# Molecular epidemiology of verotoxigenic *Escherichia coli* O157:H7

Robert Söderlund
*Faculty of Veterinary Medicine and Animal Science*
*Department of Animal Breeding and Genetics*
*Uppsala*

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2015

Acta Universitatis agriculturae Sueciae
2015:110

Cover illustration by Anders Sundström

ISSN 1652-6880
ISBN (print version) 978-91-576-8418-9
ISBN (electronic version) 978- 91-576-8419-6
© 2015 Robert Söderlund, Uppsala
Print: SLU Service/Repro, Uppsala 2015

# Molecular epidemiology of verotoxigenic *Escherichia coli* O157:H7

## Abstract

Verotoxigenic *E. coli* (VTEC) O157:H7 is a zoonotic disease agent with asymptomatic ruminants such as cattle and sheep as reservoir hosts. Infection in humans can cause bloody diarrhoea with occasionally fatal complications such as haemolytic-uraemic syndrome (HUS), particularly in children. In this thesis, the genetic variation among representative sets of O157:H7 isolates from Swedish cattle, sheep and human clinical cases is studied to find markers to predict the relative pathogenicity of different genotypes, and to improve methods for tracing spread of the bacteria. The results show that the same types occur in cattle, sheep and human patients infected within the country. However, certain types are common among ruminants while extremely rare among patient isolates, while others are overrepresented among patients. Better molecular tracing tools together with interventions targeted against highly pathogenic types of O157:H7 will allow more efficient and cost-effective efforts to reduce the burden of disease caused by VTEC O157:H7.

*Keywords: Escherichia coli*, O157:H7, VTEC, STEC, EHEC, zoonosis, cattle, MLVA

*Author's address:* Robert Söderlund, SVA, Department of Bacteriology,
Ulls väg 2B 75189 Uppsala, Sweden
*E-mail:* robert.soderlund@sva.se

# Contents

**List of Publications** **7**

**Abstract**

**Abbreviations** **9**

**1 Introduction** **11**

1.1 Verotoxigenic *E. coli* (VTEC) 11
- 1.1.1 *Escherichia coli* 11
- 1.1.2 The genetic basis of virulence in certain *E. coli* 11
- 1.1.3 Pathotypes 13
- 1.1.4 Verotoxigenic and enterohaemorrhagic *E. coli* 13
- 1.1.5 The serotype O157:H7 14

1.2 VTEC virulence factors 14
- 1.2.1 Verotoxin 14
- 1.2.2 The locus of enterocyte attachment and effacement (LEE) 16
- 1.2.3 The O157:H7 large virulence plasmid (pO157) 18
- 1.2.4 Other virulence factors 18

1.3 Evolution and phylogenetics of VTEC O157:H7 19
- 1.3.1 The evolutionary history of O157:H7 19
- 1.3.2 Lineages within the O157:H7 serotype 20

1.4 Human VTEC O157:H7 infection 21
- 1.4.1 Symptoms 21
- 1.4.2 Haemolytic-uraemic syndrome (HUS) 21
- 1.4.3 Infection routes 22

1.5 Animal reservoirs of O157:H7 24
- 1.5.1 Cattle 24
- 1.5.2 Prevalence in Swedish cattle 25
- 1.5.3 Other ruminants 27
- 1.5.4 Non-ruminant animals 28

1.6 Genotypic characterization of O157:H7 28
- 1.6.1 Molecular epidemiology 28
- 1.6.2 Virulence gene detection and subtyping 29
- 1.6.3 Pulsed-field gel electrophoresis (PFGE) 30
- 1.6.4 Multilocus variable-number tandem repeat analysis (MLVA) 30
- 1.6.5 Phage typing 31
- 1.6.6 Next generation sequencing 31

**2 Aims of the thesis** **33**

**3 Considerations on materials and methods 35**
3.1 Bacterial isolates 35
3.1.1 Isolates from samples taken on cattle dairy farms 35
3.1.2 Isolates from samples taken at slaughterhouses 36
3.1.3 Isolates from samples taken on farms with links to cases of illness in humans 36
3.1.4 Isolates from human patients 37
3.2 Comparison of cattle, sheep and human isolates 37
3.3 Molecular characterization methods 37
3.3.1 Single nucleotide polymorphism typing by real-time PCR 38
3.3.2 Subtyping of verotoxin 2 genes 38
3.3.3 Next generation sequencing 39
3.4 Bioinformatics 39
3.4.1 Microarray analysis 39
3.4.2 MLVA and LSPA-6 analysis 39
3.4.3 Sequence analysis 40
3.4.4 Clustering and visualization 40

**4 Results and discussion 41**
4.1 Evaluation and comparison of molecular typing methods for O157:H7 41
4.1.1 MLVA 41
4.1.2 Phage typing and LSPA-6 42
4.1.3 Next generation sequencing SNP typing 44
4.2 Molecular markers for virulence among Swedish O157:H7 46
4.2.1 Verotoxin genes 46
4.2.2 Other possible virulence-associated traits 47
4.2.3 Clinical relevance of clade 8 47
4.3 Sheep as a reservoir for O157:H7 48
4.4 Spatio-temporal dynamics 51
4.4.1 Geographical distribution 51
4.4.2 Changes in the clonal composition of the O157:H7 population over time 52

**5 Future perspectives 55**

**6 Populärvetenskaplig sammanfattning 57**

**Acknowledgements 61**

**References 63**

# List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I **Söderlund R**, Aspán A, Ragione RM, Eriksson E, Boqvist S (2010). Microarray-based detection of virulence genes in verotoxigenic *Escherichia coli* O157:H7 strains from Swedish cattle. *Epidemiol Infect.* 139(7), 1088-96.

II Eriksson E, **Söderlund R**, Boqvist S, Aspán A. (2011). Genotypic characterization to identify markers associated with putative hypervirulence in Swedish *Escherichia coli* O157:H7 cattle strains. *J Appl Microbiol.* 110(1), 323-32.

III **Söderlund R**, Hedenström I, Nilsson A, Eriksson E, Aspán A (2012). Genetically similar strains of *Escherichia coli* O157:H7 isolated from sheep, cattle, and human patients in the same geographic regions. *BMC Vet Res.* 8:200.

IV **Söderlund R**, Jernberg C, Ivarsson S, Hedenström I, Eriksson E, Bongcam-Rudloff E, Aspán A (2014). Molecular typing of *E. coli* O157:H7 from Swedish cattle and human cases: population dynamics and virulence. *J Clin Microbiol.* 52(11), 3906-3912.

V Loftsdóttir H, **Söderlund R**, Jinnerot T, Eriksson E, Bongcam-Rudloff E, Aspán A. IS629-mediated inactivation of verotoxin 2 occurs in diverse patterns and is relatively common among cattle isolates of *E. coli* O157:H7 (manuscript).

The contributions of Robert Söderlund (RS) to the papers included in this thesis were as follows:

I RS performed the PCR and microarray assays, analysed the data and wrote the first draft of the manuscript.

II RS planned and performed MLVA, sequencing and real-time PCR assays, analysed the data and contributed to writing the final manuscript.

III RS planned and performed fragment analysis assays and sequencing, analysed the typing data together with Ingela Hedenström and Anna Aspán, and wrote the first manuscript draft.

IV RS conceived the study together with Cecilia Jernberg and Anna Aspán, performed the molecular assays with Ingela Hedenström, and wrote the first draft of the manuscript.

V RS and Heiður Lóftsdottir conceived the study together with Anna Aspán. RS analysed and interpreted the data together with Heiður Lóftsdottir and Tomas Jinnerot. RS contributed to the final manuscript.

# Abbreviations

| | |
|---|---|
| A/E lesions | Attaching and effacing lesions |
| EAEC (EAggEC) | Enteroaggregative *E. coli* |
| EHEC | Enterohaemorrhagic *E. coli* |
| FoHM | Public Health Agency of Sweden |
| HUS | Haemolytic uraemic syndrome |
| IS-element | Insertion sequence element |
| LEE | Locus of enterocyte effacement |
| LSPA-6 | Lineage-specific polymorphism assay |
| MLVA | Multi-locus variable number tandem repeat analysis |
| NGS | Next generation sequencing |
| PT | Phage type |
| PCR | Polymerase chain reaction |
| PFGE | Pulsed-Field Gel Electrophoresis |
| RAJ | Rectal-anal junction |
| RFLP | Restriction fragment length polymorphism |
| rRNA | Ribosomal RNA |
| SNP | Single nucleotide polymorphism |
| STEC | Shiga toxin-producing *E. coli* (a.k.a. VTEC) |
| SVA | National Veterinary Institute |
| Tir | Translocated intimin receptor |
| tRNA | Transfer RNA |
| T3SS | Type III secretion system |
| VNTR | Variable-number tandem repeat |
| VT | Verotoxin |
| VTEC | Verotoxin-producing *E. coli* |

# 1 Introduction

## 1.1 Verotoxigenic *E. coli* (VTEC)

### 1.1.1 *Escherichia coli*

*Escherichia coli* (*E. coli*) is a species of gram-negative facultatively aerobic gammaproteobacterium in the family *Enterobacteriaceae*. *E. coli* bacteria are shaped as cylindrical rods (see Fig 1), and usually motile. (Greenwood *et al.*, 2002) Commensal *E. coli* bacteria are often found among the gut flora of warm-blooded animals including humans (Eckburg *et al.*, 2005). *E. coli* are divided into serotypes based on the different surface antigens of the bacteria, which can be detected by agglutination assays or predicted by detecting the genes involved in the antigen synthesis (Madic *et al.*, 2010; Gannon *et al.*, 1997). The O-antigen is the outer polysaccharide portion of lipopolysaccharide (LPS) found in the cell wall of gram-negative bacteria. The H-antigen represents flagellar proteins. *E. coli* can also have K- (capsule) and F- (fimbrial) antigens. (Greenwood *et al.*, 2002)

### 1.1.2 The genetic basis of virulence in certain *E. coli*

A significant force in the evolution of *E. coli* is the acquisition and loss of mobile genetic elements, which can encode a variety of survival and virulence factors in addition to the genes involved in the relocation of the elements themselves. These elements can be transferred horizontally, i.e. they can move between different strains or even species of bacteria. The key chromosomal element types in *E. coli* are prophages and pathogenicity islands. Elements that have been integrated for many generations in the bacterial chromosome often gradually lose their mobility due to the accumulation of mutations and become a static part of the host genome. (Craig et al., 2002) *E. coli* virulence factors can also be found outside of the chromosome encoded on plasmids, which can also be transferred horizontally. Most human-pathogenic strains of *E. coli* are dependent one or more of these virulence plasmids to cause their disease phenotype. (Johnson & Nolan, 2009) The acquisition of mobile elements in a strain of *E. coli* is a continuous process as more elements are added and old elements degrade and are lost if they do not provide a selective advantage.

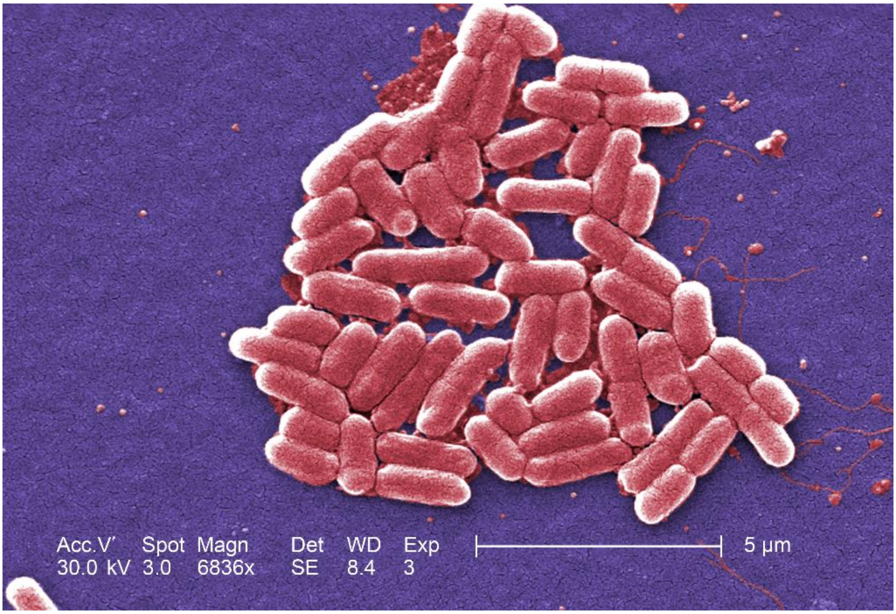

*Figure 1.* Colorized scanning electron micrograph (SEM) of *Escherichia coli* of the serotype O157:H7. Image source CDC/ National Escherichia, Shigella, Vibrio Reference Unit via the Public Health Image Library (PHIL)

Mobile genetic elements that do not encode any virulence or survival factors can still be a potent force in *E. coli* evolution. Insertion sequence (IS) elements are small mobile genetic elements that frequently occur in the genomes of bacteria. They only encode a transposase, but their insertion in the genome can cause gene inactivation, gene expression changes, deletions, or duplications (Ooka *et al.*, 2009; Chandler & Mahillon, 2002). Insertional inactivation of genes by IS-elements can be reversible (Kusumoto *et al.*, 2001).

### 1.1.3 Pathotypes

Certain combinations of virulence factors give *E. coli* the ability to cause a specific disease in healthy animals or humans. *E. coli* with these combinations are referred to as belonging to "pathotypes", characterized by the disease they cause (Kaper *et al.*, 2004), although these groupings are somewhat artificial and individual strains can have virulence characteristics of more than one pathotype (Karch *et al.*, 2012). There are six widely recognized *E. coli* pathotypes that cause diarrhoeal disease: enterohaemorrhagic (EHEC, which are the subject of this thesis), enteropathogenic (EPEC), enterotoxigenic (ETEC), enteroaggregative (EAEC/EAggEC), enteroinvasive (EIEC) and diffusely adherent *E. coli* (DAEC). In addition, extraintestinal infections are caused by ExPEC, including uropathogenic *E. coli* (UPEC) and meningitis-associated *E. coli* (MNEC). (Kaper *et al.*, 2004)

### 1.1.4 Verotoxigenic and enterohaemorrhagic *E. coli*

Enterohaemorrhagic *E. coli* (EHEC) is arguably the most severe pathotype, as it is capable of causing fatal complications in healthy infected humans. VTEC are characterized by the presence of genes encoding verotoxins. The subset of VTEC capable of causing haemorrhagic colitis in humans form the EHEC pathotype, but due to the difficulty of proving whether a strain is pathogenic, all VTEC are sometimes referred to as EHEC. In Sweden, any finding of verotoxin in a human patient is considered an EHEC case by the definitions of the National board of health and welfare (Anonymous, 2014a).

In addition to the verotoxin itself, most EHEC carry the locus of enterocyte attachment and effacement (LEE), have some form of large virulence plasmid, and an additional repertoire of adhesion factors, toxins and secreted effector proteins (described in section 1.2). VTEC appear to have evolved independently multiple times within the *E. coli* species, meaning that there is no phylogenetic lineage containing all VTEC and only the acquired DNA can be used as a marker for determination of the pathotype. (Ogura *et al.*, 2009) Most VTEC have asymptomatic ruminants as their natural hosts (section 1.5), and genetic factors associated with virulence in the human host frequently also play a role in colonization of cattle (Sheng *et al.*, 2006).

