PE	100
			FHI51	0	1	0	1	1	0	1	stx1a		6-3-0-8-3-7-1-6-7	Human faeces	ND	0	0	F	3	ERS480178	Illumina PE	189
			FHI64	0	1	0	1	1	0	1	stx1a		5-3-0-5-4-5-1-6-10	Human faeces	Bloody diarrhoea	0	0	F	2	ERS480189	Illumina PE	111
			FHI74	0	1	0	1	1	0	1	stx1a		6-3-0-8-3-5-1-6-0-13	Human faeces	ND	0	0	M	3	ERS480199	Illumina PE, MP	23
			FHI90	0	1	0	1	1	0	1	stx1a		6-0-0-8-3-6-1-6-10-11	Human faeces	ND	0	0	F	1	ERS480214	Illumina PE, MP	31
			FHI1	26	1	0	1	1	0	1	stx1a		6-0-0-8-3-5-1-6-23	Human faeces	ND	0	0	M	2	ERS480133	Illumina PE	151
			FHI20	26	1	0	1	1	0	1	stx1a		6-1-0-8-3-5-1-6-23	Human faeces	Gastroenteritis	0	0	F	1	ERS480150	Illumina PE, MP	26
			FHI46	26	1	0	1	1	0	1	stx1a		6-0-0-8-3-6-1-6-40-13	Human faeces	Bloody diarrhoea	0	0	F	22	ERS480173	Illumina PE	162
			FHI47	26	1	0	1	1	0	1	stx1a		6-1-0-5-3-9-1-6-21	Human faeces	Bloody diarrhoea	0	0	F	24	ERS480174	Illumina PE	211
			FHI50	26	1	0	1	1	0	1	stx1a		6-0-0-8-3-6-1-6-40	Human faeces	Bloody diarrhoea	0	0	F	42	ERS480177	Illumina PE	194
			FHI60	26	1	0	1	1	0	1	stx1a		6-0-0-8-3-6-1-6-22	Human faeces	Bloody diarrhoea	0	0	M	32	ERS480185	Illumina PE	252
			FHI61	26	1	0	1	1	0	1	stx1a		6-0-0-8-3-4-1-6-20	Human faeces	Gastroenteritis	0	0	F	1	ERS480186	Illumina PE	447
			FHI70	26	1	0	1	1	0	1	stx1a		6-0-0-8-3-15-1-0-39-16	Human faeces	Annet	0	0	M	0	ERS480195	Illumina PE, MP	40
			FHI77	26	1	0	1	1	0	1	stx1a		6-0-0-8-3-6-1-6-0-17	Human faeces	Bloody diarrhoea	0	0	M	9	ERS480202	Illumina PE	246
			FHI78	26	1	0	1	1	0	1	stx1a		6-0-0-8-3-7-1-6-21-15	Human faeces	Bloody diarrhoea	0	0	M	0	ERS480203	Illumina PE	243
			FHI56	103	1	0	1	1	0	1	stx1a		6-3-0-8-3-5-7-6-7	Human faeces	Gastroenteritis	0	0	M	0	ERS480182	Illumina PE	197
			FHI75	103	1	0	1	1	0	1	stx1a		6-3-0-8-3-5-7-6-7-0	Human faeces	ND	0	0	M	6	ERS480200	Illumina PE, MP	26
			FHI97	103	1	0	1	1	0	1	stx1a		6-3-0-8-3-6-1-6-10-15	Human faeces	Bloody diarrhoea	0	0	F	1	ERS480220	Illumina PE, MP	46
			St. Olav143	103	1	0	1	1	0	1	stx1a		6-3-0-8-3-4-7-6-7-0	Human faeces	Asymptomatic	0	0	F	1	ERS480226	Illumina PE	102
			FHI67	111	1	0	1	1	0	1	stx1a		6-3-0-5-3-5-1-95-19	Human faeces	Gastroenteritis	0	0	M	0	ERS480192	Illumina PE	129
			FHI84	145	1	0	1	1	0	1	stx1a		1-0-0-8-3-7-1-35-0-0	Human faeces	Bloody diarrhoea	0	0	F	65	ERS480208	Illumina PE	150
			FHI87	145	1	0	1	1	0	1	stx1a		1-3-0-8-3-6-1-35-0-0	Human faeces	Asymptomatic	0	0	M	1	ERS480211	Illumina PE, MP	25
			FHI31	0	0	1	0	0	1	0		stx2b	6-3-0-8-3-6-1-6-6	Human faeces	ND	0	0	F	56	ERS480160	Illumina PE	87
			FHI38	0	0	1	0	0	0	0			7-0-0-8-3-7-1-0-0	Sheep faeces	NA	0	0	NA	NA	ERS480166	Illumina PE	47
			FHI53	0	0	1	0	0	1	0		stx2a	6-3-0-8-3-2-1-6-14	Human faeces	ND	0	0	M	5	ERS480180	Illumina PE	54
			FHI89	0	0	1	0	0	1	0		stx2a	6-3-0-8-3-11-1-6-9-0	Human faeces	Gastroenteritis	0	0	M	10	ERS480213	Illumina PE, MP	22
			FHI99	0	0	1	0	0	1	0		stx2a	5-3-0-8-3-14-1-6-0-0	Human faeces	Gastroenteritis	0	0	F	90	ERS480222	Illumina PE, MP	32
			FHI100	0	0	1	0	0	1	0		stx2b	6-3-0-8-3-10-1-6-6-0	Human faeces	ND	0	0	M	58	ERS480223	Illumina PE, MP	32
			FHI32	0	1	1	0	1	1	0	stx1c	stx2b	5-0-4-8-4-4-1-6-0	Human faeces	ND	0	0	F	51	ERS480161	Illumina PE	75

LEE negative STEC n=41

Strain ID	O-type	Lab_stx1[6]	Lab_stx2[6]	Lab_eae[6]	Pred_stx1[7]	Pred_stx2[7]	Pred_eae[7]	stx1 subtype	stx2 subtype	MLVA Allele[8]	Source	Clinic	HUS-associated	Outbreak	Sex	Age	Accession No.	Sequencing method	Contigs
FHI37	0	1	1	0	1	1	0	stx1c	stx2b	12-3-0-8-3-8-1-6-8	Sheep faeces	NA	0	0	NA	NA	ERS480165	Illumina PE	103
FHI92	0	1	1	0	1	1	0	stx1c	stx2b	6-14-0-8-3-3-1-16-6-0	Human faeces	Bloody diarrhoea	0	0	M	72	ERS480216	Illumina PE, MP	24
FHI98	2	0	1	0	0	1	0		stx2b	5-0-0-8-4-2-1-16-0-0	Human faeces	Gastroenteritis	0	0	F	84	ERS480221	Illumina PE, MP	31
FHI28	8	0	1	0	0	1	0		stx2c	6-3-0-8-3-8-1-6-11	Human faeces	Gastroenteritis	0	0	F	68	ERS480157	Illumina PE, MP	21
FHI86	8	0	1	0	0	1	0		stx2e	6-3-0-8-3-4-1-6-0-0	Human faeces	Gastroenteritis	0	0	F	20	ERS480210	Illumina PE	46
FHI42	84	0	1	0	0	1	0		stx2b	6-3-0-8-3-7-1-6-6	Human faeces	ND	0	0	F	4	ERS480170	Illumina PE, MP	33
FHI59	91	0	1	0	0	1	0		stx2b	6-3-0-8-3-4-1-6-7	Human faeces	ND	0	ND	ND	ND	ERS480184	Illumina PE, MP	22
FHI81	91	1	1	0	1	1	0	stx1a	stx2b	7-3-0-8-1-6-1-6-7-0	Human faeces	Gastroenteritis	0	0	F	1	ERS480205	Illumina PE	143
FHI102[5]	104	0	1	0	0	1	0		stx2a	6-3-0-8-3-10-1-6-6-0	Human faeces	Bloody diarrhoea	0	0	M	40	ERS480225	Illumina PE, MP	31
FHI88	111	0	1	0	0	1	0		stx2a	6-3-0-10-3-6-1-6-6-0	Human faeces	Bloody diarrhoea	0	0	F	1	ERS480212	Illumina PE	129
FHI35	113	0	1	0	0	1	0		stx2d	6-3-0-8-3-8-1-6-6	Minced meat	NA	0	0	NA	NA	ERS480163	Illumina PE	49
FHI41	113	0	1	0	0	1	0		stx2d	5-0-8-8-3-6-1-6-0	Human faeces	ND	0	0	F	71	ERS480169	Illumina PE	98
FHI30	113	1	1	0	1	1	0	stx1c	stx2b	5-0-5-8-3-6-1-6-0	Human faeces	ND	0	0	F	37	ERS480159	Illumina PE, MP	28
FHI71	128	0	1	0	0	1	0		stx2b	2-3-0-1-3-6-1-6-13-0	Human faeces	ND	0	0	F	0	ERS480196	Illumina PE, MP	27
FHI49	146	1	1	0	1	1	0	stx1c	stx2b	6-3-0-8-3-5-7-0-7	Human faeces	Asymptomatic	0	0	F	4	ERS480176	Illumina PE	102
FHI22	146	1	0	0	1	1	0	stx1c	stx2b	6-3-0-8-3-2-1-0-7	Human faeces	Asymptomatic	0	0	F	31	ERS480152	Illumina PE	101
FHI65	146	1	1	0	1	1	0	stx1a	stx2b	6-3-0-8-3-6-1-6-7-0	Human faeces	Bloody diarrhoea	0	0	F	55	ERS480190	Illumina PE, MP	26
FHI40	0	1	0	0	1	0	0	stx1d		6-15-0-8-3-1-1-6-3	Human faeces	Gastroenteritis	0	0	M	2	ERS480168	Illumina PE, MP	32
FHI45	0	1	0	0	1	0	0	stx1a		6-3-0-8-3-13-1-6-7	Human faeces	Gastroenteritis	0	0	M	2	ERS480172	Illumina PE	71
FHI54	0	1	0	0	1	0	0	stx1c		6-13-0-8-3-3-1-16-6	Human faeces	Gastroenteritis	0	0	M	3	ERS480181	Illumina PE	70
FHI69	0	1	0	0	1	0	0	stx1c		8-3-0-8-3-5-1-6-0-0	Human faeces	ND	0	0	M	0	ERS480194	Illumina PE	79
FHI73	0	1	0	0	0	0	0			7-16-0-8-3-2-1-55-3-0	Human faeces	ND	0	0	F	1	ERS480198	Illumina PE	62
FHI91	0	1	0	0	0	0	0			5-3-7-8-3-3-1-0-7-0	Human faeces	Gastroenteritis	0	0	F	45	ERS480215	Illumina PE	95
FHI96	0	1	0	0	1	0	0	stx1a		7-3-0-8-3-10-1-6-9-0	Human faeces	Gastroenteritis	0	0	F	68	ERS480219	Illumina PE	123
FHI17	26	1	0	1	0	0	0			6-0-0-8-3-7-1-0-6-0	Human faeces	ND	0	0	F	1	ERS480148	Illumina PE	65
FHI23	76	1	0	0	1	0	0	stx1c		6-0-0-8-3-9-1-6-0	Human faeces	Gastroenteritis	0	0	F	2	ERS480153	Illumina PE, MP	33
St. Olav172	103	1	0	0	1	0	0	stx1a		6-3-0-8-3-4-7-6-7-0	Human faeces	ND	0	0	F	0	ERS480227	Illumina PE	101
FHI34	104	1	0	0	1	0	0	stx1c		6-3-0-8-3-5-1-6-7	Human faeces	Bloody diarrhoea	0	0	M	0	ERS480162	Illumina PE, MP	26
FHI52	104	1	0	0	1	0	0	stx1c		6-3-0-8-3-4-1-6-7	Human faeces	Gastroenteritis	0	0	F	11	ERS480179	Illumina PE	171
FHI68	117	1	0	0	1	0	0	stx1a		5-0-0-8-3-5-1-0-37-0	Human faeces	Gastroenteritis	0	0	M	44	ERS480193	Illumina PE	171
FHI72	117	1	0	0	1	0	0	stx1a		5-0-0-8-3-6-1-0-36-0	Human faeces	Bloody diarrhoea	0	0	F	22	ERS480197	Illumina PE, MP	34
FHI76	117	1	0	0	1	0	0	stx1a		5-0-0-8-3-5-1-6-36-15	Human faeces	Gastroenteritis	0	0	F	32	ERS480201	Illumina PE	178
FHI93	117	1	0	0	1	0	0	stx1a		5-0-0-8-3-6-1-0-37-0	Human faeces	Gastroenteritis	0	0	F	45	ERS480217	Illumina PE	162
FHI29	118	1	0	0	1	0	0	stx1a		5-1-0-8-4-4-1-6-7-0	Human faeces	Gastroenteritis	0	0	M	1	ERS480158	Illumina PE, MP	24

Reference strains

Strain ID	O-type	stx1	stx2	eae	*E. coli* phylogroup	HUS-associated	Accession No.	Contigs
K-12 MG1655	16	0	0	0	A	0	U00096.3	1
B str. REL606	7	0	0	0	A	0	CP000819.1	1
IAI1	8	0	0	0	B1	0	CU928160.2	1
O104:H4 str. 2011C-3493	104	0	1	0	B1	1	CP003289.1	1
O103:H2 str. 12009	103	1	1	1	B1	0	AP010958.1	1
O83:H1 str. NRG 857C	83	0	0	0	B2	0	CP001855.1	1
ED1a	81	0	0	0	B2	0	CU928162.2	1
UTI89	0	0	0	0	B2	0	CP000243.1	1
UMN026	17	0	0	0	D	0	CU928163.2	1
042	O44	0	0	0	D	0	FN554766.1	1
O157:H7 str. Sakai	157	1	1	1	E	1	BA000007	1
O55:H7 str. CB9615	55	0	0	1	E	0	CP001846.1	1
IAI39	7	0	0	0	F	0	CU928164.2	1
SMS-3-5	19	0	0	0	F	0	CP000970.1	1

[1]FHI10 was excluded from the study due to contamination and St. Olav104 was included only in parts of the study. Therefore 23 strains were regarded as HUS-associated and 33 strains were regarded as LEE and *stx2* positive STEC throughout the study.
[2]Strain FHI10 was excluded from the study after sequencing because the strain was contaminated with another *E. coli*.
[3]The strains were part of an outbreak where *stx2a* originally was detected in only two of the isolated strains. After sequencing *stx2a* was detected in one strain only.
[4]Strain St. Olav104 was sequenced using a PacBio RSII instrument. The strain was included late in the study and therefore included only in parts of the analyses, i.e. genomic comparison with FHI58 and FHI63.
[5]Strain FHI102 was positive for the aggr gene and epidemiologically linked to the 2011 German outbreak strain.
[6]Laboratory results of *stx1, stx2* and *eae*.
[7]*stx1, stx2* and *eae* predicted by sequence analysis.
[8]Regarding MLVA alleles: some strains were analysed with the seven loci method while others were analysed with the ten loci method.
[9]Ilumina PE: Paired end sequencing with Illumina.
[10]Illumina PE, MP: Paired end and mate pair sequencing with Illumina.

Table S2. Analysis of accessory genome in HUS-associated STEC (n=23, G1) compared to other LEE positive STEC not assooiated with HUS (n=31, G2) and LEE negative STEC (n=41, G3)

List of overrepresented genes in HUS-associated STEC

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR<0.01	Function Blast2GO	Hit Accession	E-value
1	Failed to assign function	23	0	13	18	20	21	2.12793608780932e-06	0.00822454641960963	4-phosphopantetheinyl transferase	WP_001435131	4.33E-17
2	Mobile element protein	22	1	11	20	27	14	3.8279007159416e-06	0.00822454641960963	integrase core domain protein	YP_001464049	0
3	FIG01069681: hypothetical protein	16	7	2	29	1	40	1.20876327972868e-06	0.00822454641960963	hypothetical protein	WP_001597900	5.50E-110
4	Gluconokinase (EC 2.7.1.12)	16	7	3	28	16	25	6.26176706919338e-06	0.00822454641960963	thermosensitive gluconokinase	WP_000896732	5.44E-136
5	L-idonate 5-dehydrogenase (EC 1.1.1.264)	16	7	3	28	17	24	6.26176706919338e-06	0.00822454641960963	l-idonate 5-dehydrogenase	WP_001197417	0
6	5-keto-D-gluconate 5-reductase (EC 1.1.1.69)	16	7	3	28	17	24	6.26176706919338e-06	0.00822454641960963	gluconate 5-dehydrogenase	NP_418687	0
7	Positive regulator of L-idonate catabolism	16	7	3	28	17	24	6.26176706919338e-06	0.00822454641960963	hth-type transcriptional regulator idnr	NP_418685	0
8	Failed to assign function	16	7	3	28	7	34	6.26176706919338e-06	0.00822454641960963	hypothetical protein	WP_001542619	7.11E-12
9	Hydrogenase-4 component B (EC 1.-.-.-) / Formate hydrogenlyase subunit 3	16	7	3	28	15	26	6.26176706919338e-06	0.00822454641960963	nadh-ubiquinone plastoquinone (complex i) various chains family protein	WP_001443474	9.64E-89
10	Hydrogenase-4 component B (EC 1.-.-.-) / Formate hydrogenlyase subunit 3	16	7	3	28	16	25	6.26176706919338e-06	0.00822454641960963	hydrogenase-4 component b	WP_001443473	0
11	FIG00638146: hypothetical protein	13	10	0	31	6	35	1.03238645712482e-06	0.00822454641960963	phage protein	WP_001234560	3.30E-93

List of overrepresented genes in LEE positive STEC not associated with HUS

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR<0.01	Function Blast2GO	Hit Accession	E-value
1	Shiga toxin A-chain precursor (EC 3.2.2.22)	1	22	24	7	19	22	3.63693061361827e-08	0.000525463735055567	shiga toxin subunit a	ADG56725	0.0
2	FIG00638188: hypothetical protein	0	23	19	12	0	41	7.6842136504226e-07	0.00370071729404352	transcriptional regulator	YP_003078570	1.51286E-52
3	unnamed protein product	0	23	19	12	2	39	7.6842136504226e-07	0.00370071729404352	cytochrome b562 family protein	YP_325578	2.18237E-69
4	FIG00639033: hypothetical protein	0	23	18	13	0	41	2.12793608780932e-06	0.00768610514916727	arc-like dna binding domain protein	YP_002271195	1.66987E-22

Table S3. Analysis of accessory genome in HUS-associated STEC (n=23, G1) compared to other LEE positive/stx2 positive STEC not assooiated with HUS (n=10, G2) and other STEC (n=62, G3)
List of overrepresented genes in HUS-associated STEC

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR	Function Blast2GO	Hit Accession	E-value
1	Failed to assign function	23	0	4	6	29	33	0.000189604611184144	0.69464251193951	4-phosphopantetheinyl transferase	WP_001435131	4.33364E-17
2	hypothetical protein	23	0	5	5	31	31	0.00106178582263121	0.862621678578996	hypothetical protein P12B_c1353	YP_006168388	3.10068E-68
3	FIG00638567: hypothetical protein	23	0	6	4	35	27	0.00513196480938416	0.862621678578996	inner membrane protein ynji	WP_000756940	0.0
4	Fimbriae-like adhesin SfmA	23	0	6	4	36	26	0.00513196480938416	0.862621678578996	fimbrial family protein	WP_000827804	6.17523E-110
5	VapB protein (antitoxin to VapC)	23	0	6	4	36	26	0.00513196480938416	0.862621678578996	antitoxin	WP_021547092	1.78312E-44
6	Probable secreted protein	23	0	6	4	38	24	0.00513196480938416	0.862621678578996	sel1 repeat protein	YP_002291725	0.0
7	orf, conserved hypothetical protein	23	0	6	4	40	22	0.00513196480938416	0.862621678578996	toxin-antitoxin antitoxin ribbon-helix-helix domain protein	YP_002385906	2.35824E-52
8	orf, conserved hypothetical protein	23	0	6	4	40	22	0.00513196480938416	0.862621678578996	acetyltransferase family protein	YP_003220405	9.4725E-129
9	Arylsulfatase (EC 3.1.6.1)	23	0	6	4	40	22	0.00513196480938416	0.862621678578996	arylsulfatase domain protein	WP_001593828	2.48927E-6
10	PTS system HrsA EIIA component / PTS system HrsA EIIB component / PTS system HrsA permease IIC component	23	0	6	4	41	21	0.00513196480938416	0.862621678578996	heat-responsive suppressor	WP_001054852	0.0
11	FIG00639112: hypothetical protein	23	0	6	4	42	20	0.00513196480938416	0.862621678578996	permease family protein	YP_002388967	0.0
12	Alpha-fimbriae major subunit	23	0	6	4	42	20	0.00513196480938416	0.862621678578996	cs1 type fimbrial major subunit	WP_001401159	6.87027E-83
13	VapC toxin protein	23	0	6	4	42	20	0.00513196480938416	0.862621678578996	trna -specific endonuclease	YP_003222886	7.33838E-91
14	Putative outer membrane protein	23	0	5	5	44	18	0.00106178582263121	0.862621678578996	fimbrial outer membrane usher protein	WP_001192545	0.0
15	Putative membrane protein	23	0	6	4	44	18	0.00513196480938416	0.862621678578996	rhomboid family protein	WP_001276725	0.0
16	Failed to assign function	23	0	6	4	45	17	0.00513196480938416	0.862621678578996	conserved protein	WP_001236648	2.25887E-77
17	COG3609: Predicted transcriptional regulators containing the CopG/Arc/MetJ DNA-binding domain	23	0	5	5	46	16	0.00106178582263121	0.862621678578996	antitoxin 1	WP_001329834	5.20356E-39
18	FIG00638046: hypothetical protein	23	0	5	5	46	16	0.00106178582263121	0.862621678578996	plasmid stabilization system family protein	YP_001879798	2.78976E-63
19	Methylmalonyl-CoA mutase (EC 5.4.99.2)	23	0	6	4	47	15	0.00513196480938416	0.862621678578996	methylmalonyl- mutase	WP_000064232	0.0
20	Type VI secretion protein VasI	23	0	6	4	47	15	0.00513196480938416	0.862621678578996	type vi secretion-associated protein	YP_001461389	6.02141E-152
21	ClpB protein	23	0	5	5	48	14	0.00106178582263121	0.862621678578996	type vi secretion 1 family	WP_000614330	0.0
22	Alpha-mannosidase (EC 3.2.1.24)	23	0	6	4	48	14	0.00513196480938416	0.862621678578996	alpha-mannosidase mngb	WP_000649006	0.0
23	Hydroxyaromatic non-oxidative decarboxylase protein C (EC 4.1.1.-)	23	0	6	4	48	14	0.00513196480938416	0.862621678578996	phenolic acid decarboxylase subunit c	WP_000863214	0.0
24	Failed to assign function	23	0	6	4	48	14	0.00513196480938416	0.862621678578996	crispr-associated protein cse1	WP_001325175	0.0
25	Transcriptional regulator of succinyl CoA synthetase operon	23	0	6	4	49	13	0.00513196480938416	0.862621678578996	mannosyl-d-glycerate transport metabolism system repressor mngr	WP_000509898	7.35495E-157
26	2,3-diketo-L-gulonate TRAP transporter small permease protein yiaM	23	0	6	4	49	13	0.00513196480938416	0.862621678578996	#NAME?	WP_000129599	5.62766E-98
27	TRAP-type C4-dicarboxylate transport system, large permease component	23	0	6	4	50	12	0.00513196480938416	0.862621678578996	trap subunit	WP_000386845	0.0
28	Mobile element protein	23	0	6	4	51	11	0.00513196480938416	0.862621678578996	transposase	WP_002433462	6.35642E-32
29	Hypothetical zinc-type alcohol dehydrogenase-like protein YphC	23	0	6	4	58	4	0.00513196480938416	0.862621678578996	zinc-binding dehydrogenase family protein	YP_001463867	0.0
30	FIG00639664: hypothetical protein	22	1	5	5	14	48	0.00542269187986652	0.862621678578996	non-lee-encoded effector	NP_287959	0.0
31	orf; Unknown function	22	1	5	5	18	44	0.00542269187986652	0.862621678578996	conserved protein	YP_003223642	1.69924E-59
32	Mobile element protein	22	1	3	7	35	27	0.000202028844965442	0.69464251193951	integrase core domain protein	WP_000001467	0.0
33	Mobile element protein	22	1	3	7	36	26	0.000202028844965442	0.69464251193951	transposase	WP_024185760	3.58134E-72
34	Transcriptional regulator	22	1	5	5	41	21	0.00542269187986652	0.862621678578996	transcriptional regulator	WP_001111350	2.12238E-82

35	Mobile element protein	22	1	5	5	42	20	0.00542269187986652	0.862621678578996	insertion sequence 2 protein	AAZ86942	1.02729E-68
36	FIG00638873: hypothetical protein	22	1	5	5	43	19	0.00542269187986652	0.862621678578996	hypothetical protein ETEC_0272	YP_006113858	2.55617E-79
37	Phage tail fiber protein	21	2	4	6	13	49	0.00402869284960497	0.862621678578996	phage tail protein	WP_000108499	0.0
38	VgrG protein	21	2	4	6	22	40	0.00402869284960497	0.862621678578996	rhs element vgr family protein	WP_001348343	0.0
39	putative antirepressor protein	21	2	4	6	36	26	0.00402869284960497	0.862621678578996	p22ar c-terminal domain protein	YP_002271804	6.73359E-167
40	FIG00639844: hypothetical protein	21	2	4	6	41	21	0.00402869284960497	0.862621678578996	phage protein	WP_001071780	1.30525E-76
41	similar bacteriphage P22 ninH	21	2	4	6	48	14	0.00402869284960497	0.862621678578996	protein ninh	NP_049498	4.282E-37
42	Failed to assign function	19	4	3	7	21	41	0.00591540458058812	0.862621678578996	type iv secretion protein rhs	WP_001289080	0.0
43	FIG00642676: hypothetical protein	19	4	3	7	37	25	0.00591540458058812	0.862621678578996	phage protein	WP_000081296	6.75059E-172
44	General secretion pathway protein F	18	5	1	9	31	31	0.000421764923989618	0.862621678578996	type ii secretion system protein f	WP_001173467	0.0
45	Accessory colonization factor AcfD precursor	18	5	2	8	34	28	0.00279940674823879	0.862621678578996	lipoprotein precursor	WP_001034473	0.0
46	General secretion pathway protein I	18	5	2	8	34	28	0.00279940674823879	0.862621678578996	type ii secretion system protein i	YP_409353	7.66787E-63
47	General secretion pathway protein J	18	5	1	9	35	27	0.000421764923989618	0.862621678578996	type ii secretion system protein j	WP_001443429	3.24329E-134
48	General secretion pathway protein H	18	5	2	8	35	27	0.00279940674823879	0.862621678578996	general secretion pathway protein h	YP_002388445	4.04855E-124
49	General secretion pathway protein K	18	5	2	8	35	27	0.00279940674823879	0.862621678578996	general secretion pathway protein k	WP_000633212	0.0
50	General secretion pathway protein L	18	5	2	8	36	26	0.00279940674823879	0.862621678578996	type ii secretion system protein l	WP_000094990	0.0
51	General secretion pathway protein M	18	5	2	8	38	24	0.00279940674823879	0.862621678578996	type ii secretion system m family protein	WP_000942769	1.96035E-115
52	Hypothetical fimbrial chaperone yqiH	18	5	2	8	44	18	0.00279940674823879	0.862621678578996	gram-negative pili assembly c-terminal domain protein	YP_002388529	0.0
53	General secretion pathway protein E	18	5	2	8	44	18	0.00279940674823879	0.862621678578996	type ii secretion system protein e	YP_002294519	0.0
54	Lipid A biosynthesis (KDO) 2-(lauroyl)-lipid IVA acyltransferase (EC 2.3.1.-)	17	6	2	8	3	59	0.00596959584723768	0.862621678578996	lipid a biosynthesis 2- -lipid iva acyltransferase	WP_000790278	0.0
55	Mobile element protein	17	6	2	8	9	53	0.00596959584723768	0.862621678578996	is629 protein	WP_000165113	3.11966E-28
56	Leader peptidase (Prepilin peptidase) (EC 3.4.23.43) / N-methyltransferase (EC 2.1.1.-)	17	6	2	8	35	27	0.00596959584723768	0.862621678578996	leader peptidase pppa	WP_000895871	2.44755E-159
57	FIG01069681: hypothetical protein	16	7	1	9	2	60	0.00218761586948461	0.862621678578996	hypothetical protein	WP_001597900	5.50046E-110
58	Regulatory protein cro	16	7	1	9	13	49	0.00218761586948461	0.862621678578996	regulatory protein cro	0611193A	1.68316E-35
59	FIG00639588: hypothetical protein	16	7	1	9	24	38	0.00218761586948461	0.862621678578996	hypothetical protein	WP_020237177	1.18324E-44
60	Eae protein	15	8	1	9	30	32	0.00441230997404524	0.862621678578996	phage protein	YP_001272549	3.49126E-79
61	hypothetical protein	14	9	1	9	12	50	0.00835187245087134	0.862621678578996	phage transcriptional regulator	YP_003221152	1.41927E-71
62	Rac prophage repressor	14	9	1	9	12	50	0.00835187245087134	0.862621678578996	gracr	NP_287262	3.78391E-103
63	FIG00639888: hypothetical protein	14	9	1	9	15	47	0.00835187245087134	0.862621678578996	ead ea22-like family protein	YP_001449244	9.00206E-128
64	Antirestriction protein klcA	14	9	0	10	17	45	0.000998022494129281	0.862621678578996	antirestriction protein klca	WP_000680587	8.09844E-51
65	FIG00640415: hypothetical protein	14	9	1	9	18	44	0.00835187245087134	0.862621678578996	dctp deaminase	NP_415974	4.31551E-87
66	Phage antitermination protein N	14	9	1	9	19	43	0.00835187245087134	0.862621678578996	antitermination protein n	YP_001449256	1.73959E-63
67	FIG00638146: hypothetical protein	13	10	0	10	6	56	0.00199604498825856	0.862621678578996	phage protein	WP_001234560	3.30394E-93
68	core protein	13	10	0	10	8	54	0.00199604498825856	0.862621678578996	type iv secretion protein rhs	WP_001513052	0.0
69	core protein	11	12	0	10	5	57	0.00698615745890496	0.862621678578996	protein rhsc	WP_016238874	0.0

List of overrepresented genes in LEE positive/stx2 positive STEC not associated with HUS

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR	Function Blast2GO	Hit Accession	E-value
1	Phage DNA invertase	3	20	10	0	34	28	3.08985292300086e-06	0.0318718329007539	dna-invertase	NP_313019	3.0561E-107
2	FIG00639588: hypothetical protein	6	17	8	2	29	33	0.00596959584723768	0.672639891490763	hypothetical protein	WP_020237177	1.44206E-28
3	Phage tail fiber protein	6	17	8	2	36	26	0.00596959584723768	0.672639891490763	host specificity protein j	WP_001657146	0.0
4	Failed to assign function	2	21	7	3	20	42	0.000814295085707763	0.672639891490763	anti-adapter protein iram	YP_003234843	5.39051E-67
5	Mu-like prophage FluMu protein gp42	3	20	7	3	21	41	0.00242149396765637	0.672639891490763	tape measure domain protein	WP_000113523	0.0
6	Mu-like prophage protein GP36	3	20	7	3	22	40	0.00242149396765637	0.672639891490763	protein gp36	NP_313003	9.93456E-92
7	FIG00639310: hypothetical protein	3	20	7	3	22	40	0.00242149396765637	0.672639891490763	hypothetical protein	WP_001002048	7.90551E-48

8	Putative virion morphogenesis protein	3	20	7	3	22	40	0.00242149396765637	0.672639891490763	phage virion morphogenesis protein	WP_000094814	1.83102E-84
9	Protein Mom	3	20	7	3	23	39	0.00242149396765637	0.672639891490763	prophage dna modification protein	WP_000528250	8.05035E-177
10	Phage tail/DNA circulation protein	3	20	7	3	23	39	0.00242149396765637	0.672639891490763	dna circulation family protein	WP_000146122	0.0
11	Phage tail tube protein	3	20	7	3	23	39	0.00242149396765637	0.672639891490763	phage tail tube protein	WP_000015474	9.52129E-71
12	Mu-like prophage FluMu I protein	3	20	7	3	23	39	0.00242149396765637	0.672639891490763	prophage protease protein	WP_000850812	0.0
13	FIG00639839: hypothetical protein	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	unk domain protein	NP_313022	3.87991E-31
14	Putative tail fiber assembly protein p37	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	tail fiber assembly protein	WP_001018933	1.46344E-135
15	FIG121501: Prophage tail protein	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	protein gp48	WP_000301579	2.64408E-136
16	Phage FluMu protein gp47	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	baseplate j-like family protein	NP_313014	0.0
17	Bacteriophage protein GP46	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	phage gp46 family protein	WP_000763328	8.55232E-95
18	FIG003269: Prophage tail protein	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	baseplate protein	NP_313011	0.0
19	Mu-like prophage FluMu protein gp41	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	mu-like prophage u protein gp41	NP_313008	1.61839E-84
20	Bacteriophage tail sheath protein	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	phage tail sheath family protein	WP_000606746	0.0
21	Mu-like prophage FluMu protein GP38	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	mu-like prophage u protein gp38	NP_313005	8.11492E-28
22	Mu-like prophage FluMu protein gp37	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	protein gp37	WP_000627426	3.7733E-133
23	Phage major capsid protein	3	20	7	3	24	38	0.00242149396765637	0.672639891490763	mu-like prophage major head subunit gpt family protein	WP_001142979	0.0
24	FIG00639396: hypothetical protein	2	21	6	4	7	55	0.00402869284960497	0.672639891490763	phage protein	YP_005456215	1.93932E-76
25	IncI1 plasmid conjugative transfer NusG-type transcription antiterminator TraB	2	21	6	4	30	32	0.00402869284960497	0.672639891490763	transcription antitermination factor	YP_002756602	8.64568E-152
26	IncI1 plasmid conjugative transfer protein TraC	2	21	6	4	32	30	0.00402869284960497	0.672639891490763	transfer protein c	WP_001370630	1.29454E-146
27	Failed to assign function	1	22	5	5	4	58	0.00542269187986652	0.672639891490763	tail length tape measure protein	WP_001365475	8.55738E-163
28	Prophage tail fiber protein	0	23	4	6	1	61	0.00513196480938416	0.672639891490763	tail fiber protein	WP_001473654	1.95262E-26
29	Uncharacterized fimbrial chaperone YehC precursor	0	23	4	6	5	57	0.00513196480938416	0.672639891490763	gram-negative pili assembly c-terminal domain protein	WP_001400456	3.6664E-133
30	Fimbrial protein Yad-like	0	23	4	6	5	57	0.00513196480938416	0.672639891490763	fimbrial protein	WP_001189052	0.0
31	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	0	23	4	6	6	56	0.00513196480938416	0.672639891490763	invasin	WP_000168879	0.0
32	FIG00637952: hypothetical protein	0	23	4	6	7	55	0.00513196480938416	0.672639891490763	yjik protein	WP_024221583	0.0
33	FIG00638183: hypothetical protein	0	23	4	6	7	55	0.00513196480938416	0.672639891490763	pentapeptide repeats family protein	WP_000535133	0.0
34	Putative fimbrial protein	0	23	4	6	7	55	0.00513196480938416	0.672639891490763	fimbrial family protein	NP_285715	2.40975E-108
35	FIG032766: hypothetical protein	0	23	4	6	8	54	0.00513196480938416	0.672639891490763	toxin-antitoxin antitoxin ribbon-helix-helix domain protein	WP_001277109	1.32683E-40
36	FIG00638467: hypothetical protein	0	23	4	6	9	53	0.00513196480938416	0.672639891490763	udp pyrophosphate phosphatase	WP_000868892	8.27183E-174
37	FIG001353: Acetyltransferase	0	23	4	6	9	53	0.00513196480938416	0.672639891490763	acetyltransferase family protein	WP_000342447	3.88018E-131
38	Pertactin precursor	0	23	4	6	11	51	0.00513196480938416	0.672639891490763	adhesin-like autotransporter	AHG10156	0.0
39	FIG00640449: hypothetical protein	0	23	4	6	12	50	0.00513196480938416	0.672639891490763	4-phosphopantetheinyl transferase	AHG11591	3.13024E-28
40	CcdA protein (antitoxin to CcdB)	0	23	4	6	13	49	0.00513196480938416	0.672639891490763	antitoxin of gyrase inhibiting toxin-antitoxin system	WP_000125563	6.67592E-33
41	Putative glycoporin	0	23	4	6	14	48	0.00513196480938416	0.672639891490763	glycoporin	WP_000749945	0.0
42	Arylsulfatase (EC 3.1.6.1)	0	23	4	6	17	45	0.00513196480938416	0.672639891490763	arylsulfatase	WP_000395865	0.0
43	FIG00638395: hypothetical protein	0	23	4	6	17	45	0.00513196480938416	0.672639891490763	hypothetical protein ECDH10B_1868	YP_001730707	1.93725E-175
44	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)	0	23	4	6	21	41	0.00513196480938416	0.672639891490763	abc transporter periplasmic-binding protein	WP_001400746	0.0

Table S4. Analysis of accessory genome in STEC in HUS-group 1 (n=18, G1) compared to other LEE positive positive STEC (n=36, G2) and LEE negative STEC (n=41, G3)
List of overrepresented genes in STEC in HUS-group 1

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR<0.01	Function Blast2GO	Hit Accession	E-value
1	FIG01069681: hypothetical protein	18	0	0	36	1	40	1.0317112061526e-14	1.49061635064927e-10	hypothetical protein	WP_001597900	5.50046E-110
2	Failed to assign function	18	0	1	35	7	34	1.96025129168995e-13	3.54021383279204e-10	hypothetical protein	WP_001542619	7.10515E-12
3	FIG01070140: hypothetical protein	18	0	15	21	8	33	1.07004786129841e-05	0.000822343164895712	prophage protein	YP_001743216	1.55921E-42
4	FIG00639506: hypothetical protein	18	0	7	29	12	29	4.95943576797556e-09	4.47837049848193e-06	mrna interferase	NP_289657	2.98362E-62
5	Phage repressor	18	0	14	22	12	29	4.86385391499277e-06	0.000512941323823471	repressor protein ci	WP_001274762	2.46962E-151
6	FIG00638890: hypothetical protein	18	0	6	30	14	27	1.38864201503315e-09	2.0063099833199e-06	flxa-like family protein	YP_311362	2.67098E-12
7	FIG00638388: hypothetical protein	18	0	8	28	14	27	1.61181662459205e-08	4.95479289193744e-06	inner membrane protein yedr	YP_310920	2.54409E-59
8	Hydrogenase-4 component B (EC 1.-.-.-) / Formate hydrogenlyase subunit 3	18	0	1	35	15	26	1.96025129168995e-13	3.54021383279204e-10	nadh-ubiquinone plastoquinone (complex i) various chains family protein	WP_001443474	9.64297E-89
9	Gluconokinase (EC 2.7.1.12)	18	0	1	35	16	25	1.96025129168995e-13	3.54021383279204e-10	thermosensitive gluconokinase	WP_000896732	5.4444E-136
10	Hydrogenase-4 component B (EC 1.-.-.-) / Formate hydrogenlyase subunit 3	18	0	1	35	16	25	1.96025129168995e-13	3.54021383279204e-10	hydrogenase-4 component b	WP_001443473	0.0
11	Phosphate starvation-inducible protein PhoH, predicted ATPase	18	0	7	29	16	25	4.95943576797556e-09	4.47837049848193e-06	protein phoh	NP_309293	0.0
12	Phage EaA protein	18	0	8	28	16	25	1.61181662459205e-08	4.95479289193744e-06	pf04448 family partial	YP_007001969	7.58484E-111
13	L-idonate 5-dehydrogenase (EC 1.1.1.264)	18	0	1	35	17	24	1.96025129168995e-13	3.54021383279204e-10	l-idonate 5-dehydrogenase	WP_001197417	0.0
14	5-keto-D-gluconate 5-reductase (EC 1.1.1.69)	18	0	1	35	17	24	1.96025129168995e-13	3.54021383279204e-10	gluconate 5-dehydrogenase	NP_418687	0.0
15	Positive regulator of L-idonate catabolism	18	0	1	35	17	24	1.96025129168995e-13	3.54021383279204e-10	hth-type transcriptional regulator idnr	NP_418685	0.0
16	Putative cytoplasmic protein	18	0	7	29	19	22	4.95943576797556e-09	4.47837049848193e-06	protein yoag	YP_006087527	1.25649E-31
17	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	18	0	8	28	19	22	1.61181662459205e-08	4.95479289193744e-06	pyridine nucleotide-disulfide oxidoreductase family protein	WP_000080323	0.0
18	HicA-like protein	18	0	11	25	19	22	3.56944117955114e-07	6.13943882882796e-05	#NAME?	YP_003080264	3.93181E-58
19	Failed to assign function	18	0	9	27	20	21	4.83544987377616e-08	1.39725159552636e-05	helix-turn-helix family protein	YP_405986	4.84912E-19
20	Phage DNA-packaging protein	18	0	19	17	20	21	0.000182330523734399	0.00750516070346038	dna packaging protein	YP_003220773	3.23174E-77
21	Failed to assign function	18	0	7	29	22	19	4.95943576797556e-09	4.47837049848193e-06	coenzyme a transferase family protein	WP_001226986	0.0
22	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	18	0	8	28	22	19	1.61181662459205e-08	4.95479289193744e-06	short chain dehydrogenase family protein	YP_002394282	0.0
23	4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)	18	0	7	29	23	18	4.95943576797556e-09	4.47837049848193e-06	xylose isomerase-like tim barrel family protein	YP_410494	0.0
24	Putative aldolase YdjI	18	0	8	28	23	18	1.61181662459205e-08	4.95479289193744e-06	aldolase	NP_416287	0.0
25	FIG00640016: hypothetical protein	18	0	8	28	23	18	1.61181662459205e-08	4.95479289193744e-06	bacterial regulatory tetr family protein	YP_006118096	1.56901E-153
26	3-hydroxybutyryl-CoA dehydratase (EC 4.2.1.55)	18	0	8	28	23	18	1.61181662459205e-08	4.95479289193744e-06	enoyl- hydratase isomerase family protein	WP_001283541	0.0
27	FIG00638352: hypothetical protein	18	0	17	19	23	18	4.68145939318053e-05	0.0026216172601811	conserved protein	WP_000123966	5.03763E-42
28	Uncharacterized protein similar to VCA0109	18	0	17	19	23	18	4.68145939318053e-05	0.0026216172601811	lysozyme family protein	YP_001461397	1.19452E-75
29	Alpha-fimbriae tip adhesin	18	0	9	27	24	17	4.83544987377616e-08	1.39725159552636e-05	cbld like pilus biogenesis initiator family protein	YP_001464557	0.0
30	Hypothetical oxidoreductase YdjG (EC 1.-.-.-)	18	0	8	28	25	16	1.61181662459205e-08	4.95479289193744e-06	oxidoreductase	NP_416285	0.0
31	FIG00638425: hypothetical protein	18	0	15	21	25	16	1.07004786129841e-05	0.000822343164895712	membrane protein	WP_000837550	2.50704E-98
32	Mobile element protein	18	0	18	18	25	16	9.36291878636101e-05	0.00406232584460492	transposase	WP_024185760	3.58134E-72
33	Mobile element protein	18	0	18	18	25	16	9.36291878636101e-05	0.00406232584460492	transposase	WP_024185760	3.58134E-72
34	Hypothetical zinc-type alcohol dehydrogenase-like protein YdjJ	18	0	8	28	26	15	1.61181662459205e-08	4.95479289193744e-06	sorbitol dehydrogenase	WP_000798131	0.0
35	Galactoside O-acetyltransferase (EC 2.3.1.18)	18	0	12	24	26	15	8.92360294887786e-07	0.000142730001693553	galactoside o-acetyltransferase	WP_001250720	2.99966E-144

36	Putative transport protein YdjK, MFS superfamily	18	0	7	29	27	14	4.95943576797556e-09	4.47837049848193e-06	inner membrane metabolite transporter ydje	YP_002556655	0.0
37	Putative HTH-type transcriptional regulator YdjF	18	0	8	28	27	14	1.61181662459205e-08	4.95479289193744e-06	transcriptional regulator	WP_000719082	7.65921E-165
38	Putative oxidoreductase YdjL	18	0	8	28	27	14	1.61181662459205e-08	4.95479289193744e-06	zn-dependent and nad -binding	WP_000645214	0.0
39	Mobile element protein	18	0	15	21	27	14	1.07004786129841e-05	0.000822343164895712	integrase core domain protein	WP_000001467	0.0
40	FIG00640692: hypothetical protein	18	0	11	25	28	13	3.56944117955114e-07	6.13943882882796e-05	hicb family protein	NP_290055	7.63796E-78
41	Uncharacterized protein ImpC	18	0	16	20	28	13	2.2738517052591e-05	0.00145365528484883	family type vi secretion protein	YP_001461398	0.0
42	FIG00639456: hypothetical protein	18	0	11	25	29	12	3.56944117955114e-07	6.13943882882796e-05	fimbrial family protein	WP_021543399	0.0
43	YafQ toxin protein	18	0	11	25	29	12	3.56944117955114e-07	6.13943882882796e-05	mrna interferase	YP_309286	4.65069E-59
44	Inner membrane protein YjgN	18	0	15	21	29	12	1.07004786129841e-05	0.000822343164895712	inner membrane protein yjgn	YP_002295838	0.0
45	Protein ImpG/VasA	18	0	17	19	29	12	4.68145939318053e-05	0.0026216172601811	type vi secretion family	NP_285921	0.0
46	YcgN (Fragment)	18	0	12	24	30	11	8.92360294887786e-07	0.000142730001693553	upf0260 protein ycgn	WP_001335675	3.68617E-109
47	FIG00637952: hypothetical protein	18	0	11	25	31	10	3.56944117955114e-07	6.13943882882796e-05	#NAME?	YP_410627	0.0
48	Uncharacterized protein ImpA	18	0	11	25	31	10	3.56944117955114e-07	6.13943882882796e-05	impa-related family protein	WP_001240513	0.0
49	D-galactonate transporter	18	0	12	24	33	8	8.92360294887786e-07	0.000142730001693553	l-galactonate transporter	YP_005459680	0.0
50	Autoinducer 2 (AI-2) modifying protein LsrG	18	0	16	20	34	7	2.2738517052591e-05	0.00145365528484883	autoinducer 2-degrading protein lsrg	3QMQ_A	8.89099E-62
51	Putative sucrose phosphorylase (EC 2.4.1.7)	18	0	12	24	35	6	8.92360294887786e-07	0.000142730001693553	glucosyltransferase	WP_000810512	0.0
52	Glutamate Aspartate transport system permease protein GltJ (TC 3.A.1.3.4)	18	0	15	21	37	4	1.07004786129841e-05	0.000822343164895712	amino abc permease 3-tm his glu gln arg opine family domain protein	YP_026209	0.0
53	FIG00639031: hypothetical protein	18	0	17	19	37	4	4.68145939318053e-05	0.0026216172601811	diguanylate cyclase domain protein	WP_000043731	0.0
54	Putative flagellin structural protein	18	0	11	25	38	3	3.56944117955114e-07	6.13943882882796e-05	outer membrane autotransporter barrel domain-containing protein	WP_000556465	0.0
55	Failed to assign function	18	0	13	23	40	1	2.12793608780932e-06	0.000272074518554594	acetyl- :acetoacetyl- transferase alpha beta subunit	WP_000805688	0.0
56	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)	18	0	13	23	40	1	2.12793608780932e-06	0.000272074518554594	galactitol-1-phosphate 5-dehydrogenase	YP_002387569	0.0
57	Chaperone protein hchA	18	0	16	20	40	1	2.2738517052591e-05	0.00145365528484883	chaperone hcha	WP_000218204	0.0
58	L-rhamnose operon regulatory protein RhaS	18	0	16	20	40	1	2.2738517052591e-05	0.00145365528484883	hth-type transcriptional activator rhas	NP_418341	0.0
59	Hypothetical lipoprotein yehR	18	0	16	20	40	1	2.2738517052591e-05	0.00145365528484883	lipoprotein	NP_754541	9.35062E-71
60	Failed to assign function	18	0	17	19	40	1	4.68145939318053e-05	0.0026216172601811	transcriptional regulator amino transferase	YP_002386877	0.0
61	Trans-aconitate 2-methyltransferase (EC 2.1.1.144)	18	0	17	19	40	1	4.68145939318053e-05	0.0026216172601811	trans-aconitate 2-methyltransferase	WP_001286605	1.4579E-178
62	metal-dependent phosphohydrolase	18	0	16	20	41	0	2.2738517052591e-05	0.00145365528484883	hd domain protein	WP_000365559	1.37417E-138
63	FIG00637874: hypothetical protein	18	0	17	19	41	0	4.68145939318053e-05	0.0026216172601811	diguanylate cyclase	WP_000592815	0.0
64	Soluble cytochrome b562	18	0	17	19	41	0	4.68145939318053e-05	0.0026216172601811	soluble cytochrome b562	3U8P_A	2.17662E-56
65	Alkanesulfonate utilization operon LysR-family regulator CbI	18	0	17	19	41	0	4.68145939318053e-05	0.0026216172601811	hth-type transcriptional regulator cbl	WP_001011011	0.0
66	Putative isomerase	18	0	17	19	41	0	4.68145939318053e-05	0.0026216172601811	glycosyl hydrolase	YP_002388562	0.0
67	SanA protein	18	0	17	19	41	0	4.68145939318053e-05	0.0026216172601811	thioredoxin-like protein	WP_024199571	2.31116E-170
68	Failed to assign function	18	0	17	19	41	0	4.68145939318053e-05	0.0026216172601811	purine permease ybby	WP_001759829	0.0
69	Inner membrane protein YqiK	18	0	18	18	41	0	9.36291878636101e-05	0.00406232584460492	inner membrane protein yqik	WP_001564111	0.0
70	Regulatory protein cro	17	1	6	30	7	34	3.26330873532791e-08	9.82255929333701e-06	regulatory protein cro	0611193A	1.68316E-35
71	Shikimate kinase I (EC 2.7.1.71)	17	1	14	22	8	33	6.5054046113028e-05	0.00356023052364026	shikimate kinase 1	NP_709163	1.34976E-146
72	Failed to assign function	17	1	12	24	12	29	1.37423485412718e-05	0.00103952592525809	integrase	WP_001218282	0.0
73	hypothetical protein	17	1	13	23	20	21	3.05461793250049e-05	0.00194419030346992	hypothetical protein	YP_005277482	7.30809E-18
74	FIG00638450: hypothetical protein	17	1	15	21	21	20	0.000133269597270802	0.00564656639697521	terminase	YP_007112075	8.40806E-96

75	General secretion pathway protein F	17	1	8	28	25	16	3.2856261962838e-07	6.13943882882796e-05	type ii secretion system protein f	WP_001173467	0.0
76	General secretion pathway protein C	17	1	11	25	26	15	5.89573215518963e-06	0.000612816821425754	type ii secretion system protein c	WP_001631803	9.99666E-177
77	Cytochrome b561 homolog 1	17	1	15	21	26	15	0.000133269597270802	0.00564656639697521	cytochrome b561	YP_407515	9.22549E-114
78	Flagellar hook-associated protein FliD	17	1	7	29	27	14	1.08512454603305e-07	2.61297990684758e-05	flagellar hook-associated protein 2	WP_000146822	0.0
79	Uncharacterized sugar kinase YdjH	17	1	8	28	27	14	3.2856261962838e-07	6.13943882882796e-05	kinase	WP_001602686	0.0
80	Putative transport protein YdjK, MFS superfamily	17	1	8	28	27	14	3.2856261962838e-07	6.13943882882796e-05	metabolite transporter	WP_023352181	0.0
81	Accessory colonization factor AcfD precursor	17	1	10	26	27	14	2.39838313739299e-06	0.000288765329742115	lipoprotein precursor	WP_001034473	0.0
82	General secretion pathway protein I	17	1	10	26	27	14	2.39838313739299e-06	0.000288765329742115	type ii secretion system protein i	YP_409353	7.66787E-63
83	Leader peptidase (Prepilin peptidase) (EC 3.4.23.43) / N-methyltransferase (EC 2.1.1.-)	17	1	9	27	28	13	9.18735476017469e-07	0.000142730001693553	leader peptidase pppa	WP_000895871	2.44755E-159
84	General secretion pathway protein J	17	1	9	27	28	13	9.18735476017469e-07	0.000142730001693553	type ii secretion system protein j	WP_001443429	3.24329E-134
85	General secretion pathway protein D	17	1	10	26	28	13	2.39838313739299e-06	0.000288765329742115	type ii secretion system protein d	YP_001464424	0.0
86	General secretion pathway protein H	17	1	10	26	28	13	2.39838313739299e-06	0.000288765329742115	general secretion pathway protein h	YP_002388445	4.04855E-124
87	General secretion pathway protein K	17	1	10	26	28	13	2.39838313739299e-06	0.000288765329742115	general secretion pathway protein k	WP_000633212	0.0
88	Hypothetical lipoprotein yghG precursor	17	1	11	25	28	13	5.89573215518963e-06	0.000612816821425754	lipoprotein	NP_417445	8.56611E-65
89	General secretion pathway protein L	17	1	10	26	29	12	2.39838313739299e-06	0.000288765329742115	type ii secretion system protein l	WP_000094990	0.0
90	Molybdate metabolism regulator	17	1	9	27	30	11	9.18735476017469e-07	0.000142730001693553	molybdate metabolism regulator	WP_001363074	0.0
91	General secretion pathway protein M	17	1	10	26	31	10	2.39838313739299e-06	0.000288765329742115	type ii secretion system m family protein	WP_000942769	1.96035E-115
92	Autoinducer 2 (AI-2) kinase LsrK (EC 2.7.1.-)	17	1	15	21	33	8	0.000133269597270802	0.00564656639697521	autoinducer 2 kinase	WP_000113166	0.0
93	General secretion pathway protein E	17	1	9	27	38	3	9.18735476017469e-07	0.000142730001693553	type ii secretion system protein e	YP_002294519	0.0
94	Mobile element protein	16	2	7	29	0	41	1.20876327972868e-06	0.000183833809110736	transposase	WP_001369069	0.0
95	hypothetical protein	16	2	9	27	0	41	8.88606903727022e-06	0.000721819266870833	beta family protein	WP_000156213	0.0
96	Failed to assign function	16	2	10	26	0	41	2.16338027352746e-05	0.00142075082690567	regulatory protein	YP_003233908	1.28028E-115
97	putative excisionase	16	2	13	23	18	23	0.000223790233337935	0.00913367596402961	excisionase	NP_311039	3.20391E-51
98	FIG00641505: hypothetical protein	16	2	13	23	28	13	0.000223790233337935	0.00913367596402961	reverse gyrase	YP_003223435	1.21837E-20
99	Failed to assign function	15	3	2	34	0	41	1.10194907335627e-08	4.95479289193744e-06	hypothetical protein	WP_000460278	4.81775E-82
100	FIG00641015: hypothetical protein	15	3	4	32	0	41	2.67739158449308e-07	6.13943882882796e-05	pf11726 domain protein	WP_000287900	9.57999E-92
101	hypothetical protein	15	3	5	31	0	41	9.85623954693032e-07	0.000151492498908563	hypothetical protein	WP_000222754	7.50187E-52
102	Lipid A biosynthesis (KDO) 2-(lauroyl)-lipid IVA acyltransferase (EC 2.3.1.-)	15	3	7	29	0	41	9.11781268932563e-06	0.000721819266870833	lipid a biosynthesis 2- -lipid iva acyltransferase	WP_000790278	0.0
103	Mobile element protein	15	3	7	29	1	40	9.11781268932563e-06	0.000721819266870833	transposase-like protein	WP_000343735	2.52131E-78
104	Phage restriction alleviation ral	15	3	2	34	5	36	1.10194907335627e-08	4.95479289193744e-06	restriction inhibitor protein ral	WP_000213977	5.76E-36
105	Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase) (GLYCOGENASE)	15	3	4	32	5	36	2.67739158449308e-07	6.13943882882796e-05	alpha catalytic domain protein	WP_001493308	0.0
106	D-serine dehydratase transcriptional activator	15	3	6	30	9	32	3.16743068837495e-06	0.000363252960254338	d-serine deaminase transcriptional activator	YP_003000013	0.0
107	Sugar-1-epimerase YihR	15	3	2	34	10	31	1.10194907335627e-08	4.95479289193744e-06	aldose 1-epimerase family protein	WP_001397472	0.0
108	TRAP-type transport system, small permease component, predicted N-acetylneuraminate transporter	15	3	6	30	13	28	3.16743068837495e-06	0.000363252960254338	tripartite atp-independent periplasmic family	WP_001375330	2.55737E-76
109	TRAP transporter solute receptor, unknown substrate 6	15	3	6	30	13	28	3.16743068837495e-06	0.000363252960254338	trap transporter solute family protein	WP_000695891	0.0
110	Alcohol dehydrogenase (EC 1.1.1.1)	15	3	2	34	14	27	1.10194907335627e-08	4.95479289193744e-06	chlorophyll synthesis pathway protein	WP_000357974	0.0
111	Shikimate transporter	15	3	2	34	14	27	1.10194907335627e-08	4.95479289193744e-06	shikimate transporter	YP_312804	0.0
112	Flagellar biosynthesis protein FlhA	15	3	10	26	32	9	0.000132114161451312	0.00564656639697521	fhipep family protein	YP_002291526	0.0
113	Fimbrial protein YadC	15	3	9	27	34	7	5.81773060028869e-05	0.00319599131988483	fimbrial-like adhesin protein	WP_001443439	0.0
114	Protein YjgL, putative CCAAT-box DNA binding protein subunit B	14	4	0	36	0	41	9.4288087130286e-10	1.51363809206486e-06	dna-binding protein	YP_001882963	8.84755E-60
115	Failed to assign function	14	4	1	35	0	41	1.28231798497189e-08	4.95479289193744e-06	hypothetical protein	WP_001084135	0.0

116	hypothetical protein	14	4	1	35	0	41	1.28231798497189e-08	4.95479289193744e-06	surface lpxtg-motif cell wall anchor domain protein	WP_001223333	9.14223E-93
117	FIG00641106: hypothetical protein	14	4	4	32	0	41	1.92076336032631e-06	0.000250010711981932	hypothetical protein E2348C_1441	YP_002328980	1.12626E-124
118	Probable transmembrane protein	14	4	2	34	2	39	9.27867306659812e-08	2.3505003727963e-05	hcp1 family type vi secretion system effector	WP_001612076	2.90788E-107
119	COG0582: Integrase	14	4	2	34	2	39	9.27867306659812e-08	2.3505003727963e-05	integrase	YP_003223659	0.0
120	Phage repressor	14	4	7	29	4	37	5.2258082196218e-05	0.00290394142911907	regulatory protein cii	0611193A	4.63359E-52
121	Phage antitermination protein N	14	4	5	31	5	36	6.54923112558189e-06	0.000675880652160051	antitermination protein n	WP_000512959	1.00049E-50
122	FIG00639473: hypothetical protein	14	4	6	30	6	35	1.95309811909893e-05	0.0012885096632302	hypothetical protein	WP_000581110	0.0
123	N-6 DNA methylase	14	4	1	35	9	32	1.28231798497189e-08	4.95479289193744e-06	n-6 dna methylase family protein	YP_005279645	0.0
124	Putative DNA binding protein	14	4	2	34	9	32	9.27867306659812e-08	2.3505003727963e-05	transcriptional regulator	WP_000270113	1.33339E-37
125	COG0042: tRNA-dihydrouridine synthase	14	4	2	34	9	32	9.27867306659812e-08	2.3505003727963e-05	trna-dihydrouridine synthase	YP_003223657	0.0
126	Failed to assign function	14	4	2	34	9	32	9.27867306659812e-08	2.3505003727963e-05	hypothetical protein ECO103_3813	YP_003223658	5.90343E-139
127	D-serine permease DsdX	14	4	4	32	9	32	1.92076336032631e-06	0.000250010711981932	permease	WP_001592429	0.0
128	hypothetical protein	14	4	1	35	10	31	1.28231798497189e-08	4.95479289193744e-06	membrane protein	WP_000449994	0.0
129	FIG00967058: hypothetical protein	14	4	1	35	10	31	1.28231798497189e-08	4.95479289193744e-06	hypothetical protein	WP_000007269	0.0
130	hypothetical protein	14	4	1	35	10	31	1.28231798497189e-08	4.95479289193744e-06	hypothetical protein	YP_005279649	0.0
131	hypothetical protein	14	4	1	35	10	31	1.28231798497189e-08	4.95479289193744e-06	hypothetical protein	YP_005279650	0.0
132	Bipolar DNA helicase	14	4	1	35	10	31	1.28231798497189e-08	4.95479289193744e-06	aaa-like domain protein	YP_005279651	0.0
133	hypothetical protein	14	4	1	35	10	31	1.28231798497189e-08	4.95479289193744e-06	hypothetical protein	YP_005279652	2.7731E-116
134	serine/threonine protein kinase with TPR repeats	14	4	1	35	10	31	1.28231798497189e-08	4.95479289193744e-06	kinase domain protein	WP_001022617	0.0
135	FIG00638448: hypothetical protein	14	4	1	35	11	30	1.28231798497189e-08	4.95479289193744e-06	aaa domain family protein	YP_001461068	0.0
136	hypothetical protein	14	4	1	35	11	30	1.28231798497189e-08	4.95479289193744e-06	hypothetical protein EFER_3103	YP_002384202	2.98479E-67
137	Ankyrin-repeat protein A	14	4	3	33	11	30	4.74367183683931e-07	8.06312596454757e-05	ankyrin repeat protein	WP_000632921	0.0
138	Antirestriction protein klcA	14	4	6	30	11	30	1.95309811909893e-05	0.0012885096632302	antirestriction protein klca	WP_000680587	8.09844E-51
139	FIG00641630: hypothetical protein	14	4	1	35	12	29	1.28231798497189e-08	4.95479289193744e-06	mcrbc 5-methylcytosine restriction system component family protein	YP_005279643	0.0
140	probable integral membrane protein Cj0014c	14	4	1	35	13	28	1.28231798497189e-08	4.95479289193744e-06	membrane protein	YP_313098	2.09259E-126
141	Putative transport protein	14	4	2	34	14	27	9.27867306659812e-08	2.3505003727963e-05	major facilitator superfamily transporter	NP_756684	0.0
142	Minor fimbrial subunit StfF	14	4	7	29	20	21	5.2258082196218e-05	0.00290394142911907	fimbrial family protein	YP_002293919	7.32378E-104
143	Flagellar motor rotation protein MotB	14	4	8	28	29	12	0.00012775355383328	0.00547710191627071	ompa family protein	WP_021553260	2.36463E-139
144	Mobile element protein	13	5	4	32	0	41	1.11776988908375e-05	0.000849757555656945	is protein	YP_003501107	0.0
145	Failed to assign function	13	5	5	31	1	40	3.52457312701664e-05	0.002223713211316	enterobacteria phage	2M7A_A	7.71285E-56
146	Mobile element protein	13	5	4	32	2	39	1.11776988908375e-05	0.000849757555656945	integrase core domain protein	WP_001369358	2.46497E-149
147	FIG00638146: hypothetical protein	13	5	0	36	6	35	7.73162314468348e-09	4.95479289193744e-06	phage protein	WP_001234560	3.30394E-93
148	Flagellar biosynthesis protein FliC	13	5	1	35	6	35	9.59852726986307e-08	2.3505003727963e-05	flagellin	WP_000079774	0.0
149	Mobile element protein	13	5	1	35	6	35	9.59852726986307e-08	2.3505003727963e-05	integrase core domain protein	BAO24607	2.88536E-79
150	FIG00638659: hypothetical protein	13	5	6	30	6	35	9.74214815834331e-05	0.00418912370808762	phage protein	WP_000189903	1.17005E-119
151	FIG00643641: hypothetical protein	13	5	1	35	8	33	9.59852726986307e-08	2.3505003727963e-05	hypothetical protein	WP_001012978	0.0
152	Failed to assign function	13	5	6	30	8	33	9.74214815834331e-05	0.00418912370808762	hypothetical protein ECED1_1027	YP_002397051	8.10207E-46
153	Per-activated serine protease autotransporter enterotoxin EspC	13	5	7	29	11	30	0.000242075268026544	0.0096349957918664	autotransporter	YP_003223650	0.0
154	Uncharacterized protein YadU in stf fimbrial cluster	13	5	7	29	19	22	0.000242075268026544	0.0096349957918664	fimbrial protein	WP_001397279	2.10091E-156
155	orf, conserved hypothetical protein	13	5	6	30	24	17	9.74214815834331e-05	0.00418912370808762	pf11682 family protein	NP_708754	3.87156E-99
156	Haemolysin expression modulating protein paralog	12	6	4	32	0	41	5.48305027160912e-05	0.00302363016504613	ydfa protein	YP_003223436	1.74231E-30
157	D-mannonate oxidoreductase (EC 1.1.1.57)	12	6	5	31	1	40	0.000159597231896699	0.00672262625785279	d-mannonate oxidoreductase	WP_001397145	0.0
158	Integrase	12	6	4	32	2	39	5.48305027160912e-05	0.00302363016504613	phage integrase family protein	WP_000023401	0.0