### 1.1.5 The serotype O157:H7

VTEC serotypes differ in their tendency to cause disease in humans, with some types rarely causing infections or associated with mild symptoms, while other types frequently cause severe infections with complications such as haemolytic-uraemic syndrome (see sections 1.4.1-1.4.2). Serotypes capable of causing severe disease in infected humans include O26:H11, O103:H2, O145:NM, O111:NM, O121:H19 and O104:H4 (BIOHAZ, 2013; Karmali *et al.*, 2003). The focus of this thesis is the serotype O157:H7, which is the most common source of EHEC outbreaks and cases in many countries (Karmali *et al.*, 2003), including Sweden (Anonymous, 2014a). Verotoxigenic *E. coli* of the serotype O157:H7 will generally be referred to simply as “O157:H7” in this thesis.

## 1.2 VTEC virulence factors

### 1.2.1 Verotoxin

The observation that certain strains of *E. coli* produced a cytotoxin that kills vero cells, a continuous cell line derived from the kidney of the African green monkey (*Chlorocebus* sp.), was first published in 1977 (Konowalchuk *et al.*, 1977). Due to its effect on this cell type, the toxin came to be called “verocytotoxin” or “verotoxin”. As other studies showed a high degree of similarity between verotoxin and the toxins produced by *Shigella*, the toxin is alternatively referred to as “shiga-like toxin” or simply “shiga toxin”. Depending on the toxin terminology used, the toxinogenic *E. coli* strains are synonymously called VTEC or STEC. The terms “verotoxin” and “VTEC” have consistently been used in this thesis.

#### *Verotoxin function*

Verotoxins belong to the AB5 family, consisting of a single A subunit encoded by the $vtx_2A$ gene, bound in a non-covalent association with five B subunits encoded by $vtx_2B$. The cytotoxic effect of verotoxin is based on ribosomal inactivation, a mechanism shared with several plant-derived toxins with similar structure, including ricin. The B subunits have binding sites for the trisaccharide side chain of glycosphingolipid globotriaosylceramide (Gb3). Gb3 is present on the outer surface of several mammalian cell types. Once the toxin has bound to

Gb3, it can enter the cell via endocytosis. The toxin makes its way to the cytoplasm via the Golgi apparatus and the endoplasmatic reticulum. After activation by host cell proteases, the toxin is able depurinate an adenine in the peptidyltransferase centre of the 28S rRNA, a critical component in the eukaryotic ribosome. This in turn results in blockage of elongation factor and tRNA binding, destroying the ribosome's protein synthesis ability. (Reviewed by (Walsh *et al.*, 2013)) Apart from the direct effects of the inhibition of protein synthesis, verotoxin also induces inflammatory pathways and apoptosis in affected cells (Tesh, 2010). The toxin is so potent that a single molecule can be sufficient to kill a cell (Tam & Lingwood, 2007). Cells with high Gb3 densities on their surfaces are more sensitive to verotoxin, while cells lacking this receptor are less sensitive. Most cells in the human intestinal epithelium express little or no Gb3. Human renal tubule cells, in contrast, express high Gb3 level, providing a likely explanation for the kidney damage associated with severe VTEC infection. (Tesh, 2010)

### *Verotoxin types*

There are two major types of verotoxin (VT), VT1 and VT2, encoded by $vtx_1$ and $vtx_2$ genes respectively. Within each of these types, there are several subtypes which have been assigned letters, (e.g. VT1a, encoded by $vtx_{1a}$). As mentioned in the previous section, each verotoxin also consists of A and B subunits, encoded in adjacent but separate genes. The VT2a subtype is frequently, especially in older literature, considered the default type and referred to simply as "VT2", with the other subtypes given letters. This somewhat unfortunate nomenclature was used in papers I-III, while the "a" nomenclature currently recommended (Scheutz *et al.*, 2012) is used in the later papers and the rest of this thesis.

### *Correlation between verotoxin types and severity of symptoms*

There is mounting evidence that the different verotoxin variants are associated with lower or higher risk of severe symptoms for humans infected by strains carrying them. Several studies on O157 and non-O157 VTEC have found HUS patients to significantly more often have been infected by strains carrying VT2a alone or in combination with VT2c than other variants or combinations of variants (Brandal *et al.*, 2015; Persson *et al.*, 2007; Friedrich *et al.*, 2002). However, detection of verotoxin genes alone is currently not considered sufficient to determine the pathogenic potential of a VTEC strain (BIOHAZ, 2013).

### *Verotoxin-encoding phages*

Most verotoxin genes found in VTEC are encoded on lambdoid prophages inserted in the bacterial chromosome (reviewed by (Herold *et al.*, 2004)). These phages allow verotoxin genes to spread by horizontal gene transfer, and experimental transduction has been successful for a number of donor/recipient combinations (Sekse *et al.*, 2008; Muniesa *et al.*, 2004; Schmidt *et al.*, 1999). Relatively recent verotoxin phage acquisition events may have led to the emergence of O157:H7 as a major human pathogen (Dallman *et al.*, 2015). Spontaneous loss of prophages and thereby the ability to produce verotoxin has also been documented (Mellmann *et al.*, 2008; Bielaszewska *et al.*, 2007). The phages play a critical role in the regulation of the expression of the *vtx* genes (Herold *et al.*, 2004). The expression of bacteriophage-encoded gene products such as verotoxin is suppressed by phage-encoded proteins under normal conditions, but induction of the lytic phase e.g. by host cell DNA damage, will trigger a regulatory cascade releasing this inhibition. The Q antiterminator protein is a key factor in this cascade. (Reviewed by (Waldor & Friedman, 2005; Schmidt, 2001)) Furthermore, phage lysogeny influences expression of LEE-encoded genes (see 1.2.2) (Xu *et al.*, 2012), suggesting possible secondary benefits for the bacteria when colonizing the gastrointestinal tract.

### 1.2.2 The locus of enterocyte attachment and effacement (LEE)

The locus of enterocyte attachment and effacement (LEE) is a pathogenicity island present in most VTEC capable of causing severe disease in humans. The LEE allows VTEC to cause attaching-effacing (A/E) lesions in the host intestines. A/E lesions are characterized by loss of enterocyte microvilli and tight attachment of VTEC bacteria to the enterocyte membranes on pedestals created by rearrangement of the host cell cytoskeleton (Jores *et al.*, 2004), see Fig 2. A/E lesions cause disruption of the intestinal epithelium due to the cytoskeletal structure changes, loss of tight junction integrity, and apoptosis of epithelial cells (Serna & Boedeker, 2008). This contributes to the diarrhoea in VTEC infections as well as infections by enteropathogenic *E. coli* (EPEC), which share a similar intimin-based mechanism of cell adhesion but lacks verotoxin genes. (Schmidt, 2010) On a genetic level, the LEE encodes the host cell adhesion protein intimin, a translocated intimin receptor (Tir), the components of a type III secretion system (T3SS), and a number of accessory genes and further effector proteins transferred to the target cell via the secretion system (Ogura *et al.*, 2009). The core LEE region encoding intimin and the T3SS is largely conserved between different serotypes of VTEC, while non-O157 often have structurally variable LEE accessory regions (Ogura *et al.*, 2009). A/E-lesions can also be induced in colonized cattle (Nart *et al.*, 2008), and isogenic deletion mutants lacking either intimin or Tir have a decreased ability to colonize and persist in the bovine terminal rectal mucosa (Sheng *et al.*, 2006).

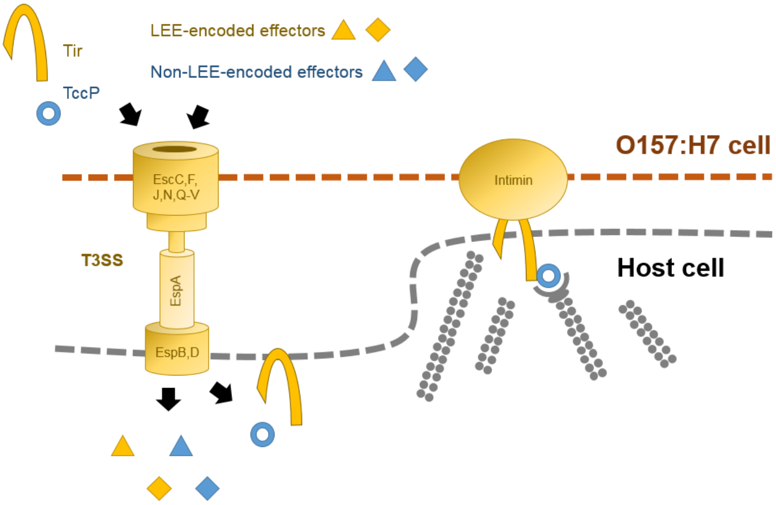

*Figure 2.* The LEE encodes the components of a type III secretion system, allowing O157:H7 to inject effector proteins that subvert host cell signalling pathways. Additionally, the translocated intimin receptor (Tir) and the Tir-cytoskeleton coupling protein (TccP) enters the host cell via the T3SS allowing an intimin complex to form which binds the bacteria and host cell closely together and initiates host cell actin polymerization, which leads to the loss of microvilli and the formation of a pedestal structure characteristic of A/E lesions (Frankel & Phillips, 2008; Garmendia *et al.*, 2005).

### *LEE-negative pathogenic VTEC*

VTEC infections can also be caused by LEE-negative VTEC that use other adhesion mechanisms (BIOHAZ, 2013), notably the strain of VTEC/EAEC O104:H4 that caused an outbreak in several European countries in 2011 with more than 4000 cases and 54 deaths (Karch *et al.*, 2012).

### 1.2.3 The O157:H7 large virulence plasmid (pO157)

*E. coli* that are pathogenic to humans frequently carry some of their genetic virulence determinants on plasmids (Johnson & Nolan, 2009). Most VTEC serotypes are associated with one or more large plasmids encoding a set of virulence factors including haemolysin, however, the plasmid backbones and genetic organization differs suggesting a complex evolutionary history (Ogura *et al.*, 2009). The O157:H7 large virulence plasmid pO157 is 92 kbp (Burland *et al.*, 1998). The pO157 plasmid carries genes encoding a type of EHEC haemolysin which can cause lysis of human microvascular endothelial cells and apoptosis, and is likely to contribute to O157:H7 pathogenesis (Bielaszewska *et al.*, 2014). The plasmid also encodes genes for putative virulence factors cytotoxin B (*toxB*), a serine protease (*espP*), and a catalase/peroxidase (*katP*) (Burland *et al.*, 1998), as well as the components of a type II secretion system (Schmidt *et al.*, 1997). Strains of O157:H7 cured of their pO157 plasmid have a decreased ability to colonize cattle (Lim *et al.*, 2010; Sheng *et al.*, 2006), but no effect was observed in this ability when haemolysin alone was deleted (Sheng *et al.*, 2006).

### 1.2.4 Other virulence factors

As previously mentioned, the LEE region encodes a number of “effector proteins”, entering the cell via the T3SS to subvert host cell functions. VTEC serotypes including O157:H7 have genes encoding additional non-LEE-encoded effectors, mostly located in prophages (Wong *et al.*, 2011; Ogura *et al.*, 2009; Tobe *et al.*, 2006). Also related to A/E-lesion formation but encoded outside the LEE is the Tir-cytoskeleton coupling protein TccP, which occurs in two copies of O157:H7 genomes, one of which is defective in most strains due to a frameshift mutation (Ogura *et al.*, 2007). TccP couples the intimin receptor Tir to the actin of the host cell cytoskeleton (Garmendia *et al.*, 2004). Cytolethal distending toxin (CDT) causes cell cycle arrest and apoptosis in a wide range of eukaryotic cell types, and has been found in VTEC as well as several other types of pathogenic bacteria (reviewed by (Heywood *et al.*, 2005)). There are also several O157:H7 adhesion factors including fimbriae encoded in genes outside the LEE, which may provide initial attachment to the host cell before pedestal formation is initiated (reviewed by (Farfan & Torres, 2012)).

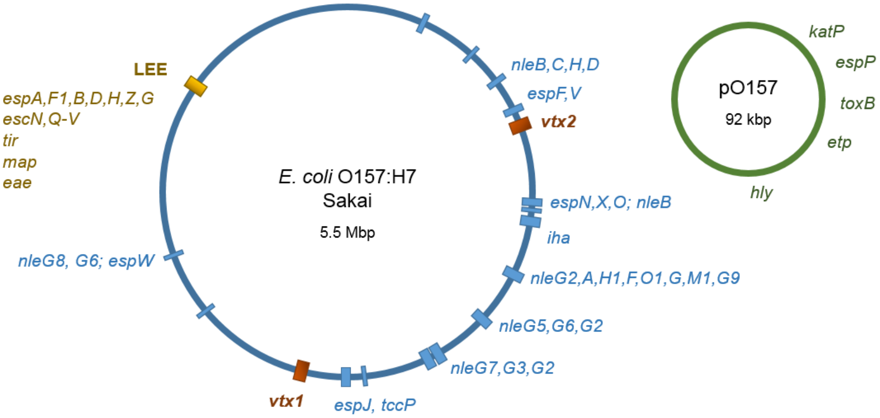
LEE
espA,F1,B,D,H,Z,G
escN,Q-V
tir
map
eae
nleB,C,H,D
espF,V
vtx2
E. coli O157:H7
Sakai
5.5 Mbp
espN,X,O; nleB
iha
nleG8, G6; espW
nleG2,A,H1,F,O1,G,M1,G9
nleG5,G6,G2
nleG7,G3,G2
vtx1
espJ, tccP
katP
espP
pO157
92 kbp
toxB
etp
hly

*Figure 3.* Schematic representation of acquired genetic elements encoding virulence factors in the O157:H7 Sakai strain chromosome, and the large virulence plasmid pO157. Adapted from (Boyd *et al.*, 2012; Tobe *et al.*, 2006) and (Burland *et al.*, 1998).

## 1.3 Evolution and phylogenetics of VTEC O157:H7

### 1.3.1 The evolutionary history of O157:H7

Genomic comparison of extant strains indicates that VTEC O157:H7 evolved from an EPEC-like O55:H7 ancestor by stepwise acquisition of the pO157 virulence plasmid, verotoxin-encoding prophages and other phage-encoded genetic regions including several T3SS effector genes (Zhou *et al.*, 2010; Wick *et al.*, 2005; Feng *et al.*, 1998), and an antigenic shift from O55 to O157 caused by the acquisition of a new *rfb* region (Wick *et al.*, 2005; Bilge *et al.*, 1996). Mutations also caused the loss of the ability to ferment sorbitol as well as loss of β-glucuronidase activity in most O157:H7 (Feng *et al.*, 1998). Strains of EPEC O55:H7 that are closely related to O157:H7 remain a cause of diarrhoea in infants in many countries (Whittam *et al.*, 1993). A subset of O157:H7 retained the ability to ferment sorbitol but lost their motility, forming an O157:$H^-$ lineage which also causes severe EHEC infections in humans (reviewed by (Karch & Bielaszewska, 2001)).

### 1.3.2 Lineages within the O157:H7 serotype

The clonal origin of all O157:H7 made early attempts to distinguish groups within the serotype based on variation in housekeeping genes difficult (Noller *et al.*, 2003). In fact, the major source of genetic variation readily observed by pulsed-field gel electrophoresis (PFGE) was found to be insertions and deletions, e.g. in the form of mobile genetic elements, rather than point mutations (Kudva *et al.*, 2002).

#### *Lineages*

Kim *et al.* suggested the existence of two separate lineages of O157:H7, lineage I (LI) and lineage II (LII), based on octamer-based genome scanning of a limited part of the genomes of isolates from the USA (Kim *et al.*, 1999). The two lineages were shown to be unevenly distributed between cattle and human sources with LII tending to be "cattle biased". A simpler assay using six fragment length polymorphisms (lineage specific polymorphism assay, LSPA-6) was shown to be a more convenient replacement for genome scanning in identifying these lineages (Yang *et al.*, 2004). Based on microarray analysis, Zhang *et al.* identified an intermediate lineage, LI/II (Zhang *et al.*, 2007). The existence of the same three lineages was largely supported by supernetwork analysis based on data from several multi-locus typing methods (Laing *et al.*, 2009), and more recent studies using next generation sequencing techniques to study genome-wide SNP variation (Dallman *et al.*, 2015; Eppinger *et al.*, 2011b; Eppinger *et al.*, 2011a; Leopold *et al.*, 2010; Leopold *et al.*, 2009).

#### *Clades*

Based on a set of 96 single nucleotide polymorphisms among clinical isolates of O157:H7, Manning et al. proposed a division of O157:H7 into nine clades. Clade 8 was associated with more severe disease as determined by comparing data from outbreaks. (Manning *et al.*, 2008). While clade typing generally is in concordance with the previously mentioned lineages, isolates representing multiple lineages are found in certain clades, in particular clade 7 (Yokoyama *et al.*, 2012; Eppinger *et al.*, 2011b; Liu *et al.*, 2009). This could be due to recombination or homoplasy, evident in the reticulated nature of the phylogenetic network in the original Manning paper (Manning *et al.*, 2008), but also to the inconsistent translation of LSPA-6 profiles into lineages by different authors (Yokoyama *et al.*, 2012). A weaknesses of the clade system is that it is based on clinical isolates, and therefore does not well represent the substantial SNP variation among strains of O157:H7 that rarely or never cause symptomatic disease in humans, many of which are clade 7 (Iyoda *et al.*, 2014; Soderlund *et al.*, 2014; Eppinger *et al.*, 2011b)). Like the lineage system, it is also based on a limited set of isolates in terms of geographic origin. As genome data from strains of O157:H7 from all over the world is increasingly made available, some of these issues can be expected to be resolved in the near future.

## 1.4 Human VTEC O157:H7 infection

### 1.4.1 Symptoms

The infectious dose of VTEC O157:H7 is considered to be less than 100 bacteria (Tilden *et al.*, 1996), meaning that it is not necessary for the bacteria to grow in contaminated material, e.g. foodstuffs, to pose a threat to exposed humans. The incubation period for an O157:H7 infection is typically 3-4 days (Bell *et al.*, 1994). Infection by VTEC can be asymptomatic, even for serotypes associated with severe disease like O157:H7 (Sartz *et al.*, 2008). Symptomatic patients will in most cases experience watery diarrhoea which can resolve spontaneously in about a week, or progress to bloody diarrhoea, stomach cramps and more rarely fever (reviewed by (Yoon & Hovde, 2008)). Treatment with antibiotics is not recommended as this is suspected to induce increased toxin production, and patients are generally given intravenous volume expansion and monitored for complications (Goldwater & Bettelheim, 2012; Karch *et al.*, 2005).

### 1.4.2 Haemolytic-uraemic syndrome (HUS)

A particular public health concern with VTEC infection is the risk of patients developing the life threatening haemolytic-uraemic syndrome (HUS). This occurs in 3-7% of sporadic cases of O157:H7 infection, but up to 20% in particular outbreaks, with young and elderly patients at greater risk (Mead & Griffin, 1998). HUS is characterized by a combination of microangiopathic haemolytic anaemia, thrombocytopenia, and acute kidney failure. Apart from directly inducing cell death, verotoxin causes local inflammation in vascular epithelium, loss of thromboresistance and complement activation. This causes thrombi to form in capillaries and arterioles, with restricted blood flow causing ischemia. Erythrocytes fragment as they pass the partially blocked vessels, causing haemolytic anaemia. Kidney damage is also caused by verotoxin action on renal endothelial cells. In addition to the kidneys, Gb3 receptors targeted by verotoxin are present in high density in the brain, and HUS causes neurological symptoms such as seizures, stroke or coma in around 25% of patients. (Mele *et al.*, 2014) Brain involvement is the main cause of death among HUS cases (Trachtman *et al.*, 2012). With appropriate healthcare interventions, only 1-4% of HUS cases currently lead to death (Mele *et al.*, 2014), but before the introduction of early dialysis therapy, the death rate could be as high as 30% (Spinale *et al.*, 2013).

20-40% of HUS patients develop long-term sequelae. The most common involve the kidneys, including proteinuria and/or microalbuminuria, hypertension, chronic kidney disease and end-stage kidney disease. In rare cases, patients with HUS may require chronic dialysis or kidney transplant. There is also a risk of long-term effects on gastro-intestinal and cognitive function. (Spinale *et al.*, 2013)

### 1.4.3 Infection routes

O157:H7 is a zoonotic disease agent, e.g. transmitted to humans from animals. Human infection by O157:H7 occurs both as sporadic cases and outbreaks, with sporadic cases making up the majority most years in Sweden (Anonymous, 2014a). The primary routes of infection are presented in figure 4.

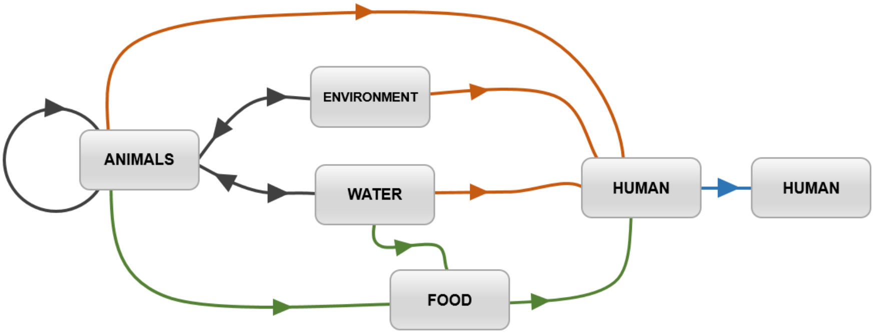

*Figure 4.* The three main routes of human infection by O157:H7; contaminated food (green), direct and indirect contact with infected animals (orange) and person-to-person (blue).

### *Food*

Historically, emphasis has been on O157:H7 as a foodborne disease, and contaminated food including ground beef, other meat products, unpasteurized dairy products and vegetables is the dominant source of major outbreaks worldwide (Heiman *et al.*, 2015; Snedeker *et al.*, 2009; Rangel *et al.*, 2005). In Sweden, the largest O157:H7 outbreaks with known sources have been caused by fermented sausages 2002 (Sartz *et al.*, 2008), lettuce contaminated by an upstream cattle farm 2005 (Soderstrom *et al.*, 2008), and locally produced meat products 2006-2008 (Erntell & Jönsson, 2010).

### *Direct and indirect contact with animals*

Contamination of the environment by ruminant faeces is increasingly recognized as another important source of O157:H7 infection (Strachan *et al.*, 2006). Direct contact with ruminants is a recurring cause of sporadic cases of O157:H7 infection in Sweden (Anonymous, 2014a). A nationwide study on risk factors for sporadic EHEC infection in Sweden found significant association with family members working with animals (especially cattle), staying at a summer house, or having a private water supply. No foodstuffs were identified as significant risk factors (Ivarsson *et al.*, 2012). Furthermore, local cattle density is significantly associated with O157:H7 incidence (Kistemann *et al.*, 2004). Direct contact with animals and their surroundings at agricultural fairs, petting zoos and similar activities have been linked to large outbreaks in the USA (MacFarquhar & Williams, 2013; CDC, 2005) and the UK (Griffin *et al.*, 2010). Recreational activities on land that is or has been used as pasture also poses a risk and has been linked to outbreaks (Howie *et al.*, 2003; Crampin *et al.*, 1999). Runoff from manure or pastures can contaminate drinking water supplies, especially private wells (Rangel *et al.*, 2005; Schets *et al.*, 2005; Said *et al.*, 2003) or surface water bodies used e.g. for swimming (Rangel *et al.*, 2005; Paunio *et al.*, 1999) or irrigation of vegetables intended to be consumed raw (Gelting *et al.*, 2011; Soderstrom *et al.*, 2008). O157:H7 can survive for long periods in water, particularly at low temperatures (Wang & Doyle, 1998).

### *Person-to-person infection*

While O157:H7 is primarily a zoonosis, secondary person-to-person infections can easily occur as a result of the low infectious dose. A statistical analysis of data from several countries found secondary infection to be the cause of around 20% of the cases in outbreaks, with infection within the home or in situations where children are in close contact (e.g. nurseries) as the most common infection settings (Snedeker *et al.*, 2009).

*Figure 5.* Direct and indirect contact with animals is considered a frequent cause of sporadic VTEC infections in Sweden. A case of O157:H7 (clade 8) infection in the summer of 2014 was linked to swimming at this pier near Egby, Öland. Photo: Kajsa Skarsgård

## 1.5 Animal reservoirs of O157:H7

### 1.5.1 Cattle

Cattle are considered the primary host of most VTEC, including the O157:H7 serotype, and the main reservoir for human O157:H7 infection (reviewed by (Ferens & Hovde, 2011)). Cattle are infected via the faecal-oral route by contamination of feed, water, or the environment or by direct contact with other animals (McGee *et al.*, 2004; Hancock *et al.*, 1998). The primary colonization site in the bovine gastrointestinal tract is the terminal mucosa region near the rectal-anal junction (RAJ) (Low *et al.*, 2005; Naylor *et al.*, 2003), where attaching/effacing lesions are induced (Nart *et al.*, 2008). Cattle infected by

O157:H7 are generally asymptomatic, although diarrhoea can occur in neonatal calves (Brown *et al.*, 1997; Dean-Nystrom *et al.*, 1997). Colonized animals most commonly shed the bacteria in their faeces for around one month, although some animals can keep shedding for far longer periods (Lim *et al.*, 2007; Grauke *et al.*, 2002). Younger animals from weaning up to one year tend to shed O157:H7 more frequently compared to adult animals and calves younger than two months (Boqvist *et al.*, 2009; Paiba *et al.*, 2003; Nielsen *et al.*, 2002). In a colonized herd, the majority of the O157:H7 can be excreted by a small subset of animals referred to as "high-shedders" or "super-shedders" (reviewed by (Chase-Topping *et al.*, 2008)). Shedding is more prevalent during the summer and early autumn (Widgren *et al.*, 2015; Schouten *et al.*, 2005; Barkocy-Gallagher *et al.*, 2003), coinciding with a seasonal increase of human cases of infection (Anonymous, 2014a; Rangel *et al.*, 2005).

### 1.5.2 Prevalence in Swedish cattle

*Prevalence at slaughter*

Nationwide slaughterhouse prevalence studies for VTEC O157:H7 have been performed regularly in Sweden since 1996. The studies include major slaughterhouses, and the number of samples collected is kept proportional to the number and age categories of animals slaughtered. In addition to a prevalence estimate, the studies have provided a collection of isolates representing the O157:H7 population in different parts of the country over the years. In the last decade, the prevalence has been fairly stable at 3.1 - 3.4 % (Table 1). Earlier studies reported lower prevalence estimates, but the numbers are not directly comparable as the analyses have been performed at different labs and methodological improvements have been made over the years (see (Eriksson, 2010) for details). For comparison, a recent meta-analysis found a European overall average of 5 % prevalence based on 53 studies in the period 1980-2012, although individual studies reported estimates ranging between 0.25 % and 10 % (Islam *et al.*, 2014).

In addition to individual faecal samples, skin samples consisting of ears have been collected in some of the Swedish slaughterhouse prevalence studies. Ear samples are considered to represent the herd status, as a single shedding animal can contaminate the skin of many animals housed as a group. Using ear samples is also a cost-effective way of collecting isolates for typing, giving a more representative picture of the clonal composition of VTEC populations. The drawback with ear samples is that the isolated strains cannot be linked to the point of origin with the same certainty, due to the higher risk of cross-contamination between animals during transport, in lairage and at the slaughterhouse.

Table 1. *Summary of all slaughterhouse prevalence studies of* E. coli *O157:H7 in Swedish cattle and sheep. Confidence intervals calculated using the binom package in R 2.15.2 with the Agresti-Coull method (Brown et al., 2001; Agresti & Coull, 1998).*

| **Year** | **No. of faecal samples** | **Prevalence [95% CI]** | **No. of ear samples** | **Prevalence [95% CI]** |
|---|---|---|---|---|
| **Cattle** | | | | |
| 1996-1997 | 3071 | 1.2 % [ 0.8-1.7 ] | - | - |
| 1997-1998 | 2308 | 0.3 % [ 0.1-0.6 ] | - | - |
| 1999 | 2057 | 0.7 % [ 0.4-1.2 ] | - | - |
| 2000 | 2001 | 1.7 % [ 1.2-2.4 ] | - | - |
| 2001 | 1998 | 1.3 % [ 0.8-1.9 ] | - | - |
| 2002 | 2032 | 1.4 % [ 0.9-2.0 ] | - | - |
| 2005-2006 | 1758 | 3.4 % [ 2.6-4.4 ] | 446 | 12.1 % [ 9.3-15 ] |
| 2008-2009 | 1993 | 3.3 % [ 2.6-4.2 ] | 500 | 8.2 % [ 6.1-11 ] |
| 2011-2012 | 2376 | 3.1 % [ 2.5-3.9 ] | - | - |
| **Sheep** | | | | |
| 2007-2008 | 492 | 1.8 % [ 0.9-3.5 ] | 105 | 1.9% [ 0.1-7.1 ] |

### *Farm-level prevalence*

A nationwide prevalence study of 371 cattle dairy farms was performed in Sweden 1998-2000 (see section 3.1.1 for details). O157:H7 was found on 8.9% of farms overall, but there was substantial regional variation with the highest prevalence in Halland county (23.3%) (Eriksson *et al.*, 2005). In a more recent study, 126 cattle farms in the southern part of Sweden were repeatedly sampled over a period of 38 months 2009-2012 with positive samples found at least once on 53% of the farms, and with 6%-72% positive sampling occasions on the farms that were positive at least once (Widgren *et al.*, 2015).

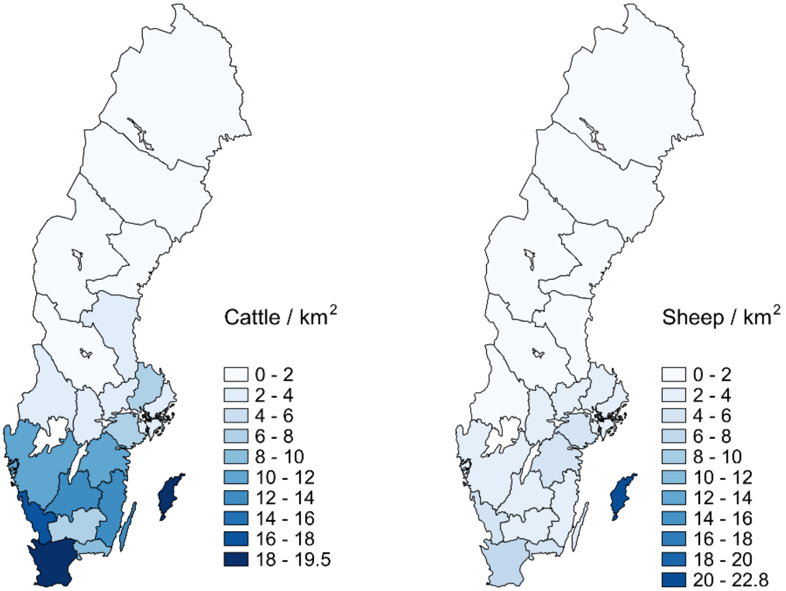

*Figure 6.* County-level cattle and sheep density in Sweden, based on data from the "Livestock in June 2014" report from the Swedish Board of Agriculture. Map and area data from Statistics Sweden (scb.se). Created in QGIS 2.4.