159	putative phage inhibition, colicin resistance and tellurite resistance protein	12	6	2	34	3	38	3.69967596281816e-06	0.000393036164049976	tellurium resistance protein	WP_001053348	0.0
160	hypothetical protein	12	6	3	33	4	37	1.59522243092247e-05	0.00112823556106331	chat domain protein	WP_000421881	0.0
161	core protein	12	6	5	31	4	37	0.000159597231896699	0.00672262625785279	type iv secretion protein rhs	WP_001513052	0.0
162	Ser/Thr and Tyr protein phosphatase (dual specificity)	11	7	0	36	0	41	3.32459795221389e-07	6.13943882882796e-05	inner membrane protein	WP_001368021	8.3907E-175
163	Ser/Thr and Tyr protein phosphatase (dual specificity)	11	7	0	36	0	41	3.32459795221389e-07	6.13943882882796e-05	dual specificity phosphatase	WP_001460866	9.26917E-98
164	D-mannonate oxidoreductase (EC 1.1.1.57)	11	7	0	36	0	41	3.32459795221389e-07	6.13943882882796e-05	mannitol dehydrogenase rossmann domain protein	WP_001370597	3.75696E-147
165	FIG01049009: hypothetical protein	11	7	0	36	0	41	3.32459795221389e-07	6.13943882882796e-05	condensation protein	WP_001443343	5.45349E-94
166	orf, hypothetical protein	11	7	0	36	0	41	3.32459795221389e-07	6.13943882882796e-05	conserved protein	WP_001443454	6.04715E-101
167	Failed to assign function	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	hypothetical protein Z6012	NP_287952	5.7526E-41
168	Failed to assign function	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	protein partial	WP_000641699	4.17756E-55
169	Penicillin G acylase precursor (EC 3.5.1.11)	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	penicillin amidase family protein	WP_000797401	0.0
170	Failed to assign function	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	type vi secretion protein	WP_024247724	3.2015E-26
171	Pertactin precursor	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	outer membrane autotransporter barrel domain-containing protein	WP_021567521	0.0
172	Transcriptional regulator, GntR family	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	transcriptional regulator	WP_001205014	4.24325E-147
173	Nitrate/nitrite transporter	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	major facilitator family protein	WP_000098091	0.0
174	putative MR-MLE-family protein	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	mandelate racemase	WP_001236010	0.0
175	Failed to assign function	11	7	1	35	0	41	3.39418256051603e-06	0.000363252960254338	hypothetical protein Z3069	NP_288435	1.67751E-36
176	Failed to assign function	11	7	3	33	0	41	7.37142379422834e-05	0.00365627832507549	hypothetical protein	WP_001187432	1.39639E-111
177	FIG00639802: hypothetical protein	11	7	4	32	0	41	0.000232559775433195	0.00930754469656176	hypothetical protein	WP_001344959	2.52182E-40
178	FIG00640631: hypothetical protein	11	7	4	32	0	41	0.000232559775433195	0.00930754469656176	t3ss effector protein	YP_003229591	1.39111E-132
179	hypothetical protein	11	7	4	32	0	41	0.000232559775433195	0.00930754469656176	phage protein	WP_001370191	3.54719E-122
180	hypothetical protein	11	7	4	32	0	41	0.000232559775433195	0.00930754469656176	phage protein	WP_000133216	1.74153E-94
181	Failed to assign function	11	7	0	36	1	40	3.32459795221389e-07	6.13943882882796e-05	conserved protein	WP_001548533	1.1472E-32
182	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	11	7	0	36	1	40	3.32459795221389e-07	6.13943882882796e-05	short-chain-fatty-acid-- ligase	WP_016247266	1.62932E-104
183	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	11	7	0	36	1	40	3.32459795221389e-07	6.13943882882796e-05	short-chain-fatty-acid- ligase	WP_000899652	0.0
184	hypothetical protein	11	7	0	36	2	39	3.32459795221389e-07	6.13943882882796e-05	serpin b5	WP_001347273	4.05016E-95
185	Mobile element protein	11	7	4	32	2	39	0.000232559775433195	0.00930754469656176	is2 orf1	WP_001443451	9.74226E-59
186	FIG00638206: hypothetical protein	11	7	0	36	3	38	3.32459795221389e-07	6.13943882882796e-05	inner membrane protein	WP_001324978	0.0
187	Failed to assign function	11	7	1	35	3	38	3.39418256051603e-06	0.000363252960254338	hypothetical protein	WP_021571853	2.5372E-79
188	core protein	11	7	2	34	3	38	1.87027963869892e-05	0.00124524424976599	protein rhsc	WP_001746441	0.0
189	Putative methyltransferase (EC 2.1.1.72)	11	7	3	33	4	37	7.37142379422834e-05	0.00365627832507549	dna methylase family protein	WP_000215753	0.0
190	Failed to assign function	11	7	0	36	5	36	3.32459795221389e-07	6.13943882882796e-05	hypothetical protein	WP_001406069	4.70107E-16
191	Protein YjgL, putative CCAAT-box DNA binding protein subunit B	11	7	1	35	5	36	3.39418256051603e-06	0.000363252960254338	dna-binding protein	WP_001324720	0.0
192	Mobile element protein	11	7	3	33	8	33	7.37142379422834e-05	0.00365627832507549	transposase	WP_000604890	1.36236E-64
193	Failed to assign function	11	7	0	36	10	31	3.32459795221389e-07	6.13943882882796e-05	transposase	WP_021547077	3.05798E-142
194	Failed to assign function	11	7	4	32	10	31	0.000232559775433195	0.00930754469656176	upf0380 protein yafz	WP_024183847	2.85501E-66
195	FIG00638199: hypothetical protein	11	7	2	34	15	26	1.87027963869892e-05	0.00124524424976599	con domain protein	YP_001458025	3.46525E-13
196	Putative lipase in cluster with Phosphatidate cytidylyltransferase	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	hydrolase	WP_024245228	0.0
197	Putative lipase in cluster with Phosphatidate cytidylyltransferase	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	ynbc protein	WP_024186023	2.58122E-125
198	Failed to assign function	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	hypothetical protein	WP_001279016	1.42357E-53
199	Retron-type RNA-directed DNA polymerase (EC 2.7.7.49)	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	reverse transcriptase family protein	WP_001408531	0.0
200	L-xylulose/3-keto-L-gulonate kinase (EC 2.7.1.-)	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	l-xylulose 3-keto-l-gulonate kinase	WP_021565439	1.84375E-89

201	Putative acetyltransferase	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	acetyltransferase	WP_024170419	3.06774E-74
202	Penicillin G acylase precursor (EC 3.5.1.11)	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	penicillin amidase family protein	WP_001542608	1.13242E-149
203	Failed to assign function	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	hth-type transcriptional activator ttdr	WP_005070978	1.214E-171
204	Failed to assign function	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	hth-type transcriptional regulator	WP_001307402	3.08914E-43
205	Allantoinase (EC 3.5.2.5)	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	allantoinase	WP_021562091	0.0
206	Allantoinase (EC 3.5.2.5)	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	allantoinase	WP_023157557	2.44775E-72
207	FIG00643775: hypothetical protein	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	membrane protein	WP_000540412	5.21772E-102
208	FIG00638277: hypothetical protein	10	8	0	36	0	41	1.82852887371764e-06	0.000242372340985985	inhibitor of apoptosis-promoting bax1 family protein	WP_001610572	2.37541E-89
209	Possible exported protein	10	8	1	35	0	41	1.67892196586802e-05	0.00112823556106331	dknyy family protein	WP_000864456	0.0
210	AidA-I adhesin-like protein	10	8	1	35	0	41	1.67892196586802e-05	0.00112823556106331	extended signal peptide of type v secretion system family protein	WP_000885159	2.31778E-90
211	Failed to assign function	10	8	1	35	0	41	1.67892196586802e-05	0.00112823556106331	hypothetical protein	WP_001278766	5.36755E-89
212	Failed to assign function	10	8	1	35	0	41	1.67892196586802e-05	0.00112823556106331	competence family protein	WP_001128527	0.0
213	FIG00640422: hypothetical protein	10	8	1	35	0	41	1.67892196586802e-05	0.00112823556106331	hypothetical protein	WP_001075213	9.56699E-179
214	Maltose/maltodextrin ABC transporter, permease protein MalF	10	8	2	34	0	41	8.37644051495008e-05	0.00365627832507549	maltose abc transporter permease	WP_001379883	5.23384E-73
215	Failed to assign function	10	8	2	34	0	41	8.37644051495008e-05	0.00365627832507549	t3ss effector protein	WP_021569001	4.46874E-26
216	hypothetical protein	10	8	2	34	0	41	8.37644051495008e-05	0.00365627832507549	dna polymerase iii subunit gamma tau	WP_024226808	9.17602E-52
217	FIG00639173: hypothetical protein	10	8	2	34	0	41	8.37644051495008e-05	0.00365627832507549	hypothetical protein ECRM13514_3978	AHG10633	1.20213E-88
218	FIG00644922: hypothetical protein	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	faimly transcriptional regulator	WP_000166117	0.0
219	PTS system, galactitol-specific IIA component (EC 2.7.1.69)	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	pts galactitol transporter subunit iia	YP_001724624	6.7191E-100
220	PTS system, galactitol-specific IIB component, putative	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	pts galactitol transporter subunit iib	WP_000723159	7.86395E-49
221	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	family sugar-phosphate isomerase	WP_000850107	1.02211E-116
222	transcriptional regulator, DeoR family	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	faimly transcriptional regulator	YP_001724628	2.24791E-180
223	PTS system, galactitol-specific IIA component (EC 2.7.1.69)	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	pts galactitol transporter subunit iia	WP_000623671	3.06879E-101
224	PTS system, galactitol-specific IIB component (EC 2.7.1.69)	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	pts galactitol transporter subunit iib	YP_006120364	1.6691E-53
225	PTS system, galactitol-specific IIC component (EC 2.7.1.69)	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	pts system galactitol-specific iic component	WP_000425554	0.0
226	Ribulose-phosphate 3-epimerase (EC 5.1.3.1)	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	ribulose-phosphate 3-epimerase	WP_001062775	4.37947E-134
227	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	short-chain dehydrogenase	WP_000064393	3.08355E-145
228	6-phospho-3-hexuloisomerase	10	8	1	35	1	40	1.67892196586802e-05	0.00112823556106331	6-phospho 3-hexuloisomerase	YP_001724634	1.89368E-115
229	Failed to assign function	10	8	2	34	1	40	8.37644051495008e-05	0.00365627832507549	hypothetical protein	WP_021557819	3.48095E-39
230	Maltose/maltodextrin ABC transporter, permease protein MalF	10	8	2	34	1	40	8.37644051495008e-05	0.00365627832507549	abc transporter permease	WP_001342235	1.22983E-147
231	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	10	8	2	34	1	40	8.37644051495008e-05	0.00365627832507549	pf11924 family protein	WP_024185293	0.0
232	probable DNA methylase	10	8	0	36	2	39	1.82852887371764e-06	0.000242372340985985	dna methylase family protein	WP_024215834	3.69181E-60
233	Penicillin G acylase precursor (EC 3.5.1.11)	10	8	2	34	2	39	8.37644051495008e-05	0.00365627832507549	penicillin g acylase	WP_021530817	0.0
234	restriction modification system DNA specificity domain	10	8	1	35	3	38	1.67892196586802e-05	0.00112823556106331	restriction modification system dna specificity domain protein	YP_005279646	0.0
235	putative tail protein	10	8	2	34	4	37	8.37644051495008e-05	0.00365627832507549	tail protein	YP_007001977	0.0

236	Failed to assign function	10	8	1	35	5	36	1.67892196586802e-05	0.00112823556106331	hypothetical protein	WP_001207141	1.03973E-96
237	Failed to assign function	10	8	1	35	5	36	1.67892196586802e-05	0.00112823556106331	hypothetical protein	WP_001278659	1.0159E-100
238	putative recet	10	8	1	35	5	36	1.67892196586802e-05	0.00112823556106331	exodeoxyribonuclease viii	YP_007001958	0.0
239	RecT protein	10	8	1	35	5	36	1.67892196586802e-05	0.00112823556106331	family protein	YP_007001957	0.0
240	Phage endopeptidase (EC 3.4.-.-) Rz	10	8	2	34	5	36	8.37644051495008e-05	0.00365627832507549	phage lysis accessory protein	YP_007002004	2.00584E-72
241	Failed to assign function	10	8	1	35	6	35	1.67892196586802e-05	0.00112823556106331	hypothetical protein	WP_001350948	3.61492E-48
242	probable DNA methylase	10	8	2	34	6	35	8.37644051495008e-05	0.00365627832507549	phage n-6-adenine-methyltransferase	WP_000129285	2.91108E-131
243	Phage holin	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	holin	YP_007002002	2.99166E-56
244	FIG00640376: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	membrane protein	YP_007002001	3.51256E-62
245	FIG00638263: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein	WP_023150288	6.40478E-52
246	14	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein	WP_022644860	3.10231E-69
247	13	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein ECIAI39_2671	YP_002408610	5.31848E-111
248	12	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	phage protein	YP_006101957	0.0
249	11	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein APECO1_4048	YP_853618	5.79497E-142
250	FIG00639781: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	phage protein	WP_024009230	7.89076E-62
251	FIG00640468: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	phage protein	YP_853620	6.4439E-63
252	FIG00641850: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein ECIAI39_2676	YP_002408615	5.36349E-84
253	FIG00639045: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	major capsid protein	YP_007001980	0.0
254	FIG00949334: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	phage protein	WP_023145361	1.32354E-117
255	FIG00639626: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	phage protein	WP_000133159	1.09393E-48
256	FIG00640622: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein	WP_001550154	4.12597E-12
257	terminase B protein, putative	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	phage large subunit	YP_007001973	0.0
258	Failed to assign function	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	terminase small subunit	WP_001364181	9.26669E-128
259	FIG00640949: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein	YP_007001971	6.87589E-56
260	FIG00641023: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	phage related protein gp8	WP_001231266	7.56999E-75
261	Origin specific replication binding factor	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	replication protein p	YP_853636	0.0
262	FIG00642684: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	phage related protein	YP_512296	6.9645E-46
263	putative repressor protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	xre family transcriptional regulator	YP_007001961	4.66907E-127
264	FIG00640380: hypothetical protein	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein PhiV10p39	YP_512294	1.70782E-21
265	36	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	hypothetical protein	WP_023908368	2.43663E-52
266	Transcriptional regulator in PFGI-1-like cluster	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	transcriptional family	YP_512291	2.00033E-44
267	Adenine DNA methyltransferase, phage-associated	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	adenine methylase	YP_007001954	2.77027E-165
268	Phage integrase	10	8	2	34	7	34	8.37644051495008e-05	0.00365627832507549	integrase	YP_007001952	0.0
269	Phage endolysin	10	8	2	34	8	33	8.37644051495008e-05	0.00365627832507549	endolysin	YP_007002003	1.72853E-143
270	FIG00638417: hypothetical protein	10	8	2	34	10	31	8.37644051495008e-05	0.00365627832507549	hypothetical protein	WP_000682294	9.88156E-31
271	FIG00638970: hypothetical protein	10	8	1	35	16	25	1.67892196586802e-05	0.00112823556106331	hypothetical protein, partial	WP_023143935	5.59613E-19
272	Transcriptional regulator, AraC family	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	bacterial regulatory helix-turn-helix family protein	WP_001343159	1.4235E-139
273	Failed to assign function	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	cyanate transporter family protein	WP_001507466	1.61695E-28
274	L-xylulose/3-keto-L-gulonate kinase (EC 2.7.1.-)	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	l-xylulose 3-keto-l-gulonate kinase	WP_000988232	0.0
275	Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12)	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	cellulose synthase catalytic subunit	WP_009008383	0.0
276	Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12)	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	pilz domain protein	WP_001173979	1.14706E-99
277	Leader peptidase (Prepilin peptidase) (EC 3.4.23.43) / N-methyltransferase (EC 2.1.1.-)	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	type iv leader peptidase family protein	WP_000340168	1.3281E-55
278	Failed to assign function	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	radical sam superfamily protein	WP_001443393	0.0
279	Failed to assign function	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	radical sam protein	WP_024228525	0.0
280	Putative SAM-dependent methyltransferases	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	methyltransferase domain protein	YP_005727852	9.65433E-79

281	Inner membrane protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	flagellar brake protein	WP_024239986	3.85327E-36
282	Possible exported protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	exported protein	WP_001307483	1.22208E-56
283	FIG01049009: hypothetical protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	conserved protein	WP_001442523	8.31203E-96
284	Heat shock protein C	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	heat shock protein c	WP_024224241	2.79375E-61
285	Leucine-rich repeat protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	type iii secretion system protein	YP_003221201	1.1962E-58
286	FIG00639161: hypothetical protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	dna-directed rna beta subunit 140 kd subunit	WP_024224382	1.46312E-56
287	Failed to assign function	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	aaa domain family protein	WP_001443256	0.0
288	Failed to assign function	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	anaerobic nitric oxide reductase transcription regulator norr	WP_001443255	1.00831E-98
289	Fimbriae usher protein StcC	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	outer membrane protein	WP_001326437	0.0
290	Fimbriae usher protein StcC	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	fimbrial usher protein	WP_001380558	2.37917E-172
291	tRNA-binding protein YgjH	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	trna-binding protein ygjh	WP_024191515	7.1671E-39
292	tRNA-binding protein YgjH	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	trna binding domain protein	WP_001397426	1.46726E-19
293	Ribose/xylose/arabinose/galactoside ABC-type transport systems, periplasmic sugar binding protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	periplasmic binding substrate ribose	YP_003232896	7.2643E-176
294	Ribose/xylose/arabinose/galactoside ABC-type transport systems, periplasmic sugar binding protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	periplasmic binding substrate ribose	WP_021512200	8.99187E-22
295	AidA-I adhesin-like protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	outer membrane autotransporter barrel domain protein	WP_001098985	0.0
296	AidA-I adhesin-like protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	outer membrane autotransporter barrel domain protein	WP_024230920	0.0
297	Carbamate kinase (EC 2.7.2.2)	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	carbamate kinase	WP_001601820	1.14006E-161
298	YjbI protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	pentapeptide repeats family protein	WP_001601102	0.0
299	YjbI protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	pentapeptide repeats family protein	YP_006162775	1.9809E-83
300	Phage terminase, large subunit	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	phage dna packaging protein	WP_001816655	0.0
301	Nondeblocking aminopeptidase YpdE (X-X-[^PR]- specific)	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	aminopeptidase ypde	ETJ26232	1.00589E-39
302	FIG00638165: hypothetical protein	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	helix-turn-helix family protein	WP_001443525	8.88938E-84
303	PTS system HrsA EIIA component / PTS system HrsA EIIB component / PTS system HrsA permease IIC component	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	heat-responsive suppressor hrsa	WP_001397039	4.12237E-29
304	ADA regulatory protein / Methylated-DNA--protein-cysteine methyltransferase (EC 2.1.1.63)	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	bifunctional transcriptional activator dna repair enzyme ada	AGW27650	2.23039E-86
305	ADA regulatory protein / Methylated-DNA--protein-cysteine methyltransferase (EC 2.1.1.63)	9	9	0	36	0	41	9.14264436858821e-06	0.000721819266870833	bifunctional transcriptional activator dna repair enzyme ada	WP_000213537	2.50498E-136
306	FIG00638318: hypothetical protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	cyanate transporter family protein	WP_000128481	4.84753E-173
307	Putative fimbrial protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	fimbrial family protein	WP_001345104	3.01299E-160
308	D-galactonate transporter	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	d-galactonate transport protein	WP_001439121	3.5698E-64
309	D-galactonate transporter	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	d-galactonate transporter	WP_003923349	0.0
310	FIG00638628: hypothetical protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	h repeat-associated protein partial	WP_001719810	1.8368E-26
311	Inner membrane protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	flagellar brake protein	WP_001443529	7.91843E-130
312	Primosomal protein I	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	primosomal protein	WP_023144389	8.84962E-101
313	Conjugative transfer protein PSLT093	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	hypothetical protein	WP_000870095	4.5858E-15
314	PTS system, mannose-specific IID component (EC 2.7.1.69)	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	pts system mannose fructose sorbose iid component family protein	WP_001766259	4.27824E-37

315	PTS system, mannose-specific IID component (EC 2.7.1.69)	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	pts system mannose fructose sorbose iid component family protein	WP_024248450	1.79725E-127
316	Putative cytoplasmic protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	phage protein	WP_019842688	5.50359E-38
317	Mobile element protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	transposase family protein	WP_001369751	5.72018E-131
318	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	bacterial group 1 ig-like partial	WP_001309573	1.8345E-128
319	IcmF-related protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	type vi secretion protein	WP_001399673	0.0
320	Secreted protein Hcp	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	type vi secretion system hcp1 family protein	WP_001248042	5.29933E-100
321	Two-component response regulator CreC	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	sensor protein	WP_021537126	0.0
322	Failed to assign function	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	phage terminase large subunit family partial	WP_000934107	0.0
323	Formate hydrogenlyase transcriptional activator	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	dna-binding transcriptional formate sensing	WP_001250136	0.0
324	FIG00642030: hypothetical protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	hypothetical protein	WP_001341551	1.86994E-175
325	FIG00642030: hypothetical protein	9	9	1	35	0	41	7.49696838224232e-05	0.00365627832507549	hypothetical protein	WP_001448324	0.0
326	Methyl-accepting chemotaxis protein III (ribose and galactose chemoreceptor protein)	9	9	0	36	1	40	9.14264436858821e-06	0.000721819266870833	methyl-accepting chemotaxis protein iii	WP_001098529	0.0
327	Methyl-accepting chemotaxis protein III (ribose and galactose chemoreceptor protein)	9	9	0	36	1	40	9.14264436858821e-06	0.000721819266870833	methyl-accepting chemotaxis protein iii	WP_024234970	1.82937E-14
328	FIG00638496: hypothetical protein	9	9	1	35	1	40	7.49696838224232e-05	0.00365627832507549	hypothetical protein	WP_001371395	3.08391E-111
329	PTS system, galactitol-specific IIC component (EC 2.7.1.69)	9	9	1	35	1	40	7.49696838224232e-05	0.00365627832507549	pts system galactitol-specific transporter subunit iic	WP_000420480	0.0
330	Failed to assign function	9	9	1	35	1	40	7.49696838224232e-05	0.00365627832507549	hypothetical protein	WP_000338655	0.0
331	Per-activated serine protease autotransporter enterotoxin EspC	9	9	1	35	3	38	7.49696838224232e-05	0.00365627832507549	outer membrane autotransporter barrel domain protein	WP_001421750	1.76057E-20
332	Mobile element protein	9	9	1	35	3	38	7.49696838224232e-05	0.00365627832507549	hypothetical protein	WP_024199501	4.55518E-98
333	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	9	9	0	36	4	37	9.14264436858821e-06	0.000721819266870833	enterochelin esterase	WP_000244946	0.0
334	Failed to assign function	9	9	0	36	5	36	9.14264436858821e-06	0.000721819266870833	dna-binding protein	WP_021499497	1.50344E-70
335	VgrG protein	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	rhs element vgr family protein	WP_001397128	1.50597E-153
336	FIG00638117: hypothetical protein	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	aldehyde ferredoxin tungsten cofactor-binding domain protein	WP_001273522	0.0
337	Failed to assign function	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	sugar abc transporter permease	WP_005114448	8.66839E-33
338	Failed to assign function	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	xylose transport system permease xylh	WP_000045993	6.76369E-142
339	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	enterochelin esterase	WP_021538199	1.51515E-16
340	FIG00638928: hypothetical protein	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	peptidoglycan peptidase	YP_003231552	1.84692E-100
341	Failed to assign function	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	is150 orfb partial	WP_000311075	5.32594E-30
342	L-fucose isomerase (EC 5.3.1.25)	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	l-fucose isomerase	WP_004993843	0.0
343	L-fucose isomerase (EC 5.3.1.25)	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	l-fucose isomerase	WP_001429740	0.0
344	Uncharacterized protein ImpB	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	type vi secretion protein	WP_023568072	1.62771E-50
345	Uncharacterized protein ImpB	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	type vi secretion protein	WP_001460472	3.31028E-11
346	Outer membrane usher protein SfmD	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	outer membrane usher protein sfmd	WP_005145078	1.18576E-122
347	Outer membrane usher protein SfmD	8	10	0	36	0	41	4.20561640955058e-05	0.00247003032053605	type vii secretion system usher family protein	WP_024224095	0.0
348	Autoinducer 2 (AI-2) ABC transport system, fused AI2 transporter subunits and ATP-binding component	8	10	0	36	1	40	4.20561640955058e-05	0.00247003032053605	autoinducer 2 import atp-binding protein	WP_000553967	0.0
349	Putative inner membrane protein	8	10	0	36	3	38	4.20561640955058e-05	0.00247003032053605	inner membrane protein yeea	WP_000092610	3.68786E-21

350	Failed to assign function	8	10	0	36	5	36	4.20561640955058e-05	0.00247003032053605	integrase core domain protein	YP_006960298	8.2957E-43
351	hypothetical protein	8	10	0	36	20	21	4.20561640955058e-05	0.00247003032053605	hypothetical protein	YP_005279998	1.20751E-68
352	hypothetical protein	7	11	0	36	0	41	0.000179694519317161	0.0074177897574124	16s ribosomal rna	WP_000585153	6.54987E-36
353	Mobile element protein	7	11	0	36	0	41	0.000179694519317161	0.0074177897574124	reverse transcriptase family partial	WP_001466808	0.0
354	Mobile element protein	7	11	0	36	0	41	0.000179694519317161	0.0074177897574124	is602 transposase	WP_000336971	6.82093E-92
355	Phage tail fibers	7	11	0	36	0	41	0.000179694519317161	0.0074177897574124	tail fiber protein	WP_000385848	8.96209E-59
356	Glyoxylate carboligase (EC 4.1.1.47)	7	11	0	36	0	41	0.000179694519317161	0.0074177897574124	glyoxylate carboligase	YP_668496	5.52238E-24
357	Mobile element protein	7	11	0	36	2	39	0.000179694519317161	0.0074177897574124	transposase	WP_000596979	3.84098E-43

List of overrepresented genes in LEE positive STEC other than STEC in HUS-group 1

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR<0.01	Function Blast2GO	Hit Accession	E-value
1	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	7	11	36	0	36	5	3.32459795221389e-07	0.000156276624733803	short-chain-fatty-acid-- ligase	YP_002293100	0.0
2	FIG00638277: hypothetical protein	8	10	36	0	37	4	1.82852887371764e-06	0.000267342787797123	inhibitor of apoptosis-promoting bax1 family protein	YP_001461976	9.65873E-118
3	Failed to assign function	9	9	36	0	37	4	9.14264436858821e-06	0.000805444669740015	branched-chain amino acid transport system permease component family protein	WP_001461725	0.0
4	FIG00638206: hypothetical protein	7	11	36	0	38	3	3.32459795221389e-07	0.000156276624733803	rfp- chloride channel family protein	YP_003229126	0.0
5	Failed to assign function	9	9	36	0	39	2	9.14264436858821e-06	0.000805444669740015	radical sam superfamily protein	YP_026196	0.0
6	Hypothetical transcriptional regulator yidL	7	11	36	0	40	1	3.32459795221389e-07	0.000156276624733803	bacterial regulatory helix-turn-helix family protein	YP_001732507	0.0
7	FIG00638146: hypothetical protein	9	9	36	0	40	1	9.14264436858821e-06	0.000805444669740015	phage protein	YP_003230357	0.0
8	Putative acetyltransferase	8	10	36	0	41	0	1.82852887371764e-06	0.000267342787797123	acetyltransferase	WP_001277594	2.8035E-117
9	Failed to assign function	8	10	36	0	41	0	1.82852887371764e-06	0.000267342787797123	hth-type transcriptional activator ttdr	WP_000458596	0.0
10	Failed to assign function	9	9	36	0	41	0	9.14264436858821e-06	0.000805444669740015	anaerobic nitric oxide reductase transcription regulator	WP_001351304	0.0
11	tRNA-binding protein YgjH	9	9	36	0	41	0	9.14264436858821e-06	0.000805444669740015	trna-binding protein ygjh	WP_024191515	4.42794E-70
12	Transcriptional regulator, AraC family	9	9	36	0	41	0	9.14264436858821e-06	0.000805444669740015	transcriptional regulator	WP_001343159	0.0
13	FIG00638165: hypothetical protein	9	9	36	0	41	0	9.14264436858821e-06	0.000805444669740015	helix-turn-helix family protein	YP_002927361	8.38662E-81
14	L-fucose isomerase (EC 5.3.1.25)	10	8	36	0	41	0	4.20561640955057e-05	0.00232432045748015	l-fucose isomerase	YP_003230803	0.0
15	Leader peptidase (Prepilin peptidase) (EC 3.4.23.43) / N-methyltransferase (EC 2.1.1.-)	9	9	35	1	26	15	7.49696838224231e-05	0.00346058144366252	type iv leader peptidase family protein	YP_003223898	3.97592E-96
16	Cyn operon transcriptional activator	9	9	35	1	28	13	7.49696838224231e-05	0.00346058144366252	hth-type transcriptional regulator	WP_001306927	0.0
17	FIG00638289: hypothetical protein	3	15	35	1	30	11	1.39981544739578e-09	1.83859396217947e-06	methylglyoxal detoxification protein	WP_001701731	0.0
18	Cyanate transport protein CynX	9	9	35	1	30	11	7.49696838224231e-05	0.00346058144366252	cyanate transporter	WP_001369164	0.0
19	Cyanate hydratase (EC 4.2.1.104)	9	9	35	1	31	10	7.49696838224231e-05	0.00346058144366252	cyanate hydratase	YP_001461517	1.60513E-99
20	LsrR, transcriptional repressor of lsr operon	9	9	35	1	32	9	7.49696838224231e-05	0.00346058144366252	transcriptional regulator lsrr	NP_707462	0.0
21	Methyl-accepting chemotaxis protein III (ribose and galactose chemoreceptor protein)	9	9	35	1	32	9	7.49696838224231e-05	0.00346058144366252	methyl-accepting chemotaxis protein iii	YP_001462694	0.0
22	Hydrogenase-4 component E (EC 1.-.-.-)	9	9	35	1	33	8	7.49696838224231e-05	0.00346058144366252	hydrogenase-4 component e	NP_289038	1.15019E-113
23	Hydrogenase-4 component J (EC 1.-.-.-)	9	9	35	1	33	8	7.49696838224231e-05	0.00346058144366252	formate hydrogenlyase maturation family protein	WP_001142095	6.54153E-94
24	FIG00638475: hypothetical protein	3	15	35	1	34	7	1.39981544739578e-09	1.83859396217947e-06	sugar phosphate isomerase	NP_418309	0.0
25	orf, hypothetical protein	7	11	35	1	34	7	3.39418256051603e-06	0.000378900459294946	conserved protein	WP_009008192	4.15264E-128
26	FIG00638952: hypothetical protein	7	11	35	1	34	7	3.39418256051603e-06	0.000378900459294946	inner membrane protein yidi	YP_001465161	1.62053E-67
27	Putative resistance protein	3	15	35	1	36	5	1.39981544739578e-09	1.83859396217947e-06	inner membrane protein yihn	WP_000956332	0.0
28	Hypothetical transcriptional regulator yihL	3	15	35	1	36	5	1.39981544739578e-09	1.83859396217947e-06	transcriptional regulator	YP_003231621	3.97526E-170
29	Fimbriae usher protein StcC	9	9	35	1	36	5	7.49696838224231e-05	0.00346058144366252	outer membrane protein	WP_023142854	0.0
30	Choline-sulfatase (EC 3.1.6.6)	7	11	35	1	38	3	3.39418256051603e-06	0.000378900459294946	sulfatase family protein	YP_003231800	0.0

31	Ser/Thr and Tyr protein phosphatase (dual specificity)	4	14	35	1	39	2	1.28231798497189e-08	1.0898194262867e-05	dual specificity catalytic domain protein	WP_001375409	0.0
32	Carbamate kinase (EC 2.7.2.2)	9	9	35	1	41	0	7.49696838224231e-05	0.00346058144366252	carbamate kinase	YP_003227628	0.0
33	Inner membrane protein	9	9	35	1	41	0	7.49696838224231e-05	0.00346058144366252	flagellar brake protein	YP_310143	4.1395E-172
34	ADA regulatory protein / Methylated-DNA--protein-cysteine methyltransferase (EC 2.1.1.63)	9	9	35	1	41	0	7.49696838224231e-05	0.00346058144366252	bifunctional transcriptional activator dna repair enzyme ada	YP_002293756	0.0
35	FIG00638318: hypothetical protein	9	9	35	1	41	0	7.49696838224231e-05	0.00346058144366252	cyanate transporter family protein	WP_001396688	0.0
36	Two-component response regulator CreC	9	9	35	1	41	0	7.49696838224231e-05	0.00346058144366252	sensor protein	YP_410697	0.0
37	Putative outer membrane protein	7	11	34	2	1	40	1.87027963869892e-05	0.00143732979893202	lpxr	WP_000628028	0.0
38	Hydrogenase-4 component B (EC 1.-.-.-) / Formate hydrogenlyase subunit 3	0	18	34	2	16	25	1.96025129168993e-12	1.41608553311681e-08	hydrogenase-4 component b	WP_024237058	0.0
39	CRISPR-associated protein, Cas2	8	10	34	2	28	13	8.37644051495006e-05	0.00384199404952376	crispr-associated endoribonuclease subtype i-e	NP_311635	3.26125E-60
40	Protein YjgL, putative CCAAT-box DNA binding protein subunit B	7	11	34	2	29	12	1.87027963869892e-05	0.00143732979893202	protein ccaat-box dna binding protein subunit b	YP_003232297	0.0
41	Formate hydrogenlyase transcriptional activator	6	12	34	2	33	8	3.69967596281817e-06	0.000408037544357228	dna-binding transcriptional activator	YP_003230476	0.0
42	Sodium-Choline Symporter	7	11	34	2	36	5	1.87027963869892e-05	0.00143732979893202	symporter	YP_001465163	0.0
43	FIG00639909: hypothetical protein	0	18	34	2	37	4	1.96025129168993e-12	1.41608553311681e-08	family transcriptional regulator	WP_001318027	4.88819E-31
44	Putative lipase in cluster with Phosphatidate cytidylyltransferase	7	11	34	2	37	4	1.87027963869892e-05	0.00143732979893202	hydrolase	YP_003229027	0.0
45	Allantoinase (EC 3.5.2.5)	8	10	34	2	41	0	8.37644051495006e-05	0.00384199404952376	allantoinase	YP_002291812	0.0
46	COG3541: Predicted nucleotidyltransferase	0	18	33	3	23	18	1.37217590418296e-11	4.95629936590884e-08	nucleotidyltransferase family protein	WP_001697550	0.0
47	Failed to assign function	0	18	33	3	23	18	1.37217590418296e-11	4.95629936590884e-08	spfh domain band 7 family protein	WP_001495680	0.0
48	FIG01049009: hypothetical protein	7	11	33	3	25	16	7.37142379422834e-05	0.00346058144366252	conserved protein	YP_002389294	0.0
49	Putative SAM-dependent methyltransferases	4	14	33	3	29	12	4.74367183683931e-07	0.000201578149113689	methyltransferase domain protein	NP_287402	1.70593E-161
50	YeeU protein (antitoxin to YeeV)	1	17	32	4	16	25	2.05140297675352e-09	2.46988918401124e-06	phage protein	YP_001458813	4.14444E-31
51	FIG00639142: hypothetical protein	4	14	32	4	28	13	1.92076336032631e-06	0.000267342787797123	intergenic-region protein	WP_001462762	1.64652E-18
52	FIG00642443: hypothetical protein	3	15	32	4	38	3	2.67739158449308e-07	0.000143270198565763	dna-binding protein	YP_002404309	3.76999E-144
53	FIG00638089: hypothetical protein	6	12	31	5	6	35	0.000159597231896699	0.00660705102132811	hypothetical protein ECO26_2447	YP_003229432	1.6526E-36
54	Mobile element protein	3	15	31	5	20	21	9.85623954693032e-07	0.000267342787797123	insertion element is1 protein insb	WP_005039619	1.90582E-45
55	FIG00641476: hypothetical protein	3	15	31	5	23	18	9.85623954693032e-07	0.000267342787797123	hypothetical protein EC55989_3343	YP_002404308	7.50484E-45
56	Failed to assign function	3	15	31	5	25	16	9.85623954693032e-07	0.000267342787797123	hypothetical protein	YP_005276598	1.8705E-13
57	D-serine permease DsdX	3	15	31	5	29	12	9.85623954693032e-07	0.000267342787797123	permease	WP_021567503	1.17163E-27
58	Sucrose-6-phosphate hydrolase (EC 3.2.1.26)	3	15	31	5	31	10	9.85623954693032e-07	0.000267342787797123	sucrose-6-phosphate hydrolase	YP_001463698	0.0
59	Sucrose specific transcriptional regulator CscR, LacI family	3	15	31	5	31	10	9.85623954693032e-07	0.000267342787797123	sucrose operon repressor	YP_002387820	0.0
60	PsiA protein	6	12	31	5	34	7	0.000159597231896699	0.00660705102132811	plasmid sos inhibition protein a	YP_004119756	2.13834E-141
61	FIG00641264: hypothetical protein	3	15	30	6	4	37	3.16743068837495e-06	0.000363198718933661	conserved protein	WP_001470661	1.92527E-33
62	FIG00641704: hypothetical protein	3	15	30	6	4	37	3.16743068837495e-06	0.000363198718933661	hypothetical protein ECED1_5037	YP_002400782	4.33157E-26
63	orf; Unknown function	3	15	30	6	4	37	3.16743068837495e-06	0.000363198718933661	conserved protein	WP_001004880	2.33927E-29
64	hypothetical protein	3	15	30	6	10	31	3.16743068837495e-06	0.000363198718933661	transposase	YP_006133018	5.92025E-18
65	FIG00638000: hypothetical protein	3	15	30	6	14	27	3.16743068837495e-06	0.000363198718933661	family transcriptional regulator	WP_001387282	9.90156E-39
66	FIG00638721: hypothetical protein	4	14	30	6	18	23	1.95309811909893e-05	0.00148654341346532	pf11682 family protein	YP_002391774	1.51994E-95
67	Glycosyltransferase IroB	0	18	30	6	22	19	1.38864201503315e-09	1.83859396217947e-06	glucosyl-transferase	YP_003228369	0.0
68	Putative exported protein precursor	0	18	30	6	26	15	1.38864201503315e-09	1.83859396217947e-06	outer membrane protein	AAN81391	1.49824E-140
69	Aldose-ketose isomerase YihS	3	15	30	6	26	15	3.16743068837495e-06	0.000363198718933661	glucosamine isomerase	CDL26975	0.0
70	Xanthosine operon regulatory protein XapR, LysR family	0	18	30	6	27	14	1.38864201503315e-09	1.83859396217947e-06	hth-type transcriptional regulator xapr	WP_000442952	0.0
71	Sugar-1-epimerase YihR	3	15	30	6	27	14	3.16743068837495e-06	0.000363198718933661	aldose-1-epimerase	WP_000430815	0.0

72	Glucuronide transport protein YihO	3	15	30	6	27	14	3.16743068837495e-06	0.000363198718933661	inner membrane symporter yihp	WP_024241265	0.0
73	Outer membrane sugar transport protein YshA	3	15	30	6	27	14	3.16743068837495e-06	0.000363198718933661	porin ompl	AAB03009	3.42808E-157
74	putative phage inhibition, colicin resistance and tellurite resistance protein	3	15	29	7	2	39	9.11781268932561e-06	0.000805444669740015	terf	YP_002269834	4.54196E-84
75	orf; Unknown function	3	15	29	7	4	37	9.11781268932561e-06	0.000805444669740015	conserved protein	NP_286721	6.68149E-52
76	FIG00642259: hypothetical protein	3	15	29	7	8	33	9.11781268932561e-06	0.000805444669740015	shia domain protein	NP_309389	8.66179E-40
77	FIG00643174: hypothetical protein	4	14	29	7	10	31	5.22580821962179e-05	0.00256811146792842	colicin m immunity protein	YP_002756569	1.96575E-66
78	Flagellar hook-associated protein FliD	1	17	29	7	14	27	1.08512454603305e-07	6.81646932221109e-05	flagellar hook-associated protein 2	YP_002398129	0.0
79	Failed to assign function	0	18	29	7	16	25	4.95943576797556e-09	4.47837049848193e-06	hypothetical protein	WP_024245286	2.59084E-75
80	Hydrogenase-4 component D (EC 1.-.-.-)	3	15	29	7	16	25	9.11781268932561e-06	0.000805444669740015	hydrogenase-4 component d	YP_003230470	0.0
81	DNA helicase	3	15	29	7	16	25	9.11781268932561e-06	0.000805444669740015	rep helicase-like protein	YP_003228378	0.0
82	putative membrane protein	3	15	29	7	17	24	9.11781268932561e-06	0.000805444669740015	membrane protein	NP_309401	8.64022E-101
83	Phosphate starvation-inducible protein PhoH, predicted ATPase	0	18	29	7	25	16	4.95943576797556e-09	4.47837049848193e-06	protein phoh	WP_001323668	4.16995E-61
84	FIG00638961: hypothetical protein	0	18	29	7	26	15	4.95943576797556e-09	4.47837049848193e-06	conserved protein	NP_417940	6.75275E-84
85	FIG00641784: hypothetical protein	3	15	29	7	26	15	9.11781268932561e-06	0.000805444669740015	hypothetical protein Z1194	NP_286729	1.06598E-111
86	Alpha-glucosyltransferase YihQ	3	15	29	7	26	15	9.11781268932561e-06	0.000805444669740015	glycosyl hydrolases 31 family protein	YP_003231615	0.0
87	Maltose-6'-phosphate glucosidase (EC 3.2.1.122)	4	14	29	7	29	12	5.22580821962179e-05	0.00256811146792842	6-phospho-alpha-glucosidase	YP_002295243	8.16711E-159
88	Outer-membrane protein yhbX precursor	4	14	29	7	30	11	5.22580821962179e-05	0.00256811146792842	inner membrane	WP_000470374	0.0
89	Fructokinase (EC 2.7.1.4)	0	18	29	7	32	9	4.95943576797556e-09	4.47837049848193e-06	fructokinase	YP_003230363	0.0
90	FIG00638493: hypothetical protein	5	13	29	7	34	7	0.000242075268026544	0.00935161356269386	impa-related family protein	YP_002385707	0.0
91	putative hemolysin activator protein	3	15	28	8	3	38	2.39472803323198e-05	0.00155851489297908	potra -type family protein	WP_001428611	0.0
92	Hydrogenase-4 component I (EC 1.-.-.-)	2	16	28	8	11	30	3.40926492997944e-06	0.000378900459294946	formate hydrogenlyase subunit 7	WP_001341609	1.18106E-103
93	FIG00639376: hypothetical protein	0	18	28	8	16	25	1.61181662459205e-08	1.10892983771933e-05	conserved protein	YP_003230659	0.0
94	Putative cytoplasmic protein	0	18	28	8	21	20	1.61181662459205e-08	1.10892983771933e-05	yoag domain protein	YP_002293245	2.54985E-18
95	COG0148: Enolase	0	18	28	8	23	18	1.61181662459205e-08	1.10892983771933e-05	zeta toxin family protein	YP_002405395	0.0
96	Xanthosine phosphorylase (EC 2.4.2.1)	0	18	28	8	27	14	1.61181662459205e-08	1.10892983771933e-05	xanthosine phosphorylase	NP_416902	0.0
97	YqeJ protein	4	14	28	8	31	10	0.00012775355383328	0.00542877454642125	protein	YP_311837	1.86382E-109
98	hypothetical protein	2	16	27	9	1	40	8.88606903727022e-06	0.000805444669740015	non-lee-encoded type iii secreted effector	YP_003230601	0.0
99	FIG00638542: hypothetical protein	0	18	27	9	19	22	4.83544987377616e-08	3.17557180801445e-05	helix-turn-helix family protein	YP_405986	1.23551E-45
100	TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins	3	15	27	9	28	13	5.81773060028868e-05	0.00284930751569393	outer membrane insertion c-terminal signal domain protein	YP_003230995	0.0
101	Mobile element protein	1	17	26	10	1	40	2.39838313739298e-06	0.000296169568966272	is4 familly	WP_001375545	1.09183E-159
102	putative membrane protein	3	15	26	10	4	37	0.000132114161451312	0.0055487947809551	membrane protein	WP_000282206	0.0
103	Phosphoesterase (EC 3.1.-.-)	0	18	26	10	8	33	1.35392596465733e-07	7.52366243744965e-05	phosphodiesterase yaei	YP_003227266	0.0
104	Integrase	3	15	26	10	8	33	0.000132114161451312	0.0055487947809551	integrase	NP_286655	0.0
105	FIG00640819: hypothetical protein	3	15	26	10	9	32	0.000132114161451312	0.0055487947809551	hypothetical protein Z1205	NP_286740	4.01503E-142
106	FIG00638496: hypothetical protein	1	17	26	10	13	28	2.39838313739298e-06	0.000296169568966272	hypothetical protein	WP_001700672	2.70544E-161
107	Failed to assign function	0	18	26	10	15	26	1.35392596465733e-07	7.52366243744965e-05	hypothetical phage protein	YP_002294180	1.04955E-80
108	FIG00642194: hypothetical protein	3	15	26	10	22	19	0.000132114161451312	0.0055487947809551	conserved protein	WP_000350594	6.8239E-70
109	Phage DNA-packaging protein	0	18	25	11	8	33	3.56944117955114e-07	0.000156276624733803	dna packaging protein	WP_000012985	1.37754E-81
110	Ankyrin-repeat protein A	0	18	25	11	10	31	3.56944117955114e-07	0.000156276624733803	enterotoxin	WP_001345136	0.0
111	Uncharacterized protein yjfZ	0	18	25	11	20	21	3.56944117955114e-07	0.000156276624733803	pf10887 family protein	YP_003232247	0.0
112	Type III secretion inner membrane protein (YscU,SpaS,EscU,HrcU,SsaU, homologous to flagellar export components)	0	18	24	12	1	40	8.92360294887787e-07	0.000267342787797123	family type iii secretion protein	WP_001613736	4.01737E-130
113	TRAP dicarboxylate transporter, DctM subunit, unknown substrate 8	0	18	24	12	7	34	8.92360294887787e-07	0.000267342787797123	c4-dicarboxylate transporter large subunit	YP_002332098	0.0

114	TRAP-type transport system, small permease component, predicted N-acetylneuraminate transporter	0	18	24	12	7	34	8.92360294887787e-07	0.000267342787797123	tripartite atp-independent periplasmic transporter (dctq-like) protein	YP_003232421	1.17927E-77
115	Regulatory protein Cro of bacteriophage BP-933W	1	17	24	12	9	32	1.37423485412718e-05	0.00107907310719726	kda cro protein	NP_287826	1.8578E-49
116	putative regulator; Regulation (Phage or Prophage Related)	1	17	24	12	9	32	1.37423485412718e-05	0.00107907310719726	helix-turn-helix family protein	WP_000233323	6.25815E-75
117	Arabinose 5-phosphate isomerase (EC 5.3.1.13)	0	18	24	12	11	30	8.92360294887787e-07	0.000267342787797123	phosphosugar isomerase	WP_001377601	4.98318E-123
118	Cystathionine beta-lyase (EC 4.4.1.8) (CBL) (Beta-cystathionase) (Cysteine lyase) / Maltose regulon modulator	0	18	24	12	11	30	8.92360294887787e-07	0.000267342787797123	cystathionine beta-lyase	YP_003230817	0.0
119	PTS system, maltose and glucose-specific IIC component (EC 2.7.1.69) / PTS system, maltose and glucose-specific IIB component (EC 2.7.1.69)	0	18	24	12	11	30	8.92360294887787e-07	0.000267342787797123	phosphotransferase system enzyme:iib iic component	YP_003230818	0.0
120	Beta-glucoside bgl operon antiterminator, BglG family	0	18	24	12	11	30	8.92360294887787e-07	0.000267342787797123	antiterminator	WP_000464369	1.0133E-175
121	YcgN (Fragment)	0	18	24	12	11	30	8.92360294887787e-07	0.000267342787797123	upf0260 protein ycgn	WP_000512275	1.41352E-136
122	FIG00642859: hypothetical protein	0	18	24	12	15	26	8.92360294887787e-07	0.000267342787797123	hypothetical protein ECO111_p2-092	YP_003237996	1.55961E-31
123	FIG00639255: hypothetical protein	0	18	24	12	18	23	8.92360294887787e-07	0.000267342787797123	peptidase	WP_001700873	0.0
124	FIG00637883: hypothetical protein	0	18	24	12	18	23	8.92360294887787e-07	0.000267342787797123	lipoprotein	WP_000011863	0.0
125	FIG00638269: hypothetical protein	0	18	24	12	22	19	8.92360294887787e-07	0.000267342787797123	phage protein	WP_001342092	7.29895E-19
126	Phosphate starvation-inducible protein PhoH, predicted ATPase	0	18	24	12	24	17	8.92360294887787e-07	0.000267342787797123	protein phoh	WP_001556530	0.0
127	FIG00638895: hypothetical protein	1	17	24	12	25	16	1.37423485412718e-05	0.00107907310719726	phage protein	YP_003227663	5.40025E-110
128	Protein NinE	1	17	24	12	26	15	1.37423485412718e-05	0.00107907310719726	protein nine	WP_000518235	5.41651E-27
129	Putative sucrose phosphorylase (EC 2.4.1.7)	0	18	23	13	0	41	2.12793608780932e-06	0.000267342787797123	sucrose phosphorylase	WP_023143154	8.99427E-98
130	Failed to assign function	0	18	23	13	0	41	2.12793608780932e-06	0.000267342787797123	acetyl- :acetoacetyl- transferase alpha beta subunit	WP_024237352	1.46471E-163
131	Acetyl-CoA:acetoacetyl-CoA transferase, alpha subunit (EC 2.8.3.8)	0	18	23	13	0	41	2.12793608780932e-06	0.000267342787797123	coenzyme a transferase family protein	WP_001342179	0.0
132	Lactaldehyde dehydrogenase involved in fucose or rhamnose utilization (EC 1.2.1.22)	0	18	23	13	0	41	2.12793608780932e-06	0.000267342787797123	alcohol dehydrogenase	YP_003231587	0.0
133	Molybdate metabolism regulator	0	18	23	13	0	41	2.12793608780932e-06	0.000267342787797123	wgr domain protein	WP_000680162	0.0
134	Cell Inhibiting Factor (Cif)	0	18	23	13	1	40	2.12793608780932e-06	0.000267342787797123	cell division protein	WP_000652081	0.0
135	FIG00638116: hypothetical protein	2	16	23	13	1	40	0.000223790233337935	0.00866842169240343	phage protein	YP_003230330	2.43155E-75
136	Mobile element protein	0	18	23	13	2	39	2.12793608780932e-06	0.000267342787797123	integrase core domain partial	WP_000001467	7.65121E-107
137	Putative exported protein	0	18	23	13	3	38	2.12793608780932e-06	0.000267342787797123	dknyy family protein	WP_001324310	1.4816E-92
138	Failed to assign function	0	18	23	13	3	38	2.12793608780932e-06	0.000267342787797123	ribose abc transporter	NP_755883	6.2564E-169
139	Mobile element protein	1	17	23	13	4	37	3.0546179325005e-05	0.00197022856646282	integrase	CAB59975	1.35043E-174
140	hypothetical protein	2	16	23	13	4	37	0.000223790233337935	0.00866842169240343	phage protein	YP_007111834	2.4532E-29
141	Phage EaA protein	0	18	23	13	5	36	2.12793608780932e-06	0.000267342787797123	valyl-trna synthetase	YP_003230652	0.0
142	FIG00637932: hypothetical protein	0	18	23	13	5	36	2.12793608780932e-06	0.000267342787797123	flagellar biosynthesis anti-sigma factor	YP_003227352	6.40198E-28
143	putative membrane protein	2	16	23	13	5	36	0.000223790233337935	0.00866842169240343	transmembrane protein	YP_003229739	2.4201E-153
144	Failed to assign function	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	alpha c-terminal all-beta domain protein	WP_001440793	2.68175E-126
145	Flagellar biosynthesis protein FlhA	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar biosynthesis protein	WP_001140804	0.0
146	Flagellar biosynthesis protein FlhB	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flhb family protein	WP_000785657	0.0
147	Flagellar biosynthesis protein FliR	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar biosynthetic protein	WP_001310546	2.16041E-121
148	Flagellar biosynthesis protein FliQ	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	bacterial export 3 family protein	YP_003227338	1.25723E-26

149	Flagellar biosynthesis protein FliP	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar biosynthetic protein	WP_000830281	3.70758E-127
150	Flagellar motor switch protein FliN	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar motor switch protein	YP_006167278	2.97909E-63
151	Flagellar motor switch protein FliM	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	surface presentation of antigens protein	YP_003227341	0.0
152	Flagellar regulatory protein FleQ	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	sigma-54 interaction domain protein	WP_009008331	0.0
153	Flagellar hook-basal body complex protein FliE	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar hook-basal body complex protein	YP_003227343	2.26475E-39
154	Flagellar motor switch protein FliG	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar motor switch protein	WP_001308395	0.0
155	Flagellar assembly protein FliH	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar assembly family protein	YP_003227346	3.96152E-138
156	Flagellum-specific ATP synthase FliI	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	atp synthase	YP_003227347	0.0
157	FIG00639403: hypothetical protein	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar export protein	YP_003227348	9.69213E-71
158	Glycerol-3-phosphate cytidylyltransferase (EC 2.7.7.39)	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	glycerol-3-phosphate cytidylyltransferase	YP_003227349	1.15464E-89
159	Cps2A	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	glycosyl transferase 2 family protein	WP_000294389	0.0
160	Lysine-N-methylase (EC 2.1.1.-)	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	lateral flagellar associated protein	YP_003227350	0.0
161	FIG00641064: hypothetical protein	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	lateral flagellar chaperone protein	YP_003227351	5.79427E-82
162	Flagellar basal-body P-ring formation protein FlgA	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagella basal body p-ring formation protein	YP_003227353	2.23096E-153
163	Flagellar basal-body rod protein FlgB	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar basal-body rod protein	WP_001315265	6.12523E-65
164	Flagellar basal-body rod protein FlgC	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar basal-body rod protein	WP_024216450	3.17607E-76
165	Flagellar basal-body rod modification protein FlgD	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar hook capping family protein	YP_006137328	6.45638E-125
166	Flagellar basal-body rod protein FlgG	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar basal-body rod protein	YP_002411046	1.61292E-154
167	Flagellar L-ring protein FlgH	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar l-ring protein	YP_003227360	4.55402E-175
168	Flagellar P-ring protein FlgI	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar p-ring protein 1	YP_003227361	0.0
169	Flagellar protein FlgJ [peptidoglycan hydrolase] (EC 3.2.1.-)	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	lateral flagellar peptidoglycan hydrolase	WP_001589748	2.02543E-59
170	Flagellar hook-associated protein FlgL	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar hook-associated protein 3	YP_003227364	6.62713E-156
171	FIG00639166: hypothetical protein	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	lateral flagellar hook associated protein	WP_001195933	1.63169E-173
172	probable regulatory protein YPO0736	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	lateral flagellar transmembrane regulator	YP_003227366	0.0
173	Flagellin protein FlaA	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	bacterial flagellin family protein	YP_002411056	9.42555E-126
174	Flagellar hook-associated protein FliD	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	lateral flagellar filament capping protein	WP_000609684	0.0
175	Flagellar hook-length control protein FliK	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar hook-length control protein	YP_003227371	1.17986E-134
176	FIG00639010: hypothetical protein	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar basal body-associated family protein	WP_001393689	5.50792E-79
177	Flagellar motor rotation protein MotA	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	flagellar motor stator protein	YP_003232831	0.0
178	Flagellar motor rotation protein MotB	0	18	23	13	6	35	2.12793608780932e-06	0.000267342787797123	chemotaxis protein	YP_003227375	0.0
179	FIG00641110: hypothetical protein	0	18	23	13	7	34	2.12793608780932e-06	0.000267342787797123	hypothetical protein ECO26_5378	YP_003232254	0.0
180	Ribose ABC transport system, ATP-binding protein RbsA (TC 3.A.1.2.1)	0	18	23	13	7	34	2.12793608780932e-06	0.000267342787797123	sugar uptake abc transporter atp-binding protein	YP_003231270	0.0
181	Transcriptional repressor of aga operon	0	18	23	13	7	34	2.12793608780932e-06	0.000267342787797123	family transcriptional regulator	YP_542672	0.0
182	Putative sugar kinase, PfkB family protein (EC 2.7.1.-)	0	18	23	13	7	34	2.12793608780932e-06	0.000267342787797123	pfkb family carbohydrate kinase	WP_001275272	0.0
183	TIORF127 protein	0	18	23	13	7	34	2.12793608780932e-06	0.000267342787797123	tiorf127 protein	YP_003231267	0.0
184	Fructokinase (EC 2.7.1.4)	0	18	23	13	7	34	2.12793608780932e-06	0.000267342787797123	carbohydrate kinase	YP_003231268	0.0
185	Ribose ABC transport system, permease protein RbsC (TC 3.A.1.2.1)	0	18	23	13	7	34	2.12793608780932e-06	0.000267342787797123	ribose transport system permease	YP_003231269	9.82216E-168

186	TRAP transporter solute receptor, unknown substrate 6	0	18	23	13	7	34	2.12793608780932e-06	0.000267342787797123	c4-dicarboxylate-binding periplasmic protein	YP_543995	0.0
187	PTS system, mannose-specific IID component (EC 2.7.1.69)	1	17	23	13	7	34	3.0546179325005e-05	0.00197022856646282	pts system mannose fructose sorbose iid component family protein	YP_002388403	0.0
188	Failed to assign function	0	18	23	13	10	31	2.12793608780932e-06	0.000267342787797123	hypothetical protein	YP_005280112	3.15197E-31
189	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	0	18	23	13	11	30	2.12793608780932e-06	0.000267342787797123	d-isomer specific 2-hydroxyacid nad binding domain protein	YP_003230815	0.0
190	Shiga toxin A-chain precursor (EC 3.2.2.22)	2	16	23	13	19	22	0.000223790233337935	0.00866842169240343	shiga toxin subunit a	ADG56725	0.0
191	Holliday junction resolvase / Crossover junction endodeoxyribonuclease rusA (EC 3.1.22.-)	0	18	23	13	26	15	2.12793608780932e-06	0.000267342787797123	crossover junction endodeoxy-ribonuclease rusa	NP_415082	3.28072E-81
192	Putative sucrose phosphorylase (EC 2.4.1.7)	0	18	22	14	0	41	4.86385391499277e-06	0.000474817306512267	sucrose phosphorylase	WP_000459435	0.0
193	Failed to assign function	1	17	22	14	0	41	6.5054046113028e-05	0.00315402972564104	antirepressor protein cro	YP_001449261	4.71416E-36
194	Molybdate metabolism regulator	0	18	22	14	1	40	4.86385391499277e-06	0.000474817306512267	twice split molybdate metabolism regulator	WP_024228772	0.0
195	hypothetical protein	0	18	22	14	2	39	4.86385391499277e-06	0.000474817306512267	sukh-3 immunity family protein	YP_002386019	8.90435E-98
196	Failed to assign function	0	18	22	14	2	39	4.86385391499277e-06	0.000474817306512267	c3-like protein	NP_958208	9.07044E-18
197	Failed to assign function	0	18	22	14	3	38	4.86385391499277e-06	0.000474817306512267	hypothetical protein ECO26_0638	YP_003227708	5.06507E-78
198	hypothetical protein	0	18	22	14	5	36	4.86385391499277e-06	0.000474817306512267	hypothetical protein	WP_001341484	3.26709E-14
199	Flagellar M-ring protein FliF	0	18	22	14	6	35	4.86385391499277e-06	0.000474817306512267	flagellar m-ring protein	YP_003227344	0.0
200	Flagellar basal-body rod protein FlgF	0	18	22	14	6	35	4.86385391499277e-06	0.000474817306512267	flagellar hook-basal body protein	YP_003227358	9.38634E-148
201	Flagellar hook-associated protein FlgK	0	18	22	14	6	35	4.86385391499277e-06	0.000474817306512267	flagellar hook-associated protein	YP_003227363	0.0
202	Flagellar biosynthesis protein FliS	0	18	22	14	6	35	4.86385391499277e-06	0.000474817306512267	flagellar protein	YP_001726302	1.75287E-80
203	FIG00639370: hypothetical protein	0	18	22	14	6	35	4.86385391499277e-06	0.000474817306512267	lateral flagellar chaperone protein	WP_001316130	1.11926E-34
204	RNA polymerase sigma factor for flagellar operon	0	18	22	14	6	35	4.86385391499277e-06	0.000474817306512267	flagellar biosynthesis sigma factor	YP_006094618	1.14855E-147
205	Putative Na(+)/H(+) exchanger protein, CPA1 family precursor	0	18	22	14	7	34	4.86385391499277e-06	0.000474817306512267	na+ h+ antiporter	YP_003232423	0.0
206	FIG00641267: hypothetical protein	0	18	22	14	14	27	4.86385391499277e-06	0.000474817306512267	hypothetical protein ECO26_3713	YP_003230650	2.99035E-32
207	COG2932: Predicted transcriptional regulator	0	18	22	14	14	27	4.86385391499277e-06	0.000474817306512267	helix-turn-helix family protein	WP_019842147	1.64065E-180
208	Failed to assign function	1	17	22	14	16	25	6.5054046113028e-05	0.00315402972564104	antirepressor protein cro	WP_023228645	4.63559E-43
209	FIG00638399: hypothetical protein	0	18	22	14	19	22	4.86385391499277e-06	0.000474817306512267	endoribonuclease	YP_002384181	3.25767E-68
210	hypothetical protein	0	18	22	14	21	20	4.86385391499277e-06	0.000474817306512267	inner membrane protein	BAB33410	1.91132E-64
211	Shikimate kinase I (EC 2.7.1.71)	1	17	22	14	33	8	6.5054046113028e-05	0.00315402972564104	shikimate kinase 1	WP_001503195	1.59222E-112
212	Failed to assign function	0	18	21	15	0	41	1.07004786129841e-05	0.000858891750002188	hypothetical protein ECO26_0588	YP_003227662	1.31447E-83
213	hypothetical protein	0	18	21	15	0	41	1.07004786129841e-05	0.000858891750002188	hepn domain protein	YP_003227661	7.73976E-91
214	FIG00642656: hypothetical protein	0	18	21	15	1	40	1.07004786129841e-05	0.000858891750002188	type iii restriction res subunit	YP_003231836	0.0
215	Phage protein	0	18	21	15	2	39	1.07004786129841e-05	0.000858891750002188	ead ea22-like family protein	YP_003230651	4.92587E-65
216	Invasin	0	18	21	15	2	39	1.07004786129841e-05	0.000858891750002188	adhesin invasin	WP_001341730	1.0883E-130
217	Mobile element protein	0	18	21	15	4	37	1.07004786129841e-05	0.000858891750002188	transposase	WP_000839180	7.4307E-84
218	Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)	0	18	21	15	4	37	1.07004786129841e-05	0.000858891750002188	type i restriction enzyme r protein	YP_672515	0.0
219	Phage NinC	0	18	21	15	6	35	1.07004786129841e-05	0.000858891750002188	phosphoadenosine phosphosulfate reductase family protein	YP_003230648	0.0
220	Flagellar hook protein FlgE	0	18	21	15	6	35	1.07004786129841e-05	0.000858891750002188	flagellar hook protein	YP_003227357	0.0
221	Maltose/maltodextrin ABC transporter, permease protein MalF	1	17	21	15	6	35	0.000133269597270802	0.00554893124313703	sugar transporter subunit	WP_023981372	0.0
222	dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133)	0	18	21	15	11	30	1.07004786129841e-05	0.000858891750002188	dtdp-4-dehydrorhamnose reductase	YP_003229914	0.0