### 1.5.3 Other ruminants

Small domestic ruminants (sheep and goats) form a known reservoir for O157:H7. The bacteria show a different tissue tropism compared to when infecting cattle, with a more diffuse colonization pattern that is not focused on the RAJ. (Reviewed by (La Ragione *et al.*, 2009)). O157:H7 in sheep is further discussed in section 4.3. Bovines other than cattle can also be colonized by O157:H7 including the American bison (*Bison bison*) (Reinstein *et al.*, 2007), and the water buffalo (*Bubalus bubalis*) (Galiero *et al.*, 2005).

Wildlife can introduce infectious diseases to farm animals, with grazing animals particularily at risk (Magnusson, 2012). Wild ruminants infected with O157:H7 can contaminate pastures, feed or water supplies to infect the domestic animals, or directly contaminate food or the environment to infect humans. For instance, O157:H7 can be isolated from deer and various deer products (Laidler *et al.*, 2013; Ahn *et al.*, 2009; Fischer *et al.*, 2001), and experimentally infected deer can shed the bacteria for several weeks (Fischer *et al.*, 2001).

### 1.5.4 Non-ruminant animals

Non-ruminant animals are generally considered to only serve as transient hosts for O157:H7 (Caprioli *et al.*, 2005), but can nonetheless play important epidemiological roles as vectors. O157:H7 has been found in Swedish domestic pigs (Eriksson *et al.*, 2003), and feral pigs are believed to have caused a major outbreak in the USA by contaminating spinach fields (Jay *et al.*, 2007). Companion and sport animals such as dogs and horses have been found to occasionally carry O157:H7 (Hancock *et al.*, 1998). Birds that inhabit agricultural areas or migrate can spread O157:H7, e.g. gulls (Wallace *et al.*, 1997) and starlings (Swirski *et al.*, 2014). One study found surprisingly long-lasting colonization and shedding in domestic chickens after experimental inoculation (Schoeni & Doyle, 1994). Other animals frequently occuring in the farm environment have also been shown to carry O157:H7 include manure-associated flies in the *Muscidae* and *Calliphoridae* families (Talley *et al.*, 2009; Alam & Zurek, 2004), and rodents (Kilonzo *et al.*, 2013; Blanco Crivelli *et al.*, 2012).

## 1.6 Genotypic characterization of O157:H7

### 1.6.1 Molecular epidemiology

Molecular biology has increasingly been recognized as a powerful tool to answer epidemiological questions. In the context of microbiology, the goals of this rapidly developing field of “molecular epidemiology” has been defined as to:

“*identify the microparasites (viruses, bacteria, fungi, and protozoa) responsible for infectious diseases and determine their physical sources, their biological (phylogenetic) relationships, and their routes of transmission and those of the genes (and accessory elements) responsible for their virulence, vaccine-relevant antigens, and drug resistance*” (Levin *et al.*, 1999).

Many molecular methods for characterization (often colloquially referred to as “typing”) have been used over the years to distinguish between different genetic groups of O157:H7 (see (Karama & Gyles, 2010; Hopkins & Hilton, 2000) for reviews of more recent and older methodology). The key methods of relevance for this thesis project will be briefly presented in the following sections.

### 1.6.2 Virulence gene detection and subtyping

As the presence of specific DNA sequences encoding virulence factors gives certain *E. coli* the ability to cause disease (see section 1.2), the detection of these sequences in a strain is frequently of interest. The polymerase chain reaction (PCR) is the central technique in current molecular biology, and the starting point for most assays. In PCR, two short artificial DNA fragments referred to as "primers" bind to the region of interest in the target DNA, which is then amplified exponentially by a DNA polymerase enzyme through repeated cycles of heating and cooling of the sample. If the sample is positive, i.e. contains the target sequence, the end result will be a high number of copies of the target region DNA. The products can be visualized and size-determined by agarose gel electrophoresis. (Saiki *et al.*, 1988), detected in real-time by the use of intercalating dyes or fluorophore-labelled probes (Reviewed by (Espy *et al.*, 2006)), or detected by other means, e.g. by hybridization to a microarray (Anjum *et al.*, 2007).  Digital PCR is a modification of the standard PCR technique in which the reaction volume is divided into hundreds to millions of separate microvolumes, each of which can be individually evaluated for the presence of one or more targets. This allows improved quantification of target DNA molecules as well as new opportunities in terms of genotyping (reviewed by (Baker, 2012)).

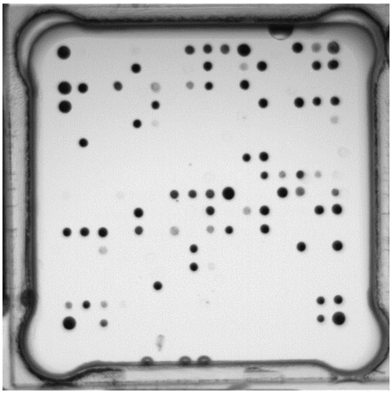

*Figure 7.* A low-density microarray presenting signals corresponding to virulence genes present in the O157:H7 isolate PN1400/00 from paper I. The array surface is ca 3 x 3 mm.

### 1.6.3 Pulsed-field gel electrophoresis (PFGE)

Pulsed-field gel electrophoresis (PFGE) has long been considered the “gold standard” for typing of VTEC, providing high discriminatory power and epidemiological concordance (Karama & Gyles, 2010). In PFGE, genomic DNA is cleaved by restriction enzymes which target rare sequence motifs only expected to occur 10-20 times in the bacterial genome. The restriction products are analyzed by gel electrophoresis, with a system of alternating voltage orientations allowing separation of the large DNA fragments. Related bacteria, e.g. a set of O157:H7 isolates, will differ in the number and position of restriction sites due to insertions and deletions, recombinations, mobile genetic element movements and point mutations in the restriction sites. The number and size of the resulting DNA restriction fragments will therefore, once separated, form a band pattern which can be used as a “fingerprint” for the bacterial strain. To improve the discriminatory ability of the method, secondary and tertiary restriction enzymes with different target motifs can be used in parallel. (Caprioli *et al.*, 2014; Anonymous, 2013)

PFGE is a laborious method, requiring substantial manual effort and taking 2-3 days to complete (Anonymous, 2013). Harmonized methodology at all stages of the process is needed to perform PFGE to a level where the data is comparable between laboratories (Caprioli *et al.*, 2014). In the European Union, PFGE typing data for VTEC from human sources is collected in a framework initiated by the European Centre for Disease Prevention and Control (ECDC), with a similar and connected scheme for collection of VTEC PFGE data from food, feed and animal samples currently under development by the European Food Safety Authority (EFSA) (Caprioli *et al.*, 2014).

### 1.6.4 Multilocus variable-number tandem repeat analysis (MLVA)

Tandem repeat sequences, i.e. the same short sequence motif occurring multiple times in direct succession in the genome, occur in both eukaryotic and prokaryotic genomes. Errors during DNA replication frequently occur in such loci, leading to repeat copies being added or lost (Bustamante *et al.*, 2013; Vogler *et al.*, 2006) and creating a variable-number tandem repeat (VNTR) locus. In multilocus VNTR analysis (MLVA), the copy number at several such loci is determined and used as an identifier for a bacterial isolate or clone (reviewed by (Lindstedt, 2005)). To determine the copy numbers, the loci are amplified by PCR and the size of the PCR products determined by capillary gel electrophoresis with fluorophore labels identifying the amplified loci.

### 1.6.5 Phage typing

Phage typing is a characterization method based on variable susceptibility to lytic infection by bacteriophages. In the O157:H7 phage typing system, each strain is exposed to a panel of 14 T4 and 2 T7 phages, and the resulting susceptibility profile translated to a phage type (PT) (Khakhria *et al.*, 1990; Ahmed *et al.*, 1987). Phage typing has been widely used as a categorization tool for O157:H7, and phage types have repeatedly been shown to be relevant markers for differing virulence (see e.g. (Matthews *et al.*, 2013; Aspan & Eriksson, 2010)).

### 1.6.6 Next generation sequencing

In recent years, a rapid technological development has made high throughput sequencing increasingly affordable and accessible for molecular epidemiology applications (Didelot *et al.*, 2012). These new sequencing technologies have been referred to as “next generation sequencing” (NGS) or “whole genome sequencing” (WGS). NGS data can be used for whole genome SNP analysis, a powerful subtyping tool which is an attractive alternative to PFGE and MLVA for O157:H7 phylogenetics, source tracing and outbreak investigation (Dallman *et al.*, 2015; Eppinger *et al.*, 2011b). NGS can also be used for targeted detection and characterisation of genetic regions of interest, e.g. virulence factors (Joensen *et al.*, 2014).

# 2 Aims of the thesis

The specific aims of this thesis project were to:

- Establish methods and reference databases for improved molecular source tracing and outbreak investigation of O157:H7 originating from Swedish cattle and sheep

- Investigate the prevalence of O157:H7 in Swedish sheep, and compare sheep and cattle isolates

- Compare the prevalence of genotypes of O157:H7 among sheep, cattle and human patients to find clinically over- or underrepresented genotypes, to provide clues to the relative virulence of the different types

- Develop methodology for detection of genetic signatures for high-virulence types

- Compare the repertoire of putative genetic virulence factors in different types of O157:H7 to explain differences in virulence

- Compare historical and recent collections of Swedish O157:H7 from animal reservoirs to study population dynamics of possible public health relevance

# 3 Considerations on materials and methods

## 3.1 Bacterial isolates

The selection of isolates included in the different studies in this project and the sampling strategies that were used to acquire those isolates are briefly described in the following sections.

### 3.1.1 Isolates from samples taken on cattle dairy farms

An on-farm prevalence study for O157:H7 was performed on Swedish dairy cattle during 1998-2000 in collaboration with the Swedish Dairy Association. A number of faecal and manure samples were taken proportionally to the number of cows in each regional livestock association in Sweden, on a random selection of 371 farms. As participation in the milk-recording scheme used to select the farms was 73% at the time, the findings of O157:H7 should be representative of dairy cattle during the study period. At total of 88 isolates were found on 33 different farms. (Eriksson *et al.*, 2005) The same set of isolates from this study was used in papers I and II, consisting of one isolate each from 32 of these 33 positive farm, as one of the isolates was unfortunately lost in storage.

### 3.1.2 Isolates from samples taken at slaughterhouses

*Cattle*

Nationwide slaughterhouse prevalence studies for VTEC O157:H7 were conducted every year in Sweden 1996-2002 (Albihn *et al.*, 2003), and every three years since then, with the latest in 2014-2015 (Dresch, 2015). The studies include major slaughterhouses, and the number of samples collected is kept proportional to the number and age categories of animals slaughtered. Sampling is performed during a one-year period to account for seasonal variations. Sample numbers and the results of each study is summarized in Table 1. All 381 isolates found in these studies 1996-2009 were used to represent the cattle population of O157:H7 in paper IV and for matching against sheep MLVA data in paper III. The subset from the 2005-2006 study (n=116) (Boqvist *et al.*, 2009) was used for comparison to the sheep isolates in study III by PFGE. Isolates from the study 2011-2012 were studied in paper V.

*Sheep*

The first slaughterhouse prevalence study was performed on Swedish sheep during a one-year period 2007-2008, using the same basic sampling strategy described for the cattle prevalence studies in the previous section. Both faecal (n=492) and ear (n=105) samples were collected, and a total of 11 isolates of O157:H7 were found. The results of this prevalence study are reported in paper III.

### 3.1.3 Isolates from samples taken on farms with links to cases of illness in humans

When an outbreak or a sporadic case of O157:H7 illness in humans occurs in Sweden, an investigation to trace the source of the infection is routinely done. If the patient has recently visited a farm where ruminants are kept, the animals are sampled, and any findings of O157:H7 compared to the patient isolate. Traditionally, this has been done with pulsed field gel electrophoresis (PFGE, see section 1.6.3). 13 isolates from 6 different cattle farms and a single goat farm linked in this way between 1998 and 2000 were compared to the prevalence study isolates in papers I and II. Similarly, three isolates each representing a farm where sheep were confirmed as the source 2004-2009 were included in paper III.

### 3.1.4 Isolates from human patients

Isolates of O157:H7 from regional laboratories nationwide are routinely sent to the Public Health Agency of Sweden (FoHM) for confirmation and characterization. Thereby, the FoHM isolate collection is representative of O157:H7 causing symptomatic infections in the country. In paper III, 850 isolates from 2001-2009 were used for matching against sheep isolates by PFGE and 110 isolates from 2008-2009 by MLVA. 197 isolates from 2008-2011 were compared to cattle isolates in paper IV, with access to additional information regarding if the infection had been considered domestic or acquired abroad, and information about which isolates had caused HUS. A subset of the human O157:H7 isolates with the LSPA-6 profile 213111 from 2011 in paper IV were also re-used for IS-element testing in paper V.

## 3.2 Comparison of cattle, sheep and human isolates

In paper III, MLVA data from cattle, sheep and human isolates could be directly compared, while PFGE analysis of the sheep isolates was performed independently at both SVA and FoHM with the resulting data matched against local databases. For both methods, the matching was performed as sheep isolates versus all cattle and human isolates, not as all versus all. The study described in paper III was carried out largely based on pre-existing data, and since MLVA had recently been introduced at that point while PFGE was being phased out at the collaborating partners, data from both methods was not available for all included isolates. In contrast, in paper IV all versus all comparison of MLVA and other typing data was performed between cattle and human isolates.

## 3.3 Molecular characterization methods

The molecular methods used in this thesis project are described in each paper. The following sections will briefly discuss cases where different assays have been used to produce the same types of data in the different papers.

### 3.3.1 Single nucleotide polymorphism typing by real-time PCR

For screening large sets of samples for one or a few SNPs, real-time PCR is an attractive option. However, standard real-time PCR systems are generally fairly tolerant of a single base difference in one of the primers or the probe, requiring specialized techniques to differentiate the SNP alleles. Two different approaches have been used in this project. In papers II and III, a system based on hairpin primers is used. With this approach, an intercalating dye provides the real-time PCR signal, and two separate reactions are performed with different primers targeting the two SNP variants. The primers are designed to form hairpins when in solution, meaning that this state will be thermodynamically more favourable. This will make the primer less likely to bind if there is an SNP mismatch. The cycle number required to reach a pre-determined fluorescens threshold is compared between the two reactions, with the fastest representing the SNP state in the assayed isolate (Riordan *et al.*, 2008). For paper IV, a new probe-based PCR assay was designed. The probes are modified with duplex stabilizers, allowing them to be very short and increasing the effect of a single base mismatch. Only the probe matching the SNP state will give a signal. With this approach, the analysis time is substantially shorter, the interpretation simpler and the reagent consumption is halved.

### 3.3.2 Subtyping of verotoxin 2 genes

As discussed elsewhere in this thesis (1.2.1 and 4.2.1), subtypes of verotoxin 2 can be relevant markers for the virulence of strains of O157:H7. Several methods for verotoxin 2 subtyping were used in this project as available technologies improved during the project period. In paper I, a PCR product restriction fragment length polymorphism analysis (PCR-RFLP) of parts of the verotoxin 2 B-subunit protocol was used (Pierard *et al.*, 1998). An improved re-analysis of the same isolates by dye-terminator sequencing of parts of the A and B subunit encoding sequences (Persson *et al.*, 2007) was performed in paper II. In contrast with PCR-RFLP, sequencing is suitable for characterizing new variants, e.g. the previously not described SNP variation in $vtx_{2a}$ in the PT2 isolates in paper II, and the presence of IS-elements in certain $vtx_{2c}$ genes in papers II, III and V. The same sequencing protocol was used in paper III and paper V. All these methods are based on PCR targeted by primers, and will fail if a primer target region is lost or altered in the investigated isolate. In paper V, we used Illumina MiSeq NGS data to investigate why verotoxin 2 PCR failed in certain isolates, and found that one of the primer sites together with a significant portion of the verotoxin 2 gene had been lost in this isolate, likely as a result of IS-element activity. Thus, NGS can provide even more unbiased information about the presence and state of genes of interest.

### 3.3.3 Next generation sequencing

Two different next generation sequencing technologies were used in this project on a low number of isolates and for specific purposes; Roche/454 GS FLX for SNP typing in paper IV (see section 4.1.3) and Illumina MiSeq for investigation of failed verotoxin gene sequencing in paper V (section 3.3.2). The two primary reasons for the change of technology were the in-house availability of a MiSeq instrument when paper V was being produced and the far lower cost per sequence unit of Illumina sequencing.

## 3.4 Bioinformatics

### 3.4.1 Microarray analysis

The microarray system used in paper I is commercial, but the official data analysis functionality was under development during the project period. Therefore, an analysis pipeline was written in the R software environment (https://www.r-project.org/), with additional tools from Bioconductor (https://www.bioconductor.org/) to normalize, filter and summarize the probe-level intensity data produced by Alere Iconoclust. The performance of each individual probe across all arrays in the project was also assessed, and the results used to eliminate probes producing ambiguous results.

### 3.4.2 MLVA and LSPA-6 analysis

To semi-automate MLVA and LSPA-6 data analysis, a script was written to identify fragments based on size and fluorophore label, perform size correction using calibration data, assign copy numbers and report possible errors. This script was written in the VBA language as a series of linked Microsoft Excel macros to provide a familiar user interface and allow later customization by the diagnostics staff at the SVA. The Bionumerics VNTR plugin was evaluated to automate peak selection, but did not perform reliably. Bionumerics was used to calculate diversity indices for the various isolate categories in paper IV, and as a database to store the multiple types of data generated from each isolate during the project.

### 3.4.3 Sequence analysis

Sanger sequence data was assembled and analysed in Bionumerics. Assembly of NGS read data was performed using Newbler for Roche/454 GS FLX data in paper IV and MIRA for Illumina MiSeq data from the single sequenced isolate in paper V. Targeted extraction of regions of interest in the NGS data in papers IV and V was performed using the BLAST+ suite from NCBI with local databases created from the assemblies.

### 3.4.4 Clustering and visualization

Clustering and creation of dendrograms for papers I-III was performed in Bionumerics. The NeighborNet algorithm (Bryant & Moulton, 2004) in the SplitsTree software was used to visualize data from NGS SNP typing performed in paper IV, since this algorithm produces a phylogenetic network which better represents the data structure originally observed by Manning and co-workers (Manning *et al.*, 2008). This figure was not included in paper IV, but is presented in this thesis (Fig. 8). MLVA data clustering was performed in Bionumerics using the minimum spanning tree algorithm (see e.g. Fig 9), except in paper III where an UPGMA analysis was performed instead to facilitate a side-by-side comparison with the clustering based on PFGE data.

To visualize geospatial information in this thesis (Fig. 6, Fig. 10) and in paper III, Supp. file 1, the QGIS open source geographic information system was used. Statistical testing for local spatial clusters in the most recent data sets presented in paper IV was performed using the Bernoulli case-control model (Kulldorf, 1997) implemented in SaTScan, with the results presented in Fig. 10.

# 4 Results and discussion

## 4.1 Evaluation and comparison of molecular typing methods for O157:H7

### 4.1.1 MLVA

After the favourable outcome in the direct comparisons with PFGE in papers II, and III, MLVA was used as the typing method of choice for O157:H7 in the later prevalence studies and in paper IV. The percentage of isolates that could be assigned to a type (“typeability”) of MLVA was 100% in all studies during this project. In the largest dataset in paper IV, the discriminatory power was also very high, with 379 unique profiles identified among the 578 isolates analysed and Simpson’s index of diversity of 0.98-0.99 in the different isolate sets. However, certain profiles seem to remain stable over extended periods of time (e.g. 5-7-14-4-6-6-7-5, which has been found repeatedly in Swedish cattle between 1996 and 2011), so the value of any match in for example a source tracing context should still take the profile frequency in prevalence study datasets into account.

Unfortunately, the reproducibility of this particular MLVA assay is dependent on careful PCR optimization to avoid false negative marker assignments in particularly unfavourable profiles (e.g. when the most efficiently amplified markers have a low copy number). PCR difficulties caused a high fail rate for this particular MLVA scheme when a laboratory proficiency test was carried out (Hyytia-Trees *et al.*, 2010). Calibration using a panel of isolates with known copy numbers is necessary to make data comparable between labs for any MLVA assay. A consistent nomenclature for reporting profiles is also critical. (Larsson *et al.*, 2013). In this project, extensive PCR multiplex optimization was performed, one of the primers was replaced, and standard panels of isolates were exchanged between the SVA and the Public Health Institute to ensure the validity of the data.

An obvious drawback of MLVA in general is the high assay specificity, meaning that each species or even serotype of a disease agent requires a separate set of VNTR markers to be selected and validated (Lindstedt *et al.*, 2013). This is in contrast to alternatives like PFGE and whole genome SNP analysis which can be applied to a broad spectrum of bacteria. However, this is somewhat less of an issue for major pathogens such as O157:H7 which may well motivate the effort of maintaining a separate assay.

The major advantages of MLVA is the low-cost, high-throughput nature of the assay, and the portable, simple data produced (Lindstedt *et al.*, 2013). This is evident in how MLVA data could be compared more directly between collaborating laboratories in papers III and IV. During the period of the project described in this thesis, approximately 1500 isolates of O157:H7 were analysed by MLVA at the SVA, a volume of samples that would not have been feasible to process with any of the other widely used methods. MLVA analysis also scales well, i.e. the fixed effort of the assay is low making it well suited for analysing small “case” batches of samples which can then be compared and cumulatively added to a larger database.

### 4.1.2 Phage typing and LSPA-6

Typing methods used for tracing, outbreak investigation and similar tasks tend to have high discriminatory power and generate diverse profiles. This means the phylogenetic signal in the typing data (i.e. the tendency for more related isolates to be more similar than randomly expected (Diniz-Filho *et al.*, 2012)) is eventually lost over generations, and that clusters can be diffuse and merged. Higher level categorization of strains or clones of a bacterial pathogen is therefore useful e.g. for communicating results and monitoring trends over time, as well as identifying clones that are more likely to cause severe infections. Phage typing is the traditional high-level typing method for O157:H7, using sensitivity to lytic infection by a panel of bacteriophages to assign a type (Khakhria *et al.*, 1990; Ahmed *et al.*, 1987). Phage typing is informative, but requires a skilled operator and significant manual labour. Phage typing was used for the isolates in paper I and II, via an external service provider.

As more sequence data from O157:H7 isolates has become available in recent years, genetic markers for high-level phylogenetic classification have been described. The LSPA-6 system consists of a set of lineage-specific length polymorphisms in six different genetic loci (Yang *et al.*, 2004). Lineages have been correlated to unequal representation between patient and cattle, and to differing expression levels of verotoxin and other virulence factors (Abu-Ali *et al.*, 2010a; Abu-Ali *et al.*, 2010b; Zhang *et al.*, 2010; Dowd & Williams, 2008). For this project, the original LSPA-6 assay was adapted for capillary gel electrophoresis, producing an assay highly similar to MLVA. This allowed far better resolution and higher sample throughput compared to slab gel electrophoresis. For instance, we describe a number of informative 1-2 bp insertions and deletions in *yhcG* that are likely to have been overlooked in previous studies. In this project, we show that the LI/II classification alone is not in itself predictive of the virulence of a strain, as the dominant LI/II strain among Swedish cattle is virtually never found among patient isolates (paper IV).

Both phage typing and LSPA-6 profiles were available from the cattle slaughterhouse prevalence studies between 1996 and 2002 in paper IV, and a comparison is presented in table 2. In more recent and ongoing studies (e.g. the 2014-2015 O157:H7 slaughterhouse prevalence study), scripts are in use that automate extraction of clade classification SNP states and LSPA-6 profiles from next generation sequencing data, providing backward compatible information.

Table 2. *Comparison between LSPA-6/clade 8 PCR typing and phage typing. Isolates are from the 1996-2002 prevalence studies in paper IV.*

| **LSPA-6** | **Phage typing (no. of isolates)** |
|---|---|
| LI/II 211111 (clade 8) | PT4 (43), PT14 (5), PT1 (1), PT8 (1), PT24 (1) |
| LI/II 211111 (non-clade 8) | PT14 (20), RDNC (3), PT49 (2), PT8 (1) |
| LI/II-like 213111 | PT8 (42), RDNC (13), PT14 (3) |
| LII 223323 | PT34 (3) |
| LII 212113 | PT2 (2) |
| LII 262123 | PT2 (1), PT54 (1) |
| LII 221123 | PT34 (1) |

### 4.1.3 Next generation sequencing SNP typing

The next generation sequencing in paper IV was performed with an early form of NGS technology, and only a low number of isolates could be sequenced due to the high cost. The low coverage (8-12x) meant that we did not perform SNP discovery in this dataset, and instead used the data for clade typing and further phylogenetic analysis based on extracting the states on known informative positions. This allowed us to confirm the high degree of similarity between Swedish clade 8 and clade 8 from the USA, and confirm the LSPA-6 based identification of Swedish isolates as lineage II. However, using a pre-defined SNP set severely limits the usefulness for genotypes that were not included in the original SNP selection process. This is illustrated in figure 5, where the inability of the clade typing SNP set to separate the “cattle biased” strains in clade 7/lineage II is evident.

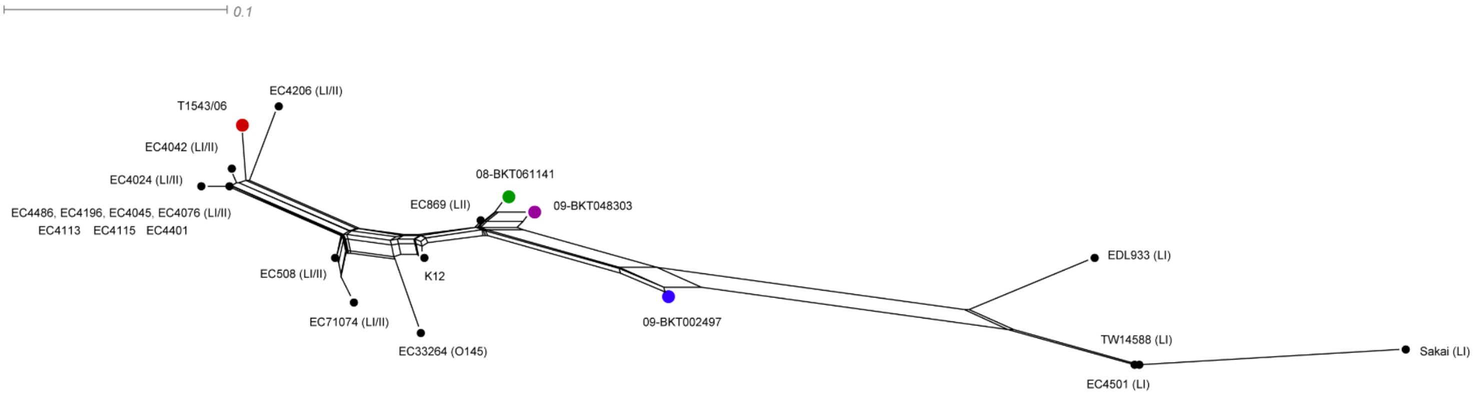
0.1
EC4206 (LI/II)
T1543/06
EC4042 (LI/II)
EC4024 (LI/II)
EC4486, EC4196, EC4045, EC4076 (LI/II)
EC4113 EC4115 EC4401
08-BKT061141
EC869 (LII)
09-BKT048303
EC508 (LI/II)
K12
EC71074 (LI/II)
EC33264 (O145)
09-BKT002497
EDL933 (LI)
TW14588 (LI)
Sakai (LI)
EC4501 (LI)

*Figure 8.* NeighborNet visualization created in SplitsTree 4.12 of SNP data from the four isolates analysed by NGS in paper IV. 92 SNP clade typing dataset (Manning *et al.*, 2008), excluding 4 positions for which the SNP contexts produced multiple or no hits in one or more isolate. Swedish strains in colour, reference data from (Laing *et al.*, 2009) in black. Swedish clade 8 (red) clusters with but is non-identical to several US clade 8 strains. The different Swedish clade 7 isolates (green, purple) are poorly separated by this SNP set.

## 4.2 Molecular markers for virulence among Swedish O157:H7

### 4.2.1 Verotoxin genes

*Verotoxin 2 subtypes*

In papers I-III, we found virtually only isolates that belonged to clade 8 to be $vtx_{2a}$-positive, frequently in combination with $vtx_{2c}$. Non-clade 8 phage type 2 isolates with $vtx_{2a}$ were found at a low frequency in the early slaughterhouse and dairy farm prevalence studies referred to in papers I-II and IV. Although phage typing is no longer used at the SVA, no isolates with LSPA-6 profiles consistent with this group (212113) have been found in any study after the year 2000. All other cattle and sheep isolates studied carried $vtx_{2c}$ and/or $vtx_1$.

The presence of $vtx_{2a}$ in high-virulence types of O157:H7 in Sweden is consistent with a mounting body of evidence that patients infected by a $vtx_{2a}$-carrying strain of O157:H7 have an increased risk of symptomatic illness and HUS (see section 1.2.1). Q gene variation has been described among O157:H7, with $vtx_{2c}$ genes in lineage II isolates more likely to carry an altered Q gene, and express lower levels of verotoxin (Zhang *et al.*, 2010; Ahmad & Zurek, 2006; Lejeune *et al.*, 2004).

*IS-element inactivation of verotoxin 2*

It has previously been suggested that IS-elements are likely to play an important role in VTEC evolution, and have a particular tendency to be associated with genetic changes in prophage-related sequence elements and plasmids in O157:H7 (Ooka *et al.*, 2009). The most common IS-element group in O157:H7 is synonymously referred to as IS*629* or as variants of IS*1203* (Ooka *et al.*, 2009). In this project, we have by chance found IS*629*/IS*1203*v elements in verotoxin encoding genes in a single isolate from cattle in paper II, an isolate from sheep in paper III, and several cattle isolates in paper V. The insertions found in papers II and III were in the sequence encoding the B-subunit, at different positions. The inserts observed in paper V were in the A subunit. Digital PCR confirmed that at least one of the insertion patterns could revert spontaneously as previously reported (Kusumoto *et al.*, 2001), while the "C" pattern in paper V involves a deletion and is likely to be irreversible. We note that all these changes occurred in isolates which also carried verotoxin 1 genes, perhaps suggesting redundancy or alternative verotoxin function plays a role in the persistence of these strains. IS-inactivated virulence factors can pose a challenge both for diagnostic tests and for the definitions of what we consider to be a VTEC, especially if the inactivation can be presumed to be reversible. In future prevalence studies, a more systematic investigation of the presence of IS-elements in O157:H7 virulence factor genes is advisable and can be facilitated by NGS characterization of isolates.