223	Flagellar biosynthesis protein FliC	0	18	21	15	11	30	1.07004786129841e-05	0.000858891750002188	flagellin	YP_003229794	0.0
224	iron aquisition yersiniabactin synthesis enzyme (Irp3)	0	18	21	15	13	28	1.07004786129841e-05	0.000858891750002188	thiazolinyl-s-hmwp1 reductase	YP_003229841	0.0
225	iron aquisition yersiniabactin synthesis enzyme (YbtT,resembles thioesterases)	0	18	21	15	13	28	1.07004786129841e-05	0.000858891750002188	yersiniabactin thioesterase component	YP_006139150	0.0
226	iron aquisition outermembrane yersiniabactin receptor (FyuA,Psn,pesticin receptor)	0	18	21	15	13	28	1.07004786129841e-05	0.000858891750002188	siderophore receptor protein	WP_009484410	0.0
227	IncI1 plasmid conjugative transfer NusG-type transcription antiterminator TraB	1	17	21	15	16	25	0.000133269597270802	0.00554893124313703	transcription antitermination factor	YP_002756602	8.64568E-152
228	IncI1 plasmid conjugative transfer protein TraC	1	17	21	15	18	23	0.000133269597270802	0.00554893124313703	transfer protein c	WP_001370630	1.29454E-146
229	Uncharacterized protein ImpJ/VasE	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	type vi secretion protein	WP_023563512	1.28988E-104
230	Uncharacterized protein ImpJ/VasE	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	type vi secretion protein	WP_000011867	0.0
231	Type III secretion inner membrane protein (YscQ,homologous to flagellar export components)	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	type iii secretion apparatus family	WP_009008252	1.4739E-28
232	Phage repressor	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	repressor protein ci	YP_001449260	5.93348E-171
233	Failed to assign function	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	cp4-like integrase domain protein	WP_001342206	6.2903E-29
234	Chaperone protein hchA	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	chaperone hcha	WP_000607839	3.16072E-170
235	Inner membrane protein YqiK	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	inner membrane protein yqik	WP_000308451	6.62741E-64
236	Inner membrane protein YqiK	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	spfh domain band 7 family protein	WP_000341041	3.05454E-144
237	Failed to assign function	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	conserved protein	YP_003230166	3.72754E-79
238	Failed to assign function	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	hypothetical protein ECSE_0571	YP_002291846	3.98348E-53
239	Failed to assign function	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	hypothetical protein ECSE_0570	YP_002291845	1.72953E-49
240	Phage Rha protein	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	bro n-terminal domain protein	WP_001161783	3.09158E-140
241	hypothetical protein	0	18	20	16	0	41	2.2738517052591e-05	0.00148654341346532	achain structure from the mobile metagenome of vibrio integron cassette protein vch_cass14	WP_000152742	6.37866E-72
242	Uncharacterized protein ImpA	0	18	20	16	1	40	2.2738517052591e-05	0.00148654341346532	type vi secretion-associated vc_a0119 family	WP_001306910	2.83692E-153
243	Phage terminase, small subunit	0	18	20	16	1	40	2.2738517052591e-05	0.00148654341346532	terminase small subunit	YP_003230299	4.15575E-92
244	Phage Rha protein	0	18	20	16	1	40	2.2738517052591e-05	0.00148654341346532	rha family phage regulatory protein	YP_003230302	4.08234E-104
245	FIG00637907: hypothetical protein	0	18	20	16	1	40	2.2738517052591e-05	0.00148654341346532	outer membrane autotransporter barrel domain protein	WP_000127325	9.14033E-98
246	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	0	18	20	16	2	39	2.2738517052591e-05	0.00148654341346532	adhesin invasin	WP_016243659	0.0
247	Co-activator of prophage gene expression IbrA	0	18	20	16	2	39	2.2738517052591e-05	0.00148654341346532	pp-loop family protein	WP_001535422	1.6638E-64
248	Co-activator of prophage gene expression IbrA	0	18	20	16	3	38	2.2738517052591e-05	0.00148654341346532	immunoglobulin-binding regulator a	WP_001285511	0.0
249	Mobile element protein	0	18	20	16	3	38	2.2738517052591e-05	0.00148654341346532	integrase	YP_003232308	0.0
250	Failed to assign function	0	18	20	16	4	37	2.2738517052591e-05	0.00148654341346532	hypothetical protein EFER_3110	YP_002384207	4.33569E-34
251	Type I restriction-modification system, DNA-methyltransferase subunit M (EC 2.1.1.72)	0	18	20	16	4	37	2.2738517052591e-05	0.00148654341346532	eco57i restriction-modification methylase family protein	WP_001341289	0.0
252	orf, hypothetical protein	0	18	20	16	7	34	2.2738517052591e-05	0.00148654341346532	protein yiba	WP_001326523	2.20834E-45
253	dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13)	0	18	20	16	10	31	2.2738517052591e-05	0.00148654341346532	rmlc	YP_003229912	4.06229E-130
254	iron aquisition 2,3-dihydroxybenzoate-AMP ligase (EC 2.7.7.58,Irp5)	0	18	20	16	12	29	2.2738517052591e-05	0.00148654341346532	(-dihydroxybenzoyl)adenylate synthase	YP_003229843	0.0
255	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	0	18	20	16	12	29	2.2738517052591e-05	0.00148654341346532	adherence and invasion outer-membrane protein (peyer s patches colonization)	WP_001342202	0.0

256	Anthranilate synthase, aminase component (EC 4.1.3.27)	0	18	20	16	13	28	2.2738517052591e-05	0.00148654341346532	salicylate synthase	WP_000703039	0.0
257	AmpG permease	0	18	20	16	13	28	2.2738517052591e-05	0.00148654341346532	major facilitator superfamily protein	YP_003229835	8.87217E-179
258	Inner membrane ABC-transporter YbtQ	0	18	20	16	13	28	2.2738517052591e-05	0.00148654341346532	yersiniabactin-iron abc transporter permease atp-binding protein	YP_003229836	0.0
259	iron aquisition regulator (YbtA,AraC-like,required for transcription of FyuA/psn,Irp2)	0	18	20	16	13	28	2.2738517052591e-05	0.00148654341346532	family transcriptional regulator	YP_003229838	0.0
260	Pilin transcriptional activator	0	18	20	16	16	25	2.2738517052591e-05	0.00148654341346532	histidine kinase	WP_000486074	0.0
261	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	purine permease ybby	WP_023568101	1.58555E-62
262	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	purine permease ybby	WP_001342080	1.8989E-173
263	Uncharacterized protein ImpA	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	impa-related family protein	WP_001240527	6.68917E-137
264	Puative phophotriesterase	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	phosphotriesterase family protein	WP_024198032	3.88192E-166
265	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	hypothetical protein ECO26_3412	YP_003230353	0.0
266	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	conserved protein	YP_003230354	0.0
267	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	hypothetical phage protein	YP_003230352	3.06682E-105
268	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	hypothetical protein	WP_000990407	3.19857E-70
269	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	hypothetical protein ECO26_3409	YP_003230350	0.0
270	Phosphoenolpyruvate-protein phosphotransferase of PTS system (EC 2.7.3.9)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	phosphoenolpyruvate-dependent sugar phosphotransferase eiia 2 family protein	CDK78899	3.03879E-79
271	Phosphoenolpyruvate-protein phosphotransferase of PTS system (EC 2.7.3.9)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	phosphoenolpyruvate-protein phosphotransferase	WP_001185114	0.0
272	hypothetical protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	membrane protein	YP_003230656	0.0
273	Alkanesulfonate utilization operon LysR-family regulator CbI	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	hth-type transcriptional regulator cbl	WP_021500360	9.47806E-63
274	Alkanesulfonate utilization operon LysR-family regulator CbI	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	hth-type transcriptional regulator cbl	WP_000138455	4.32424E-134
275	Putative flagellin structural protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	flagellin structural protein	WP_000556533	2.19878E-45
276	FIG00638425: hypothetical protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	membrane protein	YP_003231669	1.54579E-45
277	FIG00638425: hypothetical protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	membrane protein	WP_024173847	6.44218E-50
278	FIG00639456: hypothetical protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	fimbrial family protein	WP_021543399	4.40843E-99
279	FIG00639456: hypothetical protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	fimbrial family protein	WP_001269786	1.42854E-102
280	Alpha-fimbriae tip adhesin	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	cbld like pilus biogenesis initiator family protein	WP_001352530	0.0
281	SanA protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	thioredoxin-like protein	WP_024261644	5.50674E-85
282	SanA protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	thioredoxin-like protein	WP_001582007	3.6272E-72
283	Putative isomerase	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	glycosyl hydrolase domain protein	WP_000899672	0.0
284	Putative isomerase	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	glycosyl hydrolase	WP_000695499	0.0
285	L-rhamnose operon regulatory protein RhaS	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	cupin domain protein	WP_000217145	3.03906E-70
286	Fimbrial protein YadC	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	fimbrial protein	WP_024246517	1.04169E-121
287	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	family transcriptional regulator	WP_024182968	8.37809E-140
288	Transcriptional regulator, GntR family domain / Aspartate aminotransferase (EC 2.6.1.1)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	bacterial regulatory gntr family protein	WP_001341356	0.0
289	Putative uncharacterized protein YhcG	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	pf06250 domain protein	WP_001341898	5.98562E-75
290	Autoinducer 2 (AI-2) kinase LsrK (EC 2.7.1.-)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	autoinducer 2 kinase	WP_001341534	0.0
291	Autoinducer 2 (AI-2) kinase LsrK (EC 2.7.1.-)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	autoinducer-2 (ai-2) kinase	EIQ75546	4.6761E-140
292	Trans-aconitate 2-methyltransferase (EC 2.1.1.144)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	trans-aconitate 2-methyltransferase	WP_001286590	1.82577E-94

293	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	t3ss secreted effector -like protein	WP_001448330	3.08291E-146
294	PTS system, sorbose-specific IID component (EC 2.7.1.69)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	pts system mannose-specific transporter subunit iid	WP_001341741	8.46793E-100
295	PTS system, sorbose-specific IID component (EC 2.7.1.69)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	mannose-specific enzyme iid component for pts	WP_000437911	6.41375E-57
296	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	galactitol-1-phosphate 5-dehydrogenase	WP_021541923	1.81986E-58
297	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	galactitol-1-phosphate 5-dehydrogenase	WP_001512915	3.83667E-155
298	Mobile element protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	transposase	WP_001115254	2.4434E-141
299	Failed to assign function	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	hypothetical protein ECO26_4939	YP_003231835	1.90474E-86
300	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	type i restriction enzyme specificity protein (s protein)	YP_003232400	0.0
301	Putative flagellin structural protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	flagellin structural partial	WP_001427953	1.33312E-129
302	hypothetical protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	type i restriction enzyme family protein	WP_000839828	0.0
303	hypothetical protein	0	18	19	17	0	41	4.68145939318053e-05	0.00232432045748015	membrane protein	YP_003230345	0.0
304	COG3547: Transposase and inactivated derivatives	0	18	19	17	1	40	4.68145939318053e-05	0.00232432045748015	transposase is116 is110 is902 family protein	YP_003229857	0.0
305	L-rhamnose operon regulatory protein RhaS	0	18	19	17	1	40	4.68145939318053e-05	0.00232432045748015	cupin domain protein	WP_001341797	7.82368E-108
306	Propanediol utilization protein PduV	0	18	19	17	1	40	4.68145939318053e-05	0.00232432045748015	ethanolamine utilization-propanediol utilization	WP_001282181	1.95331E-55
307	hypothetical protein; Unknown protein. Putative secreted protein	0	18	19	17	1	40	4.68145939318053e-05	0.00232432045748015	membrane protein	YP_003232403	1.98208E-177
308	Putative flagellin structural protein	0	18	19	17	1	40	4.68145939318053e-05	0.00232432045748015	outer membrane autotransporter barrel domain protein	WP_001444349	0.0
309	Failed to assign function	0	18	19	17	2	39	4.68145939318053e-05	0.00232432045748015	lipoprotein	WP_000749284	2.96746E-85
310	Phage capsid and scaffold	0	18	19	17	2	39	4.68145939318053e-05	0.00232432045748015	head assembly protein	YP_003230290	3.48517E-110
311	Haemolysin expression modulating protein paralog	0	18	19	17	2	39	4.68145939318053e-05	0.00232432045748015	hemolysin expression modulating protein hha	WP_021542549	1.15506E-29
312	Fimbrial protein YadC	0	18	19	17	2	39	4.68145939318053e-05	0.00232432045748015	fimbrial family protein	WP_000713127	3.98287E-96
313	FIG00638152: hypothetical protein	0	18	19	17	2	39	4.68145939318053e-05	0.00232432045748015	pf03235 family protein	WP_000648228	0.0
314	FIG00639817: hypothetical protein	0	18	19	17	2	39	4.68145939318053e-05	0.00232432045748015	toxin-antitoxin antitoxin ribbon-helix-helix domain protein	YP_003232404	7.62172E-35
315	unnamed protein product	0	18	19	17	2	39	4.68145939318053e-05	0.00232432045748015	cytochrome b562 family protein	YP_325578	2.18237E-69
316	Anaerobic dimethyl sulfoxide reductase chain A (EC 1.8.99.-)	0	18	19	17	3	38	4.68145939318053e-05	0.00232432045748015	anaerobic dimethyl sulfoxide a family	WP_001375676	0.0
317	DNA stabilization, phage-associated	0	18	19	17	3	38	4.68145939318053e-05	0.00232432045748015	packaged dna stabilization protein p27	NP_958183	3.2184E-106
318	unknown	0	18	19	17	3	38	4.68145939318053e-05	0.00232432045748015	phage protein	YP_003230294	2.297E-29
319	Phage major capsid protein	0	18	19	17	3	38	4.68145939318053e-05	0.00232432045748015	major head protein	YP_003230295	0.0
320	Phage capsid and scaffold	0	18	19	17	3	38	4.68145939318053e-05	0.00232432045748015	scaffold protein gp8	YP_003230296	4.24965E-134
321	Inner membrane protein YjgN	0	18	19	17	4	37	4.68145939318053e-05	0.00232432045748015	inner membrane protein yjgn	WP_000079609	1.72952E-81
322	Inner membrane protein YjgN	0	18	19	17	4	37	4.68145939318053e-05	0.00232432045748015	inner membrane protein yjgn	WP_024171715	3.6781E-143
323	Failed to assign function	0	18	19	17	8	33	4.68145939318053e-05	0.00232432045748015	pf04447 domain protein	WP_001289868	2.36348E-115
324	ORF25	0	18	19	17	8	33	4.68145939318053e-05	0.00232432045748015	hnh endonuclease family protein	YP_003232367	1.53611E-157
325	Putative ABC iron siderophore transporter, fused permease and ATPase domains	0	18	19	17	12	29	4.68145939318053e-05	0.00232432045748015	yersiniabactin abc transporter atp-binding protein permease	YP_003229837	0.0
326	FIG01221203: hypothetical protein	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	rha family transcriptional regulator	YP_003230875	3.26467E-172

327	2-deoxy-D-gluconate 3-dehydrogenase (EC 1.1.1.125)	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	short chain dehydrogenase reductase family protein	YP_002294304	1.12132E-168
328	Failed to assign function	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	hypothetical protein ECO26_3408	YP_003230349	9.34397E-163
329	Failed to assign function	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	his glu gln arg opine family amino abc 3-tm region	WP_001391715	9.77461E-72
330	putative rhamnosyl transferase	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	wbua	YP_003229909	1.14378E-159
331	Failed to assign function	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	wzy	YP_003229910	1.06795E-168
332	Failed to assign function	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	wzx	YP_003229911	0.0
333	FIG01069761: hypothetical protein	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	diguanylate cyclase domain protein	WP_023910215	0.0
334	FIG00639031: hypothetical protein	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	diguanylate cyclase domain protein	WP_001494277	0.0
335	FIG00637952: hypothetical protein	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	#NAME?	WP_001598225	4.79055E-143
336	metal-dependent phosphohydrolase	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	phosphohydrolase	WP_001609122	7.57937E-88
337	Failed to assign function	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	is3 orf2 protein	WP_001361539	1.70089E-79
338	Phage integrase	0	18	18	18	0	41	9.362918786361e-05	0.00399042627213403	integrase	WP_001131470	5.92877E-114
339	hypothetical protein	0	18	18	18	1	40	9.362918786361e-05	0.00399042627213403	hypothetical protein	WP_000852869	6.24601E-151
340	Failed to assign function	0	18	18	18	1	40	9.362918786361e-05	0.00399042627213403	hypothetical protein	WP_000566882	1.48445E-21
341	FIG00640945: hypothetical protein	0	18	18	18	1	40	9.362918786361e-05	0.00399042627213403	hypothetical protein	WP_001014980	2.68388E-40
342	Failed to assign function	0	18	18	18	1	40	9.362918786361e-05	0.00399042627213403	major pilu subunit operon regulatory protein domain protein	WP_000160236	5.80699E-27
343	FIG107037: Phage late gene regulator	0	18	18	18	2	39	9.362918786361e-05	0.00399042627213403	ogr delta-like zinc finger family protein	YP_003232314	9.00923E-47
344	possible integrase	0	18	18	18	3	38	9.362918786361e-05	0.00399042627213403	integrase	YP_003230878	0.0
345	Failed to assign function	0	18	18	18	5	36	9.362918786361e-05	0.00399042627213403	hypothetical protein	WP_000128178	3.24209E-26
346	FIG00640454: hypothetical protein	0	18	18	18	9	32	9.362918786361e-05	0.00399042627213403	esterase-like activity of phytase family protein	YP_002394351	0.0
347	hypothetical protein	0	18	18	18	10	31	9.362918786361e-05	0.00399042627213403	biofilm development family protein	YP_003047326	9.46629E-41
348	iron aquisition yersiniabactin synthesis enzyme (Irp2)	0	18	18	18	11	30	9.362918786361e-05	0.00399042627213403	siderophore biosynthetic protein	YP_003229839	0.0
349	Phage EaA protein	0	18	17	19	0	41	0.000182330523734399	0.00713905530329158	eaa protein	YP_003227898	2.5784E-129
350	Phage tail fibers	0	18	17	19	0	41	0.000182330523734399	0.00713905530329158	tailspike protein	YP_003230284	0.0
351	Failed to assign function	0	18	17	19	0	41	0.000182330523734399	0.00713905530329158	acyltransferase	WP_001375798	0.0
352	FIG00637874: hypothetical protein	0	18	17	19	0	41	0.000182330523734399	0.00713905530329158	diguanylate cyclase	WP_000592825	4.01084E-148
353	Hypothetical lipoprotein yehR	0	18	17	19	0	41	0.000182330523734399	0.00713905530329158	lipoprotein yehr precursor	2JOE_A	6.92174E-49
354	FIG00637907: hypothetical protein	0	18	17	19	0	41	0.000182330523734399	0.00713905530329158	outer membrane autotransporter barrel domain-containing protein	WP_000168143	2.69205E-166
355	Failed to assign function	0	18	17	19	0	41	0.000182330523734399	0.00713905530329158	mannitol operon repressor domain protein	YP_003232309	1.96422E-148
356	Phage integrase	0	18	17	19	0	41	0.000182330523734399	0.00713905530329158	integrase	WP_001131472	1.08419E-141
357	Putative uncharacterized protein YhcG	0	18	17	19	1	40	0.000182330523734399	0.00713905530329158	pf06250 family protein	WP_001521407	3.01172E-118
358	FIG118045: Phage immunity repressor protein	0	18	17	19	1	40	0.000182330523734399	0.00713905530329158	repressor protein	WP_021546757	1.49179E-126
359	FIG070121: Phage capsid and scaffold protein	0	18	17	19	1	40	0.000182330523734399	0.00713905530329158	head size determination protein	YP_002043731	8.2019E-156
360	FIG054316: Phage polarity suppression protein	0	18	17	19	1	40	0.000182330523734399	0.00713905530329158	polarity suppression protein	YP_003232315	2.09614E-116
361	Phage portal	0	18	17	19	3	38	0.000182330523734399	0.00713905530329158	portal protein	WP_001432480	0.0
362	Catalase (EC 1.11.1.6) / Peroxidase (EC 1.11.1.7)	0	18	17	19	3	38	0.000182330523734399	0.00713905530329158	catalase peroxidase hpi	YP_325577	0.0
363	FIG01069755: hypothetical protein	0	18	17	19	4	37	0.000182330523734399	0.00713905530329158	hypothetical protein ECO26_3945	YP_003230870	0.0
364	Failed to assign function	0	18	17	19	7	34	0.000182330523734399	0.00713905530329158	hypothetical protein	WP_001332869	1.72106E-12
365	iron aquisition yersiniabactin synthesis enzyme (Irp1,polyketide synthetase)	0	18	17	19	12	29	0.000182330523734399	0.00713905530329158	polyketide synthase	YP_003229840	0.0

Table S5. Analysis of accessory genome in STEC in HUS-group 2 (n=23, G1) compared to other LEE positive positive STEC (n=31, G2) and LEE negative STEC (n=41, G3)
List of overrepresented genes in STEC in HUS-group 2

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR<0.01	Function Blast2GO	Hit Accession	E-value
1	Putative sucrose phosphorylase (EC 2.4.1.7)	23	0	0	31	0	41	9.20869683596171e-16	8.86981679239832e-13	sucrose phosphorylase	WP_023143154	8.99427E-98
2	Acetyl-CoA:acetoacetyl-CoA transferase, alpha subunit (EC 2.8.3.8)	23	0	0	31	0	41	9.20869683596171e-16	8.86981679239832e-13	coenzyme a transferase family protein	WP_001342179	0.0
3	Failed to assign function	23	0	0	31	0	41	9.20869683596171e-16	8.86981679239832e-13	acetyl- :acetoacetyl- transferase alpha beta subunit	WP_024237352	1.46471E-163
4	Lactaldehyde dehydrogenase involved in fucose or rhamnose utilization (EC 1.2.1.22)	23	0	0	31	0	41	9.20869683596171e-16	8.86981679239832e-13	alcohol dehydrogenase	YP_003231587	0.0
5	Putative sucrose phosphorylase (EC 2.4.1.7)	22	1	0	31	0	41	2.94678298750776e-14	1.41917068678374e-11	sucrose phosphorylase	WP_000459435	0.0
6	Failed to assign function	21	2	0	31	0	41	4.86219192938783e-13	1.35708196923694e-10	hypothetical protein ECO26_0588	YP_003227662	1.31447E-83
7	hypothetical protein	21	2	0	31	0	41	4.86219192938783e-13	1.35708196923694e-10	hepn domain protein	YP_003227661	7.73976E-91
8	Inner membrane protein YqiK	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	spfh domain band 7 family protein	WP_000341041	3.05454E-144
9	Inner membrane protein YqiK	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	inner membrane protein yqik	WP_000308451	6.62741E-64
10	Failed to assign function	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	cp4-like integrase domain protein	WP_001342206	6.2903E-29
11	Chaperone protein hchA	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	chaperone hcha	WP_000607839	3.16072E-170
12	Failed to assign function	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	hypothetical protein ECSE_0570	YP_002291845	1.72953E-49
13	Failed to assign function	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	hypothetical protein ECSE_0571	YP_002291846	3.98348E-53
14	Phage Rha protein	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	bro n-terminal domain protein	WP_001161783	3.09158E-140
15	hypothetical protein	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	achain structure from the mobile metagenome of vibrio integron cassette protein vch_cass14	WP_000152742	6.37866E-72
16	Failed to assign function	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	conserved protein	YP_003230166	3.72754E-79
17	Uncharacterized protein ImpJ/VasE	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	type vi secretion protein	WP_000011867	0.0
18	Uncharacterized protein ImpJ/VasE	20	3	0	31	0	41	5.51048418663954e-12	9.25761343355442e-10	type vi secretion protein	WP_023563512	1.28988E-104
19	Trans-aconitate 2-methyltransferase (EC 2.1.1.144)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	trans-aconitate 2-methyltransferase	WP_001286590	1.82577E-94
20	Autoinducer 2 (AI-2) kinase LsrK (EC 2.7.1.-)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	autoinducer-2 (ai-2) kinase	EIQ75546	4.6761E-140
21	Autoinducer 2 (AI-2) kinase LsrK (EC 2.7.1.-)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	autoinducer 2 kinase	WP_001341534	0.0
22	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	t3ss secreted effector -like protein	WP_001448330	3.08291E-146
23	Putative flagellin structural protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	flagellin structural domain protein	WP_001427953	5.63888E-111
24	Putative flagellin structural protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	flagellin structural protein	WP_000556533	2.19878E-45
25	Phosphoenolpyruvate-protein phosphotransferase of PTS system (EC 2.7.3.9)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	phosphoenolpyruvate-protein phosphotransferase	WP_001185114	0.0
26	Phosphoenolpyruvate-protein phosphotransferase of PTS system (EC 2.7.3.9)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	phosphoenolpyruvate-dependent sugar phosphotransferase eiia 2 family protein	CDK78899	3.03879E-79
27	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	hypothetical protein ECO26_3409	YP_003230350	0.0
28	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	hypothetical protein	WP_000990407	3.19857E-70
29	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	hypothetical phage protein	YP_003230352	3.06682E-105
30	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	hypothetical protein ECO26_3412	YP_003230353	0.0
31	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	conserved protein	YP_003230354	0.0
32	FIG00638425: hypothetical protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	membrane protein	YP_003231669	1.54579E-45
33	FIG00638425: hypothetical protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	membrane protein	WP_024173847	6.44218E-50
34	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	hypothetical protein ECO26_4939	YP_003231835	1.86173E-87
35	Alpha-fimbriae tip adhesin	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	cbld like pilus biogenesis initiator family protein	WP_001352530	0.0
36	FIG00639456: hypothetical protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	fimbrial family protein	WP_001269786	1.42854E-102
37	FIG00639456: hypothetical protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	fimbrial family protein	WP_021543399	4.40843E-99

38	Putative isomerase	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	glycosyl hydrolase	WP_000695499	0.0
39	Putative isomerase	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	glycosyl hydrolase domain protein	WP_000899672	0.0
40	SanA protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	thioredoxin-like protein	WP_001582007	3.6272E-72
41	SanA protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	thioredoxin-like protein	WP_024261644	5.50674E-85
42	Alkanesulfonate utilization operon LysR-family regulator CbI	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	hth-type transcriptional regulator cbl	WP_021500360	9.47806E-63
43	Alkanesulfonate utilization operon LysR-family regulator CbI	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	hth-type transcriptional regulator cbl	WP_000138455	4.32424E-134
44	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	galactitol-1-phosphate 5-dehydrogenase	WP_001512915	3.83667E-155
45	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	galactitol-1-phosphate 5-dehydrogenase	WP_021541923	1.81986E-58
46	Mobile element protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	transposase	WP_001115254	2.4434E-141
47	Puative phophotriesterase	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	phosphotriesterase family protein	WP_024198032	3.88192E-166
48	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	type i restriction enzyme specificity protein (s protein)	YP_003232400	0.0
49	L-rhamnose operon regulatory protein RhaS	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	cupin domain protein	WP_000217145	3.03906E-70
50	Fimbrial protein YadC	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	fimbrial protein	WP_024246517	1.04169E-121
51	PTS system, sorbose-specific IID component (EC 2.7.1.69)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	mannose-specific enzyme iid component for pts	WP_000437911	6.41375E-57
52	PTS system, sorbose-specific IID component (EC 2.7.1.69)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	pts system mannose-specific transporter sub	WP_001341741	8.46793E-100
53	hypothetical protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	type i restriction enzyme family protein	WP_000839828	0.0
54	hypothetical protein	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	membrane protein	YP_003230345	0.0
55	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	family transcriptional regulator	WP_024182968	8.37809E-140
56	Transcriptional regulator, GntR family domain / Aspartate aminotransferase (EC 2.6.1.1)	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	bacterial regulatory gntr family protein	WP_001341356	0.0
57	Uncharacterized protein ImpA	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	impa-related family protein	WP_001240527	6.68917E-137
58	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	purine permease ybby	WP_001342080	1.8989E-173
59	Failed to assign function	19	4	0	31	0	41	4.82167366330959e-11	4.70699601942547e-09	purine permease ybby	WP_023568101	1.58555E-62
60	2-deoxy-D-gluconate 3-dehydrogenase (EC 1.1.1.125)	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	short chain dehydrogenase reductase family protein	YP_002294304	1.12132E-168
61	Failed to assign function	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	hypothetical protein ECO26_3408	YP_003230349	9.34397E-163
62	Failed to assign function	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	his glu gln arg opine family amino abc 3-tm region	WP_001391715	9.77461E-72
63	FIG01221203: hypothetical protein	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	rha family transcriptional regulator	YP_003230875	3.26467E-172
64	Failed to assign function	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	wzx	YP_003229911	0.0
65	Failed to assign function	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	wzy	YP_003229910	1.06795E-168
66	putative rhamnosyl transferase	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	wbua	YP_003229909	1.14378E-159
67	FIG00639031: hypothetical protein	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	diguanylate cyclase domain protein	WP_001494277	0.0
68	FIG01069761: hypothetical protein	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	diguanylate cyclase domain protein	WP_023910215	0.0
69	FIG00637952: hypothetical protein	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	#NAME?	WP_001598225	4.79055E-143
70	metal-dependent phosphohydrolase	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	phosphohydrolase	WP_001609122	7.57937E-88
71	Failed to assign function	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	is3 orf2 protein	WP_001361539	1.70089E-79
72	Phage integrase	18	5	0	31	0	41	3.47160503758288e-10	2.80210891525126e-08	integrase	WP_001131470	5.92877E-114
73	FIG00637907: hypothetical protein	17	6	0	31	0	41	2.14082310650945e-09	1.41235672341774e-07	outer membrane autotransporter barrel domain-containing protein	WP_000168143	2.69205E-166
74	Hypothetical lipoprotein yehR	17	6	0	31	0	41	2.14082310650945e-09	1.41235672341774e-07	lipoprotein yehr precursor	2JOE_A	6.92174E-49
75	FIG00637874: hypothetical protein	17	6	0	31	0	41	2.14082310650945e-09	1.41235672341774e-07	diguanylate cyclase	WP_000592825	4.01084E-148

76	Failed to assign function	17	6	0	31	0	41	2.14082310650945e-09	1.41235672341774e-07	mannitol operon repressor domain protein	YP_003232309	1.96422E-148
77	Phage integrase	17	6	0	31	0	41	2.14082310650945e-09	1.41235672341774e-07	integrase	WP_001131472	1.08419E-141
78	Failed to assign function	17	6	0	31	0	41	2.14082310650945e-09	1.41235672341774e-07	acyltransferase	WP_001375798	0.0
79	Phage tail fibers	17	6	0	31	0	41	2.14082310650945e-09	1.41235672341774e-07	tailspike protein	YP_003230284	0.0
80	FIG00639653: hypothetical protein	16	7	0	31	0	41	1.16216111496227e-08	6.8815179463012e-07	nlea11 protein	WP_024230391	5.5249E-56
81	hypothetical protein	16	7	0	31	0	41	1.16216111496227e-08	6.8815179463012e-07	hypothetical protein ECO26_5437	YP_003232310	0.0
82	generated by GeneMarkS	14	9	0	31	0	41	2.51801574908491e-07	1.17008764408956e-05	eaa protein	YP_003230335	2.38587E-137
83	Failed to assign function	14	9	0	31	0	41	2.51801574908491e-07	1.17008764408956e-05	hypothetical protein	WP_000060377	1.38446E-35
84	Phage protein	14	9	0	31	0	41	2.51801574908491e-07	1.17008764408956e-05	conserved protein	YP_003230333	9.75167E-135
85	FIG00639362: hypothetical protein	13	10	0	31	0	41	1.03238645712482e-06	4.18986503723017e-05	hypothetical protein	WP_000789024	2.47194E-92
86	Failed to assign function	13	10	0	31	0	41	1.03238645712482e-06	4.18986503723017e-05	hypothetical protein	WP_001299314	3.66383E-12
87	FIG00638514: hypothetical protein	13	10	0	31	0	41	1.03238645712482e-06	4.18986503723017e-05	hypothetical protein	WP_000020598	1.88688E-23
88	Failed to assign function	13	10	0	31	0	41	1.03238645712482e-06	4.18986503723017e-05	hypothetical protein ECO26_3377	YP_003230324	2.68979E-116
89	Failed to assign function	13	10	0	31	0	41	1.03238645712482e-06	4.18986503723017e-05	hypothetical protein ECO26_3374	YP_003230321	1.44618E-18
90	Uncharacterized protein ImpI/VasC	13	10	0	31	0	41	1.03238645712482e-06	4.18986503723017e-05	type vi secretion system fha domain-containing protein	YP_002401356	1.52159E-143
91	FIG00639653: hypothetical protein	12	11	0	31	0	41	3.94183919993113e-06	0.000142736072081717	nlea11 protein	WP_001342109	0.0
92	putative cytoplasmic protein	12	11	0	31	0	41	3.94183919993113e-06	0.000142736072081717	cytoplasmic protein	WP_001317900	0.0
93	Mobile element protein	12	11	0	31	0	41	3.94183919993113e-06	0.000142736072081717	transposase family protein	WP_001424809	2.56531E-161
94	Hyphotheical protein	11	12	0	31	0	41	1.41249237997532e-05	0.000445543757564593	malate partial	WP_009425176	0.0
95	Exodeoxyribonuclease VIII (EC 3.1.11.-)	10	13	0	31	0	41	4.78074343991646e-05	0.00138699159076131	exodeoxyribonuclease 8	WP_001431973	0.0
96	Soluble cytochrome b562	10	13	0	31	0	41	4.78074343991646e-05	0.00138699159076131	soluble cytochrome b562	YP_003232282	3.00618E-61
97	His-Xaa-Ser repeat protein	9	14	0	31	0	41	0.000153666753425886	0.00405882496068959	his-xaa-ser repeat protein	YP_003229465	1.97632E-87
98	His-Xaa-Ser system radical SAM maturase HxsC	9	14	0	31	0	41	0.000153666753425886	0.00405882496068959	radical sam superfamily protein	YP_003229466	0.0
99	His-Xaa-Ser system radical SAM maturase HxsB	9	14	0	31	0	41	0.000153666753425886	0.00405882496068959	radical sam superfamily protein	YP_003229467	0.0
100	Failed to assign function	9	14	0	31	0	41	0.000153666753425886	0.00405882496068959	his-xaa-ser system protein	WP_000004182	1.82473E-49
101	Hyphotheical protein	9	14	0	31	0	41	0.000153666753425886	0.00405882496068959	malate partial	WP_021518061	4.49567E-95
102	hypothetical protein	20	3	1	30	0	41	1.05995784060654e-10	9.6316169063417e-09	transcriptional regulator	YP_003227393	2.23311E-142
103	Type III secretion inner membrane protein (YscQ,homologous to flagellar export components)	19	4	1	30	0	41	8.59635533115762e-10	6.33674192982476e-08	type iii secretion apparatus family	WP_009008252	1.4739E-28
104	Putative uncharacterized protein YhcG	18	5	1	30	0	41	5.72814831201178e-09	3.64582761286106e-07	pf06250 domain protein	WP_001341898	5.98562E-75
105	Failed to assign function	16	7	1	30	0	41	1.63314219839435e-07	7.9179994907388e-06	repressor protein c2	WP_024187092	4.65627E-82
106	Resolvase	10	13	1	30	0	41	0.000384632540393279	0.00940812693711364	site-specific recombinase	ETJ70385	3.13786E-69
107	Phage integrase	20	3	2	29	0	41	1.06517364649443e-09	7.77253982048055e-08	integrase	WP_001189092	0.0
108	Phage repressor	18	5	2	29	0	41	4.95447633220757e-08	2.64141232648469e-06	repressor protein ci	YP_001449260	5.93348E-171
109	Failed to assign function	18	5	2	29	0	41	4.95447633220757e-08	2.64141232648469e-06	excisionase	YP_005727532	3.52145E-58
110	hypothetical protein	17	6	2	29	0	41	2.61325069204052e-07	1.17008764408956e-05	membrane protein	YP_003230656	0.0
111	Phage EaA protein	14	9	2	29	0	41	1.87456587843415e-05	0.000584650628573655	phage protein	YP_003230334	1.04363E-115
112	Failed to assign function	13	10	2	29	0	41	6.44549081372011e-05	0.00185506875053044	hypothetical protein	WP_000264368	3.22187E-63
113	Eaa protein	12	11	2	29	0	41	0.000204906965239626	0.00520298037571549	conserved hypothetical plasmid protein	WP_001427476	3.76254E-80
114	Failed to assign function	12	11	2	29	0	41	0.000204906965239626	0.00520298037571549	ygaa protein	WP_001495195	5.75066E-34
115	Mobile element protein	12	11	2	29	0	41	0.000204906965239626	0.00520298037571549	transposase	YP_002756741	3.7107E-43
116	FIG00640314: hypothetical protein	12	11	2	29	0	41	0.000204906965239626	0.00520298037571549	conjugal transfer protein	WP_001073806	2.56893E-55
117	FIG00241420: hypothetical protein	12	11	2	29	0	41	0.000204906965239626	0.00520298037571549	abc transporter nucleotide-binding protein	YP_002296005	1.8393E-125
118	FIG00640631: hypothetical protein	20	3	3	28	0	41	7.43970985725224e-09	4.6532003470814e-07	secretion protein	YP_003232130	1.16701E-121
119	Molybdate metabolism regulator	19	4	4	27	0	41	2.64014792340253e-07	1.17008764408956e-05	wgr domain protein	WP_000680162	0.0

120	Failed to assign function	19	4	4	27	0	41	2.64014792340253e-07	1.17008764408956e-05	antirepressor protein cro	YP_001449261	4.71416E-36
121	FIG00643638: hypothetical protein	19	4	4	27	0	41	2.64014792340253e-07	1.17008764408956e-05	hypothetical protein ECO26_0305	YP_003227392	5.77176E-22
122	FIG00642343: hypothetical protein	14	9	4	27	0	41	0.000289229979279216	0.00719241779797954	conserved protein	ETJ57613	8.33468E-144
123	other or unknown (Phage or Prophage Related)	19	4	8	23	0	41	3.83688648588194e-05	0.00112217279247009	non-lee-encoded type iii secreted effector	YP_003228678	0.0
124	Mobile element protein	19	4	9	22	0	41	0.000102892392928388	0.00279434077637095	mutator family	WP_000343688	6.33332E-52
125	Failed to assign function	22	1	12	19	0	41	1.03627109799984e-05	0.000338734045789632	hypothetical protein Z6012	NP_287952	2.27206E-47
126	FIG00641471: hypothetical protein	21	2	13	18	0	41	0.00017272761215618	0.00450463635457127	t3ss effector protein 8	YP_003228845	9.58028E-151
127	FIG00637907: hypothetical protein	20	3	0	31	1	40	5.51048418663954e-12	9.25761343355442e-10	outer membrane autotransporter barrel dom	WP_000127325	9.14033E-98
128	Putative flagellin structural protein	19	4	0	31	1	40	4.82167366330959e-11	4.70699601942547e-09	outer membrane autotransporter barrel dom	WP_001444349	0.0
129	hypothetical protein; Unknown protein. Putative secreted protein	19	4	0	31	1	40	4.82167366330959e-11	4.70699601942547e-09	membrane protein	WP_000432642	1.21625E-178
130	L-rhamnose operon regulatory protein RhaS	19	4	0	31	1	40	4.82167366330959e-11	4.70699601942547e-09	cupin domain protein	WP_001341797	7.82368E-108
131	COG3547: Transposase and inactivated derivatives	19	4	0	31	1	40	4.82167366330959e-11	4.70699601942547e-09	transposase is116 is110 is902 family protein	YP_003229857	0.0
132	hypothetical protein	18	5	0	31	1	40	3.47160503758288e-10	2.80210891525126e-08	hypothetical protein	WP_000852869	6.24601E-151
133	Failed to assign function	18	5	0	31	1	40	3.47160503758288e-10	2.80210891525126e-08	hypothetical protein	WP_000566882	1.48445E-21
134	Failed to assign function	18	5	0	31	1	40	3.47160503758288e-10	2.80210891525126e-08	major pilu subunit operon regulatory protein domain protein	WP_000160236	5.80699E-27
135	FIG118045: Phage immunity repressor protein	17	6	0	31	1	40	2.14082310650945e-09	1.41235672341774e-07	repressor protein	WP_021546757	1.49179E-126
136	FIG070121: Phage capsid and scaffold protein	17	6	0	31	1	40	2.14082310650945e-09	1.41235672341774e-07	head size determination protein	YP_002043731	8.2019E-156
137	FIG054316: Phage polarity suppression protein	17	6	0	31	1	40	2.14082310650945e-09	1.41235672341774e-07	polarity suppression protein	YP_003232315	2.09614E-116
138	Mobile element protein	14	9	0	31	1	40	2.51801574908491e-07	1.17008764408956e-05	transposase insf for insertion sequence is3a	WP_021548339	2.15822E-126
139	FIG00638514: hypothetical protein	13	10	0	31	1	40	1.03238645712482e-06	4.18986503723017e-05	hypothetical protein	WP_001429765	9.07837E-106
140	Failed to assign function	13	10	0	31	1	40	1.03238645712482e-06	4.18986503723017e-05	hypothetical protein ECO26_2448	YP_003229433	1.71889E-33
141	Phage tail length tape-measure protein 1	9	14	0	31	1	40	0.000153666753425886	0.00405882496068959	tail length tape measure protein	WP_001431289	0.0
142	Type III secretion inner membrane protein (YscU,SpaS,EscU,HrcU,SsaU, homologous to flagellar export components)	23	0	1	30	1	40	2.21008724063083e-14	1.27725361810537e-11	family type iii secretion protein	WP_001613736	4.01737E-130
143	Uncharacterized protein ImpA	19	4	1	30	1	40	8.59635533115762e-10	6.33674192982476e-08	type vi secretion-associated vc_a0119 family	WP_001306910	2.83692E-153
144	Propanediol utilization protein PduV	18	5	1	30	1	40	5.72814831201178e-09	3.64582761286106e-07	ethanolamine utilization- propanediol utilization	WP_001282181	1.95331E-55
145	FIG00640945: hypothetical protein	17	6	1	30	1	40	3.26330873532791e-08	1.85623167748101e-06	hypothetical protein	WP_001014980	2.68388E-40
146	Putative uncharacterized protein YhcG	16	7	1	30	1	40	1.63314219839435e-07	7.9179994907388e-06	pf06250 family protein	WP_001521407	3.01172E-118
147	FIG00642656: hypothetical protein	19	4	2	29	1	40	8.01921314913931e-09	4.95135006746858e-07	type iii restriction res subunit	YP_003231836	0.0
148	Mobile element protein	18	5	2	29	1	40	4.95447633220757e-08	2.64141232648469e-06	transposase-like protein	WP_024239172	1.97444E-55
149	Phage DNA transfer protein	17	6	2	29	1	40	2.61325069204052e-07	1.17008764408956e-05	dna injection protein	WP_000246937	0.0
150	Mobile element protein	15	8	2	29	1	40	4.99851612182719e-06	0.000176142831532096	transposase is66 family protein	WP_001680141	7.47611E-99
151	Mobile element protein	15	8	2	29	1	40	4.99851612182719e-06	0.000176142831532096	transposase is66 family partial	WP_001568115	1.01426E-134
152	FIG00641170: hypothetical protein	20	3	3	28	1	40	7.43970985725224e-09	4.6532003470814e-07	conserved protein	YP_003227398	5.61102E-73
153	Phage terminase, small subunit	17	6	3	28	1	40	1.46141346920191e-06	5.51292475274914e-05	terminase small subunit	YP_003230299	4.15575E-92
154	Phage Rha protein	17	6	3	28	1	40	1.46141346920191e-06	5.51292475274914e-05	rha family phage regulatory protein	YP_003230302	4.08234E-104
155	Phage DNA transfer protein	16	7	3	28	1	40	6.26176706919338e-06	0.000211873561160904	dna transfer protein gp16	WP_000978776	0.0
156	Superinfection exclusion protein (Protein gp17)	14	9	3	28	1	40	8.28989912094081e-05	0.00227704301329568	superinfection exclusion protein	WP_001073248	1.28822E-66
157	FIG00637920: hypothetical protein	23	0	4	27	1	40	1.61612629471128e-11	2.43227007354048e-09	rhum domain protein	WP_023143361	6.12573E-23
158	FIG00642654: hypothetical protein	20	3	4	27	1	40	4.05461721131233e-08	2.19404904378429e-06	protein g7	WP_000898165	3.26281E-33
159	Failed to assign function	20	3	4	27	1	40	4.05461721131233e-08	2.19404904378429e-06	conserved protein	WP_000909240	2.29793E-66

160	hypothetical protein	20	3	4	27	1	40	4.05461721131233e-08	2.19404904378429e-06	conserved protein	YP_003227395	5.50892E-93
161	Failed to assign function	20	3	4	27	1	40	4.05461721131233e-08	2.19404904378429e-06	conserved protein	YP_003227394	1.20896E-58
162	FIG00639751: hypothetical protein	19	4	4	27	1	40	2.64014792340253e-07	1.17008764408956e-05	phage major capsid e family protein	YP_003227399	0.0
163	FIG00640415: hypothetical protein	19	4	4	27	1	40	2.64014792340253e-07	1.17008764408956e-05	dctp deaminase	NP_308270	3.0604E-24
164	Molybdate metabolism regulator	18	5	4	27	1	40	1.40856276898182e-06	5.41247736336417e-05	twice split molybdate metabolism regulator	WP_024228772	0.0
165	Failed to assign function	17	6	4	27	1	40	6.40295393978137e-06	0.0002146400893781	hypothetical protein ECO26_0313	YP_003227400	3.02349E-36
166	FIG00639888: hypothetical protein	19	4	6	25	1	40	4.05702217299542e-06	0.000144374030432113	conserved protein	YP_003222266	7.52598E-121
167	FIG00638116: hypothetical protein	19	4	6	25	1	40	4.05702217299542e-06	0.000144374030432113	phage protein	YP_003230330	2.43155E-75
168	Failed to assign function	16	7	6	25	1	40	0.000254647977943617	0.00639852867013805	hypothetical protein	WP_002431037	9.0015E-10
169	Mobile element protein	20	3	7	24	1	40	2.48964902150791e-06	9.19960334085584e-05	is4 familly	WP_001375545	1.09183E-159
170	hypothetical protein	20	3	9	22	1	40	2.2433904485363 3e-05	0.000688163592366303	non-lee-encoded type iii secreted effector	YP_003230601	0.0
171	putative epimerase	18	5	9	22	1	40	0.000386795916434489	0.00940812693711364	fnl2	WP_000118459	0.0
172	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14)	18	5	9	22	1	40	0.000386795916434489	0.00940812693711364	udp-n-acetylglucosamine 2-epimerase	YP_003229906	0.0
173	Putative glycosyltransferase	18	5	9	22	1	40	0.000386795916434489	0.00940812693711364	wbub	YP_003229905	0.0
174	Tryptophan synthase beta chain like (EC 4.2.1.20)	18	5	9	22	1	40	0.000386795916434489	0.00940812693711364	wbuc	YP_003229904	3.11144E-94
175	IncF plasmid conjugative transfer surface exclusion protein TraT	19	4	10	21	1	40	0.000255763520970133	0.00641540165100084	conjugal transfer surface exclusion protein	WP_001115381	3.63047E-38
176	Failed to assign function	23	0	15	16	1	40	1.42470387593329e-05	0.000445543757564593	hypothetical protein E2348C_0721	YP_002328288	3.95229E-80
177	FIG00642056: hypothetical protein	22	1	15	16	1	40	0.000152218361481295	0.00405882496068959	hypothetical protein E2348C_0722	YP_002328289	2.26929E-48
178	FIG00638956: hypothetical protein	23	0	18	13	1	40	0.000186119403400109	0.0047678247168879	t3ss secreted effector -like protein	WP_001456460	3.53781E-139
179	Putative outer membrane protein	23	0	18	13	1	40	0.000186119403400109	0.0047678247168879	lpxr	WP_000628028	0.0
180	hypothetical protein	22	1	0	31	2	39	2.94678298750776e-14	1.41917068678374e-11	sukh-3 immunity family protein	YP_002386019	8.90435E-98
181	Co-activator of prophage gene expression IbrA	20	3	0	31	2	39	5.51048418663954e-12	9.25761343355442e-10	pp-loop family protein	WP_001535422	1.6638E-64
182	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	20	3	0	31	2	39	5.51048418663954e-12	9.25761343355442e-10	adhesin invasin	WP_016243659	0.0
183	FIG00639817: hypothetical protein	19	4	0	31	2	39	4.82167366330959e-11	4.70699601942547e-09	toxin-antitoxin antitoxin ribbon-helix-helix domain protein	YP_003232404	7.62172E-35
184	FIG00638152: hypothetical protein	19	4	0	31	2	39	4.82167366330959e-11	4.70699601942547e-09	pf03235 family protein	WP_000648228	0.0
185	Fimbrial protein YadC	19	4	0	31	2	39	4.82167366330959e-11	4.70699601942547e-09	fimbrial family protein	WP_000713127	3.98287E-96
186	hypothetical protein	16	7	0	31	2	39	1.16216111496227e-08	6.8815179463012e-07	phage-related protein	YP_002394337	0.0
187	Integrase	15	8	0	31	2	39	5.66553543544107e-08	2.96578463663959e-06	integrase	WP_001341225	0.0
188	Invasin	20	3	1	30	2	39	1.05995784060654e-10	9.6316169063417e-09	adhesin invasin	WP_001341730	1.0883E-130
189	FIG01068028: hypothetical protein	20	3	1	30	2	39	1.05995784060654e-10	9.6316169063417e-09	hypothetical protein	WP_000788778	2.12072E-14
190	FIG107037: Phage late gene regulator	17	6	1	30	2	39	3.26330873532791e-08	1.85623167748101e-06	ogr delta-like zinc finger family protein	YP_003232314	9.00923E-47
191	FIG00643125: hypothetical protein	13	10	1	30	2	39	1.11799899259371e-05	0.000362985380786379	ead ea22-like family protein	YP_003229753	7.14856E-143
192	Phage capsid and scaffold	17	6	2	29	2	39	2.61325069204052e-07	1.17008764408956e-05	head assembly protein	YP_003230290	3.48517E-110
193	Failed to assign function	20	3	3	28	2	39	7.43970985725224e-09	4.6532003470814e-07	dctp deaminase	WP_023566101	6.80065E-30
194	hypothetical protein	17	6	3	28	2	39	1.46141346920191e-06	5.51292475274914e-05	hypothetical protein ECED1_1795	YP_002397754	2.72647E-23
195	FIG00640419: hypothetical protein	17	6	3	28	2	39	1.46141346920191e-06	5.51292475274914e-05	regulatory protein cro	NP_059607	6.43072E-27
196	Failed to assign function	16	7	3	28	2	39	6.26176706919338e-06	0.000211873561160904	lipoprotein	WP_000749284	2.96746E-85
197	Haemolysin expression modulating protein paralog	16	7	3	28	2	39	6.26176706919338e-06	0.000211873561160904	hemolysin expression modulating protein hha	WP_021542549	1.15506E-29
198	Putative stability/partitioning protein encoded within prophage CP-933T	15	8	3	28	2	39	2.39472803323198e-05	0.000728400644718646	plasmid segregation protein parm	YP_002756739	0.0
199	Failed to assign function	15	8	3	28	2	39	2.39472803323198e-05	0.000728400644718646	plasmid stable inheritance protein	YP_002756740	6.62543E-78

200	CIII protein	14	9	3	28	2	39	8.28989912094081e-05	0.00227704301329568	regulatory protein ciii	NP_597896	3.09054E-18
201	unknown	20	3	4	27	2	39	4.05461721131233e-08	2.194049043784 29e-06	hypothetical protein	WP_000870971	5.64417E-30
202	Mobile element protein	19	4	4	27	2	39	2.64014792340253e-07	1.17008764408956e-05	integrase core domain protein	WP_001757863	4.50322E-91
203	FIG00640501: hypothetical protein	15	8	4	27	2	39	9.02679550690437e-05	0.00245610435939274	conserved protein	YP_003227401	0.0
204	Phage protein	16	7	5	26	2	39	8.64559095631686e-05	0.00235682071956351	ead ea22-like family protein	YP_003230651	4.92587E-65
205	orf; Unknown function	21	2	6	25	2	39	9.77012990204762e-08	4.81770774146021e-06	hypothetical protein, partial	WP_000223921	1.59004E-57
206	putative phage inhibition, colicin resistance and tellurite resistance protein	23	0	9	22	2	39	2.58292895812523e-08	1.4867791867328e-06	terf	YP_002269834	4.54196E-84
207	orf; Unknown function	21	2	11	20	2	39	2.87789944514314e-05	0.000866247732988086	restriction endonuclease	NP_286671	1.05376E-53
208	Urease accessory protein UreD	23	0	13	18	2	39	2.12793608780932e-06	7.90345002485066e-05	urease accessory protein ured	WP_000886249	0.0
209	Mobile element protein	23	0	13	18	2	39	2.12793608780932e-06	7.90345002485066e-05	transposase	YP_002269815	2.4162E-98
210	IncF plasmid conjugative transfer surface exclusion protein TraT	22	1	13	18	2	39	2.65992010976167e-05	0.000803985894264363	enterobacterial complement resistance family protein	WP_003873216	4.12867E-15
211	FIG00642900: hypothetical protein	21	2	13	18	2	39	0.000172727 61215618	0.00450463635457127	restriction enzyme beta subunit domain prot	YP_003223638	1.27421E-58
212	Phage EaA protein	23	0	14	17	2	39	5.62383108921037e-06	0.000192087734224377	pf04448 family protein	WP_000207990	0.0
213	Mobile element protein	23	0	14	17	2	39	5.62383108921037e-06	0.000192087734224377	transposase is66 family protein	NP_286697	0.0
214	Putative exported protein	23	0	0	31	3	38	9.20869683596171e-16	8.86981679239832e-13	dknyy family protein	WP_001324310	1.4816E-92
215	Failed to assign function	23	0	0	31	3	38	9.20869683596171e-16	8.86981679239832e-13	ribose abc transporter	NP_755883	6.2564E-169
216	Failed to assign function	22	1	0	31	3	38	2.94678298750776e-14	1.41917068678374e-11	hypothetical protein ECO26_0638	YP_003227708	5.06507E-78
217	Co-activator of prophage gene expression IbrA	20	3	0	31	3	38	5.51048418663954e-12	9.25761343355442e-10	immunoglobulin-binding regulator a	WP_001285511	0.0
218	Anaerobic dimethyl sulfoxide reductase chain A (EC 1.8.99.-)	19	4	0	31	3	38	4.82167366330959e-11	4.70699601942547e-09	anaerobic dimethyl sulfoxide a family	WP_001375676	0.0
219	possible integrase	18	5	0	31	3	38	3.47160503758288e-10	2.80210891525126e-08	integrase	YP_003230878	0.0
220	Mobile element protein	11	12	0	31	3	38	1.41249237997532e-05	0.000445543757564593	transposase	WP_024212962	4.00535E-50
221	Failed to assign function	9	14	0	31	3	38	0.000153666753425886	0.00405882496068959	antirepressor	YP_003229256	7.01632E-123
222	Mobile element protein	11	12	1	30	3	38	0.000126138854397796	0.00341923858975489	integrase core domain protein	WP_000510360	7.76703E-95
223	Integrase	10	13	1	30	3	38	0.000384632540393279	0.00940812693711364	phage integrase family protein	YP_002329318	0.0
224	DNA stabilization, phage-associated	17	6	2	29	3	38	2.61325069204052e-07	1.17008764408956e-05	packaged dna stabilization protein p27	NP_958183	3.2184E-106
225	unknown	17	6	2	29	3	38	2.61325069204052e-07	1.17008764408956e-05	phage protein	YP_003230294	2.297E-29
226	Phage major capsid protein	17	6	2	29	3	38	2.61325069204052e-07	1.17008764408956e-05	major head protein	YP_003230295	0.0
227	Phage capsid and scaffold	17	6	2	29	3	38	2.61325069204052e-07	1.17008764408956e-05	scaffold protein gp8	YP_003230296	4.24965E-134
228	Phage portal	15	8	2	29	3	38	4.99851612182719e-06	0.000176142831532096	portal protein	WP_001432480	0.0
229	Mobile element protein	17	6	3	28	3	38	1.46141346920191e-06	5.51292475274914e-05	integrase	YP_003232308	0.0
230	Phage holin	17	6	4	27	3	38	6.40295393978137e-06	0.0002146400893781	holin	NP_958234	2.58751E-62
231	FIG00640847: hypothetical protein	23	0	6	25	3	38	4.37431517101856e-10	3.41622192383115e-08	conserved protein	YP_310003	1.55638E-153
232	regulatory protein	17	6	6	25	3	38	7.3972904488057e-05	0.00207930063043472	regulatory protein	YP_006168161	2.76998E-37
233	Failed to assign function	20	3	8	23	3	38	7.80508840481334e-06	0.000261036845538757	prophage protein	WP_000268111	2.42043E-78
234	putative hemolysin activator protein	22	1	9	22	3	38	4.34254931084805e-07	1.88411869198597e-05	potra -type family protein	WP_001428611	0.0
235	Failed to assign function	23	0	15	16	3	38	1.42470387593329e-05	0.000445543757564593	pentapeptide repeat-containing protein	WP_001205482	1.22552E-89
236	Inner membrane protein YjgN	19	4	0	31	4	37	4.82167366330959e-11	4.70699601942547e-09	inner membrane protein yjgn	WP_000079609	1.72952E-81
237	Inner membrane protein YjgN	19	4	0	31	4	37	4.82167366330959e-11	4.70699601942547e-09	inner membrane protein yjgn	WP_024171715	3.6781E-143
238	FIG01069755: hypothetical protein	17	6	0	31	4	37	2.14082310650945e-09	1.41235672341774e-07	hypothetical protein ECO26_3945	YP_003230870	0.0
239	Failed to assign function	9	14	0	31	4	37	0.000153666753425886	0.00405882496068959	hypothetical protein ECO26_2261	YP_003229252	2.62725E-22
240	Mobile element protein	19	4	2	29	4	37	8.01921314913931e-09	4.95135006746858e-07	transposase	WP_000839180	7.4307E-84
241	Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)	19	4	2	29	4	37	8.01921314913931e-09	4.95135006746858e-07	type i restriction enzyme r protein	YP_672515	0.0
242	Type I restriction-modification system, DNA-methyltransferase subunit M (EC 2.1.1.72)	18	5	2	29	4	37	4.95447633220757e-08	2.64141232648469e-06	eco57i restriction-modification methylase family protein	WP_001341289	0.0
243	Failed to assign function	17	6	3	28	4	37	1.46141346920191e-06	5.51292475274914e-05	hypothetical protein EFER_3110	YP_002384207	4.33569E-34