### 4.2.2 Other possible virulence-associated traits

While the LEE and effector gene profiles were shown to be remarkably similar between low- and high-virulence groups of Swedish O157:H7 in paper I, again differences in expression levels could be relevant, as clade 8 has been shown to express LEE genes and virulence plasmid genes at a higher level compared to isolates from lineage I (Abu-Ali *et al.*, 2010a; Abu-Ali *et al.*, 2010b). Several unique genetic traits in LII isolates have been identified, some of which could hypothetically improve fitness in the bovine host and/or reduce the ability to cause disease in the human host (Eppinger *et al.*, 2011a). A T255A base substitution in the translocated intimin receptor gene (*tir*) has been repeatedly shown to be unevenly distributed between cattle and human isolates of O157:H7, with the "T" form frequently occurring in both sources while the "A" form is mostly found in cattle isolates (Haugum *et al.*, 2011; Clawson *et al.*, 2009; Bono *et al.*, 2007). The "A" variant has been found to be associated with lineage II (Eppinger *et al.*, 2011a), and lineage II isolates have also been shown to have an extra copy of an 18bp tandem repeat in the *tir* receptor domain (Eppinger *et al.*, 2011a; Besser *et al.*, 2007). In paper IV, we use next generation sequencing to show the presence of the T255A as well as the 18bp insertion in the *tir* gene of the Swedish lineage II clone. It is unclear if these differences result in any significant change in Tir function or simply serve as clonal markers, but *tir* allele typing could in any case be useful as an identifier for these low-virulence strains.

### 4.2.3 Clinical relevance of clade 8

The suggestion that clade 8 was a more virulent subgroup within O157:H7 was first made based on a comparison of the rates of bloody diarrhoea, hospitalization and HUS in O157:H7 outbreaks in the USA and Japan. Clade 8 was found to be associated with HUS and infections in younger patients, while clade 7 was associated with the absence of bloody diarrhoea or HUS. Clade 8 was also found to be significantly associated with the presence of both $vtx_{2a}$ and

$vtx_{2c}$. (Manning *et al.*, 2008) A significant association between clade 8 and HUS in young patients was confirmed in a much larger study in Japan comparing 269 HUS cases and 387 asymptomatic carriers. Clade 6 was also found to be associated with HUS, with both clades carrying $stx_{2a}$ alone or in combination with $stx_{2c}$ in most but not all HUS cases. Clade 7 was significantly associated with asymptomatic carriership. However, the majority of HUS cases in the study were caused by lineage I clades 1,2,3, generally with $stx_{1a}$ and $stx_{2a}$. (Iyoda *et al.*, 2014) In Argentina, the childhood HUS rate is 10-fold higher than other industrialized countries, with O157:H7 as the primary cause, and it has been suggested that this is due to the predominance of $vtx_{2a}$-positive clade 8 in the Argentinian cattle reservoir (Amigo *et al.*, 2015; Mellor *et al.*, 2012; Leotta *et al.*, 2008). It should be noted that the uneven distribution of O157:H7 subtypes among cattle and patients in different countries means that clade 8 is being compared to different "baselines" in these studies compared to one another and compared to studies performed by us on Swedish isolates.

Our study presented in paper IV included 11 O157:H7 HUS cases reported during the study period, 10 of which were caused by clade 8. Clade 8 was significantly overrepresented among all human cases of O157:H7 compared to the cattle reservoir, and human infections with clade 8 were significantly more likely to progress to HUS. In the independent data set presented in paper II, clade 8 was significantly overrepresented among isolates from farms linked to human cases of disease compared to isolates from a dairy farm prevalence study. As further evidence of the clinical relevance of clade 8, this subgroup has been responsible for most of the major outbreaks of O157:H7 in Sweden. Since 2012, clade 8 has a unique status among VTEC in Sweden in that it is notifiable to the Swedish Board of Agriculture when found in animals, regardless of proven links to human cases of illness from that particular source (SJVFS 2012:24). With the real-time PCR protocol presented in paper IV, clade 8 isolates of O157:H7 can reliably be distinguished from non-clade 8 isolates with a minimum of effort and in less than an hour.

## 4.3 Sheep as a reservoir for O157:H7

The potential of small ruminants, primarily sheep, as a reservoir for O157:H7 is of interest both as a direct public health concern and as a source of infection for cattle. In addition to the production of meat and dairy products, grazing sheep in Sweden are used to maintain the landscape character of rural and peri-urban recreational and nature preserve areas, in close contact with the public. There is a growing interest in artisanal cheesemaking using sheep milk. While the Swedish cattle population is declining, the number of sheep increased by 47% between 1980 and 2013 (Anonymous, 2014b).

In paper III, we estimate the O157:H7 prevalence to 1.8% in Swedish sheep at slaughter. This is substantially less than the approximately 3% found in Swedish cattle, however the low number of total and positive samples means that the confidence interval is wide, and overlaps that of studies performed on cattle during the same period (see comparison in Table 1). The overall number and density of sheep is lower than cattle in Sweden, with the exception of the island of Gotland (see Fig 1). In addition, sheep herds tend to consist of a lower number of animals compared to cattle herds (Anonymous, 2014b). Herd size has been shown to be significant contributor to the risk of O157:H7 being found on a Swedish cattle farm (Widgren *et al.*, 2015; Eriksson *et al.*, 2005), and the less intensive management practices could be a contributing factor to the lower prevalence estimated among Swedish sheep.

All four major genetic groups of O157:H7 found in Swedish cattle during the project period (i.e. LSPA-6 233323, 213111 and 211111 clade 8 / non-clade 8) were also recovered from sheep in paper III, proving that sheep form a reservoir for infection of both cattle and humans and vice versa (Fig. 5). The number of isolates recovered in this study was too low to make a quantitative comparison of clonal representation between sheep and cattle meaningful.

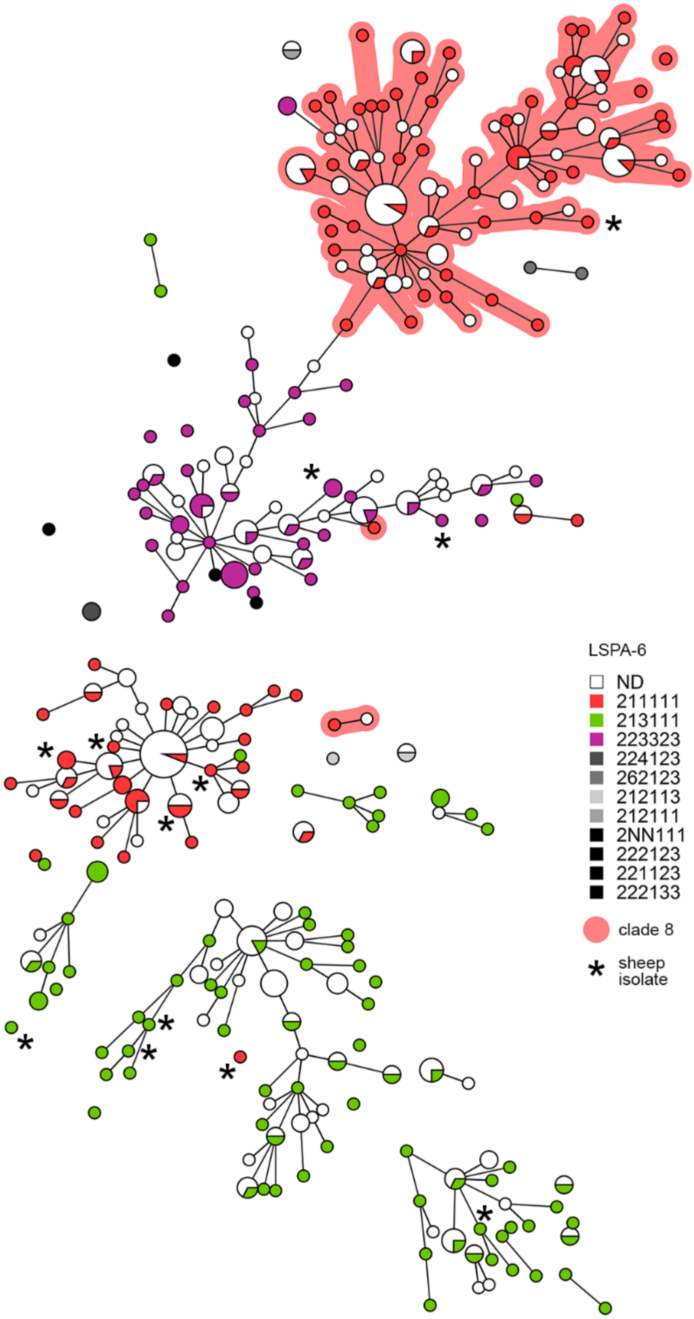
LSPA-6
ND
211111
213111
223323
224123
262123
212113
212111
2NN111
222123
221123
222133
clade 8
sheep isolate
*

*Figure 9.* MLVA minimum spanning tree of isolates of O157:H7 from Swedish cattle and sheep, 1996-2012 (n=470), merging the data from papers III, IV and V. This shows the high correlation between MLVA clusters and LSPA-6 profiles, as well as the close association between cattle isolates and isolates from sheep (marked with asterisks). Created in Bionumerics 7.5.

## 4.4 Spatio-temporal dynamics

### 4.4.1 Geographical distribution

All studies performed on O157:H7 in papers I-V consistently show that O157:H7 is extremely rare in the part of Sweden north of Gävleborg county, where cattle density is low (Fig. 6). The same phenomenon has been described in Finland, where cattle density is also low in the north (Lahti *et al.*, 2001). The importance of other infected farms nearby for the risk of O157:H7 being found in a Swedish farm was shown in a recent study (Widgren *et al.*, 2015). A tendency for the same strains to be found in multiple neighbouring farms over time was also shown in the same study, as has been shown in Scotland (Herbert *et al.*, 2014). It is conceivable that a certain "critical mass" of ruminant populations in an area is necessary for O157:H7 to establish a permanent presence.

Clade 8, a strain of which was then known as the "Halland type", was initially found in Swedish cattle in 1996 in the south-western county of Halland, where early prevalence studies found this type to be the locally dominant (Aspan & Eriksson, 2010; Eriksson, 2010). Multiple clade 8 outbreaks occurred in Halland, e.g. 2005 and 2008-2009 (see section 1.4.3), and as evident in paper I-II, several sporadic cases of clade 8 infection linked to local farms occurred. Halland remained one of the counties with the highest reported O157:H7 incidence in the country for several years (Eriksson, 2010). Clade 8 isolates have been found in cattle in Halland in the more recent prevalence studies as well, but in low numbers. In contrast, clade 8 was extremely rare on the south-eastern coast in the prevalence studies performed in the late 90's but appears to have increased dramatically in recent years (i.e. in the sample sets analysed in papers IV and V), forming a hot-spot area including the island of Öland and the nearby mainland around Kalmar (Fig 10). Further striking regional differences can be found in the data from papers III and IV, for instance clade 8 has not been found in any animal from the island of Gotland, where both cattle and sheep densities are high (Fig. 1), since 2002. This marked regionality of clones suggests local interventions could be useful to reduce the prevalence of high-virulence subtypes such as clade 8.

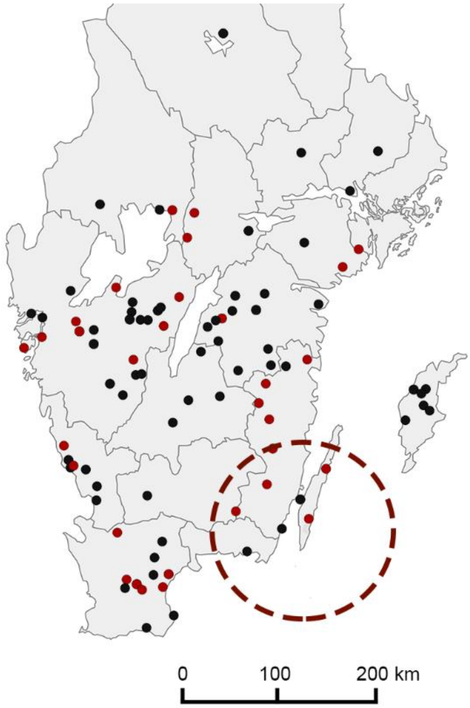

*Figure 10.* Isolates of O157:H7 clade 8 (red) and non-clade 8 (black) from slaughterhouse prevalence studies 2005-2009 (paper IV, n=222), mapped using the farm of origin postal code. Each marker can represent several isolates found in the same postal code area. A single significant cluster of clade 8 isolates (p=0.02) was identified by spatial scan analysis using the Bernoulli model (Kulldorf, 1997) in SaTScan 9.1, with both high and low rate areas considered. This cluster is located on Öland and the adjacent mainland region around Kalmar. Map from Statistics Sweden (scb.se) Created in QGIS 1.7.

### 4.4.2 Changes in the clonal composition of the O157:H7 population over time

The dataset in paper IV is unique in providing repeated, representative sampling of the O157:H7 flora in cattle on a nationwide scale. The data goes back to 1996 when O157:H7 was first being brought to attention in Sweden. We note that the currently most common type of O157:H7 is a lineage II clone with the LSPA-6 profile 223323 (Fig. 10), and that this type very rarely causes severe disease in humans. This clone was completely absent in the slaughterhouse prevalence studies 1996-1998, and very rare before 2005 (Fig. 10). It was also absent in the dairy farm prevalence study 1998-2000 providing the sample material for papers I-II, but was found in several sheep samples in 2008-2009 in paper III, consistent with a recent expansion. An alternative explanation for the apparent expansion of LII (223323) in the cattle reservoir could be that this clone is more difficult to isolate from samples, e.g. due to inducing lower shedding levels in infected animals, and that the difference is due to improved isolation methodology over the years. However, the expansion has continued well after the introduction of improved enrichment protocols. Additionally, while the increase appears to have slowed between 2008-2009 and 2011-2012, preliminary results from the 2014-2015 prevalence study indicates that 223323 and closely associated profiles have become even more dominant (Dresch, 2015). Several other lineage II LSPA-6 profiles have occurred at low levels in the prevalence studies over the years, but appear to be transient or remain at low levels as they usually are not found again. It is tempting to speculate that the 223323 clone has some form of selective advantage over other types of O157:H7 in Sweden. Since human infection by O157:H7 is generally considered to be incidental, advantages in ruminant hosts or in the form of e.g. ability to survive in the farm environment would provide a selective advantage, whereas the ability to cause severe disease in an infected human patient is of little or no selective advantage.

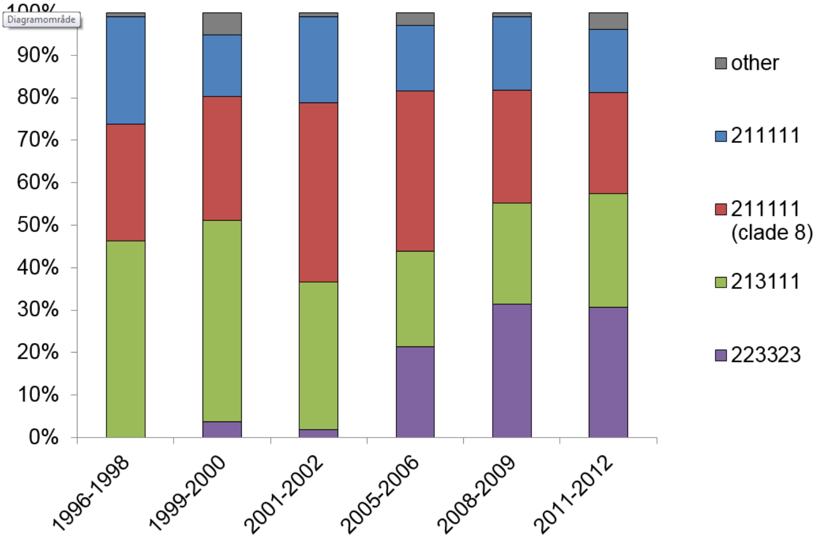

*Figure 11.* Comparison of the composition in terms of LSPA-6 profile and clade 8/non-clade 8 status of O157:H7 from cattle slaughterhouse prevalence studies 1996-2012.