244	FIG00640506: hypothetical protein	23	0	4	27	4	37	1.61612629471128e-11	2.43227007354048e-09	conserved domain protein	WP_001185377	8.28435E-45
245	hypothetical protein	21	2	4	27	4	37	4.79119287678257e-09	3.11816012088984e-07	phage protein	YP_007111834	2.4532E-29
246	Mobile element protein	20	3	4	27	4	37	4.05461721131233e-08	2.19404904378429e-06	integrase	CAB59975	1.35043E-174
247	Phage antitermination protein N	20	3	5	26	4	37	1.83566089058487e-07	8.87010988199673e-06	antitermination protein n	YP_003230322	1.9711E-23
248	FIG01068391: hypothetical protein	19	4	5	26	4	37	1.11319554920335e-06	4.41853002606868e-05	endonuclease	WP_001229299	1.4215E-86
249	Orf80	17	6	6	25	4	37	7.3972904488057e-05	0.00207930063043472	pf10834 family protein	YP_001571652	1.82939E-43
250	putative membrane protein	22	1	7	24	4	37	3.63693061361827e-08	2.03668114362623e-06	membrane protein	WP_000282206	0.0
251	Phage Kil	18	5	8	23	4	37	0.000154511215384042	0.00406626236770244	phage protein	WP_001243379	3.76135E-19
252	orf; Unknown function	23	0	9	22	4	37	2.58292895812523e-08	1.4867791867328e-06	conserved protein	NP_286721	6.68149E-52
253	FIG00641264: hypothetical protein	23	0	10	21	4	37	8.52366556181328e-08	4.23195601501987e-06	conserved protein	WP_001470661	1.92527E-33
254	orf; Unknown function	23	0	10	21	4	37	8.52366556181328e-08	4.23195601501987e-06	conserved protein	WP_001004880	2.33927E-29
255	FIG00641704: hypothetical protein	23	0	10	21	4	37	8.52366556181328e-08	4.23195601501987e-06	hypothetical protein ECED1_5037	YP_002400782	4.33157E-26
256	Failed to assign function	18	5	0	31	5	36	3.47160503758288e-10	2.80210891525126e-08	hypothetical protein	WP_000128178	3.24209E-26
257	FIG00637932: hypothetical protein	22	1	1	30	5	36	6.57500954087667e-13	1.35708196923694e-10	flagellar biosynthesis anti-sigma factor	YP_003227352	6.40198E-28
258	hypothetical protein	19	4	1	30	5	36	8.59635533115762e-10	6.33674192982476e-08	transcription termination factor nun	NP_597897	1.71202E-63
259	hypothetical protein	14	9	1	30	5	36	2.98384866266564e-06	0.000109976136423962	transcriptional repressor	YP_002402448	1.34032E-79
260	hypothetical protein	20	3	2	29	5	36	1.06517364649443e-09	7.77253982048055e-08	hypothetical protein	WP_001341484	3.26709E-14
261	Uncharacterized protein YadU in stf fimbrial cluster	23	0	4	27	5	36	1.61612629471128e-11	2.43227007354048e-09	pf11245 family protein	YP_003230267	0.0
262	Phage EaA protein	18	5	5	26	5	36	5.52893548489145e-06	0.000192087734224377	valyl-trna synthetase	YP_003230652	0.0
263	DNA stabilization, phage-associated	17	6	6	25	5	36	7.3972904488057e-05	0.00207930063043472	packaged dna stabilization protein gp26	YP_003230291	2.17336E-159
264	Phage terminase, large subunit	17	6	6	25	5	36	7.3972904488057e-05	0.00207930063043472	terminase large subunit	YP_003230298	0.0
265	Phage protein	21	2	7	24	5	36	3.63473268185746e-07	1.58654434403252e-05	phage protein	YP_003228940	2.27602E-50
266	hypothetical protein	18	5	8	23	5	36	0.000154511215384042	0.00406626236770244	o-polysaccharide acetyltransferase inhibitor	YP_003222265	1.28184E-22
267	FIG00641998: hypothetical protein	21	2	12	19	5	36	7.24434084885381e-05	0.00206442281231242	hypothetical protein ECO26_0051	YP_003227152	2.42276E-87
268	Recombinational DNA repair protein RecT (prophage associated)	20	3	12	19	5	36	0.000348984697390217	0.00857505256444533	protein rect	NP_287834	0.0
269	Failed to assign function	23	0	0	31	6	35	9.20869683596171e-16	8.86981679239832e-13	alpha c-terminal all-beta domain protein	WP_001440793	2.68175E-126
270	Flagellar biosynthesis protein FliS	22	1	0	31	6	35	2.94678298750776e-14	1.41917068678374e-11	flagellar protein	YP_001726302	1.75287E-80
271	Flagellar M-ring protein FliF	22	1	0	31	6	35	2.94678298750776e-14	1.41917068678374e-11	flagellar m-ring protein	YP_003227344	0.0
272	Failed to assign function	15	8	0	31	6	35	5.66553543544107e-08	2.96578463663959e-06	family transcriptional regulator	WP_001307541	9.10718E-56
273	Flagellar motor rotation protein MotB	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	chemotaxis protein	YP_003227375	0.0
274	Flagellar motor rotation protein MotA	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar motor stator protein	YP_003232831	0.0
275	FIG00639010: hypothetical protein	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar basal body-associated family protein	WP_001393689	5.50792E-79
276	Flagellar hook-length control protein FliK	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar hook-length control protein	YP_003227371	1.17986E-134
277	Flagellar hook-associated protein FliD	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	lateral flagellar filament capping protein	WP_000609684	0.0
278	Flagellin protein FlaA	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	bacterial flagellin family protein	YP_002411056	9.42555E-126
279	probable regulatory protein YPO0736	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	lateral flagellar transmembrane regulator	YP_003227366	0.0
280	FIG00639166: hypothetical protein	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	lateral flagellar hook associated protein	WP_001195933	1.63169E-173
281	Flagellar hook-associated protein FlgL	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar hook-associated protein 3	YP_003227364	6.62713E-156
282	Flagellar protein FlgJ [peptidoglycan hydrolase] (EC 3.2.1.-)	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	lateral flagellar peptidoglycan hydrolase	WP_001589748	2.02543E-59
283	Flagellar P-ring protein FlgI	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar p-ring protein 1	YP_003227361	0.0
284	Flagellar L-ring protein FlgH	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar l-ring protein	YP_003227360	4.55402E-175
285	Flagellar basal-body rod protein FlgG	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar basal-body rod protein	YP_002411046	1.61292E-154

286	Flagellar basal-body rod modification protein FlgD	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar hook capping family protein	YP_006137328	6.45638E-125
287	Flagellar basal-body rod protein FlgC	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar basal-body rod protein	WP_024216450	3.17607E-76
288	Flagellar basal-body rod protein FlgB	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar basal-body rod protein	WP_001315265	6.12523E-65
289	Flagellar basal-body P-ring formation protein FlgA	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagella basal body p-ring formation protein	YP_003227353	2.23096E-153
290	FIG00641064: hypothetical protein	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	lateral flagellar chaperone protein	YP_003227351	5.79427E-82
291	Lysine-N-methylase (EC 2.1.1.-)	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	lateral flagellar associated protein	YP_003227350	0.0
292	Cps2A	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	glycosyl transferase 2 family protein	WP_000294389	0.0
293	Glycerol-3-phosphate cytidylyltransferase (EC 2.7.7.39)	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	glycerol-3-phosphate cytidylyl-transferase	YP_003227349	1.15464E-89
294	FIG00639403: hypothetical protein	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar export protein	YP_003227348	9.69213E-71
295	Flagellum-specific ATP synthase FliI	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	atp synthase	YP_003227347	0.0
296	Flagellar assembly protein FliH	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar assembly family protein	YP_003227346	3.96152E-138
297	Flagellar motor switch protein FliG	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar motor switch protein	WP_001308395	0.0
298	Flagellar hook-basal body complex protein FliE	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar hook-basal body complex protein	YP_003227343	2.26475E-39
299	Flagellar regulatory protein FleQ	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	sigma-54 interaction domain protein	WP_009008331	0.0
300	Flagellar motor switch protein FliM	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	surface presentation of antigens protein	YP_003227341	0.0
301	Flagellar motor switch protein FliN	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar motor switch protein	YP_006167278	2.97909E-63
302	Flagellar biosynthesis protein FliP	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar biosynthetic protein	WP_000830281	3.70758E-127
303	Flagellar biosynthesis protein FliQ	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	bacterial export 3 family protein	YP_003227338	1.25723E-26
304	Flagellar biosynthesis protein FliR	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar biosynthetic protein	WP_001310546	2.16041E-121
305	Flagellar biosynthesis protein FlhB	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flhb family protein	WP_000785657	0.0
306	Flagellar biosynthesis protein FlhA	22	1	1	30	6	35	6.57500954087667e-13	1.35708196923694e-10	flagellar biosynthesis protein	WP_001140804	0.0
307	RNA polymerase sigma factor for flagellar operon	21	2	1	30	6	35	1.00779978172766e-11	1.56566572541948e-09	flagellar biosynthesis sigma factor	YP_006094618	1.14855E-147
308	FIG00639370: hypothetical protein	21	2	1	30	6	35	1.00779978172766e-11	1.56566572541948e-09	lateral flagellar chaperone protein	WP_001316130	1.11926E-34
309	Flagellar hook-associated protein FlgK	21	2	1	30	6	35	1.00779978172766e-11	1.56566572541948e-09	flagellar hook-associated protein	YP_003227363	0.0
310	Flagellar basal-body rod protein FlgF	21	2	1	30	6	35	1.00779978172766e-11	1.56566572541948e-09	flagellar hook-basal body protein	YP_003227358	9.38634E-148
311	Maltose/maltodextrin ABC transporter, permease protein MalF	21	2	1	30	6	35	1.00779978172766e-11	1.56566572541948e-09	sugar transporter subunit	WP_023981372	0.0
312	Flagellar hook protein FlgE	20	3	1	30	6	35	1.05995784060654e-10	9.6316169063417e-09	flagellar hook protein	YP_003227357	0.0
313	Failed to assign function	13	10	1	30	6	35	1.11799899259371e-05	0.000362985380786379	phage repressor protein ci	YP_003230320	4.42627E-114
314	hypothetical protein	15	8	2	29	6	35	4.99851612182719e-06	0.000176142831532096	terminase small subunit	WP_000577998	1.68051E-117
315	FIG00638052: hypothetical protein	16	7	3	28	6	35	6.26176706919338e-06	0.000211873561160904	dna-binding protein	WP_021499851	4.95728E-54
316	Phage NinC	16	7	5	26	6	35	8.64559095631686e-05	0.0023568207195635 1	phosphoadenosine phosphosulfate reductase family protein	YP_003230648	0.0
317	Phage DNA transfer protein	17	6	7	24	6	35	0.000208022152535731	0.0052635797895556	dna transfer protein	WP_021546256	5.69327E-65
318	FIG00245341: hypothetical protein	20	3	9	22	6	35	2.24339044853633e-05	0.000688163592366303	head protein prohead protease	WP_000973527	0.0
319	FIG00640385: hypothetical protein	20	3	9	22	6	35	2.24339044853633e-05	0.000688163592366303	gp80	WP_000006074	1.49137E-22
320	Phage terminase small subunit	20	3	9	22	6	35	2.24339044853633e-05	0.000688163592366303	terminase small subunit	WP_000223329	1.18509E-85
321	FIG00638463: hypothetical protein	20	3	9	22	6	35	2.24339044853633e-05	0.000688163592366303	transcription antitermination protein	YP_003230067	8.24063E-15
322	FIG00642058: hypothetical protein	20	3	9	22	6	35	2.24339044853633e-05	0.000688163592366303	hypothetical protein ECO26_3114	YP_003230066	4.5864E-31
323	Putative bacteriophage protein	21	2	10	21	6	35	1.07674236611247e-05	0.000351168706672528	rtr1 rpap2 family protein	WP_023143851	4.75613E-23
324	Mobile element protein	23	0	12	19	6	35	7.68421365042259e-07	3.19946740119036e-05	acetaldehyde dehydrogenase 2	WP_001342119	6.8285E-130
325	PTS system, mannose-specific IIA component (EC 2.7.1.69)	23	0	12	19	6	35	7.68421365042259e-07	3.19946740119036e-05	pts system fructose iia component family protein	YP_001464286	4.53124E-92
326	hypothetical protein	21	2	12	19	6	35	7.24434084885381e-05	0.00206442281231242	hypothetical protein	WP_000422899	2.57184E-46
327	FIG01047205: hypothetical protein	21	2	13	18	6	35	0.00017272761215618	0.00450463635457127	anti- protein	WP_024008430	4.06953E-56
328	FIG00638089: hypothetical protein	23	0	14	17	6	35	5.62383108921037e-06	0.000192087734224377	hypothetical protein ECO26_2447	YP_003229432	1.6526E-36

329	FIG00639855: hypothetical protein	21	2	14	17	6	35	0.000391920228744026	0.00950077762566054	phage protein	YP_003228946	2.16022E-74
330	FIG00641110: hypothetical protein	23	0	0	31	7	34	9.20869683596171e-16	8.86981679239832e-13	hypothetical protein ECO26_5378	YP_003232254	0.0
331	Putative sugar kinase, PfkB family protein (EC 2.7.1.-)	23	0	0	31	7	34	9.20869683596171e-16	8.86981679239832e-13	pfkb family carbohydrate kinase	WP_001275272	0.0
332	Transcriptional repressor of aga operon	23	0	0	31	7	34	9.20869683596171e-16	8.86981679239832e-13	family transcriptional regulator	YP_542672	0.0
333	Ribose ABC transport system, ATP-binding protein RbsA (TC 3.A.1.2.1)	23	0	0	31	7	34	9.20869683596171e-16	8.86981679239832e-13	sugar uptake abc transporter atp-binding protein	YP_003231270	0.0
334	Ribose ABC transport system, permease protein RbsC (TC 3.A.1.2.1)	23	0	0	31	7	34	9.20869683596171e-16	8.86981679239832e-13	ribose transport system permease	YP_003231269	9.82216E-168
335	Fructokinase (EC 2.7.1.4)	23	0	0	31	7	34	9.20869683596171e-16	8.86981679239832e-13	carbohydrate kinase	YP_003231268	0.0
336	TIORF127 protein	23	0	0	31	7	34	9.20869683596171e-16	8.86981679239832e-13	tiorf127 protein	YP_003231267	0.0
337	PTS system, mannose-specific IID component (EC 2.7.1.69)	23	0	1	30	7	34	2.21008724063083e-14	1.27725361810537e-11	pts system mannose fructose sorbose iid component family protein	YP_002388403	0.0
338	orf, hypothetical protein	19	4	1	30	7	34	8.59635533115762e-10	6.33674192982476e-08	protein yiba	WP_001326523	2.20834E-45
339	Failed to assign function	15	8	1	30	7	34	7.32161502426228e-07	3.14829446043278e-05	hypothetical protein	WP_001595802	2.05644E-19
340	Failed to assign function	15	8	2	29	7	34	4.99851612182719e-06	0.000176142831532096	hypothetical protein	WP_001332869	1.72106E-12
341	Failed to assign function	13	10	2	29	7	34	6.44549081372011e-05	0.00185506875053044	hypothetical protein	WP_001331630	8.50402E-42
342	Putative Na(+)/H(+) exchanger protein, CPA1 family precursor	19	4	3	28	7	34	5.20614633325567e-08	2.75525282867685e-06	na+ h+ antiporter	YP_003232423	0.0
343	TRAP transporter solute receptor, unknown substrate 6	19	4	4	27	7	34	2.64014792340253e-07	1.17008764408956e-05	c4-dicarboxylate-binding periplasmic protein	YP_543995	0.0
344	TRAP-type transport system, small permease component, predicted N-acetylneuraminate transporter	19	4	5	26	7	34	1.11319554920335e-06	4.41853002606868e-05	tripartite atp-independent periplasmic transporter (dctq-like) protein	YP_003232421	1.17927E-77
345	TRAP dicarboxylate transporter, DctM subunit, unknown substrate 8	19	4	5	26	7	34	1.11319554920335e-06	4.41853002606868e-05	c4-dicarboxylate transporter large subunit	YP_002332098	0.0
346	FIG01070003: hypothetical protein	19	4	5	26	7	34	1.11319554920335e-06	4.41853002606868e-05	membrane protein	WP_000558268	1.54299E-8
347	Failed to assign function	21	2	6	25	7	34	9.77012990204762e-08	4.81770774146021e-06	phage protein	YP_003222263	1.57314E-71
348	DNA stabilization, phage-associated	17	6	6	25	7	34	7.3972904488057e-05	0.00207930063043472	dna stabilization protein	WP_001122374	0.0
349	Maltose/maltodextrin ABC transporter, substrate binding periplasmic protein MalE	22	1	11	20	7	34	3.8279007159416e-06	0.000140013948212467	bacterial extracellular solute-binding family protein	WP_001369318	0.0
350	L-seryl-tRNA(Sec) selenium transferase-related protein	21	2	11	20	7	34	2.87789944514314e-05	0.000866247732988086	l-seryl-trna selenium transferase family protein	YP_001464281	0.0
351	Integrase	20	3	11	20	7	34	0.000148856133053476	0.00402747829654798	integrase	YP_003228935	0.0
352	Failed to assign function	23	0	12	19	7	34	7.68421365042259e-07	3.19946740119036e-05	2-dehydro-3-deoxyphosphogluconate aldolase	YP_001464280	4.48678E-157
353	PTS system, mannose-specific IIC component (EC 2.7.1.69)	23	0	12	19	7	34	7.68421365042259e-07	3.19946740119036e-05	pts system sorbose-specific iic component family protein	YP_001464284	3.67931E-135
354	PTS system, mannose-specific IIA component (EC 2.7.1.69) / PTS system, mannose-specific IIB component (EC 2.7.1.69)	23	0	12	19	7	34	7.68421365042259e-07	3.19946740119036e-05	pts system sorbose subiib component family protein	YP_001464285	2.48344E-108
355	FIG00643164: hypothetical protein	23	0	12	19	7	34	7.68421365042259e-07	3.19946740119036e-05	hypothetical protein ECO26_0052	YP_003227153	4.42682E-146
356	Transcriptional regulatory protein zraR	22	1	12	19	7	34	1.03627109799984e-05	0.000338734045789632	aaa domain family protein	YP_002388407	0.0
357	Neopullulanase (EC 3.2.1.135)	22	1	12	19	7	34	1.03627109799984e-05	0.000338734045789632	neopullulanase	YP_003229675	0.0
358	Maltose/maltodextrin ABC transporter, permease protein MalG	22	1	12	19	7	34	1.03627109799984e-05	0.000338734045789632	binding--dependent transport system inner membrane component family protein	WP_000357161	1.14669E-162
359	FIG00641874: hypothetical protein	22	1	12	19	7	34	1.03627109799984e-05	0.000338734045789632	cyma protein	YP_003229679	0.0
360	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (EC 2.3.1.-)	22	1	12	19	7	34	1.03627109799984e-05	0.000338734045789632	bacterial transferase hexapeptide family protein	YP_003229680	0.0

361	Glycosidase	22	1	12	19	7	34	1.03627109799984e-05	0.000338734045789632	glycosidase	YP_002387351	5.03687E-112
362	Maltose/maltodextrin transport ATP-binding protein MalK (EC 3.6.3.19)	22	1	12	19	7	34	1.03627109799984e-05	0.000338734045789632	abc transporter family protein	YP_003229682	0.0
363	Putative glycosyl hydrolase of unknown function (DUF1680)	22	1	12	19	7	34	1.03627109799984e-05	0.000338734045789632	glycosyhydrolase	YP_003229684	0.0
364	ydaQ protein	20	3	12	19	7	34	0.000348984697390217	0.00857505256444533	excisionase	NP_287837	5.57378E-46
365	Phage protein	21	2	14	17	7	34	0.000391920228744026	0.00950077762566054	gydac	WP_000230535	6.70724E-26
366	ORF25	18	5	1	30	8	33	5.72814831201178e-09	3.64582761286106e-07	hnh endonuclease family protein	YP_003232367	1.53611E-157
367	Failed to assign function	18	5	1	30	8	33	5.72814831201178e-09	3.64582761286106e-07	pf04447 domain protein	WP_001289868	2.36348E-115
368	Phosphoesterase (EC 3.1.-.-)	22	1	4	27	8	33	3.87870310730705e-10	3.11330569413179e-08	phosphodiesterase yaei	YP_003227266	0.0
369	FIG00639870: hypothetical protein	18	5	4	27	8	33	1.40856276898182e-06	5.41247736336417e-05	icd-like protein	WP_001342315	7.40978E-126
370	Integrase	23	0	6	25	8	33	4.37431517101856e-10	3.41622192383115e-08	integrase	NP_286655	0.0
371	putative regulator; Regulation (Phage or Prophage Related)	22	1	7	24	8	33	3.63693061361827e-08	2.03668114362623e-06	repressor protein	NP_288007	3.33475E-70
372	Failed to assign function	19	4	7	24	8	33	1.31205277881187e-05	0.00042408363642671	hypothetical protein ECO26_5503	YP_003232366	2.02338E-30
373	FIG00642259: hypothetical protein	23	0	9	22	8	33	2.58292895812523e-08	1.4867791867328e-06	shia domain protein	NP_309389	8.66179E-40
374	FIG00641009: hypothetical protein	23	0	12	19	8	33	7.68421365042259e-07	3.19946740119036e-05	nif3 family protein	YP_003229672	0.0
375	putative esterase	23	0	12	19	8	33	7.68421365042259e-07	3.19946740119036e-05	esterase family protein	YP_003229673	0.0
376	FIG00640408: hypothetical protein	23	0	12	19	8	33	7.68421365042259e-07	3.19946740119036e-05	amidohydrolase family protein	WP_000497897	0.0
377	CymJ protein	21	2	12	19	8	33	7.24434084885381e-05	0.00206442281231242	sugar-specific transcriptional regulator family protein	YP_002387353	8.00502E-169
378	Alpha-mannosidase (EC 3.2.1.24)	21	2	12	19	8	33	7.24434084885381e-05	0.00206442281231242	alpha-mannosidase	YP_003229685	0.0
379	Mobile element protein	21	2	13	18	8	33	0.00017272761215618	0.00450463635457127	transposase	YP_408256	5.2894E-40
380	hypothetical protein	15	8	1	30	9	32	7.32161502426228e-07	3.14829446043278e-05	phage protein	WP_001312978	2.52892E-128
381	FIG00640454: hypothetical protein	16	7	2	29	9	32	1.20876327972868e-06	4.75864083529156e-05	esterase-like activity of phytase family protein	YP_002394351	0.0
382	Failed to assign function	19	4	5	26	9	32	1.11319554920335e-06	4.41853002606868e-05	repressor protein	YP_003222279	2.25144E-66
383	Helix-turn-helix motif	19	4	5	26	9	32	1.11319554920335e-06	4.41853002606868e-05	transcriptional regulator	YP_003222280	2.2204E-72
384	putative membrane protein	19	4	5	26	9	32	1.11319554920335e-06	4.41853002606868e-05	toxin	YP_003222281	1.13584E-60
385	Regulatory protein Cro of bacteriophage BP-933W	19	4	6	25	9	32	4.05702217299542e-06	0.000144374030432113	kda cro protein	NP_287826	1.8578E-49
386	putative regulator; Regulation (Phage or Prophage Related)	19	4	6	25	9	32	4.05702217299542e-06	0.000144374030432113	helix-turn-helix family protein	WP_000233323	6.25815E-75
387	gene 66 protein	17	6	6	25	9	32	7.3972904488057e-05	0.00207930063043472	phage family protein	YP_003230301	5.37425E-79
388	FIG00640819: hypothetical protein	22	1	7	24	9	32	3.63693061361827e-08	2.03668114362623e-06	hypothetical protein Z1205	NP_286740	4.01503E-142
389	hypothetical protein	21	2	12	19	9	32	7.24434084885381e-05	0.00206442281231242	yeda protein	WP_001445027	7.56499E-99
390	Failed to assign function	23	0	0	31	10	31	9.20869683596171e-16	8.86981679239832e-13	hypothetical protein	YP_005280112	3.15197E-31
391	dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13)	19	4	1	30	10	31	8.59635533115762e-10	6.33674192982476e-08	rmlc	YP_003229912	4.06229E-130
392	hypothetical protein	16	7	2	29	10	31	1.20876327972868e-06	4.75864083529156e-05	biofilm development family protein	YP_003047326	9.46629E-41
393	Citrate:6-N-acetyl-6-N-hydroxy-L-lysine ligase, alpha subunit (EC 6.3.2.27), aerobactin biosynthesis protein IucA @ Siderophore synthetase superfamily, group C @ Siderophore synthetase component, ligase	22	1	6	25	10	31	9.11064263205247e-09	5.48460686449559e-07	aerobactin siderophore biosynthesis protein iucc	NP_755500	0.0
394	Ankyrin-repeat protein A	19	4	6	25	10	31	4.05702217299542e-06	0.000144374030432113	enterotoxin	WP_001345136	0.0
395	hypothetical protein	23	0	10	21	10	31	8.52366556181328e-08	4.23195601501987e-06	transposase	YP_006133018	5.92025E-18
396	orf; Unknown function	22	1	16	15	10	31	0.000341928930223992	0.0084592280545825 9	membrane protein	WP_000280753	1.12456E-128
397	Flagellar biosynthesis protein FliC	21	2	0	31	11	30	4.86219192938783e-13	1.35708196923694e-10	flagellin	YP_003229794	0.0
398	Beta-glucoside bgl operon antiterminator, BglG family	23	0	1	30	11	30	2.21008724063083e-14	1.27725361810537e-11	antiterminator	WP_000464369	1.0133E-175

399	PTS system, maltose and glucose-specific IIC component (EC 2.7.1.69) / PTS system, maltose and glucose-specific IIB component (EC 2.7.1.69)	23	0	1	30	11	30	2.21008724063083e-14	1.27725361810537e-11	phosphotransferase system enzyme: iib iic component	YP_003230818	0.0
400	Cystathionine beta-lyase (EC 4.4.1.8) (CBL) (Beta-cystathionase) (Cysteine lyase) / Maltose regulon modulator	23	0	1	30	11	30	2.21008724063083e-14	1.27725361810537e-11	cystathionine beta-lyase	YP_003230817	0.0
401	Arabinose 5-phosphate isomerase (EC 5.3.1.13)	23	0	1	30	11	30	2.21008724063083e-14	1.27725361810537e-11	phosphosugar isomerase	WP_001377601	4.98318E-123
402	YcgN (Fragment)	23	0	1	30	11	30	2.21008724063083e-14	1.27725361810537e-11	upf0260 protein ycgn	WP_000512275	1.41352E-136
403	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	22	1	1	30	11	30	6.57500954087667e-13	1.35708196923694e-10	d-isomer specific 2-hydroxyacid nad binding domain protein	YP_003230815	0.0
404	dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133)	20	3	1	30	11	30	1.05995784060654e-10	9.6316169063417e-09	dtdp-4-dehydrorhamnose reductase	YP_003229914	0.0
405	iron aquisition yersiniabactin synthesis enzyme (Irp2)	17	6	1	30	11	30	3.26330873532791e-08	1.85623167748101e-06	siderophore biosynthetic protein	YP_003229839	0.0
406	Failed to assign function	17	6	4	27	11	30	6.40295393978137e-06	0.0002146400893781	integrase core domain protein	WP_000589012	0.0
407	N6-hydroxylysine O-acetyltransferase (EC 2.3.1.102), aerobactin biosynthesis protein IucB @ Siderophore synthetase small component, acetyltransferase	22	1	6	25	11	30	9.11064263205247e-09	5.48460686449559e-07	aerobactin siderophore biosynthesis protein	NP_755501	0.0
408	Z1226 protein	19	4	6	25	11	30	4.05702217299542e-06	0.000144374030432113	restriction methylase	WP_024237418	3.4767E-16
409	Hydrogenase-4 component I (EC 1.-.-.-)	23	0	7	24	11	30	1.87470650186509e-09	1.34087918509637e-07	formate hydrogenlyase subunit 7	WP_001341609	1.18106E-103
410	Failed to assign function	22	1	11	20	11	30	3.8279007159416e-06	0.000140013948212467	dead deah box helicase family protein	NP_286664	0.0
411	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	20	3	0	31	12	29	5.51048418663954e-12	9.25761343355442e-10	adherence and invasion outermembrane protein (peyer s patches colonization)	WP_001342202	0.0
412	iron aquisition 2,3-dihydroxybenzoate-AMP ligase (EC 2.7.7.58,Irp5)	19	4	1	30	12	29	8.59635533115762e-10	6.33674192982476e-08	(-dihydroxybenzoyl)adenylate synthase	YP_003229843	0.0
413	Putative ABC iron siderophore transporter, fused permease and ATPase domains	18	5	1	30	12	29	5.72814831201178e-09	3.64582761286106e-07	yersiniabactin abc transporter atp-binding protein permease	YP_003229837	0.0
414	iron aquisition yersiniabactin synthesis enzyme (Irp1,polyketide synthetase)	16	7	1	30	12	29	1.63314219839435e-07	7.9179994907388e-06	polyketide synthase	YP_003229840	0.0
415	COG3311: Predicted transcriptional regulator	20	3	4	27	12	29	4.05461721131233e-08	2.19404904378429e-06	regulatory protein	YP_003230056	4.84939E-34
416	Prophage CP4-57 integrase	20	3	4	27	12	29	4.05461721131233e-08	2.19404904378429e-06	integrase	YP_003230054	0.0
417	FIG00639855: hypothetical protein	19	4	5	26	12	29	1.11319554920335e-06	4.41853002606868e-05	regulatory protein	YP_003222277	6.0766E-80
418	FIG00639052: hypothetical protein	22	1	6	25	12	29	9.11064263205247e-09	5.48460686449559e-07	membrane transport protein	WP_021523249	0.0
419	Citrate:6-N-acetyl-6-N-hydroxy-L-lysine ligase, alpha subunit (EC 6.3.2.27), aerobactin biosynthesis protein IucA @ Siderophore synthetase superfamily, group A @ Siderophore synthetase large component, acetyltransferase	22	1	6	25	12	29	9.11064263205247e-09	5.48460686449559e-07	aerobactin siderophore biosynthesis protein iuca	NP_755502	0.0
420	hypothetical protein	16	7	6	25	12	29	0.000254647977943617	0.00639852867013805	inner membrane metabolite transport protein	YP_002387059	1.71623E-15
421	iron aquisition yersiniabactin synthesis enzyme (Irp3)	20	3	1	30	13	28	1.05995784060654e-10	9.6316169063417e-09	thiazolinyl-s-hmwp1 reductase	YP_003229841	0.0
422	iron aquisition yersiniabactin synthesis enzyme (YbtT,resembles thioesterases)	20	3	1	30	13	28	1.05995784060654e-10	9.6316169063417e-09	yersiniabactin thioesterase component	YP_006139150	0.0

423	iron aquisition outermembrane yersiniabactin receptor (FyuA,Psn,pesticin receptor)	20	3	1	30	13	28	1.05995784060654e-10	9.6316169063417e-09	siderophore receptor protein	WP_009484410	0.0
424	Anthranilate synthase, aminase component (EC 4.1.3.27)	19	4	1	30	13	28	8.59635533115762e-10	6.33674192982476e-08	salicylate synthase	WP_000703039	0.0
425	AmpG permease	19	4	1	30	13	28	8.59635533115762e-10	6.33674192982476e-08	major facilitator superfamily protein	YP_003229835	8.87217E-179
426	Inner membrane ABC-transporter YbtQ	19	4	1	30	13	28	8.59635533115762e-10	6.33674192982476e-08	yersiniabactin-iron abc transporter permease atp-binding protein	YP_003229836	0.0
427	iron aquisition regulator (YbtA,AraC-like, required for transcription of FyuA/psn,Irp2)	19	4	1	30	13	28	8.59635533115762e-10	6.33674192982476e-08	family transcriptional regulator	YP_003229838	0.0
428	hypothetical protein	15	8	1	30	13	28	7.32161502426228e-07	3.14829446043278e-05	type iv secretion protein rhs	WP_001331434	5.69913E-79
429	FIG00638496: hypothetical protein	22	1	5	26	13	28	2.02339012097853e-09	1.41235672341774e-07	hypothetical protein	WP_001700672	2.70544E-161
430	Aerobactin siderophore receptor IutA @ TonB-dependent siderophore receptor	22	1	6	25	13	28	9.11064263205247e-09	5.48460686449559e-07	ferric siderophore receptor	WP_001299628	0.0
431	Putative transcriptional regulator of sorbose uptake and utilization genes	23	0	10	21	13	28	8.52366556181328e-08	4.23195601501987e-06	sorbitol operon regulator	YP_003232022	0.0
432	L-sorbose 1-phosphate reductase (EC 1.1.1.-)	23	0	14	17	13	28	5.62383108921037e-06	0.000192087734224377	zinc-binding dehydrogenase family protein	NP_290649	0.0
433	PTS system, sorbose-specific IIB component (EC 2.7.1.69)	23	0	14	17	13	28	5.62383108921037e-06	0.000192087734224377	sorbose-specific phosphotransferase enzyme iib component	YP_410308	5.23361E-107
434	PTS system, sorbose-specific IIA component (EC 2.7.1.69)	23	0	14	17	13	28	5.62383108921037e-06	0.000192087734224377	pts mannose fructose sorbose iia component domain protein	YP_003232020	2.25103E-87
435	Sorbitol-6-phosphate 2-dehydrogenase (EC 1.1.1.140)	23	0	14	17	13	28	5.62383108921037e-06	0.000192087734224377	sorbitol-6-phosphate 2-dehydrogenase	YP_002415160	3.27152E-177
436	Failed to assign function	23	0	17	14	13	28	8.17109575902916e-05	0.00225297693752774	hypothetical protein	WP_001323836	8.67831E-104
437	COG2932: Predicted transcriptional regulator	21	2	1	30	14	27	1.00779978172766e-11	1.56566572541948e-09	helix-turn-helix family protein	WP_019842147	5.73389E-161
438	Protein flxA	17	6	3	28	14	27	1.46141346920191e-06	5.51292475274914e-05	phage protein	WP_001413281	1.28007E-32
439	Failed to assign function	17	6	4	27	14	27	6.40295393978137e-06	0.0002146400893781	antitoxin	YP_003602621	1.94772E-37
440	FIG00641267: hypothetical protein	17	6	5	26	14	27	2.33838839204998e-05	0.000714271363389814	hypothetical protein ECO26_3713	YP_003230650	2.99035E-32
441	Flagellar hook-associated protein FliD	23	0	7	24	14	27	1.87470650186509e-09	1.34087918509637e-07	flagellar hook-associated protein 2	YP_002398129	0.0
442	FIG00638000: hypothetical protein	23	0	10	21	14	27	8.52366556181328e-08	4.23195601501987e-06	family transcriptional regulator	WP_001387282	9.90156E-39
443	Mobile element protein	23	0	13	18	14	27	2.12793608780932e-06	7.90345002485066e-05	iscro3 transposase	P30192	1.1331E-34
444	Uncharacterized lipoprotein yaeF precursor	22	1	13	18	14	27	2.65992010976167e-05	0.000803985894264363	nlpc p60 family protein	WP_001317489	6.92026E-132
445	FIG00639368: hypothetical protein	23	0	17	14	14	27	8.17109575902916e-05	0.00225297693752774	pf11745 family protein	WP_000522548	1.0283E-115
446	plasmid stabilization system	22	1	5	26	15	26	2.02339012097853e-09	1.41235672341774e-07	plasmid stabilization protein	YP_003222261	1.89903E-62
447	FIG00644297: hypothetical protein	22	1	5	26	15	26	2.02339012097853e-09	1.41235672341774e-07	addiction module antidote protein	WP_001260978	1.80631E-38
448	Failed to assign function	21	2	5	26	15	26	2.32851476542001e-08	1.37315842166483e-06	hypothetical phage protein	YP_002294180	1.04955E-80
449	FIG00642859: hypothetical protein	18	5	6	25	15	26	1.87761552919557e-05	0.000584650628573655	hypothetical protein ECO111_p2-092	YP_003237996	1.55961E-31
450	FIG00640898: hypothetical protein	21	2	7	24	15	26	3.63473268185746e-07	1.58654434403252e-05	conserved protein	NP_065381	7.22569E-24
451	hypothetical protein	21	2	7	24	15	26	3.63473268185746e-07	1.58654434403252e-05	conserved protein	WP_000378683	2.31054E-20
452	orf; Unknown function	20	3	10	21	15	26	5.97666431229269e-05	0.0017270169196801	hypothetical protein Z5490	NP_290572	8.4218E-52
453	Failed to assign function	23	0	12	19	15	26	7.68421365042259e-07	3.19946740119036e-05	trap transporter solute family protein	YP_002409295	0.0
454	TRAP-type transport system, small permease component, predicted N-acetylneuraminate transporter	23	0	12	19	15	26	7.68421365042259e-07	3.19946740119036e-05	tripartite atp-independent periplasmic family	WP_001583905	6.73832E-79
455	Transcriptional regulator, ArsR family	23	0	16	15	15	26	3.47271569758741e-05	0.00101979260973055	bacterial regulatory luxr family protein	WP_001570265	5.679E-89
456	Failed to assign function	21	2	2	29	16	25	1.08993214880759e-10	9.78095632669076e-09	antirepressor protein cro	WP_023228645	4.63559E-43
457	FIG00639376: hypothetical protein	23	0	5	26	16	25	9.0503072503832e-11	8.65952577175739e-09	conserved protein	YP_003230659	0.0
458	Pilin transcriptional activator	15	8	5	26	16	25	0.000284682275868697	0.00710378155742821	histidine kinase	WP_000486074	0.0
459	FIG01069681: hypothetical protein	23	0	6	25	16	25	4.37431517101856e-10	3.41622192383115e-08	hypothetical protein	WP_024245286	2.59084E-75
460	Hydrogenase-4 component D (EC 1.-.-.-)	23	0	9	22	16	25	2.58292895812523e-08	1.4867791867328e-06	hydrogenase-4 component d	YP_003230470	0.0

461	DNA helicase	23	0	9	22	16	25	2.58292895812523e-08	1.4867791867328e-06	rep helicase-like protein	YP_003228378	0.0
462	YeeU protein (antitoxin to YeeV)	23	0	10	21	16	25	8.52366556181328e-08	4.23195601501987e-06	phage protein	YP_001458813	4.14444E-31
463	Hydrogenase-4 component B (EC 1.-.-.-) / Formate hydrogenlyase subunit 3	23	0	11	20	16	25	2.63458753728775e-07	1.17008764408956e-05	hydrogenase-4 component b	WP_024237058	0.0
464	Mobile element protein	23	0	14	17	16	25	5.62383108921037e-06	0.000192087734224377	transposase	WP_001444225	0.0
465	FIG00639484: hypothetical protein	23	0	16	15	16	25	3.47271569758741e-05	0.00101979260973055	conserved protein	NP_418581	4.33314E-70
466	plasmid stabilization system	23	0	10	21	17	24	8.52366556181328e-08	4.23195601501987e-06	plasmid stabilization protein	WP_023151714	3.85069E-56
467	putative membrane protein	22	1	10	21	17	24	1.33279134239262e-06	5.16251188066719e-05	membrane protein	NP_309401	8.64022E-101
468	FIG00639556: hypothetical protein	20	3	12	19	17	24	0.000348984697390217	0.0085750525644453 3	bacteriophage regulatory protein cii	WP_001182773	9.81612E-121
469	Putative Type III secretion apparatus protein	23	0	16	15	17	24	3.47271569758741e-05	0.00101979260973055	type iii secretion apparatus protein	WP_001307374	1.0034E-49
470	FIG00639836: hypothetical protein	23	0	16	15	17	24	3.47271569758741e-05	0.00101979260973055	inner membrane protein yjeo	NP_418582	3.60029E-43
471	putative fimbrial protein precursor	23	0	17	14	17	24	8.17109575902916e-05	0.00225297693752774	fimbrial family protein	YP_001882263	3.28735E-115
472	FIG00637883: hypothetical protein	23	0	1	30	18	23	2.21008724063083e-14	1.27725361810537e-11	lipoprotein	WP_000011863	0.0
473	FIG00639255: hypothetical protein	23	0	1	30	18	23	2.21008724063083e-14	1.27725361810537e-11	peptidase	WP_001700873	0.0
474	FIG00638399: hypothetical protein	19	4	3	28	19	22	5.20614633325567e-08	2.75525282867685e-06	endoribonuclease	YP_002384181	3.25767E-68
475	Failed to assign function	18	5	4	27	19	22	1.40856276898182e-06	5.41247736336417e-05	mrna interferase	2KC8_A	2.12788E-49
476	orf, hypothetical protein	19	4	7	24	19	22	1.31205277881187e-05	0.00042408363642671	protein yiba	WP_001455496	0.0
477	FIG00638542: hypothetical protein	19	4	8	23	19	22	3.83688648588194e-05	0.00112217279247009	helix-turn-helix family protein	YP_405986	1.23551E-45
478	FIG01069793: hypothetical protein	21	2	13	18	19	22	0.00017272761215618	0.00450463635457127	fimbrial family protein	YP_001461187	0.0
479	Ribose 5-phosphate isomerase B (EC 5.3.1.6)	23	0	15	16	19	22	1.42470387593329e-05	0.000445543757564593	ribose-5-phosphate isomerase b domain protein	YP_312996	3.89265E-68
480	Type III secretion cytoplasmic protein (YscF)	23	0	17	14	19	22	8.17109575902916e-05	0.00225297693752774	type iii secretion protein	YP_002388319	3.57146E-23
481	FIG00640206: hypothetical protein	23	0	18	13	19	22	0.000186119403400109	0.0047678247168879	toxin-antitoxin system protein	NP_286854	1.54294E-25
482	Uncharacterized protein yjfZ	23	0	2	29	20	21	2.76260905078851e-13	1.28755405050943e-10	pf10887 family protein	YP_003232247	0.0
483	FIG00639051: hypothetical protein	14	9	4	27	20	21	0.000289229979279216	0.00719241779797954	phage protein	WP_016243203	1.73414E-58
484	Minor fimbrial subunit StfF	22	1	9	22	20	21	4.34254931084805e-07	1.88411869198597e-05	fimbrial family protein	WP_023982031	7.11232E-99
485	Phage NinB DNA recombination	20	3	9	22	20	21	2.24339044853633e-05	0.000688163592366303	recombination protein	YP_003227912	1.03513E-78
486	Mobile element protein	23	0	11	20	20	21	2.63458753728775e-07	1.17008764408956e-05	insertion element is1 protein insb	WP_005039619	1.90582E-45
487	prophage Kil protein	20	3	12	19	20	21	0.000348984697390217	0.0085750525644453 3	prophage kil protein	YP_003227652	1.97035E-39
488	hypothetical protein	21	2	1	30	21	20	1.00779978172766e-11	1.56566572541948e-09	inner membrane protein	BAB33410	1.91132E-64
489	Putative cytoplasmic protein	23	0	5	26	21	20	9.0503072503832e-11	8.65952577175739e-09	yoag domain protein	YP_002293245	2.54985E-18
490	hypothetical bacteriophage protein	16	7	6	25	21	20	0.000254647977943617	0.00639852867013805	protein from phage origin	YP_003228601	1.13593E-44
491	Putative fimbrial protein	23	0	17	14	21	20	8.17109575902916e-05	0.00225297693752774	long polar fimbrial protein	YP_002389016	0.0
492	ORF_f143	23	0	19	12	21	20	0.000411421839094978	0.00977668212375699	ygen	WP_021823940	3.20514E-93
493	FIG00638558: hypothetical protein	23	0	19	12	21	20	0.000411421839094978	0.00977668212375699	conserved protein	YP_003231485	0.0
494	FIG00638269: hypothetical protein	21	2	3	28	22	19	8.18417406078e-10	6.33674192982476e-08	phage protein	WP_001342092	7.29895E-19
495	FIG00642194: hypothetical protein	23	0	6	25	22	19	4.37431517101856e-10	3.41622192383115e-08	conserved protein	WP_000350594	6.8239E-70
496	Glycosyltransferase IroB	23	0	7	24	22	19	1.87470650186509e-09	1.34087918509637e-07	glucosyl-transferase	YP_003228369	0.0
497	Type III secretion protein EprH	22	1	16	15	22	19	0.000341928930223992	0.00845922805458259	type iii secretion apparatus protein	YP_006164032	0.0
498	COG0148: Enolase	22	1	6	25	23	18	9.11064263205247e-09	5.48460686449559e-07	zeta toxin family protein	YP_002405395	0.0
499	Failed to assign function	23	0	10	21	23	18	8.52366556181328e-08	4.23195601501987e-06	spfh domain band 7 family protein	WP_001495680	0.0
500	COG3541: Predicted nucleotidyltransferase	23	0	10	21	23	18	8.52366556181328e-08	4.23195601501987e-06	nucleotidyltransferase family protein	WP_001697550	0.0
501	Putative fimbrial protein	22	1	10	21	23	18	1.33279134239262e-06	5.16251188066719e-05	fimbrial family protein	YP_002295290	0.0
502	FIG00641476: hypothetical protein	23	0	11	20	23	18	2.63458753728775e-07	1.17008764408956e-05	hypothetical protein EC55989_3343	YP_002404308	7.50484E-45
503	Type III secretion bridge between inner and outermembrane lipoprotein (YscJ,HrcJ,EscJ, PscJ)	23	0	14	17	23	18	5.62383108921037e-06	0.000192087734224377	type iii secretion apparatus family	WP_001544061	1.55284E-162
504	Failed to assign function	23	0	17	14	23	18	8.17109575902916e-05	0.00225297693752774	hpch aldolase citrate lyase family protein	YP_689734	4.71616E-73
505	Oxygen-regulated invasion protein OrgA	23	0	18	13	23	18	0.000186119403400109	0.0047678247168879	type iii secretion apparatus protein	YP_002404119	2.32244E-133

506	Hypothetical response regulatory protein ygeK	23	0	18	13	23	18	0.000186119403400109	0.0047678247168879	two-component system response regulator	WP_000516293	4.54001E-101
507	Type III secretion inner membrane protein (YscU,SpaS,EscU,HrcU,SsaU, homologous to flagellar export components)	23	0	18	13	23	18	0.000186119403400109	0.0047678247168879	type iii secretion protein	WP_024182618	1.82008E-76
508	Fimbriae-like adhesin SfmA	23	0	19	12	23	18	0.000411421839094978	0.00977668212375699	fimbrial family protein	WP_000827804	6.17523E-110
509	Phosphate starvation-inducible protein PhoH, predicted ATPase	23	0	1	30	24	17	2.21008724063083e-14	1.27725361810537e-11	protein phoh	WP_001556530	0.0
510	YjbI protein	23	0	5	26	24	17	9.0503072503832e-11	8.65952577175739e-09	pentapeptide repeats family protein	WP_001345141	0.0
511	Uncharacterized protein YgcG	23	0	16	15	24	17	3.47271569758741e-05	0.00101979260973055	repair family protein	WP_001341838	1.49138E-149
512	FIG00638895: hypothetical protein	21	2	4	27	25	16	4.79119287678257e-09	3.11816012088984e-07	phage protein	YP_003227663	5.40025E-110
513	Phosphate starvation-inducible protein PhoH, predicted ATPase	23	0	6	25	25	16	4.37431517101856e-10	3.41622192383115e-08	protein phoh	WP_001323668	4.16995E-61
514	Maltose-6'-phosphate glucosidase (EC 3.2.1.122)	23	0	8	23	25	16	7.26448769472722e-09	4.60339114971135e-07	6-phospho-alpha-glucosidase	WP_001740797	1.0174E-30
515	Failed to assign function	21	2	9	22	25	16	3.76753516658154e-06	0.000138507247040128	yrhd protein	YP_002388907	4.42379E-23
516	Failed to assign function	23	0	11	20	25	16	2.63458753728775e-07	1.17008764408956e-05	hypothetical protein	YP_005276598	1.8705E-13
517	Ribose/xylose/arabinose/galactoside ABC-type transport systems, periplasmic sugar binding protein	23	0	13	18	25	16	2.12793608780932e-06	7.90345002485066e-05	periplasmic binding substrate ribose	WP_000981360	0.0
518	FIG00638628: hypothetical protein	23	0	16	15	25	16	3.47271569758741e-05	0.00101979260973055	pf08786 domain protein	WP_001719810	2.86788E-94
519	Hydrogenase-4 component F (EC 1.-.-.-)	23	0	18	13	25	16	0.000186119403400109	0.0047678247168879	nadh-ubiquinone plastoquinone (complex i) various chains family protein	WP_000122582	0.0
520	Holliday junction resolvase / Crossover junction endodeoxyribonuclease rusA (EC 3.1.22.-)	21	2	2	29	26	15	1.08993214880759e-10	9.78095632669076e-09	crossover junction endodeoxyribo-nuclease rusa	NP_415082	3.28072E-81
521	Protein NinE	21	2	4	27	26	15	4.79119287678257e-09	3.11816012088984e-07	protein nine	WP_000518235	5.41651E-27
522	Putative exported protein precursor	23	0	7	24	26	15	1.87470650186509e-09	1.34087918509637e-07	outer membrane protein	AAN81391	1.49824E-140
523	FIG00638961: hypothetical protein	22	1	7	24	26	15	3.63693061361827e-08	2.03668114362623e-06	conserved protein	NP_417940	6.75275E-84
524	FIG00641784: hypothetical protein	23	0	9	22	26	15	2.58292895812523e-08	1.4867791867328e-06	hypothetical protein Z1194	NP_286729	1.06598E-111
525	Aldose-ketose isomerase YihS	23	0	10	21	26	15	8.52366556181328e-08	4.23195601501987e-06	glucosamine isomerase	CDL26975	0.0
526	Alpha-glucosyltransferase YihQ	22	1	10	21	26	15	1.33279134239262e-06	5.16251188066719e-05	glycosyl hydrolases 31 family protein	YP_003231615	0.0
527	Putative fimbrial chaperone protein	23	0	16	15	26	15	3.47271569758741e-05	0.00101979260973055	gram-negative pili assembly c-terminal domain protein	YP_001465217	9.51159E-168
528	type 1 fimbriae anchoring protein FimD	23	0	16	15	26	15	3.47271569758741e-05	0.00101979260973055	type vii secretion system usher family protein	YP_003231750	0.0
529	Failed to assign function	23	0	19	12	26	15	0.000411421839094978	0.00977668212375699	family type iii secretion apparatus protein	WP_024229026	5.54128E-81
530	Xanthosine phosphorylase (EC 2.4.2.1)	21	2	7	24	27	14	3.63473268185746e-07	1.58654434403252e-05	xanthosine phosphorylase	NP_416902	0.0
531	Xanthosine operon regulatory protein XapR, LysR family	22	1	8	23	27	14	1.31229455130557e-07	6.44899036641593e-06	hth-type transcriptional regulator xapr	WP_000442952	0.0
532	Sugar-1-epimerase YihR	23	0	10	21	27	14	8.52366556181328e-08	4.23195601501987e-06	aldose-1-epimerase	WP_000430815	0.0
533	Glucuronide transport protein YihO	23	0	10	21	27	14	8.52366556181328e-08	4.23195601501987e-06	inner membrane symporter yihp	WP_024241265	0.0
534	Outer membrane sugar transport protein YshA	23	0	10	21	27	14	8.52366556181328e-08	4.23195601501987e-06	porin ompl	AAB03009	3.42808E-157
535	Secreted protein Hcp	23	0	14	17	27	14	5.62383108921037e-06	0.000192087734224377	type vi secretion system hcp1 family protein	WP_001284197	5.4358E-108
536	Hypothetical response regulatory protein ygeK	23	0	18	13	27	14	0.000186119403400109	0.0047678247168879	two-component system response regulator	YP_001459625	1.09901E-28
537	Colicin I receptor precursor	20	3	10	21	28	13	5.97666431229269e-05	0.0017270169196801	#NAME?	NP_286713	0.0

538	CRISPR-associated protein, Cas2	23	0	19	12	28	13	0.000411421839094978	0.00977668212375699	crispr-associated endoribonuclease subtype i-e	NP_311635	3.26125E-60
539	Phenylacetate-CoA oxygenase, PaaI subunit	23	0	19	12	28	13	0.000411421839094978	0.00977668212375699	phenylacetic acid degradation protein paac	YP_003229012	4.15219E-173
540	Maltose-6'-phosphate glucosidase (EC 3.2.1.122)	23	0	10	21	29	12	8.52366556181328e-08	4.23195601501987e-06	6-phospho-alpha-glucosidase	YP_002295243	8.16711E-159
541	YjbI protein	22	1	10	21	29	12	1.33279134239262e-06	5.16251188066719e-05	pentapeptide repeats family protein	WP_000623317	1.47615E-19
542	D-serine permease DsdX	23	0	11	20	29	12	2.63458753728775e-07	1.17008764408956e-05	permease	WP_021567503	1.17163E-27
543	FIG00638409: hypothetical protein	23	0	13	18	29	12	2.12793608780932e-06	7.90345002485066e-05	hypothetical protein ECBD_3997	YP_003038167	2.30356E-25
544	Putative SAM-dependent methyltransferases	23	0	14	17	29	12	5.62383108921037e-06	0.000192087734224377	methyltransferase domain protein	NP_287402	1.70593E-161
545	3-hydroxyacyl-CoA dehydrogenase PaaC (EC 1.1.1.-)	22	1	15	16	29	12	0.000152218361481295	0.00405882496068959	3-hydroxyacyl- dehydrogenase	YP_002292755	0.0
546	Phenylacetate degradation enoyl-CoA hydratase PaaB (EC 4.2.1.17)	22	1	16	15	29	12	0.000341928930223992	0.00845922805458259	acyl- hydratase	YP_001462674	2.49748E-165
547	Transcriptional activator feaR	23	0	17	14	29	12	8.17109575902916e-05	0.00225297693752774	transcriptional activator fear	YP_003229006	0.0
548	Protein YjgL, putative CCAAT-box DNA binding protein subunit B	23	0	18	13	29	12	0.000186119403400109	0.0047678247168879	protein ccaat-box dna binding protein subunit b	YP_003232297	0.0
549	Failed to assign function	23	0	13	18	30	11	2.12793608780932e-06	7.90345002485066e-05	invasion protein	WP_024169097	2.81981E-41
550	FIG00638289: hypothetical protein	23	0	15	16	30	11	1.42470387593329e-05	0.000445543757564593	methylglyoxal detoxification protein	WP_001701731	0.0
551	YqeJ protein	22	1	10	21	31	10	1.33279134239262e-06	5.16251188066719e-05	protein	YP_311837	1.86382E-109
552	Sucrose-6-phosphate hydrolase (EC 3.2.1.26)	23	0	11	20	31	10	2.63458753728775e-07	1.17008764408956e-05	sucrose-6-phosphate hydrolase	YP_001463698	0.0
553	Sucrose specific transcriptional regulator CscR, LacI family	23	0	11	20	31	10	2.63458753728775e-07	1.17008764408956e-05	sucrose operon repressor	YP_002387820	0.0
554	Hydrogenase-4 component H (EC 1.-.-.-)	23	0	16	15	31	10	3.47271569758741e-05	0.00101979260973055	4fe-4s binding domain protein	WP_024245115	6.01748E-96
555	Fructokinase (EC 2.7.1.4)	21	2	8	23	32	9	1.22212116856585e-06	4.79815397919548e-05	fructokinase	YP_003230363	0.0
556	Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	23	0	17	14	32	9	8.17109575902916e-05	0.00225297693752774	enterochelin esterase	WP_000244946	0.0
557	Shikimate kinase I (EC 2.7.1.71)	17	6	6	25	33	8	7.3972904488057e-05	0.00207930063043472	shikimate kinase 1	WP_001503195	1.59222E-112
558	Formate hydrogenlyase transcriptional activator	23	0	17	14	33	8	8.17109575902916e-05	0.00225297693752774	dna-binding transcriptional activator	YP_003230476	0.0
559	FIG00638493: hypothetical protein	22	1	12	19	34	7	1.03627109799984e-05	0.000338734045789632	impa-related family protein	YP_002385707	0.0
560	FIG00638475: hypothetical protein	23	0	15	16	34	7	1.42470387593329e-05	0.000445543757564593	sugar phosphate isomerase	NP_418309	0.0
561	orf, hypothetical protein	23	0	19	12	34	7	0.000411421839094978	0.00977668212375699	conserved protein	WP_009008192	4.15264E-128
562	FIG00638952: hypothetical protein	23	0	19	12	34	7	0.000411421839094978	0.00977668212375699	inner membrane protein yidi	YP_001465161	1.62053E-67
563	FIG00638983: hypothetical protein	23	0	16	15	35	6	3.47271569758741e-05	0.00101979260973055	lipoprotein	NP_289509	1.0413E-163
564	FIG00639945: hypothetical protein	23	0	17	14	35	6	8.17109575902916e-05	0.00225297693752774	hok gef family protein	WP_000078919	9.93608E-19
565	Putative resistance protein	23	0	15	16	36	5	1.42470387593329e-05	0.000445543757564593	inner membrane protein yihn	WP_000956332	0.0
566	Hypothetical transcriptional regulator yihL	23	0	15	16	36	5	1.42470387593329e-05	0.000445543757564593	transcriptional regulator	YP_003231621	3.97526E-170
567	FIG00639909: hypothetical protein	23	0	11	20	37	4	2.63458753728775e-07	1.17008764408956e-05	family transcriptional regulator	WP_001318027	4.88819E-31
568	L-xylulose/3-keto-L-gulonate kinase (EC 2.7.1.-)	23	0	16	15	37	4	3.47271569758741e-05	0.00101979260973055	l-xylulose 3-keto-l-gulonate kinase	YP_003231915	0.0
569	Putative lipase in cluster with Phosphatidate cytidylyltransferase	23	0	18	13	37	4	0.000186119403400109	0.0047678247168879	hydrolase	YP_003229027	0.0
570	FIG00642443: hypothetical protein	22	1	13	18	38	3	2.65992010976167e-05	0.000803985894264363	dna-binding protein	YP_002404309	1.56923E-143
571	D-galactonate transporter	23	0	15	16	38	3	1.42470387593329e-05	0.000445543757564593	d-galactonate transporter	WP_001543316	0.0
572	Ser/Thr and Tyr protein phosphatase (dual specificity)	23	0	16	15	39	2	3.47271569758741e-05	0.00101979260973055	dual specificity catalytic domain protein	WP_001375409	0.0
573	putative factor	23	0	19	12	39	2	0.000411421839094978	0.00977668212375699	inner membrane protein ycfz	WP_000076380	2.94628E-176
574	FIG00639097: hypothetical protein	23	0	19	12	39	2	0.000411421839094978	0.00977668212375699	inner membrane protein ymfa	WP_000076380	5.23583E-97
575	FIG00637868: hypothetical protein	23	0	19	12	40	1	0.000411421839094978	0.00977668212375699	orthopoxvirus pf05708 family	WP_009008108	9.31528E-149
576	Allantoinase (EC 3.5.2.5)	23	0	19	12	41	0	0.000411421839094978	0.00977668212375699	allantoinase	YP_002291812	0.0

List of overrepresented genes in LEE positive STEC other than STEC in HUS-group 2

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR<0.01	Function Blast2GO	Hit Accession	E-value
1	Failed to assign function	0	23	31	0	40	1	9.20869683596171e-16	1.33047251885975e-11	acetyl- :acetoacetyl- transferase alpha beta subunit	WP_000805688	0.0
2	Putative sucrose phosphorylase (EC 2.4.1.7)	0	23	30	1	35	6	2.21008724063082e-14	7.98283511315854e-11	glucosyltransferase	WP_000810512	0.0
3	Biofilm PGA outer membrane secretin PgaA	0	23	22	9	32	9	2.58292895812524e-08	6.32511145542261e-06	poly-beta- n-acetyl-d-glucosamine export porin	WP_001609828	0.0
4	Biofilm PGA synthesis N-glycosyltransferase PgaC (EC 2.4.-.-)	0	23	25	6	31	10	4.37431517101856e-10	1.50476441883039e-07	poly-beta- -n-acetyl-d-glucosamine synthase	WP_000610465	0.0
5	Biofilm PGA synthesis auxiliary protein PgaD	0	23	25	6	31	10	4.37431517101856e-10	1.50476441883039e-07	biofilm pga synthesis protein	WP_001061093	3.35608E-67
6	Sensory box/GGDEF domain protein	0	23	16	15	31	10	1.42470387593329e-05	0.0013542185262818 6	diguanylate cyclase	YP_003221048	0.0
7	YcgN (Fragment)	0	23	30	1	30	11	2.21008724063082e-14	7.98283511315854e-11	upf0260 protein ycgn	WP_001335675	3.68617E-109
8	Biofilm PGA synthesis deacetylase PgaB (EC 3.-)	0	23	17	14	30	11	5.62383108921037e-06	0.000650024892615291	poly-beta- -n-acetyl-d-glucosamine n-deacetylase	NP_415542	0.0
9	Uncharacterized protein YcdU	0	23	23	8	29	12	7.26448769472722e-09	1.94365404098924e-06	inner membrane protein	YP_002292369	0.0
10	FIG00640692: hypothetical protein	0	23	29	2	28	13	2.76260905078851e-13	6.65236259429873e-10	hicb family protein	NP_290055	7.63796E-78
11	Putative HTH-type transcriptional regulator YdjF	0	23	26	5	27	14	9.05030725038321e-11	4.50892548805299e-08	transcriptional regulator	WP_000719082	7.65921E-165
12	Putative oxidoreductase YdjL	0	23	26	5	27	14	9.05030725038321e-11	4.50892548805299e-08	zn-dependent and nad -binding	WP_000645214	0.0
13	Putative transport protein YdjK, MFS superfamily	0	23	25	6	27	14	4.37431517101856e-10	1.50476441883039e-07	inner membrane metabolite transporter ydje	YP_002556655	0.0
14	Uncharacterized sugar kinase YdjH	0	23	25	6	27	14	4.37431517101856e-10	1.50476441883039e-07	kinase	WP_001602686	0.0
15	Putative transport protein YdjK, MFS superfamily	0	23	25	6	27	14	4.37431517101856e-10	1.50476441883039e-07	metabolite transporter	WP_023352181	0.0
16	Flagellar hook-associated protein FliD	0	23	24	7	27	14	1.87470650186509e-09	5.41715190778936e-07	flagellar hook-associated protein 2	WP_000146822	0.0
17	Uncharacterized lipoprotein yaeF precursor	0	23	18	13	27	14	2.12793608780932e-06	0.000330585167706119	lipoprotein	WP_001396931	0.0
18	Galactoside O-acetyltransferase (EC 2.3.1.18)	0	23	30	1	26	15	2.21008724063082e-14	7.98283511315854e-11	galactoside o-acetyltransferase	WP_001250720	2.99966E-144
19	Hypothetical zinc-type alcohol dehydrogenase-like protein YdjJ	0	23	26	5	26	15	9.05030725038321e-11	4.50892548805299e-08	sorbitol dehydrogenase	WP_000798131	0.0
20	Hypothetical oxidoreductase YdjG (EC 1.-.-.-)	0	23	26	5	25	16	9.05030725038321e-11	4.50892548805299e-08	oxidoreductase	NP_416285	0.0
21	Putative aldolase YdjI	0	23	26	5	23	18	9.05030725038321e-11	4.50892548805299e-08	aldolase	NP_416287	0.0
22	FIG00640016: hypothetical protein	0	23	26	5	23	18	9.05030725038321e-11	4.50892548805299e-08	bacterial regulatory tetr family protein	YP_006118096	1.56901E-153
23	3-hydroxybutyryl-CoA dehydratase (EC 4.2.1.55)	0	23	26	5	23	18	9.05030725038321e-11	4.50892548805299e-08	enoyl- hydratase isomerase family protein	WP_001283541	0.0
24	4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)	0	23	25	6	23	18	4.37431517101856e-10	1.50476441883039e-07	xylose isomerase-like tim barrel family protei	YP_410494	0.0
25	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	0	23	26	5	22	19	9.05030725038321e-11	4.50892548805299e-08	short chain dehydrogenase family protein	YP_002394282	0.0
26	FIG00638522: hypothetical protein	0	23	25	6	22	19	4.37431517101856e-10	1.50476441883039e-07	hypothetical protein	YP_005276740	1.67935E-24
27	Failed to assign function	0	23	25	6	22	19	4.37431517101856e-10	1.50476441883039e-07	coenzyme a transferase family protein	WP_001226986	0.0
28	Minor fimbrial subunit StfF	0	23	21	10	20	21	8.52366556181328e-08	1.89461415441659e-05	fimbrial family protein	YP_002293919	7.32378E-104
29	Type III secretion inner membrane protein (YscU,SpaS,EscU,HrcU,SsaU, homologous to flagellar export components)	0	23	17	14	20	21	5.62383108921037e-06	0.000650024892615291	family type iii secretion protein	WP_001613736	9.60866E-153
30	HicA-like protein	0	23	29	2	19	22	2.76260905078851e-13	6.65236259429873e-10	#NAME?	YP_003080264	3.93181E-58
31	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	0	23	26	5	19	22	9.05030725038321e-11	4.50892548805299e-08	pyridine nucleotide-disulfide oxidoreductase family protein	WP_000080323	0.0
32	Putative cytoplasmic protein	0	23	25	6	19	22	4.37431517101856e-10	1.50476441883039e-07	protein yoag	YP_006087527	1.25649E-31
33	Uncharacterized protein YadU in stf fimbrial cluster	0	23	20	11	19	22	2.63458753728775e-07	5.44926531390282e-05	fimbrial protein	WP_001397279	2.10091E-156