It would be reasonable to expect an increasing proportion of low-virulence O157:H7 in a country to lead to a lower incidence of human infection. However, no strong such trend has been observed in Sweden (Anonymous, 2014a). Comparing human incidence over time is complicated by constant improvements in diagnostic methodology and reporting, which is likely to have reduced the number of missed cases over time (Anonymous, 2014a). Awareness of VTEC was particularly raised by the very large 2011 outbreak of O104:H4 in several European countries including Sweden (Karch *et al.*, 2012).

# 5 Future perspectives

The high prevalence of O157:H7 among ruminants in southern Sweden makes it necessary for any realistic interventions against this VTEC serotype to be targeted. Work in this thesis and elsewhere has identified certain Swedish O157:H7, in particular clade 8, to currently be the primary threat of severe complications in infected humans, and identified certain areas in the country where this subgroup is common. Several projects are currently underway and being planned targeting the clade 8 hotspot in south-eastern Sweden in particular, both to increase our understanding of local infection dynamics and to provide actionable advice to farmers and the general public. Continued surveillance is warranted to monitor Swedish ruminant populations for geographical expansion of high-virulence types of O157:H7 like clade 8, or introduction of new types into the country. As the verotoxin 2a variant seems to be present in most high-virulence VTEC and is currently rare among Swedish ruminant O157:H7, this could be a marker for such introductions. Genotypic and phenotypic characterization of a broader range of subgroups that are known as more or less likely to cause severe complications in humans will be necessary to improve our understanding of the mechanisms and genetic factors involved in VTEC O157:H7 pathogenesis.

With decreasing cost, simplified workflows and improved availability, next generation sequencing is increasingly adopted as a routine tool for molecular epidemiology investigation of bacterial infections. With whole-genome SNP analysis, high typing resolution with less reversible marker variation is achievable, making this technique an attractive alternative to PFGE and MLVA. NGS can also provide more detailed virulence gene typing compared to the methods used in the earlier parts of this project. NGS analysis of the isolates found in the 2014-2015 O157:H7 nationwide slaughterhouse prevalence study is ongoing, and the resulting data will be compared to isolates from human patients and imported foodstuffs in collaboration with the other Swedish national authorities. For historical comparison, clade and lineage typing data (e.g. LSPA-6 profiles) can be extracted from the NGS output. Sequencing of older isolates, e.g. from paper IV, could provide clues to when and how the types of O157:H7 currently present in Sweden were introduced in the country.

Improved methods for standardization, interpretation and communication of molecular typing data is another key area for future O157:H7 research. For example, setting the threshold for what is to be considered a match in an epidemiological investigation can be challenging, and will become increasingly so as methods improve making exact matches less likely. This means that the level of genetic variation normally found within a suspected source of infection, e.g. a clone of O157:H7 within a region, a cattle herd or even an individual animal should be studied more extensively.

The ultimate goal of all research on O157:H7 is a reduction of the number of severe cases of human infection. This is likely to require a combination of both reduced prevalence in ruminants and limiting the routes of transmission between ruminants and humans. As VTEC is both a zoonosis and a foodborne disease, both research and control requires a continued close collaboration between the research community, food producers, clinicians, and the veterinary, public health and food safety authorities. While other VTEC serotypes are on the rise in terms of diagnosed cases of human illness and should not be ignored, O157:H7 remains the most common cause of severe VTEC infection and will be a focus research area in the coming years.

# 6 Populärvetenskaplig sammanfattning

*Escherichia coli* är en bakterieart som utgör en ofarlig och vanlig del av människors tarmflora. Vissa grupper av *E. coli*-bakterier har dock kombinationer av genetiska virulensfaktorer som ger dem förmåga att orsaka sjukdom hos infekterade människor. Virulensfaktorerna kan t.ex. ge bakterien förmåga att bilda toxiner eller att fästa sig vid tarmväggen och skada den. Den typ av *E. coli*-bakterie som ger de allvarligaste mag-tarmsjukdomarna hos människa kallas enterohemorragisk *E. coli* (EHEC) eftersom den kan orsaka blödningar i tarmen. Bakterierna kallas även verotoxinbildande *E. coli* (VTEC) eftersom deras huvudsakliga virulensfaktor är förmågan att bilda celldödande verotoxin. VTEC lever normalt i tarmen på idisslare som nötkreatur och får, som inte blir sjuka av dem. Människor smittas genom att de får i sig avföring från smittade idisslare. Detta kan ske via kontaminerade livsmedel som bladgrönsaker, dåligt genomstekt köttfärs eller opastöriserade mjölkprodukter. Man kan även smittas via direkt kontakt med smittade djur eller indirekt via deras omgivning, t.ex. beteshagar eller badplatser som ligger vid betesmark. Infektion med VTEC kan ge mild diarré eller svårare blodig diarré. En mindre andel smittade människor, framförallt barn, utvecklar livshotande komplikationer som hemolytiskt uremiskt syndrom (HUS) som kan orsaka livslånga skador eller dödsfall.

VTEC-bakterier delas in i serotyper, där den som oftast gör människor svårt sjuka i Sverige och många andra länder kallas O157:H7. Infektion med O157:H7 uppmärksammades först i Sverige i samband med stora utbrott under 90-talet. Sedan dess har flera utbrott skett, och ungefär 20-100 människor rapporteras smittas inom Sverige årligen. Att få ner antalet människor som smittas med O157:H7 kräver åtgärder riktade mot bakterien i värddjuren, d.v.s. nötkreatur och får, men tidigare studier har visat att så många som ungefär 3% av svenska nötkreatur som slaktas bär O157:H7. I en studie som följde nötgårdar i den södra delen av landet med upprepad provtagning under 38 månader hittades O157:H7 minst en gång på 53% av gårdarna. Det vore alltså mycket resurskrävande att bekämpa förekomsten av all O157:H7 i Sverige. Samtidigt har det tidigare observerats att en undergrupp av O157:H7 som kallats ”Hallandstypen” står för en oproportionerligt stor andel av sjukdomsfallen och utbrotten bland människor. Med bättre förmåga att identifiera de farligaste typerna av O157:H7 bland får och nötkreatur skulle alltså åtgärder kunna riktas mot dem och bli mer kostnadseffektiva.

I det här projektet har vi jämfört VTEC O157:H7 som hittats bland svenska nötkreatur med O157:H7 som orsakat sjukdom hos människor. Den typen av jämförelser ger oss möjlighet att identifiera kännetecken hos de farligare formerna av O157:H7. Vi har även jämfört O157:H7 från får och nötkreatur för att se om smitta sker mellan olika djurslag och om får utgör en smittkälla för människor. Samtidigt utvecklade och utvärderade vi snabbare metoder för karaktärisering och smittspårning av O157:H7.

I den första artikeln i avhandlingen jämförde vi uppsättningar av virulensfaktorer bland svenska O157:H7. Vi kunde konstatera att både de bakteriestammar som orsakat sjukdom hos människa och de som inte gjort det hade gener för alla de virulensfaktorer som vanligtvis förknippas med O157:H7. Att en del stammar oftare orsakar sjukdom måste alltså bero på mer subtila genetiska skillnader. Vi såg t.ex. att ”Hallandstypen” ofta hade vissa varianter av verotoxingener, vilket eventuellt skulle kunna ge dem förmåga att producera mer toxin och därmed orsaka svårare sjukdom.

I den andra artikeln jämförde vi samma bakteriestammar genom att titta på släktskapsmarkörer istället för direkta virulensfaktorer. Vi kunde visa att ”Hallandstypen” tillhör en grupp av O157:H7 som är internationellt känd som ”klad 8” och har rapporterats ge hög risk för allvarlig sjukdom i flera andra länder. Med denna kunskap kunde vi snabba upp analysen för att identifiera dessa farligare stammar från dagar till timmar. Vi utvärderade även ett system för släktskapsanalys som kallas ”MLVA”, användbart för t.ex. smittspårning, och kunde visa att det hade likvärdiga prestanda som tidigare använda metoder samtidigt som det var snabbare och billigare.

I artikel tre undersökte vi hur vanlig O157:H7 är hos får, och jämförde stammar från får, nötkreatur och människor. Ungefär 1.8% av får som slaktades i Sverige bar O157:H7, och vi kunde visa att samma typer av O157:H7 cirkulerar mellan nötkreatur och får. Vi kunde även knyta stammar av O157:H7 från får till fall av sjukdom hos människor. Klad 8 visade sig även finnas hos får.

Med den förbättrade metodiken för att kunna klassificera svensk O157:H7 från de första tre artiklarna analyserade vi i artikel fyra om ett stort antal stammar från nötkreatur från nationella studier mellan 1996 och 2008. Vi analyserade även alla stammar från svenska patienter 2009-2011 som sparats av Folkhälsomyndigheten. Resultaten bekräftade att klad 8 oftare orsakar sjukdom bland människor, och även att klad 8 orsakar de flesta fallen av HUS. Som kontrast orsakar den vanligaste typen av O157:H7 bland svenska nötkreatur extremt sällan sjukdom hos människor. Vi såg även att det sker dramatiska förändringar över tid vad gäller hur vanliga de olika typerna av O157:H7 är bland nötkreatur.

I det sista manuskriptet i projektet tittade vi närmare på verotoxingenerna i stammar av O157:H7 som hittats i svenska nötkreatur 2011-2012. Vi upptäckte att vissa typer av O157:H7 ofta bär på verotoxingener som inaktiverats av att rörliga DNA-element satt sig mitt i själva generna, vilket med stor sannolikhet gör stammarna mindre farliga för människor. En mer detaljerad analys av individuella bakterier visade dock att DNA-elementen kunde hoppa ut igen av sig själva och lämna en intakt toxin-gen efter sig. Denna typ av komplex variation försvårar diagnostik av VTEC eftersom den ofta bygger på närvaron av oförändrade verotoxingener.

Sammantaget har projektet identifierat undergrupper av O157:H7 bland får och nötkreatur, där en del utgör ett mycket stort hot mot människor medan andra utgör ett begränsat hot. Vi presenterar även laboratoriemetodik för att snabbt och enkelt åtskilja dessa grupper. Detta ger möjlighet till riktade åtgärder som är mer effektiva än breda åtgärder mot alla O157:H7. Eftersom vi kunnat konstatera att nya typer av O157:H7 kan etablera sig i landet eller snabbt expandera från låga nivåer behövs fortsatt övervakning och ett nära samarbete mellan producenter, forskare och myndigheter på regional, nationell och internationell nivå.

# Acknowledgements

This thesis project was carried out at the Department of Bacteriology at the **National Veterinary Institute (SVA)**, Uppsala, Sweden, and at the Department of Animal Breeding and Genetics, **Swedish University of Agricultural Sciences (SLU)**, Uppsala, Sweden. The project was generously supported by grants from the **Swedish Board of Agriculture**, the **Swedish Civil Contingencies Agency**, the **Ivar and Elsa Sandberg Foundation** and the **SVA Research Fund**.

Scientific research is a team sport, and too many friends and colleagues have helped out with this project for all to be mentioned here. I would nonetheless especially like to thank:

My wonderful supervisors: **Erik Bongcam-Rudloff** for introducing me to academia, bioinformatics and surströmming, **Anna Aspán** for her patience and wisdom, and **Erik Eriksson** for his tireless effort in organizing studies and for sharing his encyclopedic knowledge of most things VTEC.

My equally wonderful co-authors **Roberto La Ragione, Sofia Boqvist, Anna Nilsson, Cecilia Jernberg, Sofie Ivarsson, Ingela Hedenström, Heiður Loftsdóttir** and **Tomas Jinnerot** for fruitful and enjoyable collaborations.

The staff at the SVA EHEC and PCR laboratories and the bacteriology R&D section, who have taught me much and provided invaluable support during the project, and in particular **Ulla Johansson**, **Griselda Loreto-Palma, Marianne Persson, Bo Segerman, Joann Börjesson, Christine Källman** and **Sara Åkerström**.

Various fellow VTEC researchers: R and statistics go-to guy **Stefan Widgren**, **Eelco Franz** and his colleagues at the RIVM for arranging a great O157:H7 short term scientific mission for me in Utrecht, and **Geraldine Smith, Eija Hyytiä-Trees** and **Shannon Manning** for providing advice, protocols and reference data for the molecular assays used in this project.

Our diligent **project students**: I will particularly acknowledge the contributions of **Erica Dresch, Adrien Janssens**, and **Julie Hurel** to the continued improvement of our VTEC typing methodology during the production of this thesis.

My **fellow PhD students and everyone else at SGBC** for sharing their bioinformatics wisdom and joining me on a worldwide quest to find 1€ beer while keeping leeches out of the socks, and why not **Oskar Karlsson** for helping me pass that wisdom on to the next generation of bioinformaticians.

The mostly 30-something **kids at the bacteriology department**, past and present, for creating an inspiring work environment, with extra bonus credit to my roomies **Joakim Ågren**, **Hanna Skarin** and **Tomas Jinnerot** for enduring my taste in music, helping me pick Christmas gifts and taking care of my plants, and local renaissance man **Anders Sundström** for making sure this book can safely be judged by its cover.

# References

Abu-Ali, G.S., Ouellette, L.M., Henderson, S.T., Lacher, D.W., Riordan, J.T., Whittam, T.S. & Manning, S.D. (2010a). Increased adherence and expression of virulence genes in a lineage of Escherichia coli O157:H7 commonly associated with human infections. *PLoS One,* 5(4), p. e10167.

Abu-Ali, G.S., Ouellette, L.M., Henderson, S.T., Whittam, T.S. & Manning, S.D. (2010b). Differences in adherence and virulence gene expression between two outbreak strains of enterohaemorrhagic Escherichia coli O157 : H7. *Microbiology,* 156(Pt 2), pp. 408-19.

Agresti, A. & Coull, B.A. (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. *The American Statistician,* 52, pp. 119-126.

Ahmad, A. & Zurek, L. (2006). Evaluation of the anti-terminator Q933 gene as a marker for Escherichia coli O157:H7 with high Shiga toxin production. *Curr Microbiol,* 53(4), pp. 324-8.

Ahmed, R., Bopp, C., Borczyk, A. & Kasatiya, S. (1987). Phage-typing scheme for Escherichia coli O157:H7. *J Infect Dis,* 155(4), pp. 806-9.

Ahn, C.K., Russo, A.J., Howell, K.R., Holt, N.J., Sellenriek, P.L., Rothbaum, R.J., Beck, A.M., Luebbering, L.J. & Tarr, P.I. (2009). Deer sausage: a newly identified vehicle of transmission of Escherichia coli O157:H7. *J Pediatr,* 155(4), pp. 587-9.

Alam, M.J. & Zurek, L. (2004). Association of Escherichia coli O157:H7 with houseflies on a cattle farm. *Appl Environ Microbiol,* 70(12), pp. 7578-80.

Albihn, A., Eriksson, E., Wallen, C. & Aspan, A. (2003). Verotoxinogenic Escherichia coli (VTEC) O157:H7--a nationwide Swedish survey of bovine faeces. *Acta Vet Scand,* 44(1-2), pp. 43-52.

Amigo, N., Mercado, E., Bentancor, A., Singh, P., Vilte, D., Gerhardt, E., Zotta, E., Ibarra, C., Manning, S.D., Larzabal, M. & Cataldi, A. (2015). Clade 8 and Clade 6 Strains of Escherichia coli O157:H7 from Cattle in Argentina have Hypervirulent-Like Phenotypes. *PLoS One,* 10(6), p. e0127710.