34	L-idonate 5-dehydrogenase (EC 1.1.1.264)	0	23	19	12	17	24	7.6842136504226e-07	0.000132168474787269	l-idonate 5-dehydrogenase	WP_001197417	0.0
35	5-keto-D-gluconate 5-reductase (EC 1.1.1.69)	0	23	19	12	17	24	7.6842136504226e-07	0.000132168474787269	gluconate 5-dehydrogenase	NP_418687	0.0
36	Positive regulator of L-idonate catabolism	0	23	19	12	17	24	7.6842136504226e-07	0.000132168474787269	hth-type transcriptional regulator idnr	NP_418685	0.0
37	Phosphate starvation-inducible protein PhoH, predicted ATPase	0	23	25	6	16	25	4.37431517101856e-10	1.50476441883039e-07	protein phoh	NP_309293	0.0
38	Hydrogenase-4 component B (EC 1.-.-.-) / Formate hydrogenlyase subunit 3	0	23	19	12	16	25	7.6842136504226e-07	0.000132168474787269	hydrogenase-4 component b	WP_001443473	0.0
39	Gluconokinase (EC 2.7.1.12)	0	23	19	12	16	25	7.6842136504226e-07	0.000132168474787269	thermosensitive gluconokinase	WP_000896732	5.4444E-136
40	Uncharacterized protein YmdE	0	23	17	14	16	25	5.62383108921037e-06	0.000650024892615291	acyl transferase domain protein	WP_000322270	1.2335E-117
41	Hydrogenase-4 component B (EC 1.-.-.-) / Formate hydrogenlyase subunit 3	0	23	19	12	15	26	7.6842136504226e-07	0.000132168474787269	nadh-ubiquinone plastoquinone (complex i) various chains family protein	WP_001443474	9.64297E-89
42	Failed to assign function	0	23	16	15	15	26	1.42470387593329e-05	0.00135421852628186	major tail sheath protein	WP_001286693	0.0
43	FIG00638890: hypothetical protein	0	23	24	7	14	27	1.87470650186509e-09	5.41715190778936e-07	flxa-like family protein	YP_311362	2.67098E-12
44	Alcohol dehydrogenase (EC 1.1.1.1)	0	23	17	14	14	27	5.62383108921037e-06	0.000650024892615291	chlorophyll synthesis pathway protein	WP_000357974	0.0
45	Shikimate transporter	0	23	17	14	14	27	5.62383108921037e-06	0.000650024892615291	shikimate transporter	YP_312804	0.0
46	Phage-related capsid packaging protein	0	23	16	15	14	27	1.42470387593329e-05	0.00135421852628186	presumed portal vertex protein	WP_000038202	0.0
47	Putative transport protein	0	23	16	15	14	27	1.42470387593329e-05	0.00135421852628186	major facilitator superfamily transporter	NP_756684	0.0
48	TRAP-type transport system, small permease component, predicted N-acetylneuraminate transporter	0	23	21	10	13	28	8.52366556181328e-08	1.89461415441659e-05	tripartite atp-independent periplasmic family	WP_001375330	2.55737E-76
49	TRAP transporter solute receptor, unknown substrate 6	0	23	21	10	13	28	8.52366556181328e-08	1.89461415441659e-05	trap transporter solute family protein	WP_000695891	0.0
50	Phage terminase, ATPase subunit	0	23	16	15	13	28	1.42470387593329e-05	0.00135421852628186	atpase subunit of terminase family protein	WP_000156851	0.0
51	probable integral membrane protein Cj0014c	0	23	15	16	13	28	3.47271569758741e-05	0.00261321856243452	membrane protein	YP_313098	2.09259E-126
52	FIG00639506: hypothetical protein	0	23	25	6	12	29	4.37431517101856e-10	1.50476441883039e-07	mrna interferase	NP_289657	2.98362E-62
53	Baseplate assembly protein J	0	23	16	15	12	29	1.42470387593329e-05	0.00135421852628186	baseplate assembly protein	YP_006141159	0.0
54	FIG00641630: hypothetical protein	0	23	15	16	12	29	3.47271569758741e-05	0.00261321856243452	mcrbc 5-methylcytosine restriction system component family protein	YP_005279643	0.0
55	Ankyrin-repeat protein A	0	23	17	14	11	30	5.62383108921037e-06	0.000650024892615291	ankyrin repeat protein	WP_000632921	0.0
56	Putative regulatory protein	0	23	17	14	11	30	5.62383108921037e-06	0.000650024892615291	bacterial regulatory helix-turn-helix family protein	WP_021557314	0.0
57	Putative glycosyl hydrolase of unknown function (DUF1680)	0	23	17	14	11	30	5.62383108921037e-06	0.000650024892615291	cytoplasmic protein	WP_001026876	0.0
58	FIG00638448: hypothetical protein	0	23	15	16	11	30	3.47271569758741e-05	0.00261321856243452	aaa domain family protein	YP_001461068	0.0
59	hypothetical protein	0	23	15	16	11	30	3.47271569758741e-05	0.00261321856243452	hypothetical protein EFER_3103	YP_002384202	2.98479E-67
60	UDP-glucose 4-epimerase (EC 5.1.3.2)	0	23	14	17	11	30	8.17109575902916e-05	0.00534190006907028	udp-glucose 4-epimerase	YP_003222419	0.0
61	Putative permease	0	23	17	14	10	31	5.62383108921037e-06	0.000650024892615291	sugar (glycoside-pentoside-hexuronide) transporter domain protein	WP_000204832	0.0
62	Sugar-1-epimerase YihR	0	23	17	14	10	31	5.62383108921037e-06	0.000650024892615291	aldose 1-epimerase family protein	WP_001397472	0.0
63	hypothetical protein	0	23	15	16	10	31	3.47271569758741e-05	0.00261321856243452	membrane protein	WP_000449994	0.0
64	FIG00967058: hypothetical protein	0	23	15	16	10	31	3.47271569758741e-05	0.00261321856243452	hypothetical protein	WP_000007269	0.0
65	hypothetical protein	0	23	15	16	10	31	3.47271569758741e-05	0.00261321856243452	hypothetical protein	YP_005279649	0.0
66	hypothetical protein	0	23	15	16	10	31	3.47271569758741e-05	0.00261321856243452	hypothetical protein	YP_005279650	0.0
67	Bipolar DNA helicase	0	23	15	16	10	31	3.47271569758741e-05	0.00261321856243452	aaa-like domain protein	YP_005279651	0.0
68	hypothetical protein	0	23	15	16	10	31	3.47271569758741e-05	0.00261321856243452	hypothetical protein	YP_005279652	2.7731E-116

69	serine/threonine protein kinase with TPR repeats	0	23	15	16	10	31	3.47271569758741e-05	0.00261321856243452	kinase domain protein	WP_001022617	0.0
70	D-serine dehydratase transcriptional activator	0	23	21	10	9	32	8.52366556181328e-08	1.89461415441659e-05	d-serine deaminase transcriptional activator	YP_003000013	0.0
71	D-serine permease DsdX	0	23	18	13	9	32	2.12793608780932e-06	0.000330585167706119	permease	WP_001592429	0.0
72	Failed to assign function	0	23	16	15	9	32	1.42470387593329e-05	0.00135421852628186	hypothetical protein ECO103_3813	YP_003223658	5.90343E-139
73	COG0042: tRNA-dihydrouridine synthase	0	23	16	15	9	32	1.42470387593329e-05	0.00135421852628186	trna-dihydrouridine synthase	YP_003223657	0.0
74	Putative DNA binding protein	0	23	16	15	9	32	1.42470387593329e-05	0.00135421852628186	transcriptional regulator	WP_000270113	1.33339E-37
75	FIG00642194: hypothetical protein	0	23	16	15	9	32	1.42470387593329e-05	0.00135421852628186	protein	WP_001406070	1.30244E-16
76	N-6 DNA methylase	0	23	15	16	9	32	3.47271569758741e-05	0.00261321856243452	n-6 dna methylase family protein	YP_005279645	0.0
77	FIG00643641: hypothetical protein	0	23	14	17	8	33	8.17109575902916e-05	0.00534190006907028	hypothetical protein	WP_001012978	0.0
78	Regulatory protein cro	0	23	23	8	7	34	7.26448769472722e-09	1.94365404098924e-06	regulatory protein cro	0611193A	1.68316E-35
79	FIG00639676: hypothetical protein	0	23	21	10	7	34	8.52366556181328e-08	1.89461415441659e-05	valyl-trna synthetase	YP_001272548	1.36792E-139
80	Failed to assign function	0	23	19	12	7	34	7.6842136504226e-07	0.000132168474787269	hypothetical protein	WP_001542619	7.10515E-12
81	Mobile element protein	0	23	16	15	6	35	1.42470387593329e-05	0.00135421852628186	transposase	AHG14388	0.0
82	Mobile element protein	0	23	14	17	6	35	8.17109575902916e-05	0.00534190006907028	integrase core domain protein	BAO24607	2.88536E-79
83	Flagellar biosynthesis protein FliC	0	23	14	17	6	35	8.17109575902916e-05	0.00534190006907028	flagellin	WP_000079774	0.0
84	Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase) (GLYCOGENASE)	0	23	19	12	5	36	7.6842136504226e-07	0.000132168474787269	alpha catalytic domain protein	WP_001493308	0.0
85	Phage antitermination protein N	0	23	19	12	5	36	7.6842136504226e-07	0.000132168474787269	antitermination protein n	WP_000512959	1.00049E-50
86	Maltose-6'-phosphate glucosidase (EC 3.2.1.122)	0	23	17	14	5	36	5.62383108921037e-06	0.000650024892615291	6-phospho-alpha-glucosidase	WP_000160100	0.0
87	Failed to assign function	0	23	16	15	5	36	1.42470387593329e-05	0.00135421852628186	hypothetical protein	WP_000961333	0.0
88	hypothetical protein	0	23	15	16	4	37	3.47271569758741e-05	0.00261321856243452	chat domain protein	WP_000421881	0.0
89	Gene D protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage late control d family protein	WP_000882931	0.0
90	Phage tail protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage p2 family protein	WP_000978920	7.90888E-100
91	Phage protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	tail fiber protein t (tape measure)	WP_000069973	0.0
92	Tail protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	mu-like prophage u gp41 family protein	WP_001336821	6.11894E-46
93	Failed to assign function	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage major tail tube protein	NP_046779	1.33676E-118
94	Failed to assign function	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	baseplate assembly protein w	NP_046772	7.74489E-69
95	Baseplate assembly protein V	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage baseplate assembly v family protein	YP_006100175	2.93934E-138
96	Failed to assign function	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage virion morphogenesis protein	WP_001001776	3.30443E-85
97	Phage tail protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	p2 phage tail completion r family protein	WP_001719223	5.00107E-78
98	Phage outer membrane lipoprotein Rz1	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage lysis protein	YP_002391898	2.16772E-25
99	Phage holin	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage holin 2 family protein	NP_046764	3.10821E-44
100	Phage head completion-stabilization protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	head completion stabilization protein	YP_852010	2.39322E-109
101	Phage major capsid protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	capsid protein	WP_001248555	0.0
102	Phage capsid scaffolding protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage capsid scaffolding protein	YP_002391908	0.0
103	Phage replication protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	replication a protein	WP_000268563	0.0
104	FIG00639324: hypothetical protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	phage conjugal plasmid c-4 type zinc finger family	WP_021519099	5.4255E-57
105	FIG00640077: hypothetical protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	hypothetical protein	WP_001277957	4.99464E-44
106	FIG00638705: hypothetical protein	0	23	15	16	3	38	3.47271569758741e-05	0.00261321856243452	uncharacterized protein in 5 region	YP_001462107	4.83804E-37
107	Phage terminase, endonuclease subunit	0	23	14	17	3	38	8.17109575902916e-05	0.00534190006907028	small terminase subunit	WP_021534006	3.38614E-101
108	putative phage inhibition, colicin resistance and tellurite resistance protein	0	23	14	17	3	38	8.17109575902916e-05	0.00534190006907028	tellurium resistance protein	WP_001053348	0.0
109	COG0582: Integrase	0	23	16	15	2	39	1.42470387593329e-05	0.00135421852628186	integrase	YP_003223659	0.0
110	Probable transmembrane protein	0	23	16	15	2	39	1.42470387593329e-05	0.00135421852628186	hcp1 family type vi secretion system effector	WP_001612076	2.90788E-107

111	Replication gene B protein	0	23	15	16	2	39	3.47271569758741e-05	0.00261321856243452	replication b protein	WP_001526247	1.87202E-110
112	Failed to assign function	0	23	22	9	1	40	2.58292895812524e-08	6.32511145542261e-06	rio1 family protein	WP_021569006	3.55009E-33
113	Mobile element protein	0	23	22	9	1	40	2.58292895812524e-08	6.32511145542261e-06	transposase-like protein	WP_000343735	2.52131E-78
114	FIG01069681: hypothetical protein	0	23	18	13	1	40	2.12793608780932e-06	0.000330585167706119	hypothetical protein	WP_001597900	5.50046E-110
115	Failed to assign function	0	23	18	13	1	40	2.12793608780932e-06	0.000330585167706119	enterobacteria phage	2M7A_A	7.71285E-56
116	D-mannonate oxidoreductase (EC 1.1.1.57)	0	23	17	14	1	40	5.62383108921037e-06	0.000650024892615291	d-mannonate oxidoreductase	WP_001397145	0.0
117	FIG00639755: hypothetical protein	0	23	17	14	1	40	5.62383108921037e-06	0.000650024892615291	hypothetical protein	WP_001014903	2.31112E-50
118	Failed to assign function	0	23	17	14	1	40	5.62383108921037e-06	0.000650024892615291	hypothetical protein G2583_pO550126	YP_003502739	2.7531E-106
119	Putative cytoplasmic protein	0	23	17	14	1	40	5.62383108921037e-06	0.000650024892615291	cytoplasmic protein	WP_001307350	6.22336E-126
120	Cox	0	23	16	15	1	40	1.42470387593329e-05	0.00135421852628186	cox protein	NP_046788	5.03626E-56
121	Failed to assign function	0	23	16	15	1	40	1.42470387593329e-05	0.00135421852628186	repressor protein c	NP_046787	6.75695E-56
122	Phage integrase	0	23	14	17	1	40	8.17109575902916e-05	0.00534190006907028	integrase	WP_000985256	0.0
123	Glycosyltransferase (EC 2.4.1.-)	0	23	14	17	1	40	8.17109575902916e-05	0.00534190006907028	wbtg	WP_000866254	0.0
124	FIG00638365: hypothetical protein	0	23	14	17	1	40	8.17109575902916e-05	0.00534190006907028	wzy	WP_000275678	1.90121E-161
125	Putative glycosyl transferase	0	23	14	17	1	40	8.17109575902916e-05	0.00534190006907028	wbte	WP_024199506	0.0
126	Putative glycosyl transferase	0	23	14	17	1	40	8.17109575902916e-05	0.00534190006907028	wbtd	WP_000587273	0.0
127	WzxE protein	0	23	14	17	1	40	8.17109575902916e-05	0.00534190006907028	wzx	YP_003222425	0.0
128	putative butyryltransferase	0	23	14	17	1	40	8.17109575902916e-05	0.00534190006907028	wbtb	YP_003222427	1.07321E-104
129	hypothetical protein	0	23	20	11	0	41	2.63458753728775e-07	5.44926531390282e-05	hypothetical protein	WP_000222754	7.50187E-52
130	FIG00641015: hypothetical protein	0	23	19	12	0	41	7.6842136504226e-07	0.000132168474787269	pf11726 domain protein	WP_000287900	9.57999E-92
131	Failed to assign function	0	23	17	14	0	41	5.62383108921037e-06	0.000650024892615291	hypothetical protein	WP_000460278	4.81775E-82
132	Haemolysin expression modulating protein paralog	0	23	16	15	0	41	1.42470387593329e-05	0.00135421852628186	ydfa protein	YP_003223436	1.74231E-30
133	hypothetical protein	0	23	15	16	0	41	3.47271569758741e-05	0.00261321856243452	surface lpxtg-motif cell wall anchor domain protein	WP_001223333	9.14223E-93
134	FIG00640631: hypothetical protein	0	23	15	16	0	41	3.47271569758741e-05	0.00261321856243452	t3ss effector protein	YP_003229591	1.58181E-131
135	Failed to assign function	0	23	15	16	0	41	3.47271569758741e-05	0.00261321856243452	hypothetical protein	WP_001084135	0.0
136	Mobile element protein	0	23	14	17	0	41	8.17109575902916e-05	0.00534190006907028	is911 orf1	YP_006162123	1.37158E-21
137	Protein YjgL, putative CCAAT-box DNA binding protein subunit B	0	23	14	17	0	41	8.17109575902916e-05	0.00534190006907028	dna-binding protein	YP_001882963	8.84755E-60
138	Flagellar biosynthesis protein FlhA	1	22	24	7	32	9	3.63693061361827e-08	8.75772891759278e-06	fhipep family protein	YP_002291526	0.0
139	Flagellar motor rotation protein MotB	1	22	21	10	29	12	1.33279134239262e-06	0.000221335279481478	ompa family protein	WP_021553260	2.36463E-139
140	hypothetical protein	1	22	29	2	20	21	7.64690185258261e-12	1.00438579969194e-08	hypothetical protein	YP_005277482	7.30809E-18
141	hypothetical protein	1	22	20	11	16	25	3.82790071594159e-06	0.000537760150049888	phage protein	WP_001087382	4.80289E-71
142	FIG00642948: hypothetical protein	1	22	17	14	15	26	6.50540461130282e-05	0.00467612367284095	conserved protein	YP_002291802	2.5244E-95
143	Failed to assign function	1	22	28	3	12	29	6.16982688009435e-11	4.45708293818016e-08	integrase	WP_001218282	0.0
144	Antirestriction protein klcA	1	22	19	12	11	30	1.03627109799984e-05	0.00115169575568475	antirestriction protein klca	WP_000680587	8.09844E-51
145	Phage NinB DNA recombination	1	22	21	10	9	32	1.33279134239262e-06	0.000221335279481478	protein ninb	NP_040634	6.14393E-95
146	Failed to assign function	1	22	18	13	8	33	2.65992010976167e-05	0.00246349524011773	hypothetical protein ECED1_1027	YP_002397051	8.10207E-46
147	FIG00639473: hypothetical protein	1	22	19	12	6	35	1.03627109799984e-05	0.00115169575568475	hypothetical protein	WP_000581110	0.0
148	FIG00638659: hypothetical protein	1	22	18	13	6	35	2.65992010976167e-05	0.00246349524011773	phage protein	WP_000189903	1.17005E-119
149	Phage repressor	1	22	17	14	5	36	6.50540461130282e-05	0.00467612367284095	helix-turn-helix family protein	WP_000571455	2.26334E-159
150	Phage restriction alleviation ral	1	22	16	15	5	36	0.000152218361481295	0.00964583722228836	restriction inhibitor protein ral	WP_000213977	5.76E-36
151	Phage repressor	1	22	20	11	4	37	3.82790071594159e-06	0.000537760150049888	regulatory protein cii	0611193A	4.63359E-52
152	Failed to assign function	1	22	16	15	3	38	0.000152218361481295	0.00964583722228836	pf11225 family protein	YP_794109	5.46973E-78
153	FIG00641106: hypothetical protein	1	22	17	14	0	41	6.50540461130282e-05	0.00467612367284095	hypothetical protein E2348C_1441	YP_002328980	1.12626E-124
154	Putative flagellin structural protein	2	21	27	4	38	3	4.79119287678256e-09	1.35731675850499e-06	outer membrane autotransporter barrel domain-containing protein	WP_000556465	0.0
155	Glutamate Aspartate transport system permease protein GltJ (TC 3.A.1.3.4)	2	21	31	0	37	4	4.86219192938783e-13	1.00355641422565e-09	amino abc permease 3-tm his glu gln arg opine family domain protein	YP_026209	0.0
156	Fimbrial protein YadC	2	21	22	9	34	7	3.76753516658155e-06	0.000537760150049888	fimbrial-like adhesin protein	WP_001443439	0.0

157	Molybdate metabolism regulator	2	21	24	7	30	11	3.63473268185746e-07	7.09656997128061e-05	molybdate metabolism regulator	WP_001363074	0.0
158	Eae protein	2	21	22	9	22	19	3.76753516658155e-06	0.000537760150049888	phage protein	YP_001272549	3.49126E-79
159	FIG00639949: hypothetical protein	2	21	24	7	16	25	3.63473268185746e-07	7.09656997128061e-05	hypothetical protein	WP_000848828	7.70478E-95
160	Phage EaA protein	2	21	24	7	16	25	3.63473268185746e-07	7.09656997128061e-05	pf04448 family partial	YP_007001969	7.58484E-111
161	FIG00638388: hypothetical protein	2	21	24	7	14	27	3.63473268185746e-07	7.09656997128061e-05	inner membrane protein yedr	YP_310920	2.54409E-59
162	Mobile element protein	2	21	21	10	0	41	1.07674236611247e-05	0.00118753997752618	transposase	WP_001369069	0.0
163	metal-dependent phosphohydrolase	3	20	31	0	41	0	5.51048418663954e-12	7.9615475528568e-09	hd domain protein	WP_000365559	1.37417E-138
164	Chaperone protein hchA	3	20	31	0	40	1	5.51048418663954e-12	7.9615475528568e-09	chaperone hcha	WP_000218204	0.0
165	L-rhamnose operon regulatory protein RhaS	3	20	31	0	40	1	5.51048418663954e-12	7.9615475528568e-09	hth-type transcriptional activator rhas	NP_418341	0.0
166	General secretion pathway protein E	3	20	23	8	38	3	7.80508840481335e-06	0.000880999353693307	type ii secretion system protein e	YP_002294519	0.0
167	Autoinducer 2 (AI-2) kinase LsrK (EC 2.7.1.-)	3	20	29	2	33	8	1.06517364649443e-09	3.20617267594822e-07	autoinducer 2 kinase	WP_000113166	0.0
168	FIG00637952: hypothetical protein	3	20	26	5	31	10	1.83566089058487e-07	4.01842856775306e-05	#NAME?	YP_410627	0.0
169	General secretion pathway protein M	3	20	24	7	31	10	2.48964902150791e-06	0.000367045398599452	type ii secretion system m family protein	WP_000942769	1.96035E-115
170	Uncharacterized protein ImpJ/VasE	3	20	21	10	30	11	5.97666431229269e-05	0.004361153837576	type vi secretion protein	WP_000246415	0.0
171	General secretion pathway protein L	3	20	24	7	29	12	2.48964902150791e-06	0.000367045398599452	type ii secretion system protein l	WP_000094990	0.0
172	General secretion pathway protein H	3	20	24	7	28	13	2.48964902150791e-06	0.000367045398599452	general secretion pathway protein h	YP_002388445	4.04855E-124
173	General secretion pathway protein K	3	20	24	7	28	13	2.48964902150791e-06	0.000367045398599452	general secretion pathway protein k	WP_000633212	0.0
174	Leader peptidase (Prepilin peptidase) (EC 3.4.23.43) / N-methyltransferase (EC 2.1.1.-)	3	20	23	8	28	13	7.80508840481335e-06	0.000880999353693307	leader peptidase pppa	WP_000895871	2.44755E-159
175	General secretion pathway protein J	3	20	23	8	28	13	7.80508840481335e-06	0.000880999353693307	type ii secretion system protein j	WP_001443429	3.24329E-134
176	General secretion pathway protein I	3	20	24	7	27	14	2.48964902150791e-06	0.000367045398599452	type ii secretion system protein i	YP_409353	7.66787E-63
177	Cytochrome b561 homolog 1	3	20	29	2	26	15	1.06517364649443e-09	3.20617267594822e-07	cytochrome b561	YP_407515	9.22549E-114
178	General secretion pathway protein F	3	20	22	9	25	16	2.24339044853633e-05	0.00210470812989954	type ii secretion system protein f	WP_001173467	0.0
179	FIG01070140: hypothetical protein	3	20	30	1	8	33	1.05995784060654e-10	5.1047569603611e-08	prophage protein	YP_001743216	1.55921E-42
180	FIG00637874: hypothetical protein	4	19	31	0	41	0	4.82167366330959e-11	3.66650216249984e-08	diguanylate cyclase	WP_000592815	0.0
181	Alkanesulfonate utilization operon LysR-family regulator CbI	4	19	31	0	41	0	4.82167366330959e-11	3.66650216249984e-08	hth-type transcriptional regulator cbl	WP_001011011	0.0
182	SanA protein	4	19	31	0	41	0	4.82167366330959e-11	3.66650216249984e-08	thioredoxin-like protein	WP_024199571	2.31116E-170
183	Putative isomerase	4	19	31	0	41	0	4.82167366330959e-11	3.66650216249984e-08	glycosyl hydrolase	YP_002388562	0.0
184	Failed to assign function	4	19	31	0	41	0	4.82167366330959e-11	3.66650216249984e-08	purine permease ybby	WP_001759829	0.0
185	Soluble cytochrome b562	4	19	31	0	41	0	4.82167366330959e-11	3.66650216249984e-08	soluble cytochrome b562	3U8P_A	2.17662E-56
186	Failed to assign function	4	19	31	0	40	1	4.82167366330959e-11	3.66650216249984e-08	transcriptional regulator amino transferase	YP_002386877	0.0
187	Trans-aconitate 2-methyltransferase (EC 2.1.1.144)	4	19	31	0	40	1	4.82167366330959e-11	3.66650216249984e-08	trans-aconitate 2-methyltransferase	WP_001286605	1.4579E-178
188	Hypothetical lipoprotein yehR	4	19	30	1	40	1	8.59635533115762e-10	2.70000308314272e-07	lipoprotein	NP_754541	9.35062E-71
189	Phosphoenolpyruvate-protein phosphotransferase of PTS system (EC 2.7.3.9)	4	19	30	1	40	1	8.59635533115762e-10	2.70000308314272e-07	phosphoenolpyruvate-protein phosphotransferase	WP_001546143	0.0
190	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)	4	19	27	4	40	1	2.64014792340253e-07	5.44926531390282e-05	galactitol-1-phosphate 5-dehydrogenase	YP_002387569	0.0
191	2-deoxy-D-gluconate 3-dehydrogenase (EC 1.1.1.125)	4	19	30	1	35	6	8.59635533115762e-10	2.70000308314272e-07	short chain dehydrogenase reductase family protein	YP_002294304	0.0
192	Autoinducer 2 (AI-2) modifying protein LsrG	4	19	30	1	34	7	8.59635533115762e-10	2.70000308314272e-07	autoinducer 2-degrading protein lsrg	3QMQ_A	8.89099E-62
193	D-galactonate transporter	4	19	26	5	33	8	1.11319554920335e-06	0.000189217050528117	l-galactonate transporter	YP_005459680	0.0
194	Fructuronate transporter GntP	4	19	25	6	32	9	4.05702217299542e-06	0.000537760150049888	high-affinity gluconate transporter	WP_001573417	0.0
195	Uncharacterized protein ImpA	4	19	25	6	31	10	4.05702217299542e-06	0.000537760150049888	impa-related family protein	WP_001240513	0.0
196	Inner membrane protein YjgN	4	19	29	2	29	12	8.01921314913929e-09	2.06895699247794e-06	inner membrane protein yjgn	YP_002295838	0.0
197	FIG00639456: hypothetical protein	4	19	25	6	29	12	4.05702217299542e-06	0.000537760150049888	fimbrial family protein	WP_021543399	0.0

198	YafQ toxin protein	4	19	25	6	29	12	4.05702217299542e-06	0.000537760150049888	mrna interferase	YP_309286	4.65069E-59
199	FIG00641505: hypothetical protein	4	19	25	6	28	13	4.05702217299542e-06	0.000537760150049888	reverse gyrase	YP_003223435	1.21837E-20
200	Hypothetical lipoprotein yghG precursor	4	19	24	7	28	13	1.31205277881187e-05	0.00135421852628186	lipoprotein	NP_417445	8.56611E-65
201	General secretion pathway protein D	4	19	23	8	28	13	3.83688648588193e-05	0.00282833346673582	type ii secretion system protein d	YP_001464424	0.0
202	Accessory colonization factor AcfD precursor	4	19	23	8	27	14	3.83688648588193e-05	0.00282833346673582	lipoprotein precursor	WP_001034473	0.0
203	General secretion pathway protein C	4	19	24	7	26	15	1.31205277881187e-05	0.00135421852628186	type ii secretion system protein c	WP_001631803	9.99666E-177
204	FIG00638425: hypothetical protein	4	19	29	2	25	16	8.01921314913929e-09	2.06895699247794e-06	membrane protein	WP_000837550	2.50704E-98
205	Alpha-fimbriae tip adhesin	4	19	23	8	24	17	3.83688648588193e-05	0.00282833346673582	cbld like pilus biogenesis initiator family protein	YP_001464557	0.0
206	FIG00639818: hypothetical protein	4	19	22	9	24	17	0.000102892392928388	0.00660706352457487	restriction endonuclease	WP_001317597	0.0
207	Failed to assign function	4	19	23	8	20	21	3.83688648588193e-05	0.00282833346673582	helix-turn-helix family protein	YP_405986	4.84912E-19
208	putative excisionase	4	19	25	6	18	23	4.05702217299542e-06	0.000537760150049888	excisionase	NP_311039	3.20391E-51
209	Inner membrane protein YqiK	5	18	31	0	41	0	3.47160503758288e-10	1.50476441883039e-07	inner membrane protein yqik	WP_001564111	0.0
210	FIG00639031: hypothetical protein	5	18	30	1	37	4	5.72814831201178e-09	1.59154397715281e-06	diguanylate cyclase domain protein	WP_000043731	0.0
211	Putative uncharacterized protein YhcG	5	18	25	6	28	13	1.87761552919557e-05	0.00177305811541291	pf06250 family protein	YP_002294774	0.0
212	FIG00638450: hypothetical protein	5	18	27	4	21	20	1.40856276898181e-06	0.000231260396434651	terminase	YP_007112075	8.40806E-96
213	Shikimate kinase I (EC 2.7.1.71)	6	17	25	6	8	33	7.39729044880571e-05	0.00529089368338341	shikimate kinase 1	NP_709163	1.34976E-146
214	Resolvase	7	16	26	5	20	21	8.64559095631685e-05	0.00557640616682437	site-specific recombinase	WP_001066925	1.19864E-160
215	General secretion pathway protein G	9	14	28	3	38	3	8.28989912094081e-05	0.00537096244391717	general secretion pathway protein g	YP_002388446	2.47418E-102
216	Phage DNA-packaging protein	9	14	28	3	20	21	8.28989912094081e-05	0.00537096244391717	dna packaging protein	YP_003220773	3.23174E-77
217	Mobile element protein	10	13	30	1	21	20	1.11799899259371e-05	0.00122370071552984	is602 transposase	YP_001739978	4.34004E-180
218	DNA-damage-inducible protein I	13	10	31	0	27	14	4.78074343991646e-05	0.00350620209238137	dna-damage-inducible protein i	WP_023063252	5.91635E-44

Table S6. Analysis of accessory genome in HUS-associated STEC of serogroup O26 in HUS-group 2 (n=5, G1) compared to other O26 STEC in HUS-group 2 (n=13, G2) and all other STEC (n=77, G3)

List of overrepresented genes in HUS-associated O26 in HUS-group 2

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR	Function Blast2GO	Hit Accession	E-value
1	Failed to assign function	5	0	1	12	13	64	0.000700280112044817	1	AE005344_11unknown protein encoded by prophage CP-933O	AAG56135	3.80556E-71
2	Shiga-like toxin II subunit A precursor (EC 3.2.2.22)	5	0	1	12	42	35	0.000700280112044817	1	shiga-like toxin 2 subunit a	NP_049500	0.0
3	Failed to assign function	5	0	2	11	24	53	0.00245098039215686	1	#NAME?	WP_024182351	2.7825E-21
4	Mobile element protein	5	0	3	10	53	24	0.0065359477124183	1	transposase	YP_002400901	5.52102E-71
5	FIG01046320: hypothetical protein	5	0	3	10	39	38	0.0065359477124183	1	hypothetical protein LH0105	YP_308766	2.12567E-50
6	FIG00638373: hypothetical protein	5	0	3	10	54	23	0.0065359477124183	1	ycha ta	WP_001006232	7.15259E-152
7	Putative antirestriction protein	5	0	3	10	57	20	0.0065359477124183	1	antirestriction protein	AHG18096	3.5145E-88
8	Exodeoxyribonuclease VIII (EC 3.1.11.-)	5	0	3	10	15	62	0.0065359477124183	1	exodeoxyribonuclease 8	WP_000105095	0.0
9	YdfA protein	5	0	3	10	34	43	0.0065359477124183	1	hypothetical protein	WP_000371885	2.58326E-52
10	YdeA protein	5	0	3	10	40	37	0.0065359477124183	1	hypothetical protein ECSE_P1-0022	YP_002296047	2.30439E-108
11	Failed to assign function	5	0	3	10	30	47	0.0065359477124183	1	secretion protein	WP_001443410	7.94601E-81
12	Adenine-specific methyltransferase (EC 2.1.1.72)	5	0	3	10	53	24	0.0065359477124183	1	protein psib	AHG18090	1.61957E-88
13	hypothetical protein	4	1	0	13	46	31	0.00163398692810458	1	phage protein	WP_023908640	2.58004E-29
14	Phage capsid and scaffold	4	1	1	12	54	23	0.00770308123249301	1	minor capsid protein c	WP_000123222	0.0
15	core protein	4	1	1	12	12	65	0.00770308123249301	1	protein rhsa	WP_000015033	0.0
16	Mobile element protein	4	1	1	12	55	22	0.00770308123249301	1	integrase core domain protein	WP_000878222	0.0
17	Mobile element protein	4	1	1	12	16	61	0.00770308123249301	1	transposase	WP_001542923	2.72003E-103

List of overrepresented genes in serogroup O26 in HUS-group 2 other than in HUS-associated O26 in HUS-group 2

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	G3_yes	G3_no	P-value Raw	FDR	Function Blast2GO	Hit Accession	E-value
1	Putative uncharacterized protein YhcG	0	5	12	1	6	71	0.000700280112044817	1	pf06250 family protein	WP_001521407	3.01172E-118
2	FIG00642398: hypothetical protein	0	5	12	1	30	47	0.000700280112044817	1	conserved protein	WP_001302917	1.15166E-38
3	Putative cytoplasmic protein	1	4	12	1	29	48	0.00770308123249301	1	phage protein	WP_001375698	7.01012E-133
4	IncI1 plasmid conjugative transfer NusG-type transcription antiterminator TraB	1	4	12	1	25	52	0.00770308123249301	1	transcription antitermination factor	YP_002756602	8.64568E-152
5	IncI1 plasmid conjugative transfer protein TraC	1	4	12	1	27	50	0.00770308123249301	1	transfer protein c	WP_001370630	1.29454E-146
6	unnamed protein product	0	5	10	3	11	66	0.0065359477124183	1	cytochrome b562 family protein	YP_325578	2.18237E-69
7	Failed to assign function	0	5	10	3	31	46	0.0065359477124183	1	lysis protein for colicin n	WP_001448505	6.22904E-18
8	Putative excisionase	0	5	10	3	43	34	0.0065359477124183	1	excisionase	YP_002397710	1.34617E-47
9	Putative regulatory protein	0	5	10	3	20	57	0.0065359477124183	1	xre family transcriptional regulator	NP_309096	9.11352E-74
10	Failed to assign function	0	5	10	3	30	47	0.0065359477124183	1	conserved protein	YP_003228160	2.31365E-152
11	enterobacterial exodeoxyribonuclease VIII family protein	0	5	10	3	0	77	0.0065359477124183	1	exonuclease viii	WP_024220841	0.0
12	FIG00639963: hypothetical protein	0	5	10	3	11	66	0.0065359477124183	1	hypothetical protein ECs1072	NP_309099	4.69731E-99
13	Shiga toxin A-chain precursor (EC 3.2.2.22)	0	5	10	3	34	43	0.0065359477124183	1	shiga toxin subunit a	ADG56725	0.0

Table S7. Analysis of core gene variants of 95 non-O157 STEC, in which different pfam domains were observed for 13 of the core gene variants.
A different number of protein sequences were observed for each of the 13 core gene variants.

Gene number	Core gene variant	Number of protein sequences 1	Number of protein sequences 2	Number of protein sequences 3
4404	1	Trypsin: 1, 3, 2, 5, 4, 7	Trypsin_2: 6	
1239	2	MTS: 2	Methyltransf_31: 1, 3, 5, 4, 7, 6	
181	3	DAO: 11, 4, 6, 8	FAD_binding_3: 7	NAD_binding_8: 10, 1, 3, 2, 5, 9
337	4	Glyco_trans_4_4: 1, 3, 2, 5, 4, 7, 6	Glycos_transf_1: 8	
2677	5	Fer4_11: 1, 2, 4, 6	Fer4_4: 3, 5	
477	6	MR_MLE: 10, 1, 3, 2, 4, 7, 6, 9, 8	MR_MLE_N: 5	
58	7	SIS: 10, 1, 3, 2, 4, 7, 6, 9, 8	SIS_2: 5	
2730	8	dUTPase: 3, 2, 7, 6	DCD: 1, 5, 4, 8	
472	9	HAD: 11, 10, 1, 3, 2, 5, 4, 7, 6, 9	Hydrolase_3: 8	
1028	10	Radical_SAM: 1, 5, 6, 8	Fer4_12: 11, 10, 3, 2, 4, 7, 9	
1649	11	Cytochrom_c3_2: 1, 3, 5, 4	Paired_CXXCH_1: 2	
3231	12	DUF1705: 2, 5	Sulfatase: 11, 10, 13, 12, 14, 1, 3, 4, 7, 6, 9, 8	
3283	13	Acetyltransf_1: 1, 3, 2, 4, 6	Acetyltransf_10: 5	

Table S8. Analysis of core genome variants in HUS-associated STEC (n=23, G1) compared to other LEE positive STEC not assoociated with HUS (n=31, G2).
List of overrepresented core gene variants in HUS-associated STEC

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	P-value Raw	FDR<0.01
1	Leucyl/phenylalanyl-tRNA--protein transferase (EC 2.3.2.6)	23	0	18	13	0.000186119403400109	0.00976993461960878
2	Failed to assign function	23	0	18	13	0.000186119403400109	0.00976993461960878
3	NAD(P)H-flavin reductase (EC 1.5.1.29) (EC 1.16.1.3)	23	0	18	13	0.000186119403400109	0.00976993461960878
4	Osmotically inducible lipoprotein E precursor	23	0	17	14	8.17109575902916e-05	0.006685671252218
5	Evolved beta-D-galactosidase, beta subunit	23	0	17	14	8.17109575902916e-05	0.006685671252218
6	Putative ACR protein	23	0	18	13	0.000186119403400109	0.00976993461960878
7	Failed to assign function	23	0	17	14	8.17109575902916e-05	0.006685671252218
8	Shikimate kinase III (EC 2.7.1.71)	23	0	18	13	0.000186119403400109	0.00976993461960878
9	Acid shock protein precursor	23	0	18	13	0.000186119403400109	0.00976993461960878
10	Iron binding protein SufA for iron-sulfur cluster assembly	23	0	18	13	0.000186119403400109	0.00976993461960878
11	Putative lipoprotein	23	0	18	13	0.000186119403400109	0.00976993461960878
12	Failed to assign function	23	0	18	13	0.000186119403400109	0.00976993461960878
13	Electron transfer flavoprotein, beta subunit	22	1	12	19	1.03627109799984e-05	0.0020300104995657
14	Arabinose 5-phosphate isomerase (EC 5.3.1.13)	21	2	13	18	0.00017272761215618	0.00976993461960878
15	Putative GTP-binding protein YdgA	20	3	9	22	2.24339044853633e-05	0.00355278389867058
16	Oligopeptide transport system permease protein OppB (TC 3.A.1.5.1)	18	5	4	27	1.40856276898182e-06	0.00066796674256108
17	Exoribonuclease II (EC 3.1.13.1)	18	5	2	29	4.95447633220757e-08	0.000187741680491929
18	Ureidoglycolate dehydrogenase (EC 1.1.1.154)	17	6	7	24	0.000208022152535731	0.00984762011378971
19	Failed to assign function	17	6	7	24	0.000208022152535731	0.00984762011378971
20	probable acetyltransferase YPO3809	17	6	3	28	1.46141346920191e-06	0.00066796674256108
21	Oxygen-insensitive NADPH nitroreductase (EC 1.-.-.-)	17	6	1	30	3.26330873532791e-08	0.000187741680491929
22	COG2110, Macro domain, possibly ADP-ribose binding module	17	6	3	28	1.46141346920191e-06	0.00066796674256108
23	FIG01219827: hypothetical protein	17	6	5	26	2.33838839204998e-05	0.00355278389867058
24	FIG136845: Rhodanese-related sulfurtransferase	16	7	2	29	1.20876327972868e-06	0.000616236818184758
25	FIG00639538: hypothetical protein	16	7	1	30	1.63314219839435e-07	0.000309247140567388
26	FIG005121: SAM-dependent methyltransferase (EC 2.1.1.-)	16	7	2	29	1.20876327972868e-06	0.000616236818184758
27	FIG00896075: hypothetical protein	16	7	3	28	6.26176706919338e-06	0.00165999445004317
28	DNA damage-inducible gene in SOS regulon, dependent on cyclic AMP and H-NS	16	7	4	27	2.54631814691592e-05	0.00355278389867058
29	Failed to assign function	16	7	4	27	2.54631814691592e-05	0.00355278389867058
30	FIG00638802: hypothetical protein	16	7	2	29	1.20876327972868e-06	0.000616236818184758
31	DNA-3-methyladenine glycosylase (EC 3.2.2.20)	16	7	2	29	1.20876327972868e-06	0.000616236818184758
32	6-phosphofructokinase (EC 2.7.1.11)	16	7	1	30	1.63314219839435e-07	0.000309247140567388
33	Transcriptional regulator, PadR family	16	7	2	29	1.20876327972868e-06	0.000616236818184758
34	Ribose-phosphate pyrophosphokinase (EC 2.7.6.1)	16	7	4	27	2.54631814691592e-05	0.00355278389867058

35	FIG00638401: hypothetical protein	16	7	3	28	6.26176706919338e-06	0.00165999445004317
36	FIG004064: hypothetical protein	16	7	2	29	1.20876327972868e-06	0.000616236818184758
37	Cytochrome c-type protein NrfB precursor	16	7	2	29	1.20876327972868e-06	0.000616236818184758
38	Failed to assign function	16	7	4	27	2.54631814691592e-05	0.00355278389867058
39	ATP-binding protein PhnN; Guanylate kinase (EC 2.7.4.8)	16	7	3	28	6.26176706919338e-06	0.00165999445004317
40	Sialic acid utilization regulator, RpiR family	16	7	2	29	1.20876327972868e-06	0.000616236818184758
41	Hypothetical protein, similar to phosphoserine phosphatase	16	7	4	27	2.54631814691592e-05	0.00355278389867058
42	Succinate dehydrogenase cytochrome b-556 subunit	16	7	4	27	2.54631814691592e-05	0.00355278389867058
43	Ferredoxin, 2Fe-2S	16	7	1	30	1.63314219839435e-07	0.000309247140567388
44	Transcription repressor of multidrug efflux pump acrAB operon, TetR (AcrR) family	16	7	4	27	2.54631814691592e-05	0.00355278389867058
45	Transcription termination protein NusB	16	7	3	28	6.26176706919338e-06	0.00165999445004317
46	Phosphate starvation-inducible protein PsiF	16	7	2	29	1.20876327972868e-06	0.000616236818184758
47	Phosphopantetheine adenylyltransferase (EC 2.7.7.3)	16	7	3	28	6.26176706919338e-06	0.00165999445004317
48	rRNA small subunit methyltransferase I	16	7	4	27	2.54631814691592e-05	0.00355278389867058
49	PTS system, glucose-specific IIB component (EC 2.7.1.69) / PTS system, glucose-specific IIC component (EC 2.7.1.69)	16	7	2	29	1.20876327972868e-06	0.000616236818184758
50	FIG00626295: hypothetical protein	16	7	4	27	2.54631814691592e-05	0.00355278389867058
51	PTS system fructose-like IIB component 2 precursor (EC 2.7.1.69)	16	7	3	28	6.26176706919338e-06	0.00165999445004317
52	Penicillin-binding protein 2 (PBP-2)	16	7	0	31	1.16216111496227e-08	0.000154044455788249
53	Failed to assign function	15	8	3	28	2.39472803323198e-05	0.00355278389867058
54	Oxalyl-CoA decarboxylase (EC 4.1.1.8)	15	8	4	27	9.02679550690437e-05	0.00712203419309627
55	FIG001881: hydrolase of alkaline phosphatase superfamily	15	8	1	30	7.32161502426228e-07	0.000616236818184758
56	Lysine-specific permease	15	8	4	27	9.02679550690437e-05	0.00712203419309627
57	Exodeoxyribonuclease III (EC 3.1.11.2)	15	8	3	28	2.39472803323198e-05	0.00355278389867058
58	FIG004212: hypothetical protein	15	8	1	30	7.32161502426228e-07	0.000616236818184758
59	Colicin V production protein	15	8	2	29	4.99851612182719e-06	0.00150580298170044
60	4-hydroxy-3-methylbut-2-enyl diphosphate reductase (EC 1.17.1.2)	15	8	4	27	9.02679550690437e-05	0.00712203419309627
61	UPF0246 protein YaaA	15	8	2	29	4.99851612182719e-06	0.00150580298170044
62	Ferric reductase (1.6.99.14)	15	8	4	27	9.02679550690437e-05	0.00712203419309627
63	TPR repeat containing exported protein; Putative periplasmic protein contains a protein prenylyltransferase domain	15	8	3	28	2.39472803323198e-05	0.00355278389867058
64	Failed to assign function	15	8	2	29	4.99851612182719e-06	0.00150580298170044
65	Sigma factor RpoE negative regulatory protein RseB precursor	15	8	3	28	2.39472803323198e-05	0.00355278389867058
66	Putative HTH-type transcriptional regulator ybaO	15	8	2	29	4.99851612182719e-06	0.00150580298170044
67	NADH oxidoreductase hcr (EC 1.-.-.-)	15	8	0	31	5.66553543544107e-08	0.000187741680491929
68	Cytidylate kinase (EC 2.7.4.14)	14	9	2	29	1.87456587843415e-05	0.00310592133983058

69	Ubiquinone biosynthesis monooxygenase UbiB	14	9	0	31	2.51801574908491e-07	0.000417203734426507
70	Fructose-specific phosphocarrier protein HPr (EC 2.7.1.69) / PTS system, fructose-specific IIA component (EC 2.7.1.69)	14	9	1	30	2.98384866266564e-06	0.00130622696487718
71	Regulatory protein AsnC	14	9	2	29	1.87456587843415e-05	0.00310592133983058
72	LSU ribosomal protein L6p (L9e)	14	9	3	28	8.28989912094081e-05	0.0067001593200043
73	L-rhamnose operon transcriptional activator RhaR	14	9	3	28	8.28989912094081e-05	0.0067001593200043
74	RecA protein	14	9	2	29	1.87456587843415e-05	0.00310592133983058
75	Glutamate Aspartate transport system permease protein GltK (TC 3.A.1.3.4)	14	9	2	29	1.87456587843415e-05	0.00310592133983058
76	hemimethylated DNA binding protein YccV	13	10	2	29	6.44549081372011e-05	0.00537326922867045
77	LSU m5C1962 methyltransferase RlmI	13	10	2	29	6.44549081372011e-05	0.00537326922867045
78	Anaerobic dimethyl sulfoxide reductase chain C (EC 1.8.99.-)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
79	Paraquat-inducible protein B	13	10	1	30	1.11799899259371e-05	0.0020300104995657
80	Transcriptional activator MetR	13	10	0	31	1.03238645712482e-06	0.000616236818184758
81	FIG000605: protein co-occurring with transport systems (COG1739)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
82	Xaa-Pro dipeptidase PepQ (EC 3.4.13.9)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
83	Putative ECA polymerase	13	10	1	30	1.11799899259371e-05	0.0020300104995657
84	WzxE protein	13	10	2	29	6.44549081372011e-05	0.00537326922867045
85	Undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase (EC 2.7.8.-)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
86	Uncharacterized PLP-dependent aminotransferase YfdZ	13	10	1	30	1.11799899259371e-05	0.0020300104995657
87	Methyl-accepting chemotaxis protein II (aspartate chemoreceptor protein)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
88	N-acetylglucosamine-1-phosphate uridyltransferase (EC 2.7.7.23) / Glucosamine-1-phosphate N-acetyltransferase (EC 2.3.1.157)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
89	Lipid A biosynthesis (KDO) 2-(lauroyl)-lipid IVA acyltransferase (EC 2.3.1.-)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
90	Tryptophanase (EC 4.1.99.1)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
91	Transporter, LysE family	13	10	1	30	1.11799899259371e-05	0.0020300104995657
92	Gluconate utilization system Gnt-I transcriptional repressor	13	10	2	29	6.44549081372011e-05	0.00537326922867045
93	Failed to assign function	13	10	2	29	6.44549081372011e-05	0.00537326922867045
94	Phosphoglycolate phosphatase (EC 3.1.3.18)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
95	HTH-type transcriptional regulator gadX	13	10	2	29	6.44549081372011e-05	0.00537326922867045
96	4'-phosphopantetheinyl transferase (EC 2.7.8.-)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
97	Putative receptor	13	10	2	29	6.44549081372011e-05	0.00537326922867045
98	Lipid A biosynthesis (KDO) 2-(lauroyl)-lipid IVA acyltransferase (EC 2.3.1.-)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
99	L-seryl-tRNA(Sec) selenium transferase (EC 2.9.1.1)	13	10	2	29	6.44549081372011e-05	0.00537326922867045

100	Glycerophosphoryl diester phosphodiesterase, periplasmic (EC 3.1.4.46)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
101	radical activating enzyme	13	10	2	29	6.44549081372011e-05	0.00537326922867045
102	Probable glutathione S-transferase (EC 2.5.1.18), YfcF homolog	13	10	2	29	6.44549081372011e-05	0.00537326922867045
103	Serine transporter	13	10	2	29	6.44549081372011e-05	0.00537326922867045
104	Glutathione S-transferase, omega (EC 2.5.1.18)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
105	Biosynthetic Aromatic amino acid aminotransferase alpha (EC 2.6.1.57)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
106	Cytochrome c-type heme lyase subunit nrfE, nitrite reductase complex assembly	13	10	1	30	1.11799899259371e-05	0.0020300104995657
107	Phosphonate ABC transporter ATP-binding protein (TC 3.A.1.9.1)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
108	Putative outer membrane lipoprotein	13	10	2	29	6.44549081372011e-05	0.00537326922867045
109	Failed to assign function	13	10	0	31	1.03238645712482e-06	0.000616236818184758
110	Exported zinc metalloprotease YfgC precursor	13	10	1	30	1.11799899259371e-05	0.0020300104995657
111	Dihydrodipicolinate synthase (EC 4.2.1.52)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
112	Predicted outer membrane lipoprotein YfeY	13	10	2	29	6.44549081372011e-05	0.00537326922867045
113	Probable 3-phenylpropionic acid transporter	13	10	1	30	1.11799899259371e-05	0.0020300104995657
114	Failed to assign function	13	10	2	29	6.44549081372011e-05	0.00537326922867045
115	Peptidase B (EC 3.4.11.23)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
116	FIG01200175: hypothetical protein	13	10	1	30	1.11799899259371e-05	0.0020300104995657
117	Phosphate regulon sensor protein PhoR (SphS) (EC 2.7.13.3)	13	10	1	30	1.11799899259371e-05	0.0020300104995657
118	D-lactate dehydrogenase (EC 1.1.1.28)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
119	FIG138517: Putative lipid carrier protein	13	10	2	29	6.44549081372011e-05	0.00537326922867045
120	FIG01219785: hypothetical protein	13	10	2	29	6.44549081372011e-05	0.00537326922867045
121	21 kDa hemolysin precursor	13	10	2	29	6.44549081372011e-05	0.00537326922867045
122	PTS system, N-acetylgalactosamine-specific IIB component (EC 2.7.1.69)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
123	Pantothenate:Na+ symporter (TC 2.A.21.1.1)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
124	RNA polymerase sigma-54 factor RpoN	13	10	2	29	6.44549081372011e-05	0.00537326922867045
125	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)	13	10	0	31	1.03238645712482e-06	0.000616236818184758
126	IMP cyclohydrolase (EC 3.5.4.10) / Phosphoribosyl-aminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
127	Potassium-transporting ATPase C chain (EC 3.6.3.12) (TC 3.A.3.7.1)	13	10	2	29	6.44549081372011e-05	0.00537326922867045
128	Failed to assign function	13	10	1	30	1.11799899259371e-05	0.0020300104995657
129	Glutaredoxin 2	13	10	2	29	6.44549081372011e-05	0.00537326922867045
130	Formate hydrogenlyase regulatory protein HycA	13	10	1	30	1.11799899259371e-05	0.0020300104995657
131	Failed to assign function	13	10	0	31	1.03238645712482e-06	0.000616236818184758

132	Cystathionine beta-lyase (EC 4.4.1.8) (CBL) (Beta-cystathionase) (Cysteine lyase) / Maltose regulon modulator	13	10	2	29	6.44549081372011e-05	0.00537326922867045
133	FolM Alternative dihydrofolate reductase 1	13	10	2	29	6.44549081372011e-05	0.00537326922867045
134	Mlc, transcriptional repressor of MalT (the transcriptional activator of maltose regulon) and manXYZ operon	13	10	1	30	1.11799899259371e-05	0.0020300104995657
135	Hydrogen peroxide-inducible genes activator	13	10	0	31	1.03238645712482e-06	0.000616236818184758
136	UPF0379 protein yjfY precursor	13	10	2	29	6.44549081372011e-05	0.00537326922867045
137	YjfP protein	13	10	2	29	6.44549081372011e-05	0.00537326922867045
138	Copper-sensing two-component system response regulator CusR	13	10	2	29	6.44549081372011e-05	0.00537326922867045
139	tRNA-i(6)A37 methylthiotransferase	13	10	1	30	1.11799899259371e-05	0.0020300104995657
140	Putative ATPase component of ABC transporter with duplicated ATPase domain	13	10	2	29	6.44549081372011e-05	0.00537326922867045
141	Hypothetical Zinc-finger containing protein	13	10	0	31	1.03238645712482e-06	0.000616236818184758
142	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.1)	12	11	2	29	0.000204906965239626	0.00976993461960878
143	Thioredoxin reductase (EC 1.8.1.9)	12	11	0	31	3.94183919993113e-06	0.00130622696487718
144	Outer membrane protein A precursor	12	11	1	30	3.88552721136067e-05	0.00440193702449451
145	Antiholin-like protein LrgA	12	11	0	31	3.94183919993113e-06	0.00130622696487718
146	Autolysis histidine kinase LytS	12	11	1	30	3.88552721136067e-05	0.00440193702449451
147	FIG00640962: hypothetical protein	12	11	2	29	0.000204906965239626	0.00976993461960878
148	Flagellar motor rotation protein MotA	12	11	1	30	3.88552721136067e-05	0.00440193702449451
149	Flagellar protein FlhE	12	11	2	29	0.000204906965239626	0.00976993461960878
150	ATP synthase alpha chain (EC 3.6.3.14)	12	11	1	30	3.88552721136067e-05	0.00440193702449451
151	FIG004088: inner membrane protein YebE	12	11	2	29	0.000204906965239626	0.00976993461960878
152	Ribosomal RNA large subunit methyltransferase A (EC 2.1.1.51)	12	11	2	29	0.000204906965239626	0.00976993461960878
153	Tail-specific protease precursor (EC 3.4.21.102)	12	11	1	30	3.88552721136067e-05	0.00440193702449451
154	Uncharacterized protein YidS	12	11	1	30	3.88552721136067e-05	0.00440193702449451
155	CDP-diacylglycerol--serine O-phosphatidyltransferase (EC 2.7.8.8)	12	11	0	31	3.94183919993113e-06	0.00130622696487718
156	Ribosome-associated heat shock protein implicated in the recycling of the 50S subunit (S4 paralog)	12	11	2	29	0.000204906965239626	0.00976993461960878
157	Type IV pilus biogenesis protein PilN	12	11	2	29	0.000204906965239626	0.00976993461960878
158	Failed to assign function	12	11	1	30	3.88552721136067e-05	0.00440193702449451
159	FIG01045643: hypothetical protein	12	11	2	29	0.000204906965239626	0.00976993461960878
160	Cell filamentation protein fic	12	11	1	30	3.88552721136067e-05	0.00440193702449451
161	Transcription antitermination protein NusG	12	11	2	29	0.000204906965239626	0.00976993461960878
162	FIG004614: Putative cytoplasmic protein	12	11	2	29	0.000204906965239626	0.00976993461960878
163	Transcriptional activator GadE	12	11	2	29	0.000204906965239626	0.00976993461960878
164	NAD(FAD)-utilizing dehydrogenases	12	11	2	29	0.000204906965239626	0.00976993461960878
165	Nickel responsive regulator NikR	12	11	2	29	0.000204906965239626	0.00976993461960878
166	probable exported protein YPO4070	12	11	0	31	3.94183919993113e-06	0.00130622696487718

167	BAX protein	12	11	1	30	3.88552721136067e-05	0.00440193702449451
168	Methionine ABC transporter ATP-binding protein	12	11	2	29	0.000204906965239626	0.00976993461960878
169	Cell division protein FtsA	12	11	2	29	0.000204906965239626	0.00976993461960878
170	Failed to assign function	12	11	2	29	0.000204906965239626	0.00976993461960878
171	Sensor protein PhoQ (EC 2.7.13.3)	12	11	2	29	0.000204906965239626	0.00976993461960878
172	Alcohol dehydrogenase (EC 1.1.1.1); Acetaldehyde dehydrogenase (EC 1.2.1.10)	12	11	1	30	3.88552721136067e-05	0.00440193702449451
173	GTP-binding and nucleic acid-binding protein YchF	12	11	2	29	0.000204906965239626	0.00976993461960878
174	Membrane-bound lytic murein transglycosylase E (EC 3.2.1.-)	12	11	1	30	3.88552721136067e-05	0.00440193702449451
175	Aspartate carbamoyltransferase regulatory chain (PyrI)	12	11	0	31	3.94183919993113e-06	0.00130622696487718
176	Acetate operon repressor	12	11	0	31	3.94183919993113e-06	0.00130622696487718
177	Aspartokinase (EC 2.7.2.4)	12	11	2	29	0.000204906965239626	0.00976993461960878
178	FIG000906: Predicted Permease	12	11	2	29	0.000204906965239626	0.00976993461960878
179	PhnO protein	12	11	2	29	0.000204906965239626	0.00976993461960878
180	Possible ring-opening amidohydrolase RutC in novel pyrimidine catabolism pathway	12	11	1	30	3.88552721136067e-05	0.00440193702449451
181	DNA recombination and repair protein RecO	12	11	1	30	3.88552721136067e-05	0.00440193702449451
182	Cardiolipin synthetase (EC 2.7.8.-)	12	11	1	30	3.88552721136067e-05	0.00440193702449451
183	Molybdenum ABC transporter, periplasmic molybdenum-binding protein ModA (TC 3.A.1.8.1)	12	11	2	29	0.000204906965239626	0.00976993461960878
184	Serine hydroxymethyltransferase (EC 2.1.2.1)	12	11	2	29	0.000204906965239626	0.00976993461960878
185	Succinyl-CoA ligase [ADP-forming] alpha chain (EC 6.2.1.5)	12	11	0	31	3.94183919993113e-06	0.00130622696487718
186	Ribosomal RNA large subunit methyltransferase N (EC 2.1.1.-)	12	11	0	31	3.94183919993113e-06	0.00130622696487718
187	Putative exported protein	12	11	1	30	3.88552721136067e-05	0.00440193702449451
188	Protein-export membrane protein SecF (TC 3.A.5.1.1)	12	11	2	29	0.000204906965239626	0.00976993461960878
189	Transcriptional regulator NanR	12	11	2	29	0.000204906965239626	0.00976993461960878
190	LptA, protein essential for LPS transport across the periplasm	12	11	2	29	0.000204906965239626	0.00976993461960878
191	Zinc resistance-associated protein	12	11	2	29	0.000204906965239626	0.00976993461960878
192	Failed to assign function	12	11	1	30	3.88552721136067e-05	0.00440193702449451
193	YcfL protein: an outer membrane lipoprotein that is part of a salvage cluster	12	11	1	30	3.88552721136067e-05	0.00440193702449451
194	FIG004453: protein YceG like	12	11	2	29	0.000204906965239626	0.00976993461960878
195	Ribosomal large subunit pseudouridine synthase C (EC 4.2.1.70)	12	11	2	29	0.000204906965239626	0.00976993461960878
196	Flagellar basal-body P-ring formation protein FlgA	12	11	1	30	3.88552721136067e-05	0.00440193702449451
197	Hydrogenase maturation factor HoxQ	12	11	2	29	0.000204906965239626	0.00976993461960878
198	FIG004819: Prepilin peptidase dependent protein B precursor	12	11	2	29	0.000204906965239626	0.00976993461960878
199	Failed to assign function	12	11	2	29	0.000204906965239626	0.00976993461960878
200	Coenzyme F420 hydrogenase maturation protease (EC 3.4.24.-)	12	11	2	29	0.000204906965239626	0.00976993461960878
201	Putative membrane protein	12	11	2	29	0.000204906965239626	0.00976993461960878
202	Glutamate--cysteine ligase (EC 6.3.2.2)	12	11	1	30	3.88552721136067e-05	0.00440193702449451

203	Rhodanese-related sulfurtransferases	12	11	1	30	3.88552721136067e-05	0.00440193702449451
204	5-formyltetrahydrofolate cyclo-ligase (EC 6.3.3.2)	12	11	2	29	0.000204906965239626	0.00976993461960878
205	Folate-dependent protein for Fe/S cluster synthesis/repair in oxidative stress	12	11	2	29	0.000204906965239626	0.00976993461960878
206	putative oxidoreductase, Fe-S subunit	12	11	2	29	0.000204906965239626	0.00976993461960878
207	Failed to assign function	12	11	2	29	0.000204906965239626	0.00976993461960878
208	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase (EC 6.3.2.-)	12	11	1	30	3.88552721136067e-05	0.00440193702449451
209	Uncharacterized protein YtfM precursor	12	11	2	29	0.000204906965239626	0.00976993461960878
210	Failed to assign function	12	11	1	30	3.88552721136067e-05	0.00440193702449451
211	FIG00638355: hypothetical protein	12	11	2	29	0.000204906965239626	0.00976993461960878
212	Oxygen-insensitive NAD(P)H nitroreductase (EC 1.-.-.-) / Dihydropteridine reductase (EC 1.5.1.34)	12	11	0	31	3.94183919993113e-06	0.00130622696487718
213	Failed to assign function	12	11	2	29	0.000204906965239626	0.00976993461960878
214	GTP-binding protein TypA/BipA	12	11	0	31	3.94183919993113e-06	0.00130622696487718
215	Putative membrane protein YfcA	12	11	2	29	0.000204906965239626	0.00976993461960878
216	Nucleoid-associated protein NdpA	11	12	1	30	0.000126138854397796	0.00901262308256692
217	Sodium:dicarboxylate symporter	11	12	1	30	0.000126138854397796	0.00901262308256692
218	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	11	12	0	31	1.41249237997532e-05	0.00246349822323327
219	Succinyl-CoA ligase [ADP-forming] beta chain (EC 6.2.1.5)	11	12	1	30	0.000126138854397796	0.00901262308256692
220	L,D-transpeptidase YcfS	11	12	0	31	1.41249237997532e-05	0.00246349822323327
221	Transcriptional regulator, TetR family	11	12	0	31	1.41249237997532e-05	0.00246349822323327
222	Glucans biosynthesis protein C (EC 2.1.-.-)	11	12	1	30	0.000126138854397796	0.00901262308256692
223	HTH-type transcriptional regulator hdfR	10	13	0	31	4.78074343991646e-05	0.00519416018820432
224	Cell wall endopeptidase, family M23/M37	10	13	0	31	4.78074343991646e-05	0.00519416018820432
225	Protease IV (EC 3.4.21.-)	10	13	0	31	4.78074343991646e-05	0.00519416018820432
226	Cytochrome O ubiquinol oxidase subunit I (EC 1.10.3.-)	10	13	0	31	4.78074343991646e-05	0.00519416018820432
227	Failed to assign function	10	13	0	31	4.78074343991646e-05	0.00519416018820432
228	Imidazole glycerol phosphate synthase cyclase subunit (EC 4.1.3.-)	9	14	0	31	0.000153666753425886	0.00901262308256692
229	Cytochrome c-type biogenesis protein CcmC, putative heme lyase for CcmE	9	14	0	31	0.000153666753425886	0.00901262308256692
230	Glycerate kinase (EC 2.7.1.31)	9	14	0	31	0.000153666753425886	0.00901262308256692
231	Cystathionine beta-lyase (EC 4.4.1.8)	9	14	0	31	0.000153666753425886	0.00901262308256692
232	Glutamate-ammonia-ligase adenylyltransferase (EC 2.7.7.42)	9	14	0	31	0.000153666753425886	0.00901262308256692
233	Periplasmic fimbrial chaperone StfD	9	14	0	31	0.000153666753425886	0.00901262308256692
234	DedA family inner membrane protein YdjX	9	14	0	31	0.000153666753425886	0.00901262308256692
235	Peptide methionine sulfoxide reductase MsrB (EC 1.8.4.12)	9	14	0	31	0.000153666753425886	0.00901262308256692
236	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	9	14	0	31	0.000153666753425886	0.00901262308256692
237	Putative transport protein	9	14	0	31	0.000153666753425886	0.00901262308256692
238	DNA polymerase III beta subunit (EC 2.7.7.7)	9	14	0	31	0.000153666753425886	0.00901262308256692

239	Uracil-DNA glycosylase, family 1	9	14	0	31	0.000153666753425886	0.00901262308256692
240	Gamma-glutamyltranspeptidase (EC 2.3.2.2)	9	14	0	31	0.000153666753425886	0.00901262308256692
241	Low-affinity gluconate/H+ symporter GntU	9	14	0	31	0.000153666753425886	0.00901262308256692
242	Multiple antibiotic resistance protein marC	9	14	0	31	0.000153666753425886	0.00901262308256692
243	Osmolarity sensory histidine kinase EnvZ	9	14	0	31	0.000153666753425886	0.00901262308256692
244	Nitrite reductase [NAD(P)H] large subunit (EC 1.7.1.4)	9	14	0	31	0.000153666753425886	0.00901262308256692
245	Cytoplasmic trehalase (EC 3.2.1.28)	9	14	0	31	0.000153666753425886	0.00901262308256692
246	Transporter, putative	9	14	0	31	0.000153666753425886	0.00901262308256692
247	Putative inner membrane protein	9	14	0	31	0.000153666753425886	0.00901262308256692
248	Aspartokinase (EC 2.7.2.4) / Homoserine dehydrogenase (EC 1.1.1.3)	9	14	0	31	0.000153666753425886	0.00901262308256692
249	Thymidine phosphorylase (EC 2.4.2.4)	9	14	0	31	0.000153666753425886	0.00901262308256692
250	NadR transcriptional regulator / Nicotinamide-nucleotide adenylyltransferase, NadR family (EC 2.7.7.1) / Ribosylnicotinamide kinase (EC 2.7.1.22)	9	14	0	31	0.000153666753425886	0.00901262308256692
251	Failed to assign function	9	14	0	31	0.000153666753425886	0.00901262308256692
252	Helicase PriA essential for oriC/DnaA-independent DNA replication	9	14	0	31	0.000153666753425886	0.00901262308256692
253	23S rRNA (guanine-N-2-) -methyltransferase rlmG (EC 2.1.1.-)	9	14	0	31	0.000153666753425886	0.00901262308256692
254	DNA-damage-inducible protein F	9	14	0	31	0.000153666753425886	0.00901262308256692
255	Inner membrane protein YbhQ	9	14	0	31	0.000153666753425886	0.00901262308256692
256	Putative sensor-like histidine kinase YfhK	9	14	0	31	0.000153666753425886	0.00901262308256692
257	Ribosyl nicotinamide transporter, PnuC-like	9	14	0	31	0.000153666753425886	0.00901262308256692
258	Inhibitor of vertebrate lysozyme precursor	9	14	0	31	0.000153666753425886	0.00901262308256692
259	Failed to assign function	9	14	0	31	0.000153666753425886	0.00901262308256692
260	Putative exported protein	9	14	0	31	0.000153666753425886	0.00901262308256692
261	Methylated-DNA--protein-cysteine methyltransferase (EC 2.1.1.63)	9	14	0	31	0.000153666753425886	0.00901262308256692
262	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8)	9	14	0	31	0.000153666753425886	0.00901262308256692
263	tRNA dihydrouridine synthase B (EC 1.-.-.-)	9	14	0	31	0.000153666753425886	0.00901262308256692
264	Phosphoribosylamine--glycine ligase (EC 6.3.4.13)	9	14	0	31	0.000153666753425886	0.00901262308256692
265	COG1399 protein, clustered with ribosomal protein L32p	9	14	0	31	0.000153666753425886	0.00901262308256692
266	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase (EC 5.3.1.17)	9	14	0	31	0.000153666753425886	0.00901262308256692
267	Exodeoxyribonuclease V beta chain (EC 3.1.11.5)	9	14	0	31	0.000153666753425886	0.00901262308256692
268	N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28)	9	14	0	31	0.000153666753425886	0.00901262308256692
269	Sulfite reductase [NADPH] hemoprotein beta-component (EC 1.8.1.2)	9	14	0	31	0.000153666753425886	0.00901262308256692
270	FIG00638716: hypothetical protein	9	14	0	31	0.000153666753425886	0.00901262308256692
271	Formate hydrogenlyase subunit 5	9	14	0	31	0.000153666753425886	0.00901262308256692
272	Electron transfer flavoprotein, alpha subunit	9	14	0	31	0.000153666753425886	0.00901262308256692
273	PTS system fructose-like IIB component 1 precursor (EC 2.7.1.69)	9	14	0	31	0.000153666753425886	0.00901262308256692
274	Trehalose operon transcriptional repressor	9	14	0	31	0.000153666753425886	0.00901262308256692
275	Putative membrane protein	9	14	0	31	0.000153666753425886	0.00901262308256692
276	N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28)	9	14	0	31	0.000153666753425886	0.00901262308256692

277	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase (EC 1.14.13.-)	9	14	0	31	0.000153666753425886	0.00901262308256692
278	Citrate lyase alpha chain (EC 4.1.3.6)	9	14	0	31	0.000153666753425886	0.00901262308256692
279	Co-activator of prophage gene expression IbrB	9	14	0	31	0.000153666753425886	0.00901262308256692
280	Apolipoprotein N-acyltransferase (EC 2.3.1.-) / Copper homeostasis protein CutE	9	14	0	31	0.000153666753425886	0.00901262308256692
281	3-oxoacyl-[acyl-carrier-protein] synthase, KASI (EC 2.3.1.41)	9	14	0	31	0.000153666753425886	0.00901262308256692

Table S9. Analysis of core genome variants in stx2 positive STEC of serogroup O26 in HUS-group 2 (n=8, G1) compared to stx1 positive STEC of serogroup O26 in HUS-group 2 (n=10, G2)
List of overrepresented core gene variants in stx2 positive O26 in HUS-group 2

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	P-value Raw	FDR
1	Multidrug resistance protein A	8	0	2	8	0.00102838338132456	0.0349154391707583
2	Failed to assign function	8	0	2	8	0.00102838338132456	0.0349154391707583
3	Exodeoxyribonuclease I (EC 3.1.11.1)	8	0	2	8	0.00102838338132456	0.0349154391707583
4	Aspartate ammonia-lyase (EC 4.3.1.1)	8	0	2	8	0.00102838338132456	0.0349154391707583
5	Exodeoxyribonuclease V gamma chain (EC 3.1.11.5)	8	0	3	7	0.00377073906485671	0.0532800532800532
6	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	8	0	3	7	0.00377073906485671	0.0532800532800532
7	DNA recombination and repair protein RecO	8	0	3	7	0.00377073906485671	0.0532800532800532
8	Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52)	8	0	3	7	0.00377073906485671	0.0532800532800532
9	Proposed peptidoglycan lipid II flippase MurJ	8	0	3	7	0.00377073906485671	0.0532800532800532
10	Dihydroorotase (EC 3.5.2.3)	8	0	3	7	0.00377073906485671	0.0532800532800532
11	Adenylate cyclase (EC 4.6.1.1)	8	0	3	7	0.00377073906485671	0.0532800532800532
12	regulator of length of O-antigen component of lipopolysaccharide chains	8	0	3	7	0.00377073906485671	0.0532800532800532
13	ATP-dependent DNA helicase Rep	8	0	3	7	0.00377073906485671	0.0532800532800532
14	Aspartokinase (EC 2.7.2.4)	8	0	3	7	0.00377073906485671	0.0532800532800532
15	Hydrogenase maturation factor HoxO/HyaE	8	0	3	7	0.00377073906485671	0.0532800532800532
16	DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)	8	0	3	7	0.00377073906485671	0.0532800532800532
17	Preprotein translocase subunit SecE (TC 3.A.5.1.1)	8	0	3	7	0.00377073906485671	0.0532800532800532
18	Failed to assign function	8	0	3	7	0.00377073906485671	0.0532800532800532
19	Rod shape-determining protein MreC	8	0	3	7	0.00377073906485671	0.0532800532800532
20	FIG006303: protein yraQ	8	0	3	7	0.00377073906485671	0.0532800532800532
21	Uncharacterized protein YhaL	8	0	3	7	0.00377073906485671	0.0532800532800532
22	Nickel transport system permease protein NikB (TC 3.A.1.5.3)	8	0	3	7	0.00377073906485671	0.0532800532800532
23	Cell division trigger factor (EC 5.2.1.8)	8	0	3	7	0.00377073906485671	0.0532800532800532
24	FIG01220476: hypothetical protein	8	0	3	7	0.00377073906485671	0.0532800532800532
25	Fructose-specific phosphocarrier protein HPr (EC 2.7.1.69) / PTS system, fructose-specific IIA component (EC 2.7.1.69)	8	0	3	7	0.00377073906485671	0.0532800532800532
26	1-phosphofructokinase (EC 2.7.1.56)	8	0	3	7	0.00377073906485671	0.0532800532800532
27	FIG00638355: hypothetical protein	8	0	3	7	0.00377073906485671	0.0532800532800532
28	Cytochrome c-type biogenesis protein DsbD, protein-disulfide reductase (EC 1.8.1.8)	8	0	3	7	0.00377073906485671	0.0532800532800532
29	Hypothetical Zinc-finger containing protein	8	0	3	7	0.00377073906485671	0.0532800532800532
30	FIG01201035: hypothetical protein	8	0	3	7	0.00377073906485671	0.0532800532800532
31	Putative membrane protein	8	0	3	7	0.00377073906485671	0.0532800532800532
32	Putative GTP-binding protein YdgA	8	0	3	7	0.00377073906485671	0.0532800532800532
33	Succinylarginine dihydrolase (EC 3.5.3.23)	8	0	3	7	0.00377073906485671	0.0532800532800532

34	GTPase (EC 3.6.1.-)	7	1	1	9	0.0018510900863842	0.0419580419580419
35	DNA mismatch repair protein MutL	7	1	2	8	0.00761003702180172	0.0997556660207262
36	Membrane-bound lytic murein transglycosylase C precursor (EC 3.2.1.-)	6	2	0	10	0.00150829562594268	0.0349154391707583
37	Putative alpha helix chain	6	2	0	10	0.00150829562594268	0.0349154391707583
38	Twitching motility protein PilT	6	2	0	10	0.00150829562594268	0.0349154391707583
39	Failed to assign function	6	2	0	10	0.00150829562594268	0.0349154391707583
40	Lactaldehyde reductase (EC 1.1.1.77)	6	2	0	10	0.00150829562594268	0.0349154391707583
41	ATP synthase beta chain (EC 3.6.3.14)	6	2	0	10	0.00150829562594268	0.0349154391707583
42	DNA ligase (EC 6.5.1.2), LigB	6	2	0	10	0.00150829562594268	0.0349154391707583
43	L-threonine 3-dehydrogenase (EC 1.1.1.103)	6	2	0	10	0.00150829562594268	0.0349154391707583
44	Cation transport regulator chaB	6	2	0	10	0.00150829562594268	0.0349154391707583
45	Protein SseB	6	2	0	10	0.00150829562594268	0.0349154391707583
46	Phosphoserine phosphatase (EC 3.1.3.3)	6	2	0	10	0.00150829562594268	0.0349154391707583
47	Failed to assign function	6	2	0	10	0.00150829562594268	0.0349154391707583
48	Chorismate mutase I (EC 5.4.99.5) / Cyclohexadienyl dehydrogenase (EC 1.3.1.12)(EC 1.3.1.43)	6	2	0	10	0.00150829562594268	0.0349154391707583
49	3-dehydroquinate synthase (EC 4.2.3.4)	6	2	0	10	0.00150829562594268	0.0349154391707583
50	Signal recognition particle receptor protein FtsY (=alpha subunit) (TC 3.A.5.1.1)	6	2	0	10	0.00150829562594268	0.0349154391707583
51	Cell division protein FtsX	6	2	0	10	0.00150829562594268	0.0349154391707583
52	Cytoplasmic axial filament protein CafA and Ribonuclease G (EC 3.1.4.-)	6	2	0	10	0.00150829562594268	0.0349154391707583
53	Failed to assign function	6	2	0	10	0.00150829562594268	0.0349154391707583
54	FUSARIC ACID RESISTANCE PROTEIN FUSB / FUSARIC ACID RESISTANCE PROTEIN FUSC	6	2	0	10	0.00150829562594268	0.0349154391707583
55	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)	6	2	0	10	0.00150829562594268	0.0349154391707583
56	PTS system, N-acetylgalactosamine-specific IIB component (EC 2.7.1.69)	6	2	0	10	0.00150829562594268	0.0349154391707583
57	Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)	6	2	0	10	0.00150829562594268	0.0349154391707583
58	Glycogen debranching enzyme (EC 3.2.1.-)	6	2	0	10	0.00150829562594268	0.0349154391707583
59	Glycogen synthase, ADP-glucose transglucosylase (EC 2.4.1.21)	6	2	0	10	0.00150829562594268	0.0349154391707583
60	Putative HTH-type transcriptional regulator ybaO	6	2	0	10	0.00150829562594268	0.0349154391707583
61	Adenine phosphoribosyltransferase (EC 2.4.2.7)	6	2	0	10	0.00150829562594268	0.0349154391707583
62	Adenylate kinase (EC 2.7.4.3)	6	2	0	10	0.00150829562594268	0.0349154391707583
63	AsmA protein	6	2	0	10	0.00150829562594268	0.0349154391707583
64	Failed to assign function	6	2	0	10	0.00150829562594268	0.0349154391707583
65	N-acetylglucosamine-6P-responsive transcriptional repressor NagC, ROK family	6	2	0	10	0.00150829562594268	0.0349154391707583
66	Putative exported protein	6	2	0	10	0.00150829562594268	0.0349154391707583
67	Molybdenum transport ATP-binding protein ModC (TC 3.A.1.8.1)	6	2	0	10	0.00150829562594268	0.0349154391707583

68	tRNA pseudouridine synthase A (EC 4.2.1.70)	6	2	0	10	0.00150829562594268	0.0349154391707583
69	Histidine ABC transporter, permease protein HisQ (TC 3.A.1.3.1)	6	2	0	10	0.00150829562594268	0.0349154391707583
70	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	6	2	0	10	0.00150829562594268	0.0349154391707583
71	Transport ATP-binding protein CydC	6	2	0	10	0.00150829562594268	0.0349154391707583
72	Predicted ATP-dependent endonuclease of the OLD family, YbjD subgroup	6	2	0	10	0.00150829562594268	0.0349154391707583
73	Mn-dependent transcriptional regulator MntR	6	2	0	10	0.00150829562594268	0.0349154391707583
74	Ureidoglycolate dehydrogenase (EC 1.1.1.154)	6	2	0	10	0.00150829562594268	0.0349154391707583
75	Vitamin B12 ABC transporter, permease component BtuC	6	2	0	10	0.00150829562594268	0.0349154391707583
76	Failed to assign function	6	2	0	10	0.00150829562594268	0.0349154391707583
77	Aldose 1-epimerase family protein YeaD	6	2	0	10	0.00150829562594268	0.0349154391707583
78	Cob(I)alamin adenosyltransferase (EC 2.5.1.17)	6	2	0	10	0.00150829562594268	0.0349154391707583
79	Fructose repressor FruR, LacI family	6	2	1	9	0.0090497737556561	0.113174182139699
80	Thiamin ABC transporter, substrate-binding component	6	2	1	9	0.0090497737556561	0.113174182139699
81	L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4)	6	2	1	9	0.0090497737556561	0.113174182139699
82	Failed to assign function	6	2	1	9	0.0090497737556561	0.113174182139699
83	Ethanolamine operon regulatory protein	5	3	0	10	0.00653594771241829	0.086720867208672
84	ABC transporter, periplasmic spermidine putrescine-binding protein PotD (TC 3.A.1.11.1)	5	3	0	10	0.00653594771241829	0.086720867208672
85	Acetylornithine aminotransferase (EC 2.6.1.11) / N-succinyl-L,L-diaminopimelate aminotransferase (EC 2.6.1.17)	5	3	0	10	0.00653594771241829	0.086720867208672
86	Putative transport protein	5	3	0	10	0.00653594771241829	0.086720867208672
87	Lactam utilization protein LamB	5	3	0	10	0.00653594771241829	0.086720867208672

Analysis of overrepresented core gene variants in stx1 positive O26 in HUS-group 2

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	P-value Raw	FDR
1	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)	0	8	10	0	1	1
2	Transport ATP-binding protein CydC	1	7	10	0	1	1
3	Ureidoglycolate dehydrogenase (EC 1.1.1.154)	2	6	10	0	1	1
4	Protein SseB	2	6	10	0	1	1
5	Failed to assign function	2	6	10	0	1	1
6	Failed to assign function	2	6	10	0	1	1
7	Twitching motility protein PilT	2	6	10	0	1	1
8	Putative alpha helix chain	2	6	10	0	1	1
9	Membrane-bound lytic murein transglycosylase C precursor (EC 3.2.1.-)	2	6	10	0	1	1
10	Cob(I)alamin adenosyltransferase (EC 2.5.1.17)	2	6	10	0	1	1
11	Lactaldehyde reductase (EC 1.1.1.77)	2	6	10	0	1	1
12	N-acetylglucosamine-6P-responsive transcriptional repressor NagC, ROK family	2	6	10	0	1	1

13	Failed to assign function	2	6	10	0	1	1
14	Vitamin B12 ABC transporter, permease component BtuC	2	6	10	0	1	1
15	Failed to assign function	2	6	10	0	1	1
16	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	2	6	10	0	1	1
17	Cation transport regulator chaB	2	6	10	0	1	1
18	Histidine ABC transporter, permease protein HisQ (TC 3.A.1.3.1)	2	6	10	0	1	1
19	3-dehydroquinate synthase (EC 4.2.3.4)	2	6	10	0	1	1
20	Signal recognition particle receptor protein FtsY (=alpha subunit) (TC 3.A.5.1.1)	2	6	10	0	1	1
21	Cell division protein FtsX	2	6	10	0	1	1
22	Glycogen debranching enzyme (EC 3.2.1.-)	2	6	10	0	1	1
23	Glycogen synthase, ADP-glucose transglucosylase (EC 2.4.1.21)	2	6	10	0	1	1
24	Mn-dependent transcriptional regulator MntR	2	6	10	0	1	1
25	Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)	2	6	10	0	1	1
26	PTS system, N-acetylgalactosamine-specific IIB component (EC 2.7.1.69)	2	6	10	0	1	1
27	Aldose 1-epimerase family protein YeaD	2	6	10	0	1	1
28	ATP synthase beta chain (EC 3.6.3.14)	2	6	10	0	1	1
29	FUSARIC ACID RESISTANCE PROTEIN FUSB / FUSARIC ACID RESISTANCE PROTEIN FUSC	2	6	10	0	1	1
30	Cytoplasmic axial filament protein CafA and Ribonuclease G (EC 3.1.4.-)	2	6	10	0	1	1
31	Chorismate mutase I (EC 5.4.99.5) / Cyclohexadienyl dehydrogenase (EC 1.3.1.12)(EC 1.3.1.43)	2	6	10	0	1	1
32	Adenine phosphoribosyltransferase (EC 2.4.2.7)	2	6	10	0	1	1
33	Adenylate kinase (EC 2.7.4.3)	2	6	10	0	1	1
34	AsmA protein	2	6	10	0	1	1
35	Predicted ATP-dependent endonuclease of the OLD family, YbjD subgroup	2	6	10	0	1	1
36	Putative exported protein	2	6	10	0	1	1
37	Molybdenum transport ATP-binding protein ModC (TC 3.A.1.8.1)	2	6	10	0	1	1
38	Phosphoserine phosphatase (EC 3.1.3.3)	2	6	10	0	1	1
39	DNA ligase (EC 6.5.1.2), LigB	2	6	10	0	1	1
40	L-threonine 3-dehydrogenase (EC 1.1.1.103)	2	6	10	0	1	1
41	Putative HTH-type transcriptional regulator ybaO	2	6	10	0	1	1
42	Ethanolamine operon regulatory protein	3	5	10	0	1	1
43	ABC transporter, periplasmic spermidine putrescine-binding protein PotD (TC 3.A.1.11.1)	3	5	10	0	1	1
44	Putative transport protein	3	5	10	0	1	1
45	Failed to assign function	3	5	10	0	1	1
46	Lactam utilization protein LamB	3	5	10	0	1	1

47	GTPase (EC 3.6.1.-)	1	7	9	1	0.999977147035971	1
48	tRNA pseudouridine synthase A (EC 4.2.1.70)	2	6	9	1	0.999748617395676	1
49	Failed to assign function	2	6	9	1	0.999748617395676	1
50	TldD protein, part of proposed TldE/TldD proteolytic complex (PMID 12029038)	2	6	9	1	0.999748617395676	1
51	Fructose repressor FruR, LacI family	2	6	9	1	0.999748617395676	1
52	Thiamin ABC transporter, substrate-binding component	2	6	9	1	0.999748617395676	1
53	Multidrug resistance protein A	0	8	8	2	1	1
54	DNA mismatch repair protein MutL	0	8	8	2	1	1
55	Aspartate ammonia-lyase (EC 4.3.1.1)	0	8	8	2	1	1
56	D-alanine--D-alanine ligase B (EC 6.3.2.4)	0	8	8	2	1	1
57	DNA recombination and repair protein RecO	0	8	7	3	1	1
58	Proposed peptidoglycan lipid II flippase MurJ	0	8	7	3	1	1
59	Dihydroorotase (EC 3.5.2.3)	0	8	7	3	1	1
60	FIG00638355: hypothetical protein	0	8	7	3	1	1
61	Putative GTP-binding protein YdgA	0	8	7	3	1	1
62	Putative membrane protein	0	8	7	3	1	1
63	Cytochrome c-type biogenesis protein DsbD, protein-disulfide reductase (EC 1.8.1.8)	0	8	7	3	1	1
64	FIG01201035: hypothetical protein	0	8	7	3	1	1
65	Nickel transport system permease protein NikB (TC 3.A.1.5.3)	0	8	7	3	1	1
66	Hypothetical Zinc-finger containing protein	0	8	7	3	1	1
67	FIG01220476: hypothetical protein	0	8	7	3	1	1
68	ATP-dependent DNA helicase Rep	0	8	7	3	1	1
69	regulator of length of O-antigen component of lipopolysaccharide chains	0	8	7	3	1	1
70	Adenylate cyclase (EC 4.6.1.1)	0	8	7	3	1	1
71	Uncharacterized protein YhaL	0	8	7	3	1	1
72	FIG006303: protein yraQ	0	8	7	3	1	1
73	Fructose-specific phosphocarrier protein HPr (EC 2.7.1.69) / PTS system, fructose-specific IIA component (EC 2.7.1.69)	0	8	7	3	1	1
74	Succinylarginine dihydrolase (EC 3.5.3.23)	0	8	7	3	1	1
75	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	0	8	7	3	1	1
76	Rod shape-determining protein MreC	0	8	7	3	1	1
77	Failed to assign function	0	8	7	3	1	1
78	Hydrogenase maturation factor HoxO/HyaE	0	8	7	3	1	1
79	Aspartokinase (EC 2.7.2.4)	0	8	7	3	1	1
80	Exodeoxyribonuclease I (EC 3.1.11.1)	0	8	7	3	1	1
81	Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52)	0	8	7	3	1	1

82	DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)	0	8	7	3	1	1
83	Preprotein translocase subunit SecE (TC 3.A.5.1.1)	0	8	7	3	1	1

Table S10. Analysis of core genome variants in HUS-associated STEC of serogroup O26 in HUS-group 2 (n=5, G1) compared to non-HUS STEC of serogroup O26 in HUS-group 2 (n=13, G2)
List of overrepresented core gene variants in HUS-associated O26 in HUS-group 2

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	P-value Raw	FDR
1	Proofreading thioesterase in enterobactin biosynthesis EntH	5	0	4	9	0.0147058823529412	1
2	Failed to assign function	5	0	5	8	0.0294117647058824	1
3	Fe-S-cluster-containing hydrogenase components 2	5	0	5	8	0.0294117647058824	1
4	Multidrug resistance protein A	5	0	5	8	0.0294117647058824	1
5	Aspartate ammonia-lyase (EC 4.3.1.1)	5	0	5	8	0.0294117647058824	1
6	Exodeoxyribonuclease I (EC 3.1.11.1)	5	0	5	8	0.0294117647058824	1
7	FIG01220476: hypothetical protein	5	0	6	7	0.053921568627451	1
8	Hypothetical Zinc-finger containing protein	5	0	6	7	0.053921568627451	1
9	Adenylate cyclase (EC 4.6.1.1)	5	0	6	7	0.053921568627451	1
10	regulator of length of O-antigen component of lipopolysaccharide chains	5	0	6	7	0.053921568627451	1
11	ATP-dependent DNA helicase Rep	5	0	6	7	0.053921568627451	1
12	Exodeoxyribonuclease V gamma chain (EC 3.1.11.5)	5	0	6	7	0.053921568627451	1
13	Cytochrome c-type biogenesis protein DsbD, protein-disulfide reductase (EC 1.8.1.8)	5	0	6	7	0.053921568627451	1
14	DNA recombination and repair protein RecO	5	0	6	7	0.053921568627451	1
15	Succinylarginine dihydrolase (EC 3.5.3.23)	5	0	6	7	0.053921568627451	1
16	Putative membrane protein	5	0	6	7	0.053921568627451	1
17	Putative GTP-binding protein YdgA	5	0	6	7	0.053921568627451	1
18	Uncharacterized protein YhaL	5	0	6	7	0.053921568627451	1
19	FIG006303: protein yraQ	5	0	6	7	0.053921568627451	1
20	Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52)	5	0	6	7	0.053921568627451	1
21	1-phosphofructokinase (EC 2.7.1.56)	5	0	6	7	0.053921568627451	1
22	Fructose-specific phosphocarrier protein HPr (EC 2.7.1.69) / PTS system, fructose-specific IIA component (EC 2.7.1.69)	5	0	6	7	0.053921568627451	1
23	Preprotein translocase subunit SecE (TC 3.A.5.1.1)	5	0	6	7	0.053921568627451	1
24	DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)	5	0	6	7	0.053921568627451	1
25	Biotin sulfoxide reductase (EC 1.-.-.-)	5	0	6	7	0.053921568627451	1
26	Nickel transport system permease protein NikB (TC 3.A.1.5.3)	5	0	6	7	0.053921568627451	1
27	Dihydroorotase (EC 3.5.2.3)	5	0	6	7	0.053921568627451	1
28	Proposed peptidoglycan lipid II flippase MurJ	5	0	6	7	0.053921568627451	1
29	FIG01201035: hypothetical protein	5	0	6	7	0.053921568627451	1
30	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	5	0	6	7	0.053921568627451	1
31	FIG00638355: hypothetical protein	5	0	6	7	0.053921568627451	1
32	Aspartokinase (EC 2.7.2.4)	5	0	6	7	0.053921568627451	1
33	Cell division trigger factor (EC 5.2.1.8)	5	0	6	7	0.053921568627451	1
34	Hydrogenase maturation factor HoxO/HyaE	5	0	6	7	0.053921568627451	1
35	Failed to assign function	5	0	6	7	0.053921568627451	1
36	Rod shape-determining protein MreC	5	0	6	7	0.053921568627451	1

37	Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61)	5	0	7	6	0.0924369747899159	1
38	Transaldolase (EC 2.2.1.2)	5	0	7	6	0.0924369747899159	1
39	Lipid A biosynthesis (KDO) 2-(lauroyl)-lipid IVA acyltransferase (EC 2.3.1.-)	5	0	7	6	0.0924369747899159	1
40	Failed to assign function	5	0	7	6	0.0924369747899159	1
41	Thiazole biosynthesis protein ThiG	5	0	7	6	0.0924369747899159	1
42	Transcriptional regulatory protein UhpA	5	0	7	6	0.0924369747899159	1
43	Respiratory nitrate reductase beta chain (EC 1.7.99.4)	4	1	2	11	0.0217086834733893	1
44	Hypothetical MFS-type transporter protein YcaD	4	1	2	11	0.0217086834733893	1
45	L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4)	4	1	3	10	0.0473856209150326	1
46	Failed to assign function	4	1	3	10	0.0473856209150326	1
47	Glucarate dehydratase (EC 4.2.1.40)	4	1	4	9	0.0882352941176471	1
48	GTPase (EC 3.6.1.-)	4	1	4	9	0.0882352941176471	1
49	Failed to assign function	4	1	4	9	0.0882352941176471	1
50	Failed to assign function	3	2	1	12	0.0441176470588235	1
51	Homolog of E. coli HemX protein	3	2	2	11	0.0987394957983192	1
52	Lactam utilization protein LamB	3	2	2	11	0.0987394957983192	1
53	Ethanolamine operon regulatory protein	3	2	2	11	0.0987394957983192	1
54	Putative transport protein	3	2	2	11	0.0987394957983192	1
55	ABC transporter, periplasmic spermidine putrescine-binding protein PotD (TC 3.A.1.11.1)	3	2	2	11	0.0987394957983192	1
56	Molybdopterin biosynthesis Mog protein, molybdochelatase	2	3	0	13	0.0653594771241829	1
57	Arabinose operon regulatory protein	2	3	0	13	0.0653594771241829	1
58	Acetolactate synthase small subunit (EC 2.2.1.6)	2	3	0	13	0.0653594771241829	1
59	Redox-sensing transcriptional regulator QorR	2	3	0	13	0.0653594771241829	1
60	Methylglyoxal reductase, acetol producing (EC 1.1.1.-) / 2,5-diketo-D-gluconic acid reductase B (EC 1.1.1.274)	2	3	0	13	0.0653594771241829	1
61	Failed to assign function	2	3	0	13	0.0653594771241829	1
62	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase (EC 5.3.1.17)	2	3	0	13	0.0653594771241829	1
63	Transaldolase (EC 2.2.1.2)	2	3	0	13	0.0653594771241829	1
64	RND efflux system, inner membrane transporter CmeB	2	3	0	13	0.0653594771241829	1
65	Failed to assign function	2	3	0	13	0.0653594771241829	1
66	N-ethylmaleimide reductase (EC 1.-.-.-)	2	3	0	13	0.0653594771241829	1
67	Haemolysin expression modulating protein paralog	2	3	0	13	0.0653594771241829	1
68	putative endopeptidase	2	3	0	13	0.0653594771241829	1
69	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	2	3	0	13	0.0653594771241829	1
70	Galactose/methyl galactoside ABC transport system, permease protein MglC (TC 3.A.1.2.3)	2	3	0	13	0.0653594771241829	1
71	Excinuclease ABC subunit B	2	3	0	13	0.0653594771241829	1
72	Two-component system response regulator OmpR	2	3	0	13	0.0653594771241829	1
73	Nickel transport ATP-binding protein NikE (TC 3.A.1.5.3)	2	3	0	13	0.0653594771241829	1

74	RNA polymerase sigma factor RpoH	2	3	0	13	0.0653594771241829	1
75	Aspartate aminotransferase (EC 2.6.1.1)	2	3	0	13	0.0653594771241829	1
76	ClpB protein	2	3	0	13	0.0653594771241829	1
77	Thiamine-monophosphate kinase (EC 2.7.4.16)	2	3	0	13	0.0653594771241829	1
78	Alkaline phosphatase (EC 3.1.3.1)	2	3	0	13	0.0653594771241829	1
79	Failed to assign function	2	3	0	13	0.0653594771241829	1
80	ATP-dependent DNA helicase RecG (EC 3.6.1.-)	2	3	0	13	0.0653594771241829	1
81	Guanylate kinase (EC 2.7.4.8)	2	3	0	13	0.0653594771241829	1
82	Aspartokinase (EC 2.7.2.4) / Homoserine dehydrogenase (EC 1.1.1.3)	2	3	0	13	0.0653594771241829	1
83	Cytochrome c-type protein NapC	2	3	0	13	0.0653594771241829	1
84	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)	2	3	0	13	0.0653594771241829	1

Analysis of overrepresented core gene variants in non-HUS STEC of serogroup O26 in HUS-group 2

CDS	Function MyRAST	G1_yes	G1_no	G2_yes	G2_no	P-value Raw	FDR
1	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase (EC 5.3.1.17)	3	2	13	0	1	1
2	Redox-sensing transcriptional regulator QorR	3	2	13	0	1	1
3	Aspartate aminotransferase (EC 2.6.1.1)	3	2	13	0	1	1
4	Cytochrome c-type protein NapC	3	2	13	0	1	1
5	Galactose/methyl galactoside ABC transport system, permease protein MglC (TC 3.A.1.2.3)	3	2	13	0	1	1
6	Failed to assign function	3	2	13	0	1	1
7	Alkaline phosphatase (EC 3.1.3.1)	3	2	13	0	1	1
8	Thiamine-monophosphate kinase (EC 2.7.4.16)	3	2	13	0	1	1
9	Haemolysin expression modulating protein	3	2	13	0	1	1
10	N-ethylmaleimide reductase (EC 1.-.-.-)	3	2	13	0	1	1
11	Failed to assign function	3	2	13	0	1	1
12	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	3	2	13	0	1	1
13	Failed to assign function	3	2	13	0	1	1
14	Two-component system response regulator OmpR	3	2	13	0	1	1
15	Nickel transport ATP-binding protein NikE (TC 3.A.1.5.3)	3	2	13	0	1	1
16	RNA polymerase sigma factor RpoH	3	2	13	0	1	1
17	Guanylate kinase (EC 2.7.4.8)	3	2	13	0	1	1
18	ATP-dependent DNA helicase RecG (EC 3.6.1.-)	3	2	13	0	1	1
19	ClpB protein	3	2	13	0	1	1
20	Molybdopterin biosynthesis Mog protein, molybdochelatase	3	2	13	0	1	1
21	Acetolactate synthase small subunit (EC 2.2.1.6)	3	2	13	0	1	1
22	Arabinose operon regulatory protein	3	2	13	0	1	1
23	Excinuclease ABC subunit B	3	2	13	0	1	1
24	Methylglyoxal reductase, acetol producing (EC 1.1.1.-) / 2,5-diketo-D-gluconic acid reductase B (EC 1.1.1.274)	3	2	13	0	1	1
25	Hypothetical MFS-type transporter protein YcaD	1	4	11	2	0.999299719887955	1

26	Ethanolamine operon regulatory protein	2	3	11	2	0.992296918767507	1
27	Lactam utilization protein LamB	2	3	11	2	0.992296918767507	1
28	Homolog of E. coli HemX protein	2	3	11	2	0.992296918767507	1
29	Putative transport protein	2	3	11	2	0.992296918767507	1
30	ABC transporter, periplasmic spermidine putrescine-binding protein PotD (TC 3.A.1.11.1)	2	3	11	2	0.992296918767507	1
31	Failed to assign function	2	3	11	2	0.992296918767507	1
32	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)	0	5	10	3	1	1
33	Transport ATP-binding protein CydC	1	4	10	3	0.997549019607843	1
34	Proofreading thioesterase in enterobactin biosynthesis EntH	0	5	9	4	1	1
35	Glucarate dehydratase (EC 4.2.1.40)	1	4	9	4	0.993464052287582	1
36	Transaldolase (EC 2.2.1.2)	1	4	9	4	0.993464052287582	1
37	GTPase (EC 3.6.1.-)	1	4	9	4	0.993464052287582	1
38	Failed to assign function	1	4	9	4	0.993464052287582	1
39	Failed to assign function	1	4	9	4	0.993464052287582	1
40	Multidrug resistance protein A	0	5	8	5	1	1
41	Fe-S-cluster-containing hydrogenase components 2	0	5	8	5	1	1
42	D-alanine--D-alanine ligase B (EC 6.3.2.4)	0	5	8	5	1	1
43	DNA mismatch repair protein MutL	0	5	8	5	1	1
44	Aspartate ammonia-lyase (EC 4.3.1.1)	0	5	8	5	1	1
45	DNA recombination and repair protein RecO	0	5	7	6	1	1
46	FIG01201035: hypothetical protein	0	5	7	6	1	1
47	Hydrogenase maturation factor HoxO/HyaE	0	5	7	6	1	1
48	Fructose-specific phosphocarrier protein HPr (EC 2.7.1.69) / PTS system, fructose-specific IIA component (EC 2.7.1.69)	0	5	7	6	1	1
49	Putative GTP-binding protein YdgA	0	5	7	6	1	1
50	Putative membrane protein	0	5	7	6	1	1
51	Hypothetical Zinc-finger containing protein	0	5	7	6	1	1
52	ATP-dependent DNA helicase Rep	0	5	7	6	1	1
53	regulator of length of O-antigen component of lipopolysaccharide chains	0	5	7	6	1	1
54	Adenylate cyclase (EC 4.6.1.1)	0	5	7	6	1	1
55	Failed to assign function	0	5	7	6	1	1
56	Rod shape-determining protein MreC	0	5	7	6	1	1
57	Preprotein translocase subunit SecE (TC 3.A.5.1.1)	0	5	7	6	1	1
58	DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)	0	5	7	6	1	1
59	Nickel transport system permease protein NikB (TC 3.A.1.5.3)	0	5	7	6	1	1
60	Uncharacterized protein YhaL	0	5	7	6	1	1
61	FIG006303: protein yraQ	0	5	7	6	1	1
62	FIG01220476: hypothetical protein	0	5	7	6	1	1
63	Exodeoxyribonuclease I (EC 3.1.11.1)	0	5	7	6	1	1
64	Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52)	0	5	7	6	1	1

65	Dihydroorotase (EC 3.5.2.3)	0	5	7	6	1	1
66	Proposed peptidoglycan lipid II flippase MurJ	0	5	7	6	1	1
67	FIG00638355: hypothetical protein	0	5	7	6	1	1
68	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	0	5	7	6	1	1
69	Cytochrome c-type biogenesis protein DsbD, protein-disulfide reductase (EC 1.8.1.8)	0	5	7	6	1	1
70	Aspartokinase (EC 2.7.2.4)	0	5	7	6	1	1
71	Succinylarginine dihydrolase (EC 3.5.3.23)	0	5	7	6	1	1
72	Transaldolase (EC 2.2.1.2)	0	5	6	7	1	1
73	Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61)	0	5	6	7	1	1
74	1-phosphofructokinase (EC 2.7.1.56)	0	5	6	7	1	1
75	Cell division trigger factor (EC 5.2.1.8)	0	5	6	7	1	1
76	Thiazole biosynthesis protein ThiG	0	5	6	7	1	1
77	Failed to assign function	0	5	6	7	1	1
78	Transcriptional regulatory protein UhpA	0	5	6	7	1	1

Table S11. An overview of unique genes in strains from the same outbreaks

Strain	Unique genes	Serotype	Clinic
FHI4	179	O26	HUS
FHI3	377	O26	Unknown
FHI48	153	O121	HUS
FHI43	41	O121	Gastroenteritis
FHI62	41	O121	Gastroenteritis
FHI58	54	O145	HUS
FHI63	54	O145	HUS
St. Olav104	506	O145	Asymptomatic

Table S12. Analysis of genes differently present in strain FHI4 from a HUS-patient and FHI3 from the patient's mother

Unique genes present in FHI4

Contig Id	Start	Stop	Length	Type	Function myRAST	Function Blast2GO	Hit Accession	E-value
NODE_0	116	3	113	cds	Failed to assign function	escherichia coli chi7122 chi7122 genomic chromosome chi7122	WP_001701179	2.57266E-15
NODE_0	2847	2137	710	cds	Glutamate Aspartate transport system permease protein GltJ (TC 3.A.1.3.4)	amino abc permease 3-tm his glu gln arg opine family domain protein	YP_001460069	3.63364E-146
NODE_0	3860	2973	887	cds	Mobile element protein	integrase core domain protein	WP_001345376	0.0
NODE_0	162509	162165	344	cds	Threonine catabolic operon transcriptional activator TdcR	dna-binding transcriptional activator	YP_002388603	5.71087E-68
NODE_0	460086	460292	206	cds	Failed to assign function	hypothetical protein	WP_000151767	1.76496E-31
NODE_1	31228	30917	311	cds	Failed to assign function	cell surface protein	WP_024181921	2.05352E-64
NODE_1	257301	257732	431	cds	Mobile element protein	transposase	YP_003229545	1.11252E-104
NODE_1	305974	304661	1313	cds	Phage tail fiber protein	tail fiber protein	WP_000279017	9.02479E-134
NODE_1	307656	306706	950	cds	Phage tail fiber protein	fibronectin type iii family protein	WP_001432888	9.91113E-152
NODE_1	317606	317103	503	cds	Failed to assign function	tail assembly protein	WP_001408918	5.41731E-77
NODE_1	321902	321057	845	cds	Phage tail length tape-measure protein 1	tail protein	WP_001369645	6.56965E-177
NODE_1	350333	351118	785	cds	Recombinational DNA repair protein RecT (prophage associated)	phage recombination protein bet	NP_049474	0.0
NODE_1	531656	530790	866	cds	Mobile element protein	integrase core domain protein	WP_000878222	0.0
NODE_1	599536	601008	1472	cds	Betaine aldehyde dehydrogenase (EC 1.2.1.8)	betaine aldehyde dehydrogenase	YP_003227432	0.0
NODE_1	708180	707833	347	cds	Mobile element protein	transposase dde domain partial	WP_001307055	2.91539E-79
NODE_1	712770	710149	2621	cds	core protein	type iv secretion protein rhs	WP_000509077	0.0
NODE_1	737612	737286	326	cds	membrane protein, putative	hypothetical protein, partial	WP_001295789	3.70754E-52
NODE_10	1012	1464	452	cds	Phage tail fiber protein	tail fiber protein	WP_001372101	2.22495E-27
NODE_10	127533	127279	254	cds	putative excisionase	excisionase	WP_001193435	9.09441E-51
NODE_10	132265	131441	824	cds	FIG00642676: hypothetical protein	phage protein	WP_000081322	0.0
NODE_10	132692	132330	362	cds	FIG00638873: hypothetical protein	hypothetical protein ETEC_0272	YP_006113858	2.55617E-79
NODE_10	137726	138379	653	cds	Failed to assign function	phage n-6-adenine-methyltransferase	WP_001312296	4.79369E-160
NODE_10	145928	146632	704	cds	Mobile element protein	yjhs protein	WP_000298067	2.84431E-144
NODE_11	7777	9090	1313	cds	Phage tail fiber protein	tail fiber protein	YP_003230104	6.28092E-126
NODE_11	144746	144979	233	cds	hypothetical protein	phage protein	NP_287504	1.65407E-47
NODE_11	145364	144957	407	cds	hypothetical protein	conserved protein	NP_287505	7.29541E-86
NODE_11	146649	146242	407	cds	putative regulator; Regulation (Phage or Prophage Relatec	helix-turn-helix family protein	WP_024216105	1.09424E-69
NODE_11	146726	146953	227	cds	Putative regulator of cell division encoded by prophage CP-933O	transcriptional repressor of cell division inhibition protein	NP_287509	9.32609E-43
NODE_11	146937	147488	551	cds	unknown protein encoded by prophage CP-933O	ydfx	YP_003228819	5.19611E-127
NODE_11	147460	148500	1040	cds	Primosomal protein I	primosomal protein	NP_287511	0.0
NODE_11	150581	151339	758	cds	Putative intestinal colonization factor encoded by prophage CP-933O	major antigenic peptide peb3	NP_287515	1.15934E-166
NODE_11	152378	152656	278	cds	orf, conserved hypothetical protein	conserved protein	NP_309802	3.75265E-46
NODE_12	44386	43127	1259	cds	core protein	rhs repeat-associated core domain protein	WP_001564902	0.0
NODE_12	46710	47276	566	cds	hypothetical protein	hypothetical protein	WP_001395792	1.20268E-119
NODE_13	298	2	296	cds	Phage integrase	integrase	YP_003230655	1.37794E-58
NODE_13	2876	2265	611	cds	Phage EaA protein	eaa protein	WP_000207991	5.41155E-140

NODE_13	3307	2888	419	cds	Phage EaA protein	eaa protein	WP_000207994	9.80994E-80
NODE_13	113882	113232	650	cds	FIG00641471: hypothetical protein	t3ss effector protein 8	YP_003228845	9.58028E-151
NODE_13	117834	117565	269	cds	Failed to assign function	hypothetical protein Z6012	NP_287952	2.27206E-47
NODE_13	118286	118633	347	cds	Mobile element protein	transposase	NP_287956	9.29016E-75
NODE_13	118663	118920	257	cds	Mobile element protein	mutator family	WP_000343688	6.33332E-52
NODE_13	119536	118967	569	cds	orf, hypothetical protein	type iii effector	NP_287958	1.36888E-134
NODE_13	120513	119602	911	cds	FIG00639664: hypothetical protein	non-lee-encoded effector	NP_287959	0.0
NODE_13	122338	122658	320	cds	FIG00639653: hypothetical protein	nlea11 protein	WP_024230391	5.5249E-56
NODE_13	122622	123662	1040	cds	FIG00639653: hypothetical protein	nlea11 protein	WP_001342109	0.0
NODE_13	127741	127637	104	cds	Putative excisionase	hypothetical protein	WP_001432858	1.93736E-10
NODE_14	19522	19854	332	cds	Mobile element protein	transposon tn3 partial	CDK44333	1.07324E-66
NODE_14	36320	34641	1679	cds	Failed to assign function	transposase	YP_001096405	0.0
NODE_14	106627	106262	365	cds	Mobile element protein	transposase	YP_003237884	2.12894E-72
NODE_16	36264	36121	143	cds	Failed to assign function	hypothetical protein X657_0128	EWD04843	8.78991E-19
NODE_17	3	1703	1700	cds	Putative secreted effector protein	t3ss effector protein 7	WP_001453949	0.0
NODE_17	10095	9760	335	cds	Ribosomal RNA small subunit methyltransferase F (EC 2.1.1.-)	ribosomal rna small subunit methyltransferase f	WP_023567023	2.88335E-75
NODE_17	12173	12063	110	cds	Failed to assign function	upf0759 protein yece	WP_021570441	5.00756E-16
NODE_18	1301	441	860	cds	Dienelactone hydrolase and related enzymes	dienelactone hydrolase family protein	YP_002756580	0.0
NODE_18	4852	4463	389	cds	hypothetical protein	membrane protein	YP_002756604	2.29751E-42
NODE_18	5784	5119	665	cds	IncI1 plasmid conjugative transfer protein TraC	transfer protein c	WP_001370630	1.29454E-146
NODE_18	6566	5925	641	cds	IncI1 plasmid conjugative transfer NusG-type transcription antiterminator TraB	transcription antitermination factor	YP_002756602	8.64568E-152
NODE_18	18766	9266	9500	cds	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)	toxin b	YP_003237860	0.0
NODE_18	19326	19532	206	cds	Mobile element protein	is3 orfb	YP_003377849	8.0324E-37
NODE_18	19895	19593	302	cds	Mobile element protein	transposase	AHG10675	4.00901E-69
NODE_18	21093	20554	539	cds	Mobile element protein	transposase	WP_001443815	6.9256E-50
NODE_18	21742	21236	506	cds	Mobile element protein	integrase core domain protein	WP_001448584	3.6806E-113
NODE_18	22257	21955	302	cds	Mobile element protein	is911 orf1	WP_000464929	3.40069E-49
NODE_18	22714	22238	476	cds	Mobile element protein	transposase family protein	YP_002756593	9.15349E-106
NODE_18	23732	23256	476	cds	Mobile element protein	transposase	ADZ45187	2.07718E-73
NODE_18	24517	23762	755	cds	Mobile element protein	transposase family protein	YP_002756591	8.13809E-180
NODE_18	25670	25095	575	cds	Mobile element protein	transposase	WP_001370617	2.01656E-133
NODE_18	26417	25902	515	cds	Mobile element protein	is91 transposase	YP_002756589	2.0866E-114
NODE_18	32704	31598	1106	cds	Hexosyltransferase homolog	glycosyl transferases group 1 family protein	NP_052687	0.0
NODE_18	33525	32704	821	cds	Polysaccharide deacetylase	polysaccharide deacetylase family protein	YP_325586	0.0
NODE_18	35328	35675	347	cds	Mobile element protein	transposase r14	WP_001364163	5.05876E-72
NODE_18	45999	46229	230	cds	Virulence-associated protein vagC	antidote-toxin recognition family protein	YP_003829221	3.37153E-46
NODE_18	46241	46642	401	cds	VagD	trna -specific endonuclease	NP_863394	1.99424E-79
NODE_18	47349	46804	545	cds	FIG00241420: hypothetical protein	abc transporter nucleotide-binding protein	YP_002296005	1.8393E-125
NODE_18	47417	47719	302	cds	Mobile element protein	transposase family protein	WP_005107299	1.1101E-45
NODE_18	50366	50752	386	cds	Resolvase	site-specific recombinase	YP_002756581	3.46832E-77

NODE_2	482	3	479	cds	Mobile element protein	transposase is116 is110 is902 family protein	WP_000316846	4.96996E-109
NODE_2	229743	229252	491	cds	Z5092 protein	yeew protein	WP_000976831	1.51573E-107
NODE_2	237216	236122	1094	cds	Putative vimentin	50s ribosome-binding gtpase family partial	WP_001470670	0.0
NODE_2	237516	237244	272	cds	Putative vimentin	50s ribosome-binding gtpase family partial	WP_001511229	2.09434E-41
NODE_2	244942	243728	1214	cds	Mobile element protein	transposase	YP_006152018	0.0
NODE_2	245834	245232	602	cds	FIG00638808: hypothetical protein	conserved protein	WP_000804438	2.93805E-143
NODE_2	299630	299328	302	cds	Soluble cytochrome b562	soluble cytochrome b562	YP_003232282	1.30696E-50
NODE_2	302176	301970	206	cds	Soluble cytochrome b562	inner membrane protein yeer	WP_001405854	2.30134E-18
NODE_2	335829	336380	551	cds	Mobile element protein	is609 transposase	YP_003232248	1.63246E-135
NODE_2	407748	406129	1619	cds	Mobile element protein	transposase	YP_003232177	0.0
NODE_20	7864	8076	212	cds	FIG01046738: hypothetical protein	post-segregation killing protein	WP_001297820	6.56033E-44
NODE_20	12486	14705	2219	cds	involved in conjugative DNA transfer	relaxase mobilization nuclease	YP_003377894	0.0
NODE_20	16132	16473	341	cds	involved in conjugative DNA transfer	relaxase mobilization nuclease	WP_021511716	9.87063E-76
NODE_20	17014	16694	320	cds	FIG00640314: hypothetical protein	conjugal transfer protein	WP_001073806	6.58343E-50
NODE_20	23076	19174	3902	cds	Per-activated serine protease autotransporter enterotoxin EspC	serine protease espp	WP_001034095	0.0
NODE_20	23754	24017	263	cds	Mobile element protein	transposase family protein	YP_002756611	1.98034E-45
NODE_21	18349	19935	1586	cds	Failed to assign function	terminase large subunit	WP_024175621	0.0
NODE_21	20147	21337	1190	cds	Phage portal protein	portal protein	WP_023063802	0.0
NODE_23	1494	1102	392	cds	Eaa protein	conserved hypothetical plasmid protein	WP_001427476	3.76254E-80
NODE_23	1699	1436	263	cds	Failed to assign function	ygaa protein	WP_001495195	5.75066E-34
NODE_23	2894	2412	482	cds	FIG00638373: hypothetical protein	plasmid sos inhibition protein b	YP_002756733	1.7562E-112
NODE_23	7488	7264	224	cds	Mobile element protein	transposase	YP_002756741	3.7107E-43
NODE_23	7721	7473	248	cds	Mobile element protein	transposase	WP_000585967	5.75704E-51
NODE_23	7960	7721	239	cds	Mobile element protein	is629 transposase	WP_009451370	2.76575E-35
NODE_23	8848	9825	977	cds	RepFIB replication protein A	replication protein a	WP_000325436	0.0
NODE_24	4661	4170	491	cds	NgrB	gtpase	WP_021527023	8.32876E-99
NODE_25	1	156	155	cds	Failed to assign function	hypothetical protein	WP_024196141	1.41336E-20
NODE_25	526	708	182	cds	Failed to assign function	hypothetical protein	WP_023149634	3.52164E-26
NODE_25	730	1869	1139	cds	hypothetical protein	reverse transcriptase	WP_024242975	0.0
NODE_25	1874	2077	203	cds	Failed to assign function	cold shock domain family protein	WP_000427298	8.58753E-33
NODE_25	2725	2276	449	cds	Failed to assign function	jq1541hypothetical protein - salmonella typhimurium plasmid ntp16	JQ1541	1.05006E-73
NODE_25	3108	2914	194	cds	putative replication regulatory protein	rop family protein	WP_001360351	6.12583E-15
NODE_25	3393	3716	323	cds	Failed to assign function	plasmid mobilization protein	YP_002221413	2.1681E-66
NODE_27	1517	2158	641	cds	unknown protein encoded by prophage CP-933R	t3ss effector	NP_287772	1.15828E-147
NODE_27	3692	2667	1025	cds	Mobile element protein	phage integrase family protein	WP_005096324	0.0
NODE_27	3895	3692	203	cds	FIG01070140: hypothetical protein	prophage protein	WP_000067202	4.0992E-42
NODE_27	4211	3954	257	cds	Failed to assign function	exonuclease	WP_001694302	5.14047E-56
NODE_28	1534	2115	581	cds	Failed to assign function	hypothetical protein	WP_024165516	1.44747E-97
NODE_28	2275	2141	134	cds	Phage minor tail protein	tail partial	WP_001571199	2.28225E-24
NODE_3	366	1	365	cds	FIG00638423: hypothetical protein	type iii secreted effector protein	WP_001426598	4.59154E-37
NODE_3	112871	113905	1034	cds	UDP-N-acetylglucosamine 4,6-dehydratase (EC 4.2.1.-)	fnl1	YP_003229908	0.0

NODE_3	145972	146289	317	cds	Failed to assign function	hypothetical protein	WP_001430457	1.76228E-65
NODE_3	148805	148569	236	cds	FIG00639908: hypothetical protein	aec69	WP_024202074	2.86705E-20
NODE_3	167440	168156	716	cds	Mobile element protein	transposase family protein	WP_001424809	2.03764E-161
NODE_3	212194	213996	1802	cds	Putative ABC iron siderophore transporter, fused permease and ATPase domains	yersiniabactin abc transporter atp-binding protein permease	YP_003229837	0.0
NODE_3	228331	228759	428	cds	metal-dependent phosphohydrolase	phosphohydrolase	WP_001609122	7.57937E-88
NODE_3	281116	278984	2132	cds	Phage tail fiber protein	phage side tail fibrerassembly protein	YP_003229771	0.0
NODE_3	347505	347317	188	cds	Failed to assign function	hypothetical protein	WP_001242071	7.51235E-24
NODE_3	378851	379549	698	cds	Phage minor tail protein	phage minor tail protein l	YP_003221192	9.92686E-171
NODE_3	387950	388540	590	cds	Failed to assign function	t3ss effector protein 8	WP_001429305	2.30779E-137
NODE_3	483173	484912	1739	cds	core protein	rhs repeat-associated core domain protein	WP_004030074	0.0
NODE_3	525702	524302	1400	cds	Glutamate decarboxylase (EC 4.1.1.15)	glutamate decarboxylase	WP_000896043	0.0
NODE_3	540309	540635	326	cds	Mobile element protein	is629 transposase	WP_000165018	7.35469E-53
NODE_3	592227	590182	2045	cds	Dipeptidyl carboxypeptidase Dcp (EC 3.4.15.5)	peptidyl-dipeptidase dcp	YP_003229150	0.0
NODE_3	599390	599157	233	cds	Phage tail fiber assembly protein	tail fiber assembly protein	CDK56023	2.41166E-40
NODE_3	599732	599547	185	cds	Phage tail fiber assembly protein	tail fiber assembly protein	WP_024242942	9.92421E-37
NODE_3	600946	599732	1214	cds	Phage tail fiber protein	phage tail fiber repeat family protein	WP_001685797	0.0
NODE_30	1	240	239	cds	Antirestriction protein klcA	phage antirestriction protein	WP_000982319	5.63201E-52
NODE_30	1924	2109	185	cds	Z5092 protein	phage protein	YP_405943	2.02607E-35
NODE_31	1174	2	1172	cds	Phage DNA replication protein	replication-associated protein a	WP_001453088	0.0
NODE_31	1642	1238	404	cds	Phage minor capsid protein - DNA pilot protein	minor spike protein partial	WP_021402912	6.48509E-72
NODE_32	1424	1558	134	cds	Failed to assign function	hypothetical protein	WP_021548777	5.13942E-18
NODE_4	35227	36540	1313	cds	Phosphate ABC transporter, periplasmic phosphate-binding protein PstS (TC 3.A.1.7.1)	phosphate-binding protein psts	YP_006140992	0.0
NODE_4	187942	187004	938	cds	core protein	rhs family protein	WP_001347575	0.0
NODE_4	189695	188853	842	cds	core protein	rhs repeat-associated core domain protein	WP_001428338	0.0
NODE_4	337441	336275	1166	cds	core protein	type iv secretion protein rhs	WP_000059357	0.0
NODE_4	442878	443555	677	cds	Shikimate kinase I (EC 2.7.1.71)	shikimate kinase 1	NP_709163	1.34976E-146
NODE_4	490748	491230	482	cds	Translation elongation factor Tu	elongation protein tu gtp binding domain-containing protein	WP_001350818	5.97242E-83
NODE_4	491528	491932	404	cds	Translation elongation factor Tu	elongation factor partial	WP_023196752	2.13411E-85
NODE_5	1584	4427	2843	cds	Phage tail fiber protein	phage tail protein	WP_001613294	0.0
NODE_5	18546	18229	317	cds	COG3209: Rhs family protein	paar motif family protein	WP_023140043	3.46957E-68
NODE_5	24441	19705	4736	cds	Failed to assign function	type iv secretion protein rhs	WP_001289031	0.0
NODE_5	26851	24950	1901	cds	VgrG protein	rhs element vgr family protein	WP_000103160	0.0
NODE_5	119987	119247	740	cds	Glutamate Aspartate transport system permease protein GltJ (TC 3.A.1.3.4)	glutamate aspartate transport system permease gltj	YP_309609	9.70624E-166
NODE_5	166369	167802	1433	cds	core protein	protein rhsa	WP_021562708	0.0
NODE_5	265339	266940	1601	cds	Failed to assign function	phage portal lambda family	YP_003229614	0.0
NODE_6	492	1	491	cds	Failed to assign function	phage-related minor tail partial	WP_024177044	1.99099E-47
NODE_6	3688	3338	350	cds	Phage capsid and scaffold	phage head-tail adaptor	WP_001007906	3.15324E-69
NODE_6	6769	4028	2741	cds	Phage portal protein	portal protein	WP_021538399	0.0
NODE_6	8918	6981	1937	cds	Phage capsid and scaffold	phage major capsid hk97 family	WP_001210636	0.0
NODE_6	12357	12217	140	cds	hypothetical protein	protein from phage origin	WP_001433942	1.02359E-17

NODE_6	15691	15353	338	cds	Phage lysis protein S	holin	ETJ59073	6.99693E-48
NODE_6	18919	18740	179	cds	Phage protein	conserved protein	WP_000143457	1.53921E-34
NODE_6	21902	21471	431	cds	FIG00639830: hypothetical protein	tcia	WP_001432304	5.49802E-84
NODE_6	25838	25548	290	cds	Holliday junction resolvase / Crossover junction endodeoxyribonuclease rusA (EC 3.1.22.-)	endodeoxyribonuclease	WP_000904105	4.09979E-40
NODE_6	31166	30705	461	cds	Phage protein	ead ea22-like family protein	YP_003222271	2.3281E-97
NODE_6	31459	31163	296	cds	FIG00640735: hypothetical protein	conserved protein	WP_001445093	6.2525E-55
NODE_6	36377	36733	356	cds	FIG00644832: hypothetical protein	conserved protein	YP_003222282	3.98076E-81
NODE_6	207657	205786	1871	cds	Glutothione ABC transporter ATP-binding protein	glutathione import atp-binding protein	WP_009008403	0.0
NODE_6	279182	278739	443	cds	Phage tail fiber protein	prophage tail fiber family protein	WP_000216556	5.32216E-56
NODE_7	1	204	203	cds	FIG00638146: hypothetical protein	phage protein	WP_021570321	8.43822E-40
NODE_8	2602	2273	329	cds	Phage protein	membrane protein	WP_001386157	3.05312E-38
NODE_8	4140	3475	665	cds	Failed to assign function	serine threonine-protein phosphatase	NP_289188	1.15843E-168
NODE_8	9452	9288	164	cds	FIG00643006: hypothetical protein	membrane protein	WP_000352568	4.33733E-15
NODE_8	177853	177329	524	cds	Mobile element protein	transposase insf for insertion sequence is3a	WP_021548339	2.15822E-126
NODE_9	6379	6627	248	cds	Failed to assign function	integrase core domain protein	NP_286742	1.31972E-47
NODE_9	17928	17428	500	cds	Bundle-forming pilus protein BfpM	#NAME?	WP_021514533	1.89331E-114
NODE_9	18196	18041	155	cds	orf; Unknown function	conserved protein	NP_286741	2.11269E-27
NODE_9	61471	61812	341	cds	Mobile element protein	transposase family protein	NP_286685	3.70591E-68
NODE_9	71752	71213	539	cds	IncF plasmid conjugative transfer surface exclusion protein TraT	complement resistance partial	WP_024183429	1.20674E-87
NODE_9	82237	82422	185	cds	Failed to assign function	hypothetical protein	WP_021528100	3.0705E-25
NODE_9	142881	143351	470	cds	FIG00638423: hypothetical protein	tir-cytoskeleton coupling protein	BAF52034	7.33649E-48
NODE_9	146205	145861	344	cds	Phage tail fiber protein	tail fiber protein	WP_001372101	5.1844E-30

Unique genes present in FHI3

Contig Id	Start	Stop	Length	Type	Function myRAST	Function Blast2GO	Hit Accession	E-value
NODE_101	2	145	143	cds	Resolvase	phage integrase family protein	WP_001349619	1.29259E-21
NODE_108	333	73	260	cds	Failed to assign function	family transcriptional regulator	WP_001307541	9.10718E-56
NODE_109	41912	43078	1166	cds	Putative ABC iron siderophore transporter, fused permease and ATPase domains	yersiniabactin abc transporter atp-binding protein permease	WP_001429235	0.0
NODE_11	28228	28124	104	cds	Failed to assign function	abc transport membrane permease	YP_006126815	6.29891E-12
NODE_11	48191	48370	179	cds	core protein	protein partial	WP_000015117	1.75501E-23
NODE_115	120	1	119	cds	Failed to assign function	phage regulatory protein cii	WP_016232197	3.19079E-17
NODE_116	6205	5978	227	cds	Hypothetical protein yfdR	phage protein	YP_860014	2.19241E-49
NODE_121	253	2	251	cds	Failed to assign function	hypothetical protein, partial	WP_016235307	1.0941E-27
NODE_130	28387	29688	1301	cds	IncI1 plasmid conjugative transfer pilus-tip adhesin protein PilV	shufflon protein a	WP_001500355	0.0
NODE_130	30741	30607	134	cds	Failed to assign function	shufflon protein d	YP_007749304	4.61955E-7
NODE_140	181116	180472	644	cds	Phage NinI serine/threonine phosphatase	serine threonine protein partial	WP_024017475	8.65691E-78
NODE_140	181689	181303	386	cds	Protein NinG	bacteriophage lambda family protein	WP_000631012	5.70643E-36
NODE_140	187221	187015	206	cds	FIG00643006: hypothetical protein	membrane protein	WP_000352568	6.15707E-20
NODE_150	1404	1267	137	cds	Failed to assign function	pf06147 family protein	WP_001677875	2.39603E-25

NODE_156	30	257	227	cds	Failed to assign function	dsbgoxidized	WP_000074670	8.42498E-29
NODE_156	993	847	146	cds	Phage endopeptidase	endopeptidase	WP_000933212	2.2328E-26
NODE_161	534	1	533	cds	Mobile element protein	integrase core domain partial	WP_001467623	3.6483E-119
NODE_164	10998	10636	362	cds	Soluble cytochrome b562	soluble cytochrome b562	YP_003232282	4.76046E-64
NODE_165	2	157	155	cds	Failed to assign function	bacteriophage lambda tail assembly protein i	WP_001318272	3.41508E-24
NODE_165	908	1126	218	cds	Phage tail fiber protein	phage tail partial	WP_001430216	9.73365E-49
NODE_169	1	279	278	cds	hypothetical protein	small toxic polypeptide partial	WP_001317374	2.43399E-47
NODE_169	592	284	308	cds	Mobile element protein	integrase core domain protein	WP_001378447	2.21614E-64
NODE_18	1	393	392	cds	Mobile element protein	transposon tn3 partial	CDK44333	7.9605E-83
NODE_180	772	338	434	cds	Mobile element protein	site-specific phage integrase family partial	WP_001260081	1.53919E-102
NODE_180	1167	796	371	cds	Mobile element protein	site-specific phage integrase family partial	WP_000630372	1.99087E-65
NODE_193	327	1268	941	cds	Primosomal protein I	phage replication protein o	WP_021526407	0.0
NODE_198	4680	4883	203	cds	FIG00640735: hypothetical protein	conserved protein	WP_001445093	1.40263E-35
NODE_198	4868	5092	224	cds	FIG00640735: hypothetical protein	phage protein	YP_003228951	2.43313E-35
NODE_198	5089	5415	326	cds	Phage protein	protein ea22	WP_023982512	3.89567E-64
NODE_20	268	71	197	cds	FIG003276: zinc-binding protein	inhibitor	1LV3_A	2.00305E-34
NODE_20	1021	278	743	cds	FIG002842: hypothetical protein	upf0289 protein yacf	YP_001461272	6.84691E-164
NODE_20	1641	1021	620	cds	Dephospho-CoA kinase (EC 2.7.1.24)	dephospho- kinase	1VHL_A	6.68155E-126
NODE_20	1866	2909	1043	cds	GMP reductase (EC 1.7.1.7)	gmp reductase	NP_414646	0.0
NODE_20	4146	2944	1202	cds	Type IV fimbrial assembly protein PilC	type iv pilin biogenesis protein	YP_001456893	0.0
NODE_20	5521	4136	1385	cds	Type IV fimbrial assembly, ATPase PilB	type ii secretion system n-terminal domain protein	YP_001456894	0.0
NODE_20	5971	5531	440	cds	Type IV pilin PilA	prepilin peptidase-dependent protein d	WP_021568589	8.01101E-93
NODE_20	7067	6174	893	cds	Quinolinate phosphoribosyltransferase [decarboxylating] (EC 2.4.2.19)	nicotinate-nucleotide diphosphorylase	WP_001307569	0.0
NODE_20	7155	7706	551	cds	N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) AmpD	#NAME?	AAY89685	3.89919E-136
NODE_20	7703	8557	854	cds	AmpE protein	protein	WP_001442660	0.0
NODE_20	10514	11278	764	cds	Transcriptional repressor for pyruvate dehydrogenase complex	pyruvate dehydrogenase complex repressor	YP_001571832	5.81071E-138
NODE_20	11439	14102	2663	cds	Pyruvate dehydrogenase E1 component (EC 1.2.4.1)	pyruvate dehydrogenase e1 component	NP_285810	0.0
NODE_20	14117	16009	1892	cds	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)	dihydrolipoyllysine-residue acetyltransferase	WP_024224794	0.0
NODE_20	16334	17758	1424	cds	Failed to assign function	dihydrolipoyl dehydrogenase	YP_001742243	0.0
NODE_20	19586	17829	1757	cds	Putative exported protein	exported protein	YP_003220129	0.0
NODE_20	19941	22538	2597	cds	Aconitate hydratase 2 (EC 4.2.1.3) @ 2-methylisocitrate dehydratase (EC 4.2.1.99)	aconitate hydratase 2	YP_006131945	0.0
NODE_20	22713	23075	362	cds	FIG00638902: hypothetical protein	upf0231 protein yacl	WP_001332320	5.86134E-75
NODE_20	23907	23113	794	cds	S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50), prokaryotic class 1A	s-adenosylmethionine decarboxylase proenzyme	NP_285816	0.0
NODE_20	24789	23923	866	cds	Spermidine synthase (EC 2.5.1.16)	spermidine synthase	NP_414663	0.0
NODE_20	25203	24895	308	cds	Predicted chaperone lipoprotein YacC, potentially involved in protein secretion	hypothetical protein ECSE_0122	YP_002291397	8.76313E-68

NODE_20	25408	26958	1550	cds	Blue copper oxidase CueO precursor	blue copper oxidase	YP_002401256	0.0
NODE_20	29395	27005	2390	cds	Glucose dehydrogenase, PQQ-dependent (EC 1.1.5.2)	quinoprotein glucose dehydrogenase	YP_003227227	0.0
NODE_20	29601	30137	536	cds	Hypoxanthine-guanine phosphoribosyltransferase (EC 2.4.2.8)	hypoxanthine phosphoribosyltransferase	WP_001277500	7.76763E-119
NODE_20	30840	30178	662	cds	Carbonic anhydrase (EC 4.2.1.1)	carbonic anhydrase 2	NP_414668	5.23878E-149
NODE_20	30949	31875	926	cds	ABC-type multidrug transport system, ATPase component	abc transporter atp-binding protein	WP_024237799	0.0
NODE_20	31872	32642	770	cds	ABC-type multidrug transport system, permease component	inner membrane transport permease yadh	WP_000972205	3.18774E-135
NODE_20	32747	33187	440	cds	Putative PTS system IIA component yadI (EC 2.7.1.69)	pts system fructose iia component family protein	YP_002291404	6.13698E-105
NODE_20	33251	34480	1229	cds	Polysaccharide deacetylase	polysaccharide deacetylase family protein	YP_002291405	0.0
NODE_20	34864	34484	380	cds	Aspartate 1-decarboxylase (EC 4.1.1.11)	aspartate 1-decarboxylase	NP_285827	2.41502E-84
NODE_20	35081	36040	959	cds	Transposase ECs0136	transposase	WP_001375465	0.0
NODE_20	36965	36114	851	cds	Pantoate--beta-alanine ligase (EC 6.3.2.1)	pantoate--beta-alanine ligase	YP_003227236	0.0
NODE_20	37771	36977	794	cds	3-methyl-2-oxobutanoate hydroxymethyltransferase (EC 2.1.2.11)	3-methyl-2-oxobutanoate hydroxymethyltransferase	YP_003227237	2.47933E-174
NODE_20	38339	37884	455	cds	Fimbrial protein YadC	fimbrial family protein	WP_000713127	3.98287E-96
NODE_20	39158	38616	542	cds	Fimbrial protein YadC	fimbrial protein	WP_024246517	1.04169E-121
NODE_20	39807	39211	596	cds	Fimbrial protein YadK	fimbrial protein	YP_309174	8.03164E-138
NODE_20	40436	39834	602	cds	Fimbrial protein YadL	fimbrial family protein	WP_001512364	1.50818E-87
NODE_20	41020	40451	569	cds	Fimbrial protein YadM	fimbrial-like adhesin protein	WP_001307574	2.39494E-106
NODE_20	43637	41037	2600	cds	Outer membrane usher protein HtrE	outer membrane usher protein htre	WP_001585345	0.0
NODE_20	44412	43672	740	cds	Chaperone protein EcpD	chaperone ecpd	YP_002291415	1.28074E-173
NODE_20	45101	44517	584	cds	Fimbrial protein YadN	fimbrial family protein	WP_024225961	5.53631E-115
NODE_20	45950	45471	479	cds	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (EC 2.7.6.3)	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	YP_001461312	4.7574E-109
NODE_20	47311	45947	1364	cds	Poly(A) polymerase (EC 2.7.7.19)	poly polymerase	YP_002385637	0.0
NODE_20	48330	47404	926	cds	glutamyl-Q-tRNA synthetase	glutamyl-q trna synthetase	YP_002401277	0.0
NODE_20	48822	48367	455	cds	C4-type zinc finger protein, DksA/TraR family	rna polymerase-binding protein	YP_003519065	1.16202E-79
NODE_20	49704	49000	704	cds	Sugar/maltose fermentation stimulation protein homolog	sugar fermentation stimulation protein	NP_285842	5.17416E-167
NODE_20	50249	49719	530	cds	Failed to assign function	2 -5 rna ligase	WP_009008319	1.45615E-120
NODE_20	50323	52752	2429	cds	ATP-dependent helicase HrpB	atp-dependent helicase	WP_000938620	0.0
NODE_20	52884	55418	2534	cds	Multimodular transpeptidase-transglycosylase (EC 2.4.1.129) (EC 3.4.-.-)	penicillin-binding protein 1b	WP_001774752	0.0
NODE_20	55638	57881	2243	cds	Ferric hydroxamate outer membrane receptor FhuA	ferrichrome-iron receptor	YP_003227252	0.0
NODE_20	57932	58729	797	cds	Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), ATP-binding protein FhuC	iron(3+)-hydroxamate import atp-binding protein	WP_001307576	0.0
NODE_20	58729	59619	890	cds	Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), periplasmic substrate binding protein FhuD	iron(3+)-hydroxamate-binding protein fhud	YP_003227254	0.0
NODE_20	59616	61598	1982	cds	Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), permease component FhuB	iron(3+)-hydroxamate import system permease protein fhub	WP_000044062	0.0
NODE_20	62913	61633	1280	cds	Glutamate-1-semialdehyde aminotransferase (EC 5.4.3.8)	glutamate-1-semialdehyde -aminomutase	WP_001588977	0.0
NODE_20	63138	64559	1421	cds	H(+)/Cl(-) exchange transporter ClcA	h(+) cl(-) exchange transporter	WP_000845380	0.0

NODE_20	64641	64985	344	cds	ErpA, essential respiratory protein A / probable iron binding protein from the HesB_IscA_SufA family	iron-sulfur cluster insertion protein erpa	NP_752141	4.93349E-74
NODE_20	65655	65032	623	cds	Putative inner membrane protein	inner membrane protein	WP_024244632	1.13391E-110
NODE_20	66493	65693	800	cds	Vitamin B12 ABC transporter, B12-binding component Btu	vitamin b12-binding protein	WP_000020101	2.43313E-173
NODE_20	67184	66486	698	cds	5'-methylthioadenosine nucleosidase (EC 3.2.2.16) / S-adenosylhomocysteine nucleosidase (EC 3.2.2.9)	mta sah nucleosidase	WP_000689834	3.22665E-138
NODE_20	67268	68785	1517	cds	Failed to assign function	deoxyguanosinetriphosphate triphosphohydrolase	NP_835891	0.0
NODE_20	68915	70339	1424	cds	HtrA protease/chaperone protein	serine endoprotease	WP_001269776	0.0
NODE_20	70532	71647	1115	cds	Sugar diacid utilization regulator SdaR	carbohydrate diacid regulator	WP_001307578	0.0
NODE_20	72122	71736	386	cds	Chromosome segregation ATPase	upf0325 protein yaeh	WP_001408258	6.90553E-72
NODE_20	73096	72284	812	cds	Phosphoesterase (EC 3.1.-.-)	phosphodiesterase yaei	YP_003227266	0.0
NODE_20	73975	73151	824	cds	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (EC 2.3.1.117)	#NAME?	NP_414708	0.0
NODE_20	76678	74006	2672	cds	[Protein-PII] uridylyltransferase (EC 2.7.7.59)	protein-p-ii uridylyltransferase	WP_001479040	0.0
NODE_20	77534	76740	794	cds	Methionine aminopeptidase (EC 3.4.11.18)	methionine aminopeptidase	YP_003227269	0.0
NODE_20	77902	78627	725	cds	SSU ribosomal protein S2p (SAe)	30s ribosomal protein s2	WP_016255742	1.31023E-175
NODE_20	78885	79736	851	cds	Translation elongation factor Ts	elongation factor ts	3MMP_A	9.82621E-163
NODE_20	79883	80608	725	cds	Uridylate kinase (EC 2.7.4.-)	uridylate kinase	WP_001483307	2.59943E-151
NODE_20	80900	81457	557	cds	Ribosome recycling factor	ribosome recycling factor	WP_001717688	3.86296E-92
NODE_20	81549	82745	1196	cds	1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267)	1-deoxy-d-xylulose 5-phosphate reductoisomerase	1Q0L_A	0.0
NODE_20	82934	83692	758	cds	Failed to assign function	undecaprenyl pyrophosphate synthase	WP_001302924	0.0
NODE_20	83692	83844	152	cds	hypothetical protein	phosphatidate cytidylyltransferase	WP_001399911	5.76381E-8
NODE_20	83813	84562	749	cds	Phosphatidate cytidylyltransferase (EC 2.7.7.41)	phosphatidate cytidylyltransferase	WP_001463147	2.87001E-146
NODE_20	84574	85926	1352	cds	Failed to assign function	rip metalloprotease	WP_001334819	0.0
NODE_20	85956	88388	2432	cds	Outer membrane protein assembly factor YaeT precursor	outer membrane protein assembly factor yaet	WP_023280981	0.0
NODE_20	88510	88995	485	cds	Failed to assign function	chaperone skp	WP_000142627	6.09122E-68
NODE_20	88999	90024	1025	cds	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (EC 2.3.1.-)	udp-3-o-	NP_414721	0.0
NODE_20	90129	90584	455	cds	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase (EC 4.2.1.-)	#NAME?	NP_752165	6.35897E-104
NODE_20	90588	91376	788	cds	Acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase (EC 2.3.1.129)	acyl-	YP_006094512	0.0
NODE_20	91376	92524	1148	cds	Lipid-A-disaccharide synthase (EC 2.4.1.182)	lipid-a-disaccharide synthase	YP_002385676	0.0
NODE_20	92521	93117	596	cds	Ribonuclease HII (EC 3.1.26.4)	ribonuclease hii	NP_414725	8.2824E-131
NODE_20	93154	96636	3482	cds	DNA polymerase III alpha subunit (EC 2.7.7.7)	dna polymerase iii subunit alpha	YP_003227285	0.0
NODE_20	96649	97608	959	cds	Acetyl-coenzyme A carboxyl transferase alpha chain (EC 6.4.1.2)	acetyl- carboxyl alpha subunit	WP_024230905	0.0
NODE_20	99905	100294	389	cds	FIG01280259: hypothetical protein	glyoxalase bleomycin resistance dioxygenase superfamily protein	NP_706132	1.65965E-78
NODE_20	100359	101657	1298	cds	tRNA(Ile)-lysidine synthetase	trna -lysidine synthase	YP_001461357	0.0

NODE_20	101966	101706	260	cds	Rho-specific inhibitor of transcription termination (YaeO)	protein rof	NP_285883	9.96583E-50
NODE_20	102153	101953	200	cds	FIG00905232: hypothetical protein	upf0253 protein yaep	NP_706135	2.47788E-29
NODE_20	102319	102864	545	cds	YaeQ protein	yaeq family protein	WP_001185296	6.3709E-115
NODE_20	102861	103283	422	cds	Hypothetical protein YaeJ with similarity to translation release factor	ribosome-associated protein	WP_001593966	1.79572E-70
NODE_20	103297	104007	710	cds	Copper homeostasis protein CutF precursor / Lipoprotein NlpE involeved in surface adhesion	lipoprotein	YP_309232	2.27587E-167
NODE_21	706	371	335	cds	Phage minor tail protein	phage minor tail protein l	WP_024164940	1.34021E-63
NODE_21	1282	737	545	cds	Phage minor tail protein	phage minor tail protein l	WP_001163793	3.34943E-84
NODE_21	1571	1365	206	cds	Phage minor tail protein	phage minor tail protein partial	WP_001179475	5.5817E-27
NODE_21	2503	2039	464	cds	Phage tail length tape-measure protein 1	phage tail tape measure lambda family	WP_001430275	1.74955E-92
NODE_21	3551	2841	710	cds	Phage tail length tape-measure protein 1	tail protein	WP_001477268	2.60283E-91
NODE_21	4002	3751	251	cds	Failed to assign function	tail length tape measure partial	WP_001359199	5.09354E-30
NODE_21	4232	4059	173	cds	Failed to assign function	tail length tape measure family partial	WP_000960779	1.39859E-23
NODE_21	5254	4229	1025	cds	Phage tail length tape-measure protein 1	phage-related minor tail partial	WP_000212937	9.66223E-143
NODE_21	5753	5316	437	cds	Phage tail length tape-measure protein 1	phage tail tape measure protein	NP_286895	4.39133E-75
NODE_212	3104	2151	953	cds	Dipeptidyl carboxypeptidase Dcp (EC 3.4.15.5)	dipeptidyl carboxypeptidase partial	WP_024236384	0.0
NODE_215	111	1	110	cds	Phage tail fiber protein	host specificity j domain partial	WP_021501697	1.91235E-14
NODE_22	1365	1243	122	cds	Mobile element protein	bfpm-like partial	NP_309419	2.19179E-18
NODE_221	351	1	350	cds	Terminase small subunit	terminase small subunit	WP_000235434	2.44844E-71
NODE_221	1001	867	134	cds	hypothetical protein	hypothetical protein, partial	WP_000303949	1.64734E-16
NODE_226	2	343	341	cds	Phage antitermination protein Q	q anti-termination protein	WP_023307781	5.73427E-76
NODE_229	674	2947	2273	cds	Phage tail fiber protein	phage tail protein	WP_001588584	0.0
NODE_235	93600	92884	716	cds	core protein	rhs repeat-associated core domain protein	WP_024254528	3.84434E-158
NODE_235	98688	97960	728	cds	core protein	rhs repeat family partial	WP_000312597	1.67356E-136
NODE_247	986	3	983	cds	Phage tail fiber protein	tail fiber protein	WP_001375793	2.93767E-24
NODE_247	1418	1197	221	cds	Phage tail fiber protein	prophage tail fiber family protein	WP_001369610	1.56924E-37
NODE_247	1761	1408	353	cds	Phage tail fiber protein	prophage tail fiber family protein	WP_024200316	4.53873E-70
NODE_247	4282	3608	674	cds	Phage tail fiber protein	host specificity partial	WP_024200503	7.5761E-142
NODE_247	4889	4359	530	cds	Phage tail fiber protein	host specificity partial	ETJ69828	1.28007E-62
NODE_247	5677	5216	461	cds	Phage tail fiber protein	host specificity partial	WP_001419937	2.8123E-65
NODE_247	6032	5616	416	cds	Phage tail fiber protein	host specificity j domain partial	WP_000302795	1.4437E-79
NODE_247	7144	6035	1109	cds	Phage tail fiber protein	host specificity partial	WP_000396727	0.0
NODE_247	9655	9458	197	cds	Phage minor tail protein	phage minor tail protein partial	WP_021515024	1.48495E-42
NODE_247	11176	10937	239	cds	Phage tail length tape-measure protein 1	lambda family phage tail tape measure partial	WP_001727972	3.8082E-40
NODE_247	12168	11176	992	cds	Phage tail length tape-measure protein 1	lambda family phage tail tape measure protein	WP_001341548	4.33063E-126
NODE_247	12410	12243	167	cds	Phage minor tail protein	phage tail partial	WP_001152176	3.42085E-27
NODE_247	13654	12647	1007	cds	Phage tail length tape-measure protein 1	phage minor tail family protein	AHG07819	6.16092E-69
NODE_247	17137	16943	194	cds	Phage minor tail protein	minor tail protein	WP_000683154	2.41546E-27
NODE_247	17430	17224	206	cds	Phage minor tail protein	minor tail protein	WP_000683153	3.61582E-28
NODE_248	199954	199508	446	cds	Failed to assign function	hypothetical protein, partial	WP_000641971	1.64678E-36
NODE_269	623	3	620	cds	Mobile element protein	integrase core domain protein	WP_024247191	4.52513E-141
NODE_273	2	142	140	cds	Phage protein	phage protein	WP_000408193	1.89355E-20

NODE_273	668	913	245	cds	Phage EaA protein	eaa protein	WP_000207994	4.23814E-49
NODE_280	2	691	689	cds	Failed to assign function	vimentin yjda	WP_001689131	4.88868E-137
NODE_280	1590	1775	185	cds	Ferredoxin reductase	nte family protein	WP_021519334	6.30223E-16
NODE_284	144937	145077	140	cds	Failed to assign function	hypothetical protein	WP_001366574	2.42988E-19
NODE_289	1	1302	1301	cds	core protein	protein rhsa	WP_024171771	0.0
NODE_29	10210	9536	674	cds	IncF plasmid conjugative transfer surface exclusion protein TraT	complement resistance partial	WP_024183429	9.4487E-123
NODE_302	3	119	116	cds	Uncharacterized protein YeeT	phage protein	WP_021562655	7.1331E-20
NODE_326	243	1	242	cds	Failed to assign function	conserved domain partial	WP_001337247	4.30865E-47
NODE_326	3386	3706	320	cds	Antigen 43 precursor	antigen 43	WP_024165957	8.65897E-42
NODE_326	3793	5073	1280	cds	Antigen 43 precursor	autotransporter beta-domain protein	WP_001432038	0.0
NODE_326	5401	6024	623	cds	FIG00638146: hypothetical protein	phage protein	WP_001234394	4.21464E-125
NODE_326	6148	6390	242	cds	Antirestriction protein klcA	phage antirestriction protein	WP_021548403	1.61993E-23
NODE_326	6694	6344	350	cds	FIG00638371: hypothetical protein	hypothetical protein c1279	NP_753193	1.11891E-51
NODE_326	6723	6917	194	cds	Antirestriction protein klcA	antirestriction family protein	WP_000214417	8.60141E-17
NODE_326	7075	7299	224	cds	Antirestriction protein klcA	phage antirestriction protein	WP_000213727	8.17548E-36
NODE_343	988	2	986	cds	Antigen 43 precursor	antigen 43	YP_006152022	6.35079E-144
NODE_343	1490	942	548	cds	Antigen 43 precursor	antigen 43 precursor (fluffing protein)	WP_023156906	3.46559E-69
NODE_358	167	3	164	cds	Phage minor tail protein	phage minor tail protein	WP_000847293	5.78775E-30
NODE_364	2	1360	1358	cds	Phage portal protein	portal protein	YP_003233932	0.0
NODE_364	1563	2633	1070	cds	Phage portal protein	phage portal hk97 family	YP_002328963	0.0
NODE_365	152	3	149	cds	RelF inactive antibacterial toxin protein	regulatory protein	WP_024184904	1.82494E-13
NODE_365	3981	3631	350	cds	LygF	phage protein	WP_001151144	1.11146E-43
NODE_365	4521	4225	296	cds	LygF	phage protein	WP_024168731	7.42931E-62
NODE_365	7761	7865	104	cds	Failed to assign function	hypothetical protein	WP_016237926	6.3704E-14
NODE_378	1	1188	1187	cds	Phage tail fiber protein	host specificity protein j	WP_024222397	0.0
NODE_378	1142	2077	935	cds	Phage tail fiber protein	host specificity protein partial	WP_000300018	0.0
NODE_380	423	1	422	cds	Phage tail fiber protein	host specificity protein partial	WP_000515015	2.31437E-103
NODE_383	333	1	332	cds	FIG00639830: hypothetical protein	tcia	WP_001432304	6.03606E-63
NODE_392	50845	50153	692	cds	Exodeoxyribonuclease VIII (EC 3.1.11.-)	exonuclease partial	WP_000313748	1.86263E-163
NODE_392	111295	110717	578	cds	Mobile element protein	transposase is200 like family protein	WP_001499638	2.52993E-110
NODE_392	203602	203778	176	cds	FIG00638426: hypothetical protein	hypothetical protein	WP_020236084	6.38591E-31
NODE_403	42549	43637	1088	cds	Mobile element protein	transposase	YP_006152018	0.0
NODE_408	1741	1914	173	cds	Putative tail component of prophage CP-933K	tail protein	WP_001432703	5.83005E-26
NODE_408	2365	2769	404	cds	Putative tail component of prophage CP-933K	bacterial ig-like domain family protein	YP_003227936	3.42873E-56
NODE_408	3241	3834	593	cds	Phage minor tail protein	phage minor tail protein g	WP_024234758	4.50875E-90
NODE_408	3861	4118	257	cds	Phage minor tail protein	phage tail assembly protein t	WP_016241241	4.07967E-29
NODE_41	118	2	116	cds	Failed to assign function	orf61	CAC05539	1.99247E-14
NODE_413	617	66	551	cds	Ferredoxin reductase	patatin-like phospholipase family protein	WP_001555248	9.9199E-119
NODE_417	803	1069	266	cds	Failed to assign function	endopeptidase	WP_024017595	3.26118E-33
NODE_417	4983	5543	560	cds	Phage terminase, large subunit	large partial	WP_001438306	5.45231E-130
NODE_417	5730	6134	404	cds	Failed to assign function	portal partial	WP_023141508	4.1846E-55
NODE_417	6139	7053	914	cds	Failed to assign function	phage portal lambda family	WP_001430285	0.0

NODE_417	7056	7880	824	cds	Failed to assign function	phage portal lambda family	WP_001431213	5.37398E-179
NODE_417	10728	11174	446	cds	Phage major capsid protein	head partial	WP_001483950	2.36446E-96
NODE_417	11193	11600	407	cds	Phage DNA-packaging protein	dna packaging protein	WP_000012985	1.37754E-81
NODE_417	12533	12757	224	cds	Phage minor tail protein	phage minor tail u family protein	WP_001429674	7.13702E-40
NODE_417	12723	12977	254	cds	Phage minor tail protein	minor tail protein u	WP_000683072	1.69447E-33
NODE_417	12985	13164	179	cds	Putative tail component of prophage CP-933K	bacterial ig-like domain family protein	WP_024227684	8.15039E-33
NODE_427	2933	2457	476	cds	putative enzyme; Integration, recombination (Phage or Prophage Related)	partial	WP_024219402	7.79072E-77
NODE_427	3250	2930	320	cds	Recombinational DNA repair protein RecT (prophage associated)	phage recombination protein bet	WP_000292833	1.68687E-66
NODE_427	3924	3247	677	cds	Recombinational DNA repair protein RecT (prophage associated)	phage recombination protein partial	WP_001499621	5.70686E-87
NODE_451	262	2	260	cds	putative endopeptidase	bacteriophage lysis protein	WP_001431885	9.14448E-46
NODE_451	709	446	263	cds	Endopeptidase (EC 3.4.-.-)	endopeptidase	WP_001255338	1.57678E-35
NODE_456	120310	119654	656	cds	Z5092 protein	phage protein	AAL67388	4.80398E-87
NODE_46	9923	10912	989	cds	Glutothione ABC transporter ATP-binding protein	abc transporter family protein	WP_016232537	0.0
NODE_46	10909	11793	884	cds	Glutothione ABC transporter ATP-binding protein	abc transporter	WP_001429079	0.0
NODE_468	742	497	245	cds	Attachment invasion locus protein precursor	membrane partial	WP_023909234	5.88579E-36
NODE_468	1002	883	119	cds	Phage tail fiber protein	fibronectin type iii family partial	WP_001442277	1.00784E-17
NODE_481	93694	93852	158	cds	hypothetical protein ECs1673	upf0745 protein	NP_309700	3.14951E-27
NODE_481	152050	152229	179	cds	Failed to assign function	conserved protein	CAU97184	1.27105E-28
NODE_481	181694	180807	887	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-93	exonuclease family protein	WP_001684017	4.69415E-155
NODE_481	182484	181633	851	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-93	exonuclease family protein	WP_001428583	0.0
NODE_481	182830	182489	341	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-93	exonuclease ds dna exonuclease encoded by prophage	WP_001416113	3.46918E-62
NODE_481	183385	183128	257	cds	Failed to assign function	exonuclease ds dna exonuclease encoded by partial	WP_021501910	7.29824E-38
NODE_481	184906	184562	344	cds	FIG00641683: hypothetical protein	hypothetical protein	WP_001340571	1.36257E-64
NODE_481	186592	185636	956	cds	Exodeoxyribonuclease VIII (EC 3.1.11.-)	exonuclease partial	WP_000105129	0.0
NODE_481	189123	188938	185	cds	Mobile element protein	phage protein	WP_001343118	1.64439E-25
NODE_481	190031	189315	716	cds	Failed to assign function	helix-turn-helix family protein	WP_016569934	3.29242E-158
NODE_481	190081	190296	215	cds	Failed to assign function	regulatory protein cro	YP_003233552	8.90056E-39
NODE_485	65869	65567	302	cds	Mobile element protein	h repeat-associated protein yhhi	WP_024226894	8.86821E-64
NODE_488	1	549	548	cds	Failed to assign function	rhs repeat-associated core domain protein	WP_024243127	1.40148E-118
NODE_488	1713	2087	374	cds	Mobile element protein	h repeat-associated protein yhhi	WP_024232483	1.09393E-68
NODE_488	45304	44129	1175	cds	Glutamate decarboxylase (EC 4.1.1.15)	glutamate decarboxylase	WP_001326770	0.0
NODE_488	45624	45304	320	cds	Glutamate decarboxylase (EC 4.1.1.15)	glutamate decarboxylase partial	WP_001467225	3.37732E-50
NODE_488	60382	60062	320	cds	Glutamate decarboxylase (EC 4.1.1.15)	glutamate decarboxylase partial	WP_000372190	5.26121E-27
NODE_5	213	1	212	cds	Failed to assign function	escherichia coli chi7122 chi7122 genomic chromosome chi7122	WP_000696434	4.97774E-37
NODE_5	3079	2234	845	cds	Glutamate Aspartate transport system permease protein GltJ (TC 3.A.1.3.4)	amino abc permease 3-tm his glu gln arg opine family domain protein	WP_024220044	1.23199E-179
NODE_50	1	1050	1049	cds	Failed to assign function	transposase	WP_024236171	0.0
NODE_503	1902	1	1901	cds	Phage tail length tape-measure protein 1	tail length tape measure protein	WP_001431289	0.0
NODE_503	2532	2068	464	cds	Phage tail length tape-measure protein 1	tape measure domain partial	WP_021546323	3.42858E-44
NODE_503	2899	2690	209	cds	Failed to assign function	phage tail protein	WP_000320464	7.9578E-39

NODE_503	3851	3450	401	cds	Phage neck whiskers	major tail protein	YP_001449301	9.02308E-77
NODE_503	6133	5723	410	cds	Phage capsid and scaffold	phage head-tail adaptor	WP_001007902	1.53289E-45
NODE_503	6302	6144	158	cds	FIG00638841: hypothetical protein	dna packaging protein	WP_000341806	2.06888E-22
NODE_503	7476	6628	848	cds	Phage portal protein	phage portal protein	WP_001412419	0.0
NODE_504	5117	3828	1289	cds	Putative vimentin	hypothetical protein, partial	WP_000544658	0.0
NODE_504	5488	5228	260	cds	Antigen 43 precursor	antigen partial	WP_001568367	1.84522E-45
NODE_508	5684	6352	668	cds	NgrB	gtp-binding protein	YP_003230355	1.55635E-137
NODE_508	6581	7090	509	cds	NgrB	yeep protein	WP_001483999	7.63101E-73
NODE_508	10657	11259	602	cds	Antigen 43 precursor	antigen partial	WP_023140470	6.65412E-85
NODE_508	13972	14139	167	cds	FIG00639908: hypothetical protein	aec69	WP_024202074	4.90594E-21
NODE_525	312	1	311	cds	core protein	core partial	WP_009439342	1.77616E-60
NODE_525	793	254	539	cds	core protein	rhs repeat family partial	WP_001593553	2.16724E-58
NODE_525	164073	164594	521	cds	Shikimate kinase I (EC 2.7.1.71)	shikimate kinase 1	WP_001503195	1.59222E-112
NODE_525	211787	212098	311	cds	Translation elongation factor Tu	elongation factor partial	WP_001308090	1.06818E-23
NODE_528	133	435	302	cds	Failed to assign function	hypothetical protein	WP_000555557	3.24094E-27
NODE_53	2	376	374	cds	DNA repair protein RadC	phage dna repair protein	WP_001379499	2.46984E-67
NODE_53	1568	1714	146	cds	Z5092 protein	phage protein	CDL48076	1.8952E-25
NODE_535	1	243	242	cds	Failed to assign function	#NAME?	WP_000363187	5.38287E-33
NODE_539	665	940	275	cds	Phage endolysin	partial	WP_023150136	7.24976E-35
NODE_554	1	165	164	cds	Z5092 protein	yeew protein	WP_024179757	5.03768E-32
NODE_554	89194	89075	119	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-93	exodeoxyribonuclease viii	WP_001360666	1.56682E-19
NODE_557	12667	13365	698	cds	Failed to assign function	outer membrane lipoprotein rz1 lambdoid prophage rac	WP_020236743	7.38152E-63
NODE_557	13690	13397	293	cds	Lipoprotein Bor	lipoprotein bor	YP_002386242	1.60985E-57
NODE_559	50468	50136	332	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-93	exonuclease family protein	WP_016245381	1.73836E-59
NODE_56	68259	67882	377	cds	Maltoporin (maltose/maltodextrin high-affinity receptor, phage lambda receptor protein)	cryptic outer membrane porin	WP_001731343	2.00055E-77
NODE_56	69497	68298	1199	cds	Maltoporin (maltose/maltodextrin high-affinity receptor, phage lambda receptor protein)	cryptic outer membrane porin	CDL28375	0.0
NODE_56	70977	69583	1394	cds	6-phospho-beta-glucosidase (EC 3.2.1.86)	6-phospho-beta-glucosidase	YP_002405111	0.0
NODE_56	72873	70996	1877	cds	PTS system, beta-glucoside-specific IIB component (EC 2.7.1.69) / PTS system, beta-glucoside-specific IIC component (EC 2.7.1.69) / PTS system, beta-glucoside-specific IIA component (EC 2.7.1.69)	pts system beta-glucoside-specific eiibca component	YP_005280242	0.0
NODE_56	73843	73007	836	cds	Beta-glucoside bgl operon antiterminator, BglG family	cryptic beta-glucoside bgl operon antiterminator	NP_418179	0.0
NODE_56	74854	74129	725	cds	Phosphate transport system regulatory protein PhoU	phosphate transport system regulatory protein	NP_418180	1.38843E-159
NODE_56	75642	74869	773	cds	Phosphate transport ATP-binding protein PstB (TC 3.A.1.7.	phosphate import atp-binding protein	NP_290357	0.0
NODE_56	76623	75733	890	cds	Phosphate transport system permease protein PstA (TC 3.A.1.7.1)	phosphate abc permease protein	NP_312689	1.09066E-169
NODE_56	77582	76623	959	cds	Phosphate transport system permease protein PstC (TC 3.A.1.7.1)	phosphate abc permease protein	WP_001562350	1.29418E-169
NODE_56	78709	77669	1040	cds	Phosphate ABC transporter, periplasmic phosphate-binding protein PstS (TC 3.A.1.7.1)	phosphate-binding protein psts	WP_000867121	0.0
NODE_56	114355	116703	2348	cds	Putative secreted effector protein	t3ss effector protein 7	WP_001453949	0.0

NODE_560	2243	1734	509	cds	TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins	bifunctional enterobactin receptor adhesin partial	WP_001432076	6.05081E-98
NODE_574	2	208	206	cds	Failed to assign function	tail assembly partial	WP_001417262	2.73077E-40
NODE_580	2	793	791	cds	Phage tail fiber protein	host specificity partial	WP_001587011	2.58682E-157
NODE_580	1480	1148	332	cds	hypothetical protein	hypothetical protein	WP_023607108	7.82898E-53
NODE_580	1633	2127	494	cds	Phage tail fiber protein	host specificity partial	WP_000305955	2.37137E-103
NODE_580	2282	2800	518	cds	Phage tail fiber protein	host specificity partial	WP_000299137	1.50308E-103
NODE_580	3225	3470	245	cds	Attachment invasion locus protein precursor	outer membrane precursor lom	WP_000296839	3.78358E-45
NODE_580	3533	3823	290	cds	Attachment invasion locus protein precursor	membrane protein	WP_003566761	1.1616E-58
NODE_583	41830	42234	404	cds	Glutamate Aspartate transport system permease protein GltJ (TC 3.A.1.3.4)	amino abc permease 3-tm his glu gln arg opine family domain protein	WP_021501086	1.53168E-86
NODE_583	42271	42570	299	cds	Glutamate Aspartate transport system permease protein GltJ (TC 3.A.1.3.4)	amino abc permease 3-tm his glu gln arg opine family domain protein	WP_001330086	1.28469E-58
NODE_583	71177	71001	176	cds	Failed to assign function	cold shock protein	YP_001454079	5.60129E-30
NODE_583	137087	137620	533	cds	core protein	type iv secretion protein rhs	WP_000014899	5.72819E-74
NODE_583	137580	138083	503	cds	VgrG protein	rhs element vgr family partial	CDK49641	1.2557E-107
NODE_583	138592	140514	1922	cds	COG3209: Rhs family protein	paar motif family protein	WP_024237458	0.0
NODE_583	140641	143640	2999	cds	Failed to assign function	type iv secretion protein rhs	WP_024017455	0.0
NODE_583	144969	144718	251	cds	hypothetical protein	phage tail fiber repeat family protein	WP_021568668	1.43669E-35
NODE_587	58135	58635	500	cds	metal-dependent phosphohydrolase	phosphohydrolase	WP_021513658	7.34975E-93
NODE_587	134234	133899	335	cds	Failed to assign function	9-o-acetyl-n-acetylneuraminate esterase domain protein	WP_000202149	3.66628E-66
NODE_587	174994	173891	1103	cds	Integrase	int protein	WP_001692455	0.0
NODE_587	191207	190980	227	cds	putative excisionase	excisionase	WP_021538069	1.77093E-38
NODE_587	207674	208213	539	cds	Mobile element protein	9-o-acetyl-n-acetylneuraminate partial	ETJ69708	1.31401E-79
NODE_587	208237	208506	269	cds	Failed to assign function	9-o-acetyl-n-acetylneuraminate partial	WP_000329031	6.97613E-41
NODE_587	208503	209141	638	cds	Failed to assign function	9-o-acetyl-n-acetylneuraminate esterase	WP_009436089	1.05379E-138
NODE_587	209147	210256	1109	cds	unknown protein encoded within prophage CP-933V	9-o-acetyl-n-acetylneuraminate esterase domain protein	WP_000296435	0.0
NODE_587	210621	211199	578	cds	Phage protein	hypothetical protein	WP_000332739	7.67455E-79
NODE_602	2342	1701	641	cds	Mobile element protein	transposase family protein	WP_001424809	4.87077E-129
NODE_602	3052	2339	713	cds	Mobile element protein	transposase family protein	YP_003232177	3.69256E-154
NODE_602	8956	9651	695	cds	FIG00638808: hypothetical protein	conserved protein	YP_003229861	2.57338E-98
NODE_602	139745	138750	995	cds	Guanine-hypoxanthine permease	permease yjcd	WP_021545140	3.31E-144
NODE_602	140098	139730	368	cds	Guanine-hypoxanthine permease	xanthine uracil permeases family partial	WP_000106902	3.2358E-61
NODE_608	100728	100561	167	cds	Mobile element protein	is3 transposase partial	WP_001431166	3.24905E-29
NODE_608	101940	102866	926	cds	hypothetical protein	prophage protein	WP_024220649	0.0
NODE_608	102853	103401	548	cds	Failed to assign function	prophage protein	WP_001131908	4.54359E-101
NODE_608	103934	103773	161	cds	Failed to assign function	pf06147 family protein	ETJ58659	1.02111E-30
NODE_608	106641	106201	440	cds	Phage-related protein	kila-n domain protein	YP_860025	6.3922E-96
NODE_608	107279	107055	224	cds	Holliday junction resolvase / Crossover junction endodeoxyribonuclease rusA (EC 3.1.22.-)	crossover junction endodeoxyribonuclease partial	WP_023147754	4.01779E-28
NODE_608	108193	107858	335	cds	DNA adenine methylase	phage n-6-adenine-methyltransferase	YP_001272560	7.11252E-73
NODE_608	116216	115701	515	cds	FIG00643513: hypothetical protein	hypothetical protein	WP_001434468	4.23105E-118
NODE_608	116691	116912	221	cds	FIG00638873: hypothetical protein	hypothetical protein ETEC_0272	YP_006113858	8.36732E-35

NODE_608	116912	117166	254	cds	FIG00638873: hypothetical protein	hypothetical protein ETEC_0272	YP_006113858	3.36743E-52
NODE_608	117232	117726	494	cds	FIG00642676: hypothetical protein	phage protein	WP_001401081	5.01827E-89
NODE_608	117753	118310	557	cds	FIG00642676: hypothetical protein	phage protein	WP_001555696	1.77511E-89
NODE_608	118438	118746	308	cds	Hypothetical protein yfdR	phage protein	WP_000008244	1.34114E-39
NODE_608	119222	120100	878	cds	Phage NinC	phosphoadenosine phosphosulfate reductase family protein	YP_003230648	0.0
NODE_608	120293	120532	239	cds	FIG00641267: hypothetical protein	hypothetical protein ECO26_3713	YP_003230650	2.99035E-32
NODE_608	120529	120858	329	cds	Phage protein	ead ea22-like family protein	YP_003230651	4.92587E-65
NODE_608	120860	121723	863	cds	Phage EaA protein	valyl-trna synthetase	YP_003230652	0.0
NODE_608	122373	123548	1175	cds	Phage integrase	integrase	YP_003230655	0.0
NODE_608	124031	124939	908	cds	hypothetical protein	membrane protein	YP_003230656	0.0
NODE_608	132095	136399	4304	cds	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.4€	toxin b	WP_024220711	0.0
NODE_608	136933	141528	4595	cds	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.4€	toxin b	ETJ70347	0.0
NODE_608	142315	142007	308	cds	Mobile element protein	transposase family protein	WP_001549231	1.53723E-32
NODE_608	210432	211193	761	cds	Betaine aldehyde dehydrogenase (EC 1.2.1.8)	betaine aldehyde dehydrogenase	WP_001426529	8.911E-176
NODE_618	93009	92548	461	cds	core protein	rhs repeat-associated core domain partial	WP_024190865	3.36784E-98
NODE_618	95144	95689	545	cds	Failed to assign function	hypothetical protein, partial	WP_000550374	5.65941E-53
NODE_618	97020	96547	473	cds	Failed to assign function	rhs family protein	WP_021351779	7.58249E-88
NODE_626	51760	50633	1127	cds	DNA primase (EC 2.7.7.-), phage-associated	dna primase	YP_003230064	0.0
NODE_633	180	1	179	cds	Failed to assign function	caudovirus prohead protease family protein	WP_001413580	1.25324E-30
NODE_633	3255	3058	197	cds	putative Dnase	hnh nuclease	WP_001431535	1.08391E-17
NODE_633	5283	5017	266	cds	FIG00639442: hypothetical protein	conserved protein	NP_309563	1.84816E-39
NODE_633	5426	5307	119	cds	Failed to assign function	endopeptidase	WP_000320120	1.11115E-17
NODE_638	764	318	446	cds	Phage portal protein	phage portal protein	WP_001432333	4.50521E-58
NODE_638	2455	1130	1325	cds	Phage portal protein	portal protein	YP_003228271	0.0
NODE_638	3746	3408	338	cds	Failed to assign function	structural protein	WP_024241612	3.93379E-65
NODE_638	4337	3954	383	cds	Failed to assign function	phage major capsid hk97 partial	WP_001369400	2.5836E-59
NODE_638	4680	4438	242	cds	Failed to assign function	structural domain protein	WP_024228898	2.03174E-29
NODE_638	5012	4677	335	cds	Failed to assign function	head protein prohead partial	WP_001430293	1.66747E-52
NODE_638	5776	5078	698	cds	Failed to assign function	head protein prohead partial	WP_001453158	4.58282E-132
NODE_638	6239	5745	494	cds	Failed to assign function	caudovirus prohead protease family protein	WP_001429458	4.82507E-91
NODE_646	228	1	227	cds	Phage tail length tape-measure protein 1	tail protein	WP_001436222	4.39545E-28
NODE_646	1026	262	764	cds	Phage tail length tape-measure protein 1	tail length tape measure partial	WP_024230708	3.9804E-92
NODE_646	1457	1035	422	cds	Phage tail length tape-measure protein 1	prophage tail length tape measure family protein	WP_001409008	6.08372E-72
NODE_646	2256	1585	671	cds	Phage tail length tape-measure protein 1	tail protein	WP_001436177	2.32523E-126
NODE_646	2506	2237	269	cds	Phage tail length tape-measure protein 1	tail protein	WP_001428026	2.88027E-36
NODE_656	3	170	167	cds	FIG00644525: hypothetical protein	hypothetical protein	WP_001113757	5.20163E-29
NODE_656	77078	75999	1079	cds	Mobile element protein	transposase is66 family protein	WP_001415941	0.0
NODE_656	78478	78236	242	cds	Hypothetical lipoprotein yehR	ethanolamine utilization protein	WP_024171011	2.48071E-9
NODE_656	182043	182204	161	cds	UDP-N-acetylglucosamine 4,6-dehydratase (EC 4.2.1.-)	fnl1	WP_001342286	5.19469E-23
NODE_656	182174	183076	902	cds	UDP-N-acetylglucosamine 4,6-dehydratase (EC 4.2.1.-)	fnl1	YP_003229908	0.0
NODE_656	213127	212921	206	cds	Failed to assign function	phage dna repair protein	WP_001449313	3.36456E-27
NODE_656	213744	213622	122	cds	Failed to assign function	phage dna repair partial	WP_020232840	2.04774E-15
NODE_656	213906	213760	146	cds	Failed to assign function	upf0758 protein yees	WP_021520249	3.87882E-23

NODE_656	214818	214462	356	cds	hypothetical protein	upf0380 protein partial	WP_000313106	3.3153E-41
NODE_656	216131	217093	962	cds	Phage tail fiber protein	host specificity protein	WP_000302466	3.00236E-149
NODE_656	217968	218456	488	cds	Mobile element protein	transposase is66 family protein	WP_000099149	1.77443E-99
NODE_71	834	1034	200	cds	Failed to assign function	hypothetical protein	WP_024182289	1.36611E-36
NODE_71	8718	9296	578	cds	Putative cytoplasmic protein	hypothetical protein, partial	WP_024252710	2.91615E-107
NODE_71	9277	10587	1310	cds	Phage antitermination protein Q	phage protein	WP_024238166	2.14279E-124
NODE_71	10542	10874	332	cds	Failed to assign function	hypothetical protein, partial	WP_024165906	2.46954E-78
NODE_71	11081	11242	161	cds	Failed to assign function	hypothetical protein	WP_001453943	2.19202E-31
NODE_71	11255	11569	314	cds	Holliday junction resolvase / Crossover junction endodeoxyribonuclease rusA (EC 3.1.22.-)	endodeoxyribonuclease	WP_000904110	2.8976E-58
NODE_71	11523	11897	374	cds	Holliday junction resolvase / Crossover junction endodeoxyribonuclease rusA (EC 3.1.22.-)	endodeoxyribonuclease	WP_001217412	1.66446E-35
NODE_71	26041	26853	812	cds	Mobile element protein	9-o-acetyl-n-acetylneuraminate partial	WP_001410975	1.18887E-118
NODE_71	26847	27245	398	cds	Mobile element protein	5 -nucleotidase	WP_000001261	5.59513E-88
NODE_78	5378	5809	431	cds	Mobile element protein	integrase core domain protein	WP_001675491	6.75875E-102
NODE_80	17	253	236	cds	Terminase small subunit	terminase small subunit	YP_004169181	8.66398E-36
NODE_80	3176	4093	917	cds	Failed to assign function	portal protein b	WP_021553629	4.97069E-132
NODE_80	4108	4482	374	cds	Failed to assign function	phage portal lambda partial	WP_000856879	8.80842E-59
NODE_80	4434	5114	680	cds	Failed to assign function	portal protein	ETJ68185	1.37018E-141
NODE_80	5126	5371	245	cds	Failed to assign function	phage portal lambda family	WP_021573087	1.94102E-31
NODE_80	7198	7770	572	cds	Phage major capsid protein	head protein	WP_001414157	1.92544E-129
NODE_80	7767	8210	443	cds	Phage major capsid protein	major capsid protein	ETJ58577	5.38523E-104
NODE_80	9756	9929	173	cds	Failed to assign function	tail partial	WP_021533913	1.46165E-26
NODE_80	9926	10210	284	cds	Phage minor tail protein	minor tail protein	WP_000683126	1.72425E-56
NODE_89	15817	16080	263	cds	FIG00639870: hypothetical protein	prophage protein	YP_003221208	8.51017E-58

Table S13. Analysis of genes differently present in strain FHI48 from a HUS-patient and FHI43 and FHI62 from the same outbreak

Unique genes present in FHI48

Contig Id	Start	Stop	Length	Type	Function myRAST
NODE_0	388553	387549	1004	cds	Putative large exoprotein involved in heme utilization or adhesion of ShlA/HecA/FhaA family
NODE_0	390868	389684	1184	cds	Mobile element protein
NODE_0	419603	419917	314	cds	Failed to assign function
NODE_1	393947	393666	281	cds	FIG00639363: hypothetical protein
NODE_1	396094	394205	1889	cds	Prophage Clp protease-like protein
NODE_1	397485	397171	314	cds	FIG00638168: hypothetical protein
NODE_1	399518	397770	1748	cds	DNA primase (EC 2.7.7.-), phage-associated / Replicative helicase RepA
NODE_1	400047	399820	227	cds	FIG00639856: hypothetical protein
NODE_1	401733	400222	1511	cds	FIG00639870: hypothetical protein
NODE_1	403620	402391	1229	cds	Mobile element protein
NODE_1	496679	496305	374	cds	Failed to assign function
NODE_1	497969	498220	251	cds	Mobile element protein
NODE_11	101203	100979	224	cds	Mobile element protein
NODE_14	73809	73456	353	cds	Failed to assign function
NODE_15	764	3	761	cds	IncF plasmid conjugative transfer DNA-nicking and unwinding protein TraI
NODE_15	79385	79209	176	cds	YihA
NODE_16	2920	2774	146	cds	hypothetical protein
NODE_16	8623	9084	461	cds	Failed to assign function
NODE_16	37707	38360	653	cds	hypothetical protein
NODE_16	38935	39183	248	cds	Failed to assign function
NODE_16	44252	44572	320	cds	Failed to assign function
NODE_16	45043	45486	443	cds	FIG00642006: hypothetical protein
NODE_16	51145	51546	401	cds	hypothetical protein
NODE_16	51564	52316	752	cds	TraL
NODE_16	52294	53025	731	cds	Failed to assign function
NODE_16	53488	53691	203	cds	hypothetical protein
NODE_17	149	3	146	cds	Failed to assign function
NODE_17	11816	6546	5270	cds	IncF plasmid conjugative transfer DNA-nicking and unwinding protein TraI
NODE_17	14014	11816	2198	cds	IncF plasmid conjugative transfer protein TraD
NODE_17	15043	14267	776	cds	IncF plasmid conjugative transfer surface exclusion protein TraT
NODE_17	15527	15030	497	cds	IncF plasmid conjugative transfer surface exclusion protein TraS

NODE_17	18370	15548	2822	cds	IncF plasmid conjugative transfer protein TraG
NODE_17	19740	18367	1373	cds	Failed to assign function
NODE_17	20042	19740	302	cds	IncF plasmid conjugative transfer protein TrbJ
NODE_17	20577	20032	545	cds	IncF plasmid conjugative transfer protein TrbB
NODE_17	20848	20564	284	cds	IncF plasmid conjugative transfer protein TraQ
NODE_17	21314	20967	347	cds	IncF plasmid conjugative transfer protein TrbA
NODE_17	22073	21330	743	cds	IncF plasmid conjugative transfer pilus assembly protein TraF
NODE_17	22323	22066	257	cds	IncF plasmid conjugative transfer protein TrbE
NODE_17	24158	22350	1808	cds	IncF plasmid conjugative transfer protein TraN
NODE_17	24793	24155	638	cds	IncF plasmid conjugative transfer protein TrbC
NODE_17	25273	24821	452	cds	hypothetical protein
NODE_17	25764	25303	461	cds	hypothetical protein
NODE_17	26782	25790	992	cds	IncF plasmid conjugative transfer pilus assembly protein TraU
NODE_17	27411	26779	632	cds	IncF plasmid conjugative transfer pilus assembly protein TraW
NODE_17	27794	27408	386	cds	IncF plasmid conjugative transfer protein TrbI
NODE_17	30418	27791	2627	cds	IncF plasmid conjugative transfer pilus assembly protein TraC
NODE_17	30585	30481	104	cds	FIG00641102: hypothetical protein
NODE_17	30799	30578	221	cds	IncF plasmid conjugative transfer protein TraR
NODE_17	31449	30934	515	cds	IncF plasmid conjugative transfer pilus assembly protein TraV
NODE_17	31697	31446	251	cds	IncF plasmid conjugative transfer protein TrbG
NODE_17	31906	31709	197	cds	IncF plasmid conjugative transfer protein TrbD
NODE_17	32483	31893	590	cds	IncF plasmid conjugative transfer protein TraP
NODE_17	33900	32473	1427	cds	IncF plasmid conjugative transfer pilus assembly protein TraB
NODE_17	34628	33900	728	cds	IncF plasmid conjugative transfer pilus assembly protein TraK
NODE_17	35181	34615	566	cds	IncF plasmid conjugative transfer pilus assembly protein TraE
NODE_17	35514	35203	311	cds	IncF plasmid conjugative transfer pilus assembly protein TraL
NODE_17	35921	35529	392	cds	IncF plasmid conjugative transfer pilin protein TraA
NODE_17	36182	35955	227	cds	IncF plasmid conjugative transfer regulator TraY
NODE_17	36949	36302	647	cds	Failed to assign function
NODE_17	37524	37141	383	cds	IncF plasmid conjugative transfer mating signal transduction protein TraM
NODE_17	37938	38447	509	cds	X polypeptide
NODE_17	39969	39673	296	cds	FIG00642953: hypothetical protein
NODE_17	41118	40903	215	cds	Failed to assign function
NODE_17	41755	41237	518	cds	FIG00642735: hypothetical protein

NODE_17	43094	42960	134	cds	FIG00638906: hypothetical protein
NODE_17	44671	43694	977	cds	FIG00638906: hypothetical protein
NODE_17	45541	44681	860	cds	FIG00638906: hypothetical protein
NODE_18	1	420	419	cds	RepA1
NODE_18	1334	1603	269	cds	Prevent host death protein, Phd antitoxin
NODE_18	1600	1881	281	cds	YacB
NODE_18	1966	2313	347	cds	FIG01046320: hypothetical protein
NODE_18	3527	4528	1001	cds	FIG01046273: hypothetical protein
NODE_18	5048	4704	344	cds	Failed to assign function
NODE_18	5801	5547	254	cds	hypothetical protein
NODE_18	6772	5846	926	cds	FIG01048484: hypothetical protein
NODE_18	8672	9025	353	cds	FIG00643174: hypothetical protein
NODE_18	9890	9075	815	cds	Colicin-M
NODE_18	10134	10661	527	cds	Failed to assign function
NODE_18	12214	10679	1535	cds	Failed to assign function
NODE_18	12655	13167	512	cds	Failed to assign function
NODE_18	13668	13318	350	cds	stable plasmid inheritance protein B
NODE_18	14633	13671	962	cds	Putative stability/partitioning protein encoded within prophage CP-933T
NODE_19	793	236	557	cds	Failed to assign function
NODE_2	647	3	644	cds	Rhs core protein with extension
NODE_2	440025	439411	614	cds	FIG00638423: hypothetical protein
NODE_2	467929	468102	173	cds	Phage terminase small subunit
NODE_2	468511	469020	509	cds	Terminase small subunit
NODE_2	478670	479128	458	cds	Phage minor tail protein
NODE_2	524984	523908	1076	cds	Phage integrase
NODE_2	525231	524950	281	cds	Failed to assign function
NODE_2	525526	525338	188	cds	Phage protein
NODE_2	526579	525770	809	cds	Recombinational DNA repair protein RecT (prophage associated)
NODE_2	529253	526572	2681	cds	Exodeoxyribonuclease VIII (EC 3.1.11.-)
NODE_2	529630	529361	269	cds	Phage protein
NODE_2	529875	529705	170	cds	Putative bacteriophage protein
NODE_2	530096	529875	221	cds	Kil protein
NODE_2	531382	530963	419	cds	putative regulator
NODE_2	531479	531721	242	cds	Failed to assign function

NODE_2	531718	532140	422	cds	FIG00639855: hypothetical protein
NODE_2	532218	533006	788	cds	Replication protein 15
NODE_2	536050	536862	812	cds	Failed to assign function
NODE_2	537508	537903	395	cds	Phage protein
NODE_2	538619	538831	212	cds	FIG01047714: hypothetical protein
NODE_2	538903	539502	599	cds	Failed to assign function
NODE_2	539502	539792	290	cds	putative bacteriophage protein
NODE_20	11426	11031	395	cds	Failed to assign function
NODE_21	5734	4856	878	cds	Failed to assign function
NODE_21	6584	6802	218	cds	FIG00638841: hypothetical protein
NODE_24	2641	2736	95	cds	FIG00642676: hypothetical protein
NODE_25	2	481	479	cds	Phage minor capsid protein - DNA pilot protein
NODE_25	545	2086	1541	cds	Phage DNA replication protein
NODE_25	2083	2343	260	cds	Phage single stranded DNA synthesis
NODE_25	2340	2540	200	cds	Failed to assign function
NODE_26	770	1042	272	cds	FIG00638906: hypothetical protein
NODE_27	2182	2319	137	cds	FIG00642676: hypothetical protein
NODE_29	188	3	185	cds	FIG00641663: hypothetical protein
NODE_29	2133	2026	107	cds	Failed to assign function
NODE_3	95350	95448	98	cds	hypothetical protein
NODE_31	70	2	68	cds	Phage minor capsid protein - DNA pilot protein
NODE_32	886	1194	308	cds	Failed to assign function
NODE_4	228417	228596	179	cds	Failed to assign function
NODE_5	97435	97163	272	cds	Failed to assign function
NODE_5	303463	304152	689	cds	FIG00638104: hypothetical protein
NODE_5	305677	306969	1292	cds	Failed to assign function
NODE_5	307063	308919	1856	cds	Phage tail length tape-measure protein 1
NODE_5	310782	311000	218	cds	Phage tail assembly protein I
NODE_5	318167	317853	314	cds	FIG00638841: hypothetical protein
NODE_5	431672	428655	3017	cds	core protein
NODE_6	2732	3301	569	cds	Phage tail fiber protein
NODE_6	155568	155897	329	cds	Outer membrane protein C precursor
NODE_6	155851	156660	809	cds	Outer membrane protein C precursor
NODE_6	185208	184939	269	cds	FIG00639870: hypothetical protein

NODE_6	185374	185153	221	cds	Failed to assign function
NODE_6	275890	274823	1067	cds	Molybdate metabolism regulator
NODE_6	278616	276022	2594	cds	Molybdate metabolism regulator
NODE_6	452528	452031	497	cds	Failed to assign function
NODE_8	277059	274597	2462	cds	Exodeoxyribonuclease VIII (EC 3.1.11.-)
NODE_8	279628	277337	2291	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-933P
NODE_8	280662	280871	209	cds	FIG00642398: hypothetical protein
NODE_8	281510	280872	638	cds	FIG00641599: hypothetical protein
NODE_8	295907	296329	422	cds	Putative single stranded DNA-binding protein of prophage
NODE_8	296796	296990	194	cds	hypothetical protein
NODE_8	296987	297583	596	cds	Prophage Clp protease-like protein
NODE_8	306587	306742	155	cds	Phage terminase, large subunit
NODE_8	309424	309627	203	cds	FIG00638989: hypothetical protein
NODE_8	309665	310903	1238	cds	Mobile element protein
NODE_8	310917	311120	203	cds	Failed to assign function
NODE_8	311180	311761	581	cds	COG3617: Prophage antirepressor
NODE_8	312061	311882	179	cds	conserved hypothetical protein
NODE_8	312258	312461	203	cds	FIG00639870: hypothetical protein
NODE_8	324403	324519	116	cds	Phage minor tail protein
NODE_9	231559	231002	557	cds	FIG00642676: hypothetical protein

Unique genes present in FHI43 and FHI62

Contig Id	Start	Stop	Length	Type	Function myRAST
NODE_0	303788	303898	110	cds	hypothetical protein
NODE_0	310470	310991	521	cds	unknown protein encoded by prophage CP-933O
NODE_0	312397	312753	356	cds	Methyltransferase (EC 2.1.1.-)
NODE_0	313906	314085	179	cds	hypothetical protein
NODE_0	314242	314060	182	cds	hypothetical protein
NODE_0	314420	315733	1313	cds	Failed to assign function
NODE_0	316527	316195	332	cds	Protein flxA
NODE_0	317060	317299	239	cds	Failed to assign function
NODE_0	317299	317586	287	cds	Failed to assign function
NODE_0	319691	320443	752	cds	Phage antitermination protein Q
NODE_0	321077	320865	212	cds	Cold shock protein CspF

NODE_0	322314	322195	119	cds	Failed to assign function
NODE_0	322567	322878	311	cds	FIG00639051: hypothetical protein
NODE_0	323405	323902	497	cds	FIG00639844: hypothetical protein
NODE_0	325886	325476	410	cds	putative envelope protein
NODE_0	326566	327111	545	cds	Terminase small subunit
NODE_0	329008	329214	206	cds	Failed to assign function
NODE_0	330793	332112	1319	cds	Phage capsid and scaffold
NODE_0	332122	332454	332	cds	Head decoration protein
NODE_0	333577	333972	395	cds	Phage DNA-packaging protein
NODE_0	346089	349115	3026	cds	Phage tail fiber protein
NODE_0	350375	349788	587	cds	Mobile element protein
NODE_0	350925	350692	233	cds	FIG00640097: hypothetical protein
NODE_12	6913	5657	1256	cds	IncF plasmid conjugative transfer DNA-nicking and unwinding protein TraI
NODE_12	85692	85354	338	cds	YihA
NODE_2	391624	389618	2006	cds	Mobile element protein
NODE_3	38833	39177	344	cds	FIG00638104: hypothetical protein
NODE_3	44451	47693	3242	cds	Phage tail length tape-measure protein 1
NODE_3	244177	243956	221	cds	FIG00639870: hypothetical protein
NODE_3	244392	244150	242	cds	FIG00639870: hypothetical protein
NODE_3	337635	333841	3794	cds	Molybdate metabolism regulator
NODE_4	141635	145828	4193	cds	Type II secretory pathway, ATPase PulE/Tfp pilus assembly pathway, ATPase PilB
NODE_4	145825	146358	533	cds	Failed to assign function
NODE_4	146850	147281	431	cds	hypothetical protein
NODE_4	147760	148110	350	cds	Failed to assign function
NODE_4	148110	148868	758	cds	Rhs core protein with extension
NODE_5	431084	430260	824	cds	FIG00642676: hypothetical protein
NODE_5	435927	436415	488	cds	hypothetical protein
NODE_6	299548	299844	296	cds	putative endopeptidase
NODE_6	301584	302000	416	cds	Terminase small subunit
NODE_7	215844	216257	413	cds	Phage minor tail protein

Table S14. Analysis of genes differently present in strains FHI58 and FHI63 from HUS-patients and St. Olav104 from the same outbreak

Unique genes present in FHI58 and FHI63

Contig Id	Start	Stop	Length	Type	Function myRAST
NODE_3	214960	213311	1649	cds	Enterotoxin
NODE_23	7558	6578	980	cds	Modification methylase EcoRI (EC 2.1.1.72)
NODE_19	2	151	149	cds	Mobile element protein
NODE_1	539595	538936	659	cds	Flagellar basal-body P-ring formation protein FlgA
NODE_6	11015	10101	914	cds	FIG00638504: hypothetical protein
NODE_14	15341	16780	1439	cds	Microcin H47 secretion protein
NODE_1	443228	445021	1793	cds	Uptake hydrogenase large subunit (EC 1.12.99.6)
NODE_2	427057	425816	1241	cds	putative phage inhibition, colicin resistance and tellurite resistance protein
NODE_6	44006	43032	974	cds	Flagellar biosynthesis protein FlhA
NODE_8	25775	23343	2432	cds	Outer membrane protein assembly factor YaeT precursor
NODE_4	320843	321364	521	cds	hypothetical protein
NODE_12	70802	69903	899	cds	LysR family transcriptional regulator YcjZ
NODE_3	344219	345094	875	cds	Type III secretion host injection and negative regulator protein (YopD)
NODE_0	477386	477207	179	cds	hypothetical protein
NODE_23	4615	5052	437	cds	entry exclusion protein 1
NODE_0	122124	121963	161	cds	FIG00638240: hypothetical protein
NODE_14	1148	363	785	cds	Lipid A biosynthesis (KDO) 2-(lauroyl)-lipid IVA acyltransferase (EC 2.3.1.-)
NODE_9	76651	75227	1424	cds	4-aminobutyraldehyde dehydrogenase (EC 1.2.1.19)
NODE_4	423130	423032	98	cds	hypothetical protein
NODE_0	398253	399944	1691	cds	PTS system, fructose-specific IIB component (EC 2.7.1.69) / PTS system, fructose-specific IIC component (EC 2.7.1.69)
NODE_0	578171	579436	1265	cds	N-Acetylneuraminate cytidylyltransferase (EC 2.7.7.43)
NODE_2	239077	238814	263	cds	conserved hypothetical protein
NODE_7	89652	89867	215	cds	Hypothetical protein yfgJ
NODE_23	4356	2788	1568	cds	hypothetical protein
NODE_2	66907	67833	926	cds	Flagellar motor rotation protein MotB
NODE_0	291384	292142	758	cds	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase (EC 4.2.99.20)
NODE_5	317029	318522	1493	cds	General secretion pathway protein E
NODE_23	9029	8796	233	cds	mobilization protein mbeD
NODE_12	130710	128035	2675	cds	Aconitate hydratase (EC 4.2.1.3)
NODE_4	70450	69692	758	cds	Phosphoglycolate phosphatase (EC 3.1.3.18)
NODE_23	1066	1257	191	cds	RNAI modulator protein Rom
NODE_12	88660	86393	2267	cds	Maltose phosphorylase (EC 2.4.1.8) / Trehalose phosphorylase (EC 2.4.1.64)
NODE_23	5133	5561	428	cds	entry exclusion protein 2
NODE_4	418766	417447	1319	cds	Hexose phosphate uptake regulatory protein UhpC
NODE_23	1034	678	356	cds	mobilization protein MobC
NODE_2	240295	239180	1115	cds	FIG00638086: hypothetical protein
NODE_2	18083	19777	1694	cds	FIG00638997: hypothetical protein
NODE_0	580756	581982	1226	cds	Failed to assign function
NODE_0	687934	686909	1025	cds	Mobile element protein
NODE_9	79412	78399	1013	cds	Putrescine transport ATP-binding protein PotA (TC 3.A.1.11.1)
NODE_3	243737	243006	731	cds	Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.-)
NODE_1	363474	364166	692	cds	Putative inner membrane protein
NODE_0	522608	522342	266	cds	Phage EaA protein

NODE_14	13217	15337	2120	cds	Methionine ABC transporter ATP-binding protein
NODE_1	301630	302763	1133	cds	Putrescine transport ATP-binding protein PotG (TC 3.A.1.11.2)
NODE_0	582018	583178	1160	cds	Failed to assign function
NODE_6	44101	44736	635	cds	Flagellar motor rotation protein MotB
NODE_1	21630	21382	248	cds	Putative inner membrane protein
NODE_12	30641	30276	365	cds	Failed to assign function
NODE_23	8421	7588	833	cds	Type II restriction enzyme EcoRI
NODE_2	149929	151839	1910	cds	DinG family ATP-dependent helicase YoaA
NODE_10	93305	91839	1466	cds	Outer membrane component of tripartite multidrug resistance system
NODE_10	155746	156399	653	cds	Type III secretion inner membrane protein (YscR,SpaR,HrcR,EscR,homologous to flagellar export components)
NODE_2	487107	488390	1283	cds	Putative transport protein

Unique genes present in St. Olav104

Contig Id	Start	Stop	Length	Type	Function myRAST
scf7180000000168_CDS_10616-10927_5313	10616	10927	311	cds	Phage tail length tape-measure protein 1
scf7180000000169_CDS_101477-102553_4531	101477	102553	1076	cds	Outer membrane component of tripartite multidrug resistance system
scf7180000000169_CDS_102580-102942_4039	102580	102942	362	cds	Outer membrane component of tripartite multidrug resistance system
scf7180000000169_CDS_1-213_4071	1	213	212	cds	Efa1-Lymphostatin-like protein
scf7180000000169_CDS_1448-1720_1991	1448	1720	272	cds	Efa1-Lymphostatin-like protein
scf7180000000169_CDS_1806-2012_5434	1806	2012	206	cds	Efa1-Lymphostatin-like protein
scf7180000000169_CDS_2100-1993_3258	2100	1993	107	cds	Failed to assign function
scf7180000000169_CDS_2240-6970_4050	2240	6970	4730	cds	Efa1-Lymphostatin-like protein
scf7180000000169_CDS_278-529_4655	278	529	251	cds	Efa1-Lymphostatin-like protein
scf7180000000169_CDS_39036-38629_4753	39036	38629	407	cds	Type III secretion inner membrane protein (YscR,SpaR,HrcR,EscR,homologous to flagellar export components)
scf7180000000169_CDS_859-1134_4687	859	1134	275	cds	Efa1-Lymphostatin-like protein
scf7180000000169_CDS_9718-9987_4140	9718	9987	269	cds	Mobile element protein
scf7180000000170_CDS_160297-161247_2757	160297	161247	950	cds	Putrescine transport ATP-binding protein PotG (TC 3.A.1.11.2)
scf7180000000170_CDS_161202-161429_5502	161202	161429	227	cds	Putrescine transport ATP-binding protein PotG (TC 3.A.1.11.2)
scf7180000000170_CDS_222140-222556_5545	222140	222556	416	cds	Putative inner membrane protein
scf7180000000171_CDS_1009-452_1156	1009	452	557	cds	Phage terminase, large subunit
scf7180000000171_CDS_238-2_5455	238	2	236	cds	Failed to assign function
scf7180000000171_CDS_28908-28660_5188	28908	28660	248	cds	Phage tail fiber protein
scf7180000000171_CDS_4177-3833_610	4177	3833	344	cds	FIG00638659: hypothetical protein
scf7180000000171_CDS_455-270_4266	455	270	185	cds	Failed to assign function
scf7180000000172_CDS_115-2_4704	115	2	113	cds	Failed to assign function
scf7180000000172_CDS_2405-819_5384	2405	819	1586	cds	unknown protein encoded within prophage CP-933V
scf7180000000172_CDS_2764-2402_4318	2764	2402	362	cds	Mobile element protein
scf7180000000172_CDS_38169-37396_5093	38169	37396	773	cds	LysR family transcriptional regulator YcjZ
scf7180000000172_CDS_54233-53760_3529	54233	53760	473	cds	Maltose phosphorylase (EC 2.4.1.8) / Trehalose phosphorylase (EC 2.4.1.64)
scf7180000000173_CDS_4274-3594_4414	4274	3594	680	cds	Phage tail assembly protein I
scf7180000000173_CDS_4844-4329_5362	4844	4329	515	cds	Phage tail assembly protein
scf7180000000173_CDS_6315-5950_4447	6315	5950	365	cds	Phage tail length tape-measure protein 1
scf7180000000173_CDS_7542-7309_760	7542	7309	233	cds	Phage tail length tape-measure protein 1
scf7180000000174_CDS_187994-187209_4006	187994	187209	785	cds	Flagellar motor rotation protein MotB
scf7180000000174_CDS_188047-188913_5450	188047	188913	866	cds	Flagellar biosynthesis protein FlhA
scf7180000000174_CDS_221229-221843_5380	221229	221843	614	cds	FIG00638504: hypothetical protein

scf7180000000174_CDS_53279-53380_4453	53279	53380	101	cds	Failed to assign function
scf7180000000175_CDS_128695-128399_245	128695	128399	296	cds	FIG00638240: hypothetical protein
scf7180000000175_CDS_164669-164445_4214	164669	164445	224	cds	FIG00638841: hypothetical protein
scf7180000000175_CDS_165367-165666_288	165367	165666	299	cds	Failed to assign function
scf7180000000175_CDS_165666-165929_3505	165666	165929	263	cds	Failed to assign function
scf7180000000175_CDS_165926-166978_5491	165926	166978	1052	cds	Failed to assign function
scf7180000000175_CDS_167053-167301_5172	167053	167301	248	cds	Failed to assign function
scf7180000000175_CDS_167643-167882_5551	167643	167882	239	cds	Failed to assign function
scf7180000000175_CDS_167971-168555_4391	167971	168555	584	cds	Failed to assign function
scf7180000000175_CDS_168781-169080_3086	168781	169080	299	cds	Failed to assign function
scf7180000000175_CDS_169077-169196_3730	169077	169196	119	cds	Phage capsid and scaffold
scf7180000000176_CDS_138199-138483_4853	138199	138483	284	cds	Failed to assign function
scf7180000000176_CDS_138566-138766_5509	138566	138766	200	cds	Failed to assign function
scf7180000000177_CDS_1226-1567_5554	1226	1567	341	cds	Phage tail fiber protein
scf7180000000177_CDS_1687-2226_5394	1687	2226	539	cds	Phage tail fiber protein
scf7180000000177_CDS_2-154_4376	2	154	152	cds	Failed to assign function
scf7180000000177_CDS_2184-3314_1660	2184	3314	1130	cds	Phage tail fiber protein
scf7180000000177_CDS_3580-3861_4113	3580	3861	281	cds	Phage tail fiber protein
scf7180000000177_CDS_4592-5521_4226	4592	5521	929	cds	Phage tail fiber protein
scf7180000000177_CDS_502-1140_998	502	1140	638	cds	Phage tail fiber protein
scf7180000000177_CDS_8058-7687_5464	8058	7687	371	cds	FIG00643151: hypothetical protein
scf7180000000177_CDS_8892-8572_1653	8892	8572	320	cds	Failed to assign function
scf7180000000178_CDS_1439-2422_1578	1439	2422	983	cds	Phage tail fiber protein
scf7180000000178_CDS_220-2_636	220	2	218	cds	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
scf7180000000178_CDS_2544-3818_4975	2544	3818	1274	cds	Phage tail fiber protein
scf7180000000178_CDS_41727-43703_519	41727	43703	1976	cds	Aconitate hydratase (EC 4.2.1.3)
scf7180000000178_CDS_523-672_4460	523	672	149	cds	Phage tail fiber protein
scf7180000000178_CDS_6511-6846_2828	6511	6846	335	cds	FIG00638227: hypothetical protein
scf7180000000178_CDS_669-1232_5034	669	1232	563	cds	Phage tail fiber protein
scf7180000000179_CDS_1057-2_5392	1057	2	1055	cds	FIG00638997: hypothetical protein
scf7180000000179_CDS_41719-42402_5557	41719	42402	683	cds	Mobile element protein
scf7180000000179_CDS_81819-81448_4797	81819	81448	371	cds	Failed to assign function
scf7180000000179_CDS_82334-82185_2009	82334	82185	149	cds	Failed to assign function
scf7180000000179_CDS_83743-83456_4422	83743	83456	287	cds	Phage DNA transfer protein
scf7180000000179_CDS_84188-83748_2649	84188	83748	440	cds	Phage DNA transfer protein
scf7180000000180_CDS_553-53_5531	553	53	500	cds	Phenylacetaldehyde dehydrogenase (EC 1.2.1.39)
scf7180000000181_CDS_10295-10417_5544	10295	10417	122	cds	conserved hypothetical protein
scf7180000000181_CDS_123-1_4946	123	1	122	cds	Phage portal protein
scf7180000000181_CDS_1278-607_3248	1278	607	671	cds	Phage portal protein
scf7180000000181_CDS_17877-16888_5400	17877	16888	989	cds	Mobile element protein
scf7180000000181_CDS_2607-1792_2121	2607	1792	815	cds	Phage terminase, large subunit
scf7180000000181_CDS_559-335_5506	559	335	224	cds	Phage portal protein
scf7180000000182_CDS_2-1723_384	2	1723	1721	cds	Formate dehydrogenase N alpha subunit (EC 1.2.1.2) @ selenocysteine-containing
scf7180000000182_CDS_85752-84952_4025	85752	84952	800	cds	Putative transport protein
scf7180000000183_CDS_630-1_3644	630	1	629	cds	Flagellar motor rotation protein MotB
scf7180000000184_CDS_298-2_2566	298	2	296	cds	Formate dehydrogenase N alpha subunit (EC 1.2.1.2) @ selenocysteine-containing

scf7180000000184_CDS_35559-34729_1255	35559	34729	830	cds	4-aminobutyraldehyde dehydrogenase (EC 1.2.1.19)
scf7180000000185_CDS_1220-3_5437	1220	3	1217	cds	Maltose phosphorylase (EC 2.4.1.8) / Trehalose phosphorylase (EC 2.4.1.64)
scf7180000000185_CDS_40741-40595_529	40741	40595	146	cds	Failed to assign function
scf7180000000187_CDS_1558-2148_4698	1558	2148	590	cds	4-aminobutyraldehyde dehydrogenase (EC 1.2.1.19)
scf7180000000187_CDS_2-175_4768	2	175	173	cds	Spermidine Putrescine ABC transporter permease component PotB (TC 3.A.1.11.1)
scf7180000000187_CDS_658-753_5347	658	753	95	cds	Spermidine Putrescine ABC transporter permease component PotB (TC 3.A.1.11.1)
scf7180000000187_CDS_743-883_4088	743	883	140	cds	Spermidine Putrescine ABC transporter permease component potC (TC_3.A.1.11.1)
scf7180000000187_CDS_901-1536_3863	901	1536	635	cds	Spermidine Putrescine ABC transporter permease component potC (TC_3.A.1.11.1)
scf7180000000188_CDS_1256-1453_3835	1256	1453	197	cds	Uncharacterized protein ynbE; probable lipoprotein STY1424
scf7180000000188_CDS_2-136_4254	2	136	134	cds	Putative uncharacterized protein ydbH
scf7180000000188_CDS_349-621_5296	349	621	272	cds	Putative uncharacterized protein ydbH
scf7180000000188_CDS_5161-4658_5385	5161	4658	503	cds	Monoamine oxidase (1.4.3.4)
scf7180000000188_CDS_979-1272_3590	979	1272	293	cds	Putative uncharacterized protein ydbH
scf7180000000189_CDS_340-2_923	340	2	338	cds	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14)
scf7180000000189_CDS_946-812_5515	946	812	134	cds	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14)
scf7180000000190_CDS_1296-2369_4727	1296	2369	1073	cds	Formate dehydrogenase N alpha subunit (EC 1.2.1.2) @ selenocysteine-containing
scf7180000000190_CDS_2363-2668_4910	2363	2668	305	cds	Failed to assign function
scf7180000000190_CDS_375-1_1291	375	1	374	cds	Permease of the drug/metabolite transporter (DMT) superfamily
scf7180000000191_CDS_122-3_3001	122	3	119	cds	Threonyl-tRNA synthetase (EC 6.1.1.3)
scf7180000000191_CDS_1920-2231_2832	1920	2231	311	cds	Failed to assign function
scf7180000000191_CDS_704-1162_4944	704	1162	458	cds	Failed to assign function
scf7180000000192_CDS_152-310_2900	152	310	158	cds	Attachment invasion locus protein precursor
scf7180000000192_CDS_307-750_5275	307	750	443	cds	Attachment invasion locus protein precursor
scf7180000000193_CDS_2549-2034_3475	2549	2034	515	cds	Putrescine transport ATP-binding protein PotA (TC 3.A.1.11.1)
scf7180000000193_CDS_286-2_3429	286	2	284	cds	4-aminobutyraldehyde dehydrogenase (EC 1.2.1.19)
scf7180000000194_CDS_1128-1_1500	1128	1	1127	cds	Outer membrane protein assembly factor YaeT precursor
scf7180000000194_CDS_229293-228928_3106	229293	228928	365	cds	Phage tail fiber assembly protein
scf7180000000194_CDS_230111-229293_4894	230111	229293	818	cds	Phage tail fiber protein
scf7180000000194_CDS_230789-230211_4112	230789	230211	578	cds	Phage tail assembly protein
scf7180000000194_CDS_236369-236001_4498	236369	236001	368	cds	Phage minor tail protein
scf7180000000194_CDS_236611-236366_3727	236611	236366	245	cds	Failed to assign function
scf7180000000194_CDS_237301-237128_5359	237301	237128	173	cds	Phage capsid and scaffold
scf7180000000195_CDS_1449-961_4123	1449	961	488	cds	Phage tail length tape-measure protein 1
scf7180000000195_CDS_1963-1532_2182	1963	1532	431	cds	Phage tail length tape-measure protein 1
scf7180000000195_CDS_2475-2014_4707	2475	2014	461	cds	Phage tail length tape-measure protein 1
scf7180000000195_CDS_2950-2486_5428	2950	2486	464	cds	Phage tail length tape-measure protein 1
scf7180000000195_CDS_3275-3003_3998	3275	3003	272	cds	Phage tail length tape-measure protein 1
scf7180000000195_CDS_367-197_2699	367	197	170	cds	Phage minor tail protein
scf7180000000195_CDS_3700-3323_5451	3700	3323	377	cds	Failed to assign function
scf7180000000195_CDS_6139-5669_2201	6139	5669	470	cds	Phage capsid and scaffold
scf7180000000195_CDS_762-433_5297	762	433	329	cds	Phage minor tail protein
scf7180000000196_CDS_1466-957_5470	1466	957	509	cds	Phage tail length tape-measure protein 1
scf7180000000196_CDS_2059-1850_5417	2059	1850	209	cds	Phage tail length tape-measure protein 1
scf7180000000196_CDS_3520-3191_5500	3520	3191	329	cds	Failed to assign function
scf7180000000196_CDS_3967-3668_5508	3967	3668	299	cds	Phage tail assembly
scf7180000000196_CDS_4130-3975_4916	4130	3975	155	cds	Phage minor tail protein

scf7180000000196_CDS_4942-4697_5517	4942	4697	245	cds	Phage tail fiber protein
scf7180000000196_CDS_5306-4923_2781	5306	4923	383	cds	Phage capsid and scaffold
scf7180000000196_CDS_6407-6240_5277	6407	6240	167	cds	Failed to assign function
scf7180000000196_CDS_65-3_5005	65	3	62	cds	Phage tail length tape-measure protein 1
scf7180000000196_CDS_7426-7139_299	7426	7139	287	cds	Failed to assign function
scf7180000000196_CDS_883-65_3480	883	65	818	cds	Phage tail length tape-measure protein 1
scf7180000000197_CDS_136967-137125_3674	136967	137125	158	cds	Mobile element protein
scf7180000000197_CDS_296187-296357_2119	296187	296357	170	cds	Mobile element protein
scf7180000000197_CDS_296501-297205_5489	296501	297205	704	cds	Enterotoxin
scf7180000000197_CDS_297264-298148_3932	297264	298148	884	cds	Enterotoxin
scf7180000000197_CDS_342134-341166_2639	342134	341166	968	cds	General secretion pathway protein E
scf7180000000197_CDS_342658-342134_6	342658	342134	524	cds	General secretion pathway protein E
scf7180000000197_CDS_465736-465407_5243	465736	465407	329	cds	putative phage inhibition, colicin resistance and tellurite resistance protein
scf7180000000197_CDS_466647-465736_5320	466647	465736	911	cds	putative phage inhibition, colicin resistance and tellurite resistance protein
scf7180000000197_CDS_62-3_5311	62	3	59	cds	Phage minor tail protein
scf7180000000197_CDS_705368-705207_4587	705368	705207	161	cds	Failed to assign function
scf7180000000197_CDS_723583-723900_5481	723583	723900	317	cds	hypothetical protein
scf7180000000197_CDS_725338-725171_5525	725338	725171	167	cds	Failed to assign function
scf7180000000198_CDS_11182-11454_3088	11182	11454	272	cds	Failed to assign function
scf7180000000198_CDS_1-270_5276	1	270	269	cds	Glutamate Aspartate transport system permease protein GltK (TC 3.A.1.3.4)
scf7180000000198_CDS_1343-1747_1541	1343	1747	404	cds	ABC-type polar amino acid transport system, ATPase component
scf7180000000198_CDS_2054-2371_5481	2054	2371	317	cds	hypothetical protein
scf7180000000198_CDS_267-980_345	267	980	713	cds	Glutamate Aspartate transport system permease protein GltK (TC 3.A.1.3.4)
scf7180000000198_CDS_8527-8324_2632	8527	8324	203	cds	Failed to assign function
scf7180000000198_CDS_9082-8579_4492	9082	8579	503	cds	Shikimate 5-dehydrogenase I alpha (EC 1.1.1.25)
scf7180000000198_CDS_9396-9124_5513	9396	9124	272	cds	Shikimate 5-dehydrogenase I alpha (EC 1.1.1.25)
scf7180000000198_CDS_988-1350_5549	988	1350	362	cds	Failed to assign function
scf7180000000198_CDS_9971-9576_4217	9971	9576	395	cds	Failed to assign function
scf7180000000199_CDS_322206-322628_5401	322206	322628	422	cds	hypothetical protein
scf7180000000199_CDS_419167-418808_5228	419167	418808	359	cds	Hexose phosphate uptake regulatory protein UhpC
scf7180000000199_CDS_420126-419164_3913	420126	419164	962	cds	Hexose phosphate uptake regulatory protein UhpC
scf7180000000199_CDS_519414-518329_5240	519414	518329	1085	cds	Permeases of the major facilitator superfamily
scf7180000000199_CDS_519755-519507_678	519755	519507	248	cds	Permeases of the major facilitator superfamily
scf7180000000199_CDS_71615-70986_4734	71615	70986	629	cds	Phosphoglycolate phosphatase (EC 3.1.3.18)
scf7180000000200_CDS_13083-13352_3965	13083	13352	269	cds	Translation elongation factor Tu
scf7180000000200_CDS_13357-13779_1790	13357	13779	422	cds	Translation elongation factor Tu
scf7180000000200_CDS_13863-14267_90	13863	14267	404	cds	Translation elongation factor Tu
scf7180000000200_CDS_14495-14641_2102	14495	14641	146	cds	Preprotein translocase subunit SecE (TC 3.A.5.1.1)
scf7180000000200_CDS_14614-14880_806	14614	14880	266	cds	Preprotein translocase subunit SecE (TC 3.A.5.1.1)
scf7180000000200_CDS_14882-15151_2244	14882	15151	269	cds	Transcription antitermination protein NusG
scf7180000000200_CDS_15585-15755_4822	15585	15755	170	cds	LSU ribosomal protein L11p (L12e)
scf7180000000200_CDS_16014-16238_3318	16014	16238	224	cds	LSU ribosomal protein L1p (L10Ae)
scf7180000000200_CDS_16301-16600_1113	16301	16600	299	cds	LSU ribosomal protein L1p (L10Ae)
scf7180000000200_CDS_17381-17623_5271	17381	17623	242	cds	LSU ribosomal protein L10p (P0)
scf7180000000200_CDS_1-78_313	1	78	77	cds	Ribokinase (EC 2.7.1.15)
scf7180000000200_CDS_18374-18478_5516	18374	18478	104	cds	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)

scf7180000000200_CDS_314-610_4665	314	610	296	cds	Ribose operon repressor
scf7180000000200_CDS_735-1298_5530	735	1298	563	cds	Ribose operon repressor
scf7180000000200_CDS_75-296_5473	75	296	221	cds	Ribokinase (EC 2.7.1.15)
scf7180000000201_CDS_1089-1352_5478	1089	1352	263	cds	UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.1.1.158)
scf7180000000201_CDS_1343-1906_5291	1343	1906	563	cds	UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.1.1.158)
scf7180000000201_CDS_1939-2166_838	1939	2166	227	cds	Failed to assign function
scf7180000000201_CDS_2388-2885_3458	2388	2885	497	cds	Failed to assign function
scf7180000000201_CDS_907-1092_5265	907	1092	185	cds	UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.1.1.158)
scf7180000000202_CDS_1045-893_5358	1045	893	152	cds	orf; Unknown function
scf7180000000202_CDS_137003-137467_3977	137003	137467	464	cds	Type III secretion host injection and negative regulator protein (YopD)
scf7180000000202_CDS_137530-137877_5432	137530	137877	347	cds	Type III secretion host injection and negative regulator protein (YopD)
scf7180000000202_CDS_15463-16347_1866	15463	16347	884	cds	Ornithine decarboxylase (EC 4.1.1.17)
scf7180000000202_CDS_1564-1127_4534	1564	1127	437	cds	orf; Unknown function
scf7180000000202_CDS_16344-17597_3337	16344	17597	1253	cds	Ornithine decarboxylase (EC 4.1.1.17)
scf7180000000202_CDS_2025-1594_3571	2025	1594	431	cds	FIG00638856: hypothetical protein
scf7180000000202_CDS_2600-2019_5556	2600	2019	581	cds	FIG00638856: hypothetical protein
scf7180000000202_CDS_3566-3207_11	3566	3207	359	cds	Enterotoxin
scf7180000000202_CDS_36399-35791_4470	36399	35791	608	cds	Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.-)
scf7180000000202_CDS_373482-373153_5328	373482	373153	329	cds	hypothetical protein
scf7180000000202_CDS_377862-378266_1874	377862	378266	404	cds	Phage outer membrane lipoprotein Rz1
scf7180000000202_CDS_378915-378757_4242	378915	378757	158	cds	Failed to assign function
scf7180000000202_CDS_378955-379158_5420	378955	379158	203	cds	putative Dnase
scf7180000000202_CDS_462-1_5452	462	1	461	cds	Mobile element protein
scf7180000000202_CDS_4855-3569_5181	4855	3569	1286	cds	Enterotoxin
scf7180000000202_CDS_5169-4999_2119	5169	4999	170	cds	Mobile element protein
scf7180000000203_CDS_1-282_3209	1	282	281	cds	Mobile element protein
scf7180000000203_CDS_279-506_4817	279	506	227	cds	Mobile element protein
scf7180000000203_CDS_5793-6098_4767	5793	6098	305	cds	Failed to assign function
scf7180000000203_CDS_6098-6844_4972	6098	6844	746	cds	Failed to assign function
scf7180000000203_CDS_649-1188_4677	649	1188	539	cds	Mobile element protein
scf7180000000203_CDS_6920-7357_4143	6920	7357	437	cds	Failed to assign function
scf7180000000204_CDS_10041-10256_5459	10041	10256	215	cds	Failed to assign function
scf7180000000204_CDS_10312-10485_4251	10312	10485	173	cds	Failed to assign function
scf7180000000204_CDS_12341-13045_5467	12341	13045	704	cds	Phage tail length tape-measure protein 1
scf7180000000204_CDS_13132-13317_4693	13132	13317	185	cds	Phage tail length tape-measure protein 1
scf7180000000204_CDS_13278-13889_4932	13278	13889	611	cds	Failed to assign function
scf7180000000204_CDS_13886-14143_3504	13886	14143	257	cds	Phage tail length tape-measure protein 1
scf7180000000204_CDS_14283-14465_4955	14283	14465	182	cds	Failed to assign function
scf7180000000204_CDS_14653-14949_1234	14653	14949	296	cds	Failed to assign function
scf7180000000204_CDS_14889-15251_5543	14889	15251	362	cds	Phage tail length tape-measure protein 1
scf7180000000204_CDS_15894-16019_5445	15894	16019	125	cds	Phage minor tail protein
scf7180000000204_CDS_2-316_5023	2	316	314	cds	Failed to assign function
scf7180000000205_CDS_1199-1384_5540	1199	1384	185	cds	Failed to assign function
scf7180000000205_CDS_1415-1687_4931	1415	1687	272	cds	Failed to assign function
scf7180000000205_CDS_1633-1851_4815	1633	1851	218	cds	Phage portal protein
scf7180000000205_CDS_2295-2990_1539	2295	2990	695	cds	Phage portal protein

scf7180000000205_CDS_2987-3925_5389	2987	3925	938	cds	Phage portal protein
scf7180000000205_CDS_3885-4208_1673	3885	4208	323	cds	Phage portal protein
scf7180000000205_CDS_5065-5334_4602	5065	5334	269	cds	Failed to assign function
scf7180000000205_CDS_5331-5453_2608	5331	5453	122	cds	FIG00638104: hypothetical protein
scf7180000000205_CDS_5423-5674_3503	5423	5674	251	cds	FIG00638104: hypothetical protein
scf7180000000205_CDS_5740-6216_5553	5740	6216	476	cds	Phage neck whiskers
scf7180000000205_CDS_6213-6455_5138	6213	6455	242	cds	Phage neck whiskers
scf7180000000205_CDS_69-239_3121	69	239	170	cds	Failed to assign function
scf7180000000205_CDS_7188-9317_1154	7188	9317	2129	cds	Phage tail length tape-measure protein 1
scf7180000000205_CDS_9865-10239_2371	9865	10239	374	cds	Failed to assign function
scf7180000000206_CDS_1033-1491_5077	1033	1491	458	cds	Phage portal protein
scf7180000000206_CDS_11005-12474_3779	11005	12474	1469	cds	Phage tail fiber protein
scf7180000000206_CDS_12677-12967_500	12677	12967	290	cds	Phage tail fiber protein
scf7180000000206_CDS_12919-13125_3881	12919	13125	206	cds	Phage tail fiber protein
scf7180000000206_CDS_13302-13748_5485	13302	13748	446	cds	Phage tail fiber protein
scf7180000000206_CDS_13700-14479_14	13700	14479	779	cds	Phage tail fiber protein
scf7180000000206_CDS_14537-14692_5044	14537	14692	155	cds	Attachment invasion locus protein precursor
scf7180000000206_CDS_14646-15062_3171	14646	15062	416	cds	Attachment invasion locus protein precursor
scf7180000000206_CDS_15568-16134_4598	15568	16134	566	cds	Phage tail fiber protein
scf7180000000206_CDS_1571-1864_5546	1571	1864	293	cds	FIG00638841: hypothetical protein
scf7180000000206_CDS_16227-16514_5377	16227	16514	287	cds	Phage tail fiber protein
scf7180000000206_CDS_18480-18025_4878	18480	18025	455	cds	FIG00643151: hypothetical protein
scf7180000000206_CDS_2-88_3256	2	88	86	cds	Phage portal protein
scf7180000000206_CDS_3107-3346_4278	3107	3346	239	cds	Failed to assign function
scf7180000000206_CDS_3325-3822_1278	3325	3822	497	cds	Phage neck whiskers
scf7180000000206_CDS_443-625_5468	443	625	182	cds	Phage portal protein
scf7180000000206_CDS_622-846_4906	622	846	224	cds	Phage portal protein
scf7180000000206_CDS_788-1036_5453	788	1036	248	cds	Phage portal protein
scf7180000000207_CDS_1337-1480_3678	1337	1480	143	cds	Failed to assign function
scf7180000000207_CDS_14279-14037_5173	14279	14037	242	cds	DNA repair protein RecN
scf7180000000207_CDS_14650-14291_987	14650	14291	359	cds	DNA repair protein RecN
scf7180000000207_CDS_14951-14622_3211	14951	14622	329	cds	DNA repair protein RecN
scf7180000000207_CDS_15422-15213_3428	15422	15213	209	cds	DNA repair protein RecN
scf7180000000207_CDS_16284-15508_3440	16284	15508	776	cds	NAD kinase (EC 2.7.1.23)
scf7180000000207_CDS_2-337_5299	2	337	335	cds	Phage minor tail protein
scf7180000000207_CDS_2346-2167_5472	2346	2167	179	cds	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
scf7180000000207_CDS_2478-2669_1028	2478	2669	191	cds	Phage tail fiber protein
scf7180000000207_CDS_3013-3441_5050	3013	3441	428	cds	Phage tail fiber protein
scf7180000000207_CDS_3438-3764_5284	3438	3764	326	cds	Phage tail fiber protein
scf7180000000207_CDS_3999-4625_2637	3999	4625	626	cds	Phage tail fiber protein
scf7180000000207_CDS_6672-7856_453	6672	7856	1184	cds	Phage tail fiber protein
scf7180000000207_CDS_7955-8149_5043	7955	8149	194	cds	Phage tail fiber protein
scf7180000000207_CDS_975-1361_2441	975	1361	386	cds	Phage tail assembly protein I
scf7180000000208_CDS_27030-26548_5537	27030	26548	482	cds	Failed to assign function
scf7180000000211_CDS_2-721_4774	2	721	719	cds	Outer membrane protein assembly factor YaeT precursor
scf7180000000211_CDS_28474-27980_5448	28474	27980	494	cds	hypothetical protein

scf7180000000212_CDS_230457-230567_5201	230457	230567	110	cds	YfdE protein
scf7180000000212_CDS_94917-94687_3620	94917	94687	230	cds	Failed to assign function
scf7180000000213_CDS_5778-5410_4515	5778	5410	368	cds	Failed to assign function
scf7180000000213_CDS_6032-5868_5233	6032	5868	164	cds	Failed to assign function
scf7180000000213_CDS_6474-6662_5426	6474	6662	188	cds	Phage repressor
scf7180000000213_CDS_6991-7167_4907	6991	7167	176	cds	Failed to assign function
scf7180000000213_CDS_7388-7660_1115	7388	7660	272	cds	Origin specific replication initiation factor
scf7180000000214_CDS_19368-19616_1038	19368	19616	248	cds	Phage head-to-tail joining protein
scf7180000000214_CDS_20557-20844_2750	20557	20844	287	cds	Failed to assign function
scf7180000000214_CDS_20799-21170_5121	20799	21170	371	cds	Failed to assign function
scf7180000000214_CDS_21151-21348_4566	21151	21348	197	cds	Phage capsid and scaffold
scf7180000000214_CDS_21388-21645_5089	21388	21645	257	cds	Phage capsid and scaffold
scf7180000000214_CDS_21775-21999_5402	21775	21999	224	cds	Phage capsid and scaffold
scf7180000000214_CDS_21983-22396_5406	21983	22396	413	cds	Phage capsid and scaffold
scf7180000000214_CDS_22812-22267_1557	22812	22267	545	cds	Failed to assign function
scf7180000000214_CDS_22848-23009_5140	22848	23009	161	cds	Phage major capsid protein
scf7180000000214_CDS_304-2_4529	304	2	302	cds	Recombinational DNA repair protein RecT (prophage associated)
scf7180000000214_CDS_585-247_1468	585	247	338	cds	Recombinational DNA repair protein RecT (prophage associated)
scf7180000000215_CDS_318-1_4069	318	1	317	cds	Failed to assign function
scf7180000000215_CDS_4785-5969_5533	4785	5969	1184	cds	Phage tail length tape-measure protein 1
scf7180000000215_CDS_6101-7105_5194	6101	7105	1004	cds	Phage tail length tape-measure protein 1
scf7180000000215_CDS_7131-7664_403	7131	7664	533	cds	Phage minor tail protein
scf7180000000215_CDS_7664-7867_5522	7664	7867	203	cds	Phage minor tail protein
scf7180000000215_CDS_8133-8354_5164	8133	8354	221	cds	Phage minor tail protein
scf7180000000215_CDS_8359-8970_1270	8359	8970	611	cds	Phage tail assembly protein
scf7180000000215_CDS_9034-9357_2732	9034	9357	323	cds	Phage tail assembly protein I
scf7180000000216_CDS_1-117_5447	1	117	116	cds	Phage DNA-packaging protein
scf7180000000216_CDS_1288-1443_4916	1288	1443	155	cds	Phage minor tail protein
scf7180000000216_CDS_169799-170179_5431	169799	170179	380	cds	Failed to assign function
scf7180000000216_CDS_2184-2381_5376	2184	2381	197	cds	Phage minor tail protein
scf7180000000216_CDS_2649-3014_5343	2649	3014	365	cds	Failed to assign function
scf7180000000216_CDS_3007-3138_5493	3007	3138	131	cds	Phage tail length tape-measure protein 1
scf7180000000216_CDS_3146-4753_4814	3146	4753	1607	cds	Phage tail length tape-measure protein 1
scf7180000000216_CDS_4837-5565_4268	4837	5565	728	cds	Phage tail length tape-measure protein 1
scf7180000000216_CDS_48910-48710_5046	48910	48710	200	cds	Putative inner membrane protein
scf7180000000216_CDS_51000-51251_3537	51000	51251	251	cds	HokE protein
scf7180000000217_CDS_105358-104792_5084	105358	104792	566	cds	FIG00639870: hypothetical protein
scf7180000000217_CDS_1112-150_2446	1112	150	962	cds	Failed to assign function
scf7180000000217_CDS_142-2_1966	142	2	140	cds	Failed to assign function
scf7180000000217_CDS_182771-182283_4158	182771	182283	488	cds	PTS system, fructose-specific IIB component (EC 2.7.1.69) / PTS system, fructose-specific IIC component (EC 2.7.1.69)
scf7180000000217_CDS_183973-182771_5193	183973	182771	1202	cds	PTS system, fructose-specific IIB component (EC 2.7.1.69) / PTS system, fructose-specific IIC component (EC 2.7.1.69)
scf7180000000217_CDS_290478-289996_3202	290478	289996	482	cds	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase (EC 4.2.99.20)
scf7180000000217_CDS_290753-290475_4594	290753	290475	278	cds	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase (EC 4.2.99.20)
scf7180000000217_CDS_3959-3201_5458	3959	3201	758	cds	N-Acetylneuraminate cytidylyltransferase (EC 2.7.7.43)
scf7180000000217_CDS_397567-397286_1097	397567	397286	281	cds	FIG00638886: hypothetical protein
scf7180000000217_CDS_397812-397555_4832	397812	397555	257	cds	FIG00638886: hypothetical protein

scf7180000000217_CDS_398726-399400_1516	398726	399400	674	cds	Phage protein
scf7180000000217_CDS_59574-59885_4716	59574	59885	311	cds	Phage EaA protein
scf7180000000218_CDS_10448-10122_5340	10448	10122	326	cds	Phage tail assembly protein
scf7180000000218_CDS_10917-10459_2917	10917	10459	458	cds	Phage minor tail protein
scf7180000000218_CDS_11156-10920_173	11156	10920	236	cds	Phage minor tail protein
scf7180000000218_CDS_11536-11156_5336	11536	11156	380	cds	Phage minor tail protein
scf7180000000218_CDS_12047-11481_4533	12047	11481	566	cds	Phage tail length tape-measure protein 1
scf7180000000218_CDS_12975-12502_1439	12975	12502	473	cds	Phage tail length tape-measure protein 1
scf7180000000218_CDS_422-1174_4362	422	1174	752	cds	Failed to assign function
scf7180000000219_CDS_11150-10995_4916	11150	10995	155	cds	Phage minor tail protein
scf7180000000219_CDS_1796-1470_5529	1796	1470	326	cds	Phage tail fiber protein
scf7180000000219_CDS_4163-2082_5349	4163	2082	2081	cds	Phage tail fiber protein
scf7180000000219_CDS_552-310_4387	552	310	242	cds	Attachment invasion locus protein precursor
scf7180000000219_CDS_709-617_3699	709	617	92	cds	Attachment invasion locus protein precursor
scf7180000000220_CDS_1716-772_4962	1716	772	944	cds	Phage tail length tape-measure protein 1
scf7180000000220_CDS_20785-19589_86	20785	19589	1196	cds	Failed to assign function
scf7180000000220_CDS_21189-20782_4408	21189	20782	407	cds	Failed to assign function
scf7180000000220_CDS_2171-1830_1560	2171	1830	341	cds	Phage tail length tape-measure protein 1
scf7180000000220_CDS_23499-23272_2340	23499	23272	227	cds	Terminase small subunit
scf7180000000221_CDS_132656-133093_3217	132656	133093	437	cds	FIG00639870: hypothetical protein
scf7180000000221_CDS_178327-178827_4183	178327	178827	500	cds	Phage antitermination protein Q
scf7180000000221_CDS_552-4_5435	552	4	548	cds	Failed to assign function
scf7180000000222_CDS_1-222_5055	1	222	221	cds	Phage antitermination protein Q
scf7180000000222_CDS_1860-2375_4399	1860	2375	515	cds	FIG00639830: hypothetical protein
scf7180000000222_CDS_2852-4051_5514	2852	4051	1199	cds	Mobile element protein
scf7180000000222_CDS_4206-4430_5227	4206	4430	224	cds	Failed to assign function
scf7180000000223_CDS_5857-6228_4056	5857	6228	371	cds	Failed to assign function
scf7180000000224_CDS_1441-1638_5492	1441	1638	197	cds	Failed to assign function
scf7180000000224_CDS_7587-7715_1738	7587	7715	128	cds	Failed to assign function
scf7180000000224_CDS_7888-8226_4643	7888	8226	338	cds	Failed to assign function
scf7180000000224_CDS_8213-8455_5465	8213	8455	242	cds	Failed to assign function
scf7180000000224_CDS_8598-8786_5058	8598	8786	188	cds	Failed to assign function
scf7180000000224_CDS_8786-9058_2750	8786	9058	272	cds	Failed to assign function
scf7180000000224_CDS_900-1157_5369	900	1157	257	cds	Phage protein
scf7180000000224_CDS_9140-9478_5418	9140	9478	338	cds	Phage capsid and scaffold
scf7180000000224_CDS_9502-10236_5412	9502	10236	734	cds	Phage capsid and scaffold
scf7180000000225_CDS_138-1_5294	138	1	137	cds	Failed to assign function
scf7180000000225_CDS_170083-169874_1037	170083	169874	209	cds	Uptake hydrogenase large subunit (EC 1.12.99.6)
scf7180000000225_CDS_171665-170097_5476	171665	170097	1568	cds	Uptake hydrogenase large subunit (EC 1.12.99.6)
scf7180000000225_CDS_174227-175120_5386	174227	175120	893	cds	FIG00638423: hypothetical protein
scf7180000000225_CDS_176237-177211_2961	176237	177211	974	cds	Mobile element protein
scf7180000000225_CDS_75489-75959_5353	75489	75959	470	cds	Flagellar basal-body P-ring formation protein FlgA
scf7180000000226_CDS_10751-9591_3749	10751	9591	1160	cds	Methionine ABC transporter ATP-binding protein
scf7180000000226_CDS_11284-10892_401	11284	10892	392	cds	Methionine ABC transporter ATP-binding protein
scf7180000000226_CDS_1171-683_58	1171	683	488	cds	Uptake hydrogenase small subunit precursor (EC 1.12.99.6)
scf7180000000226_CDS_12368-11505_5433	12368	11505	863	cds	RTX toxins determinant A and related Ca2+-binding proteins

scf7180000000226_CDS_1683-1165_3255	1683	1165	518	cds	Uptake hydrogenase small subunit precursor (EC 1.12.99.6)
scf7180000000226_CDS_2409-2212_2003	2409	2212	197	cds	Failed to assign function
scf7180000000226_CDS_3841-4128_4841	3841	4128	287	cds	FIG00638423: hypothetical protein
scf7180000000226_CDS_5245-6378_4438	5245	6378	1133	cds	Mobile element protein
scf7180000000226_CDS_585-1_5503	585	1	584	cds	Uptake hydrogenase large subunit (EC 1.12.99.6)
scf7180000000226_CDS_6353-6784_1896	6353	6784	431	cds	Mobile element protein
scf7180000000226_CDS_686-582_2326	686	582	104	cds	Uptake hydrogenase large subunit (EC 1.12.99.6)
scf7180000000226_CDS_8161-7724_3919	8161	7724	437	cds	Microcin H47 secretion protein
scf7180000000226_CDS_9162-8158_4506	9162	8158	1004	cds	Microcin H47 secretion protein
scf7180000000226_CDS_9573-9166_1357	9573	9166	407	cds	Methionine ABC transporter ATP-binding protein
scf7180000000227_CDS_109121-109681_1668	109121	109681	560	cds	Mobile element protein
scf7180000000227_CDS_109675-110511_844	109675	110511	836	cds	Mobile element protein
scf7180000000227_CDS_1740-1648_4034	1740	1648	92	cds	Microcin H47 secretion protein
scf7180000000227_CDS_20722-21303_5466	20722	21303	581	cds	Mobile element protein
scf7180000000227_CDS_2324-2001_3605	2324	2001	323	cds	Microcin H47 secretion protein
scf7180000000227_CDS_2-601_5222	2	601	599	cds	Mobile element protein
scf7180000000227_CDS_3277-3089_5501	3277	3089	188	cds	Methionine ABC transporter ATP-binding protein
scf7180000000227_CDS_3756-3361_4823	3756	3361	395	cds	Methionine ABC transporter ATP-binding protein
scf7180000000227_CDS_5207-3753_3876	5207	3753	1454	cds	Methionine ABC transporter ATP-binding protein
scf7180000000228_CDS_1517-1867_1330	1517	1867	350	cds	IncF plasmid conjugative transfer fertility inhibition protein FinO
scf7180000000228_CDS_202-516_585	202	516	314	cds	IncF plasmid conjugative transfer pilin acetylase TraX
scf7180000000229_CDS_11825-11151_4268	11825	11151	674	cds	Phage tail length tape-measure protein 1
scf7180000000229_CDS_12585-12394_5449	12585	12394	191	cds	Phage tail length tape-measure protein 1
scf7180000000229_CDS_12806-12525_5425	12806	12525	281	cds	Phage tail length tape-measure protein 1
scf7180000000229_CDS_2544-2260_4243	2544	2260	284	cds	Phage tail fiber protein
scf7180000000229_CDS_534-767_5496	534	767	233	cds	Mobile element protein
scf7180000000229_CDS_6706-5798_5101	6706	5798	908	cds	Phage tail fiber protein
scf7180000000229_CDS_7953-6661_4695	7953	6661	1292	cds	Phage tail fiber protein
scf7180000000229_CDS_8087-8491_2308	8087	8491	404	cds	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
scf7180000000229_CDS_817-1113_4448	817	1113	296	cds	Mobile element protein
scf7180000000229_CDS_9381-8932_5054	9381	8932	449	cds	Phage tail assembly protein I
scf7180000000230_CDS_112-2_1968	112	2	110	cds	Phage tail fiber protein
scf7180000000230_CDS_1376-1065_5511	1376	1065	311	cds	Phage tail fiber protein
scf7180000000230_CDS_1986-1642_5547	1986	1642	344	cds	Phage tail fiber protein
scf7180000000230_CDS_2541-2326_5439	2541	2326	215	cds	Phage tail fiber protein
scf7180000000230_CDS_2739-2975_636	2739	2975	236	cds	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
scf7180000000230_CDS_3526-4041_5415	3526	4041	515	cds	Failed to assign function
scf7180000000230_CDS_438-184_5483	438	184	254	cds	Phage tail fiber protein
scf7180000000230_CDS_5026-4763_3978	5026	4763	263	cds	Phage minor tail protein
scf7180000000230_CDS_956-762_3479	956	762	194	cds	Phage tail fiber protein
scf7180000000231_CDS_11824-11600_5441	11824	11600	224	cds	Head decoration protein
scf7180000000231_CDS_11946-11824_3565	11946	11824	122	cds	Head decoration protein
scf7180000000231_CDS_12234-11983_1629	12234	11983	251	cds	Head-tail preconnector protein GP5
scf7180000000231_CDS_12583-12242_5364	12583	12242	341	cds	Head-tail preconnector protein GP5
scf7180000000231_CDS_13675-13472_5048	13675	13472	203	cds	Failed to assign function
scf7180000000231_CDS_1407-1078_5370	1407	1078	329	cds	Phage tail fiber protein

scf7180000000231_CDS_14608-14153_673	14608	14153	455	cds	Failed to assign function
scf7180000000231_CDS_15057-14548_1706	15057	14548	509	cds	Failed to assign function
scf7180000000231_CDS_15468-15244_2413	15468	15244	224	cds	Phage terminase, large subunit
scf7180000000231_CDS_15762-15583_4106	15762	15583	179	cds	Phage terminase, large subunit
scf7180000000231_CDS_1969-2172_5333	1969	2172	203	cds	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
scf7180000000231_CDS_2757-2149_4854	2757	2149	608	cds	Phage tail assembly protein I
scf7180000000231_CDS_315-1_5526	315	1	314	cds	Phage tail fiber protein
scf7180000000231_CDS_3551-3297_798	3551	3297	254	cds	Phage tail assembly protein
scf7180000000231_CDS_4052-3573_1570	4052	3573	479	cds	Phage minor tail protein
scf7180000000231_CDS_4252-4049_3765	4252	4049	203	cds	Phage minor tail protein
scf7180000000231_CDS_987-616_4502	987	616	371	cds	Phage tail fiber protein
scf7180000000232_CDS_16712-16852_830	16712	16852	140	cds	hypothetical protein
scf7180000000232_CDS_29827-29994_5457	29827	29994	167	cds	Kil protein
scf7180000000232_CDS_30171-30566_2081	30171	30566	395	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-933P
scf7180000000232_CDS_30521-30808_3148	30521	30808	287	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-933P
scf7180000000232_CDS_31588-31124_5454	31588	31124	464	cds	Failed to assign function
scf7180000000232_CDS_457-158_5463	457	158	299	cds	Putative tail component of prophage CP-933K
scf7180000000232_CDS_9726-9950_4214	9726	9950	224	cds	FIG00638841: hypothetical protein
scf7180000000233_CDS_1013-1252_4545	1013	1252	239	cds	plasmid stabilization system
scf7180000000233_CDS_1503-1943_5061	1503	1943	440	cds	FIG00644832: hypothetical protein
scf7180000000233_CDS_347-799_4890	347	799	452	cds	putative regulator; Regulation (Phage or Prophage Related)
scf7180000000233_CDS_4335-4460_5155	4335	4460	125	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-933P
scf7180000000233_CDS_4571-4933_287	4571	4933	362	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-933P
scf7180000000233_CDS_5211-5867_2138	5211	5867	656	cds	Exodeoxyribonuclease encoded by cryptic prophage CP-933P
scf7180000000234_CDS_3-392_2476	3	392	389	cds	FIG00644832: hypothetical protein
scf7180000000234_CDS_734-928_5006	734	928	194	cds	hypothetical protein
scf7180000000234_CDS_80673-80401_1850	80673	80401	272	cds	Putative inner membrane protein
scf7180000000235_CDS_158875-158585_5307	158875	158585	290	cds	Phage tail length tape-measure protein 1
scf7180000000235_CDS_3-233_5427	3	233	230	cds	conserved hypothetical protein
scf7180000000235_CDS_87539-87207_5335	87539	87207	332	cds	DinG family ATP-dependent helicase YoaA
scf7180000000235_CDS_88729-87512_3434	88729	87512	1217	cds	DinG family ATP-dependent helicase YoaA
scf7180000000235_CDS_89115-88735_5235	89115	88735	380	cds	DinG family ATP-dependent helicase YoaA
scf7180000000236_CDS_3-176_2184	3	176	173	cds	Flagellar motor rotation protein MotB
scf7180000000236_CDS_46535-48658_5249	46535	48658	2123	cds	Phage terminase, large subunit
scf7180000000236_CDS_54071-54463_3613	54071	54463	392	cds	Putative tail component of prophage CP-933K
scf7180000000236_CDS_54418-54822_2599	54418	54822	404	cds	Putative tail component of prophage CP-933K
scf7180000000236_CDS_55634-55750_4463	55634	55750	116	cds	Phage tail length tape-measure protein 1
scf7180000000236_CDS_55750-55872_5129	55750	55872	122	cds	Phage tail length tape-measure protein 1
scf7180000000237_CDS_10238-11212_733	10238	11212	974	cds	Phage tail fiber protein
scf7180000000237_CDS_11428-11823_5541	11428	11823	395	cds	Phage tail fiber protein
scf7180000000237_CDS_187-543_5411	187	543	356	cds	Prophage Clp protease-like protein
scf7180000000237_CDS_2099-2320_1443	2099	2320	221	cds	FIG00638599: hypothetical protein
scf7180000000237_CDS_4638-4946_5510	4638	4946	308	cds	Phage minor tail protein
scf7180000000237_CDS_681-932_5246	681	932	251	cds	Prophage Clp protease-like protein
scf7180000000237_CDS_972-1817_989	972	1817	845	cds	Prophage Clp protease-like protein
scf7180000000238_CDS_173-1861_4712	173	1861	1688	cds	Phage tail length tape-measure protein 1

scf7180000000238_CDS_3-83_3904	3	83	80	cds	Phage tail length tape-measure protein 1
scf7180000000238_CDS_4082-4213_4877	4082	4213	131	cds	Failed to assign function
scf7180000000238_CDS_4454-4735_5102	4454	4735	281	cds	Phage tail fiber protein
scf7180000000238_CDS_4803-5066_1385	4803	5066	263	cds	Phage tail fiber protein
scf7180000000239_CDS_11241-13700_5488	11241	13700	2459	cds	Phage tail fiber protein
scf7180000000239_CDS_1350-1613_5317	1350	1613	263	cds	Failed to assign function
scf7180000000239_CDS_13772-13993_4312	13772	13993	221	cds	Phage tail fiber protein
scf7180000000239_CDS_1623-1826_4359	1623	1826	203	cds	Failed to assign function
scf7180000000239_CDS_1842-1964_4570	1842	1964	122	cds	FIG00638104: hypothetical protein
scf7180000000239_CDS_2247-2402_4674	2247	2402	155	cds	Failed to assign function
scf7180000000239_CDS_2585-2959_5270	2585	2959	374	cds	Phage neck whiskers
scf7180000000239_CDS_2-718_5210	2	718	716	cds	Phage portal protein
scf7180000000239_CDS_3038-3304_5520	3038	3304	266	cds	Failed to assign function
scf7180000000239_CDS_3280-3639_3323	3280	3639	359	cds	Failed to assign function
scf7180000000239_CDS_3686-5236_5018	3686	5236	1550	cds	Phage tail length tape-measure protein 1
scf7180000000239_CDS_5233-6927_5548	5233	6927	1694	cds	Phage tail length tape-measure protein 1
scf7180000000239_CDS_8340-8492_2657	8340	8492	152	cds	Failed to assign function
scf7180000000240_CDS_10460-10882_1919	10460	10882	422	cds	Phage portal protein
scf7180000000240_CDS_10891-11283_4262	10891	11283	392	cds	Phage portal protein
scf7180000000240_CDS_188-445_5369	188	445	257	cds	Phage protein
scf7180000000240_CDS_4725-5024_288	4725	5024	299	cds	Failed to assign function
scf7180000000240_CDS_5024-5287_3505	5024	5287	263	cds	Failed to assign function
scf7180000000240_CDS_5284-6045_4650	5284	6045	761	cds	Failed to assign function
scf7180000000240_CDS_675-1223_3308	675	1223	548	cds	Failed to assign function
scf7180000000240_CDS_6917-6618_5302	6917	6618	299	cds	Failed to assign function
scf7180000000240_CDS_6982-7551_513	6982	7551	569	cds	Failed to assign function
scf7180000000240_CDS_7548-7817_5552	7548	7817	269	cds	Failed to assign function
scf7180000000240_CDS_8323-8979_4000	8323	8979	656	cds	Phage capsid and scaffold
scf7180000000240_CDS_9943-10455_3270	9943	10455	512	cds	Phage portal protein
scf7180000000241_CDS_10148-10291_2965	10148	10291	143	cds	FIG00638104: hypothetical protein
scf7180000000241_CDS_10291-10419_4731	10291	10419	128	cds	FIG00638104: hypothetical protein
scf7180000000241_CDS_10547-10696_4674	10547	10696	149	cds	Failed to assign function
scf7180000000241_CDS_11394-10939_642	11394	10939	455	cds	Failed to assign function
scf7180000000241_CDS_1320-1063_5528	1320	1063	257	cds	Failed to assign function
scf7180000000241_CDS_3-248_483	3	248	245	cds	putative endopeptidase
scf7180000000241_CDS_9031-9273_5100	9031	9273	242	cds	FIG00638841: hypothetical protein
scf7180000000241_CDS_9230-9355_3553	9230	9355	125	cds	FIG00638841: hypothetical protein
scf7180000000241_CDS_9366-9623_5422	9366	9623	257	cds	Phage capsid and scaffold
scf7180000000241_CDS_9707-10039_1105	9707	10039	332	cds	Failed to assign function
scf7180000000242_CDS_3-335_5103	3	335	332	cds	UDP-N-acetylglucosamine 4,6-dehydratase (EC 4.2.1.-)
scf7180000000242_CDS_60510-61730_1321	60510	61730	1220	cds	Phage tail fibers
scf7180000000243_CDS_1-939_5535	1	939	938	cds	Failed to assign function
scf7180000000243_CDS_5334-4615_4011	5334	4615	719	cds	hypothetical protein
scf7180000000243_CDS_5972-5340_5242	5972	5340	632	cds	Phage EaA protein
scf7180000000244_CDS_1563-1823_5356	1563	1823	260	cds	Phage terminase, large subunit
scf7180000000244_CDS_1777-1944_5277	1777	1944	167	cds	Failed to assign function

scf7180000000244_CDS_1941-2102_4554	1941	2102	161	cds	Failed to assign function
scf7180000000244_CDS_2-376_5383	2	376	374	cds	Failed to assign function
scf7180000000244_CDS_2479-2880_2016	2479	2880	401	cds	Failed to assign function
scf7180000000244_CDS_2945-3298_412	2945	3298	353	cds	Failed to assign function
scf7180000000244_CDS_3310-3498_1575	3310	3498	188	cds	Failed to assign function
scf7180000000244_CDS_3593-3766_5507	3593	3766	173	cds	Phage capsid and scaffold
scf7180000000244_CDS_3763-3987_5490	3763	3987	224	cds	Phage capsid and scaffold
scf7180000000244_CDS_3933-4160_5539	3933	4160	227	cds	Phage capsid and scaffold
scf7180000000244_CDS_660-893_5318	660	893	233	cds	Failed to assign function
scf7180000000244_CDS_890-1129_4788	890	1129	239	cds	Failed to assign function
scf7180000000245_CDS_1178-990_5499	1178	990	188	cds	Phage minor tail protein
scf7180000000245_CDS_1393-1190_5283	1393	1190	203	cds	Phage minor tail protein
scf7180000000245_CDS_217-2_354	217	2	215	cds	Phage tail length tape-measure protein 1
scf7180000000245_CDS_560-417_5132	560	417	143	cds	Phage tail length tape-measure protein 1

Table S15. Gene ontology enrichment analysis of the genes significantly overrepresented in the 23 HUS-associated STEC (Table S2)

GO-ID	Term	Category	FDR	P-Value
GO:0019521	D-gluconate metabolic process	P[1]	3,95E-07	1,74E-10
GO:0019523	L-idonate metabolic process	P	3,95E-07	2,32E-10
GO:0046183	L-idonate catabolic process	P	3,95E-07	2,32E-10
GO:0019520	aldonic acid metabolic process	P	5,91E-07	4,63E-10
GO:0046176	aldonic acid catabolic process	P	1,98E-05	1,95E-08
GO:0044275	cellular carbohydrate catabolic process	P	1,07E-03	1,25E-06
GO:0072329	monocarboxylic acid catabolic process	P	1,69E-02	2,31E-05
GO:0044262	cellular carbohydrate metabolic process	P	2,77E-02	4,34E-05
GO:0032787	monocarboxylic acid metabolic process	P	3,28E-02	5,78E-05

[1]Gene ontology category biological process.

Table S16. Gene ontology enrichment analysis of the genes significantly overrepresented in HUS-group 1 (Table S4)

GO-ID	Term	Category	FDR	P-Value
GO:0015628	protein secretion by the type II secretion system	P[1]	1,00E-02	3,96E-06
GO:0015627	type II protein secretion system complex	C[2]	1,00E-02	3,96E-06
GO:0051704	multi-organism process	P	1,21E-02	1,05E-05
GO:0019523	L-idonate metabolic process	P	1,21E-02	1,19E-05
GO:0046183	L-idonate catabolic process	P	1,21E-02	1,19E-05
GO:0003939	L-iditol 2-dehydrogenase activity	F[3]	3,96E-02	4,69E-05

[1]Gene ontology category biological process.

[2]Gene ontology category cellular component

[3]Gene ontology category molecular function.

Table S17. Gene ontology enrichment analysis of the genes significantly overrepresented in HUS-group 2 (Table S5)

GO-ID	Term	Category	FDR	P-Value
GO:0009288	bacterial-type flagellum	C[1]	9,73E-12	1,92E-15
GO:0042995	cell projection	C	1,84E-08	7,24E-12
GO:0044463	cell projection part	C	2,10E-08	1,65E-11
GO:0044461	bacterial-type flagellum part	C	2,10E-08	1,65E-11
GO:0001539	ciliary or bacterial-type flagellar motility	P[2]	8,07E-08	7,96E-11
GO:0051674	localization of cell	P	9,00E-08	1,24E-10
GO:0048870	cell motility	P	9,00E-08	1,24E-10
GO:0003774	motor activity	F[3]	1,84E-07	2,91E-10
GO:0006928	cellular component movement	P	4,41E-07	7,83E-10
GO:0009425	bacterial-type flagellum basal body	C	3,76E-05	7,42E-08
GO:0040011	locomotion	P	5,79E-05	1,26E-07
GO:0019290	siderophore biosynthetic process	P	3,73E-03	9,57E-06
GO:0009237	siderophore metabolic process	P	3,73E-03	9,57E-06
GO:0044550	secondary metabolite biosynthetic process	P	6,49E-03	1,79E-05
GO:0050486	intramolecular transferase activity, transferring hydroxy groups	F	8,57E-03	2,71E-05
GO:0008909	isochorismate synthase activity	F	8,57E-03	2,71E-05
GO:0019184	nonribosomal peptide biosynthetic process	P	1,11E-02	3,73E-05
GO:0043043	peptide biosynthetic process	P	1,67E-02	5,94E-05
GO:0030694	bacterial-type flagellum basal body, rod	C	1,73E-02	6,47E-05
GO:0044422	organelle part	C	3,14E-02	1,34E-04
GO:0043229	intracellular organelle	C	3,14E-02	1,42E-04
GO:0043226	organelle	C	3,14E-02	1,42E-04
GO:0043232	intracellular non-membrane-bounded organelle	C	3,42E-02	1,69E-04
GO:0043228	non-membrane-bounded organelle	C	3,42E-02	1,69E-04
GO:0016833	oxo-acid-lyase activity	F	3,49E-02	1,79E-04
GO:0019748	secondary metabolic process	P	3,73E-02	2,06E-04

[1]Gene ontology category cellular component.

[2]Gene ontology category biological process.

[3]Gene ontology category molecular function.