Anjum, M.F., Mafura, M., Slickers, P., Ballmer, K., Kuhnert, P., Woodward, M.J. & Ehricht, R. (2007). Pathotyping Escherichia coli by using miniaturized DNA microarrays. *Appl Environ Microbiol,* 73(17), pp. 5692-7.

Anonymous (2013). *Standard Operating Procedure for PulseNet PFGE of Escherichia coli O157:H7, Escherichia coli non-O157 (STEC), Salmonella serotypes, Shigella sonnei and Shigella flexneri*: Centers for Disease Control and Prevention (CDC).

Anonymous (2014a). *Infektion med EHEC/VTEC - Ett nationellt strategidokument*.

Anonymous (2014b). *Jordbruksstatistisk årsbok 2014*.

Aspan, A. & Eriksson, E. (2010). Verotoxigenic Escherichia coli O157:H7 from Swedish cattle; isolates from prevalence studies versus strains linked to human infections--a retrospective study. *BMC Vet Res,* 6, p. 7.

Baker, M. (2012). Digital PCR hits its stride. *Nature Methods,* 9, pp. 541-44.

Barkocy-Gallagher, G.A., Arthur, T.M., Rivera-Betancourt, M., Nou, X., Shackelford, S.D., Wheeler, T.L. & Koohmaraie, M. (2003). Seasonal prevalence of Shiga toxin-producing Escherichia coli, including O157:H7 and non-O157 serotypes, and Salmonella in commercial beef processing plants. *J Food Prot,* 66(11), pp. 1978-86.

Bell, B.P., Goldoft, M., Griffin, P.M., Davis, M.A., Gordon, D.C., Tarr, P.I., Bartleson, C.A., Lewis, J.H., Barrett, T.J., Wells, J.G. & et al. (1994). A multistate outbreak of Escherichia coli O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. *JAMA,* 272(17), pp. 1349-53.

Besser, T.E., Shaikh, N., Holt, N.J., Tarr, P.I., Konkel, M.E., Malik-Kale, P., Walsh, C.W., Whittam, T.S. & Bono, J.L. (2007). Greater diversity of Shiga toxin-encoding bacteriophage insertion sites among Escherichia coli O157:H7 isolates from cattle than in those from humans. *Appl Environ Microbiol,* 73(3), pp. 671-9.

Bielaszewska, M., Aldick, T., Bauwens, A. & Karch, H. (2014). Hemolysin of enterohemorrhagic Escherichia coli: structure, transport, biological activity and putative role in virulence. *Int J Med Microbiol,* 304(5-6), pp. 521-9.

Bielaszewska, M., Prager, R., Kock, R., Mellmann, A., Zhang, W., Tschape, H., Tarr, P.I. & Karch, H. (2007). Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic Escherichia coli O26 infection in humans. *Appl Environ Microbiol,* 73(10), pp. 3144-50.

Bilge, S.S., Vary, J.C., Jr., Dowell, S.F. & Tarr, P.I. (1996). Role of the Escherichia coli O157:H7 O side chain in adherence and analysis of an rfb locus. *Infect Immun,* 64(11), pp. 4795-801.

BIOHAZ (2013). Scientific Opinion on VTEC - seropathotype and scientific criteria regarding pathogenicity assessment. *EFSA Journal,* 11(4), p. 3138.

Blanco Crivelli, X., Rumi, M.V., Carfagnini, J.C., Degregorio, O. & Bentancor, A.B. (2012). Synanthropic rodents as possible reservoirs of shigatoxigenic Escherichia coli strains. *Front Cell Infect Microbiol,* 2, p. 134.

Bono, J.L., Keen, J.E., Clawson, M.L., Durso, L.M., Heaton, M.P. & Laegreid, W.W. (2007). Association of Escherichia coli O157:H7 tir polymorphisms with human infection. *BMC Infect Dis,* 7, p. 98.

Boqvist, S., Aspan, A. & Eriksson, E. (2009). Prevalence of verotoxigenic Escherichia coli O157:H7 in fecal and ear samples from slaughtered cattle in Sweden. *J Food Prot,* 72(8), pp. 1709-12.

Boyd, E.F., Carpenter, M.R. & Chowdhury, N. (2012). Mobile effector proteins on phage genomes. *Bacteriophage,* 2(3), pp. 139-148.

Brandal, L.T., Wester, A.L., Lange, H., Lobersli, I., Lindstedt, B.A., Vold, L. & Kapperud, G. (2015). Shiga toxin-producing escherichia coli infections in Norway, 1992-2012: characterization of isolates and identification of risk factors for haemolytic uremic syndrome. *BMC Infect Dis,* 15(1), p. 324.

Brown, C.A., Harmon, B.G., Zhao, T. & Doyle, M.P. (1997). Experimental Escherichia coli O157:H7 carriage in calves. *Appl Environ Microbiol,* 63(1), pp. 27-32.

Brown, L.D., Cai, T.T. & DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science,* 16(2), pp. 101-117.

Bryant, D. & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol,* 21(2), pp. 255-65.

Burland, V., Shao, Y., Perna, N.T., Plunkett, G., Sofia, H.J. & Blattner, F.R. (1998). The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli O157:H7. *Nucleic Acids Res,* 26(18), pp. 4196-204.

Bustamante, A.V., Sanso, A.M., Segura, D.O., Parma, A.E. & Lucchesi, P.M. (2013). Dynamic of mutational events in variable number tandem repeats of Escherichia coli O157:H7. *Biomed Res Int,* 2013, p. 390354.

Caprioli, A., Maugliani, A., Michelacci, V. & Morabito, S. (2014). *Molecular typing of Verocytotoxin-producing E. coli (VTEC) strains isolated from food, feed and animals: state of play and standard operating procedures for pulsed field gel electrophoresis (PFGE) typing, profiles interpretation and curation*: European Food Safety Authority.

Caprioli, A., Morabito, S., Brugere, H. & Oswald, E. (2005). Enterohaemorrhagic Escherichia coli: emerging issues on virulence and modes of transmission. *Vet Res,* 36(3), pp. 289-311.

CDC (2005). Outbreaks of Escherichia coli O157:H7 Associated with Petting Zoos - North Carolina, Florida, and Arizona, 2004 and 2005. *Morbidity and Mortality Weekly Report,* 54(50), pp. 1277-1280.

Chandler, M. & Mahillon, J. (2002). *Mobile DNA II*: ASM Press.

Chase-Topping, M., Gally, D., Low, C., Matthews, L. & Woolhouse, M. (2008). Super-shedding and the link between human infection and livestock carriage of Escherichia coli O157. *Nat Rev Microbiol,* 6(12), pp. 904-12.

Clawson, M.L., Keen, J.E., Smith, T.P., Durso, L.M., McDaneld, T.G., Mandrell, R.E., Davis, M.A. & Bono, J.L. (2009). Phylogenetic classification of Escherichia coli O157:H7 strains of human and bovine origin using a novel set of nucleotide polymorphisms. *Genome Biol,* 10(5), p. R56.

Craig, N.L., Craigie, R., Gellert, M. & Lambowitz, A.M. (2002). *Mobile DNA II*: ASM Press.

Crampin, M., Willshaw, G., Hancock, R., Djuretic, T., Elstob, C., Rouse, A., Cheasty, T. & Stuart, J. (1999). Outbreak of Escherichia coli O157 infection associated with a music festival. *Eur J Clin Microbiol Infect Dis,* 18(4), pp. 286-8.

Dallman, T.J., Byrne, L., Ashton, P.M., Cowley, L.A., Perry, N.T., Adak, G., Petrovska, L., Ellis, R.J., Elson, R., Underwood, A., Green, J., Hanage, W.P., Jenkins, C., Grant, K. & Wain, J. (2015). Whole-Genome Sequencing for National Surveillance of Shiga Toxin-Producing Escherichia coli O157. *Clin Infect Dis,* 61(3), pp. 305-12.

Dallman, T., Ashton, P., Byrne, L., Perry, N., Petrovska, L., Ellis, R., Allison, L., Hanson, M., Holmes, A., Gunn, G., Chase-Topping, M., Woolhouse, M., Grant, K., Gally, D., Wain, J. & Jenkins, C. (2015). Applying phylogenomics to understand the emergence of Shiga-toxin-producing Escherichia coli O157:H7 strains causing severe human disease in the UK *Microbial Genomics*, 1.

Dean-Nystrom, E.A., Bosworth, B.T. & Moon, H.W. (1997). Pathogenesis of O157:H7 Escherichia coli infection in neonatal calves. *Adv Exp Med Biol,* 412, pp. 47-51.

Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E. & Crook, D.W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet,* 13(9), pp. 601-12.

Diniz-Filho, J.A., Santos, T., Rangel, T.F. & Bini, L.M. (2012). A comparison of metrics for estimating phylogenetic signal under alternative evolutionary models. *Genet Mol Biol,* 35(3), pp. 673-9.

Dowd, S.E. & Williams, J.B. (2008). Comparison of Shiga-like toxin II expression between two genetically diverse lineages of Escherichia coli O157:H7. *J Food Prot,* 71(8), pp. 1673-8.

Dresch, E. (2015). *Genetic characterization and disinfectant susceptibility test of VTEC O157:H7 isolated from Swedish cattle.* Diss.: Uppsala University.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E. & Relman, D.A. (2005). Diversity of the human intestinal microbial flora. *Science,* 308(5728), pp. 1635-8.

Eppinger, M., Mammel, M.K., Leclerc, J.E., Ravel, J. & Cebula, T.A. (2011a). Genome signatures of Escherichia coli O157:H7 isolates from the bovine host reservoir. *Appl Environ Microbiol,* 77(9), pp. 2916-25.

Eppinger, M., Mammel, M.K., Leclerc, J.E., Ravel, J. & Cebula, T.A. (2011b). Genomic anatomy of Escherichia coli O157:H7 outbreaks. *Proc Natl Acad Sci U S A,* 108(50), pp. 20142-7.

Eriksson, E. (2010). Verotoxinogenic Escherichia coli O157:H7 in Swedish cattle and pigs. *Acta Universitatis Agriculturae Sueciae,* Doctoral Thesis No. 2010:3.

Eriksson, E., Aspan, A., Gunnarsson, A. & Vagsholm, I. (2005). Prevalence of verotoxin-producing Escherichia coli (VTEC) 0157 in Swedish dairy herds. *Epidemiol Infect,* 133(2), pp. 349-58.

Eriksson, E., Nerbrink, E., Borch, E., Aspan, A. & Gunnarsson, A. (2003). Verocytotoxin-producing Escherichia coli O157:H7 in the Swedish pig population. *Vet Rec,* 152(23), pp. 712-7.

Erntell, M. & Jönsson, B. (2010). *Slutrapport: Utredning avseende EHEC-fall kopplade till produkter från Mostorps gård.* : Smittskydd Halland.

Espy, M.J., Uhl, J.R., Sloan, L.M., Buckwalter, S.P., Jones, M.F., Vetter, E.A., Yao, J.D., Wengenack, N.L., Rosenblatt, J.E., Cockerill, F.R., 3rd & Smith, T.F. (2006). Real-time PCR in clinical microbiology: applications for routine laboratory testing. *Clin Microbiol Rev,* 19(1), pp. 165-256.

Farfan, M.J. & Torres, A.G. (2012). Molecular mechanisms that mediate colonization of Shiga toxin-producing Escherichia coli strains. *Infect Immun,* 80(3), pp. 903-13.

Feng, P., Lampel, K.A., Karch, H. & Whittam, T.S. (1998). Genotypic and phenotypic changes in the emergence of Escherichia coli O157:H7. *J Infect Dis,* 177(6), pp. 1750-3.

Ferens, W.A. & Hovde, C.J. (2011). Escherichia coli O157:H7: animal reservoir and sources of human infection. *Foodborne Pathog Dis,* 8(4), pp. 465-87.

Fischer, J.R., Zhao, T., Doyle, M.P., Goldberg, M.R., Brown, C.A., Sewell, C.T., Kavanaugh, D.M. & Bauman, C.D. (2001). Experimental and field studies of Escherichia coli O157:H7 in white-tailed deer. *Appl Environ Microbiol,* 67(3), pp. 1218-24.

Frankel, G. & Phillips, A.D. (2008). Attaching effacing Escherichia coli and paradigms of Tir-triggered actin polymerization: getting off the pedestal. *Cell Microbiol,* 10(3), pp. 549-56.

Friedrich, A.W., Bielaszewska, M., Zhang, W.L., 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(1), pp. 74-84.

Galiero, G., Conedera, G., Alfano, D. & Caprioli, A. (2005). Isolation of verocytotoxin-producing Escherichia coli O157 from water buffaloes (Bubalus bubalis) in southern Italy. *Vet Rec,* 156(12), pp. 382-3.

Gannon, V.P., D'Souza, S., Graham, T., King, R.K., Rahn, K. & Read, S. (1997). Use of the flagellar H7 gene as a target in multiplex PCR assays and improved specificity in identification of enterohemorrhagic Escherichia coli strains. *J Clin Microbiol,* 35(3), pp. 656-62.

Garmendia, J., Frankel, G. & Crepin, V.F. (2005). Enteropathogenic and enterohemorrhagic Escherichia coli infections: translocation, translocation, translocation. *Infect Immun,* 73(5), pp. 2573-85.

Garmendia, J., Phillips, A.D., Carlier, M.F., Chong, Y., Schuller, S., Marches, O., Dahan, S., Oswald, E., Shaw, R.K., Knutton, S. & Frankel, G. (2004). TccP is an enterohaemorrhagic Escherichia coli O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton. *Cell Microbiol,* 6(12), pp. 1167-83.

Gelting, R.J., Baloch, M.A., Zarate-Bermudez, M.A. & Selman, C. (2011). Irrigation water issues potentially related to the 2006 multistate E. coli O157:H7 outbreak associated with spinach. *Agricultural Water Management,* 98(9), pp. 1395-1402.

Goldwater, P.N. & Bettelheim, K.A. (2012). Treatment of enterohemorrhagic Escherichia coli (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Med,* 10, p. 12.

Grauke, L.J., Kudva, I.T., Yoon, J.W., Hunt, C.W., Williams, C.J. & Hovde, C.J. (2002). Gastrointestinal tract location of Escherichia coli O157:H7 in ruminants. *Appl Environ Microbiol*, 68(5), pp. 2269-77.

Greenwood, D., Slack, R.C.B. & Peutherer, J.F. (2002). *Medical microbiology*. 16. ed: Elsevier Science.

Griffin, G., Evans, M., Eves, D., Jones, K., Low, C. & Strachan, D. (2010). *Review of the major outbreak of E. coli O157 in Surrey, 2009*: Independent Investigation Committee.

Hancock, D.D., Besser, T.E., Rice, D.H., Ebel, E.D., Herriott, D.E. & Carpenter, L.V. (1998). Multiple sources of Escherichia coli O157 in feedlots and dairy farms in the northwestern USA. *Prev Vet Med*, 35(1), pp. 11-9.

Haugum, K., Brandal, L.T., Lobersli, I., Kapperud, G. & Lindstedt, B.A. (2011). Detection of virulent Escherichia coli O157 strains using multiplex PCR and single base sequencing for SNP characterization. *J Appl Microbiol*, 110(6), pp. 1592-600.

Heiman, K.E., Mody, R.K., Johnson, S.D., Griffin, P.M. & Gould, L.H. (2015). Escherichia coli O157 Outbreaks in the United States, 2003-2012. *Emerg Infect Dis*, 21(8).

Herbert, L.J., Vali, L., Hoyle, D.V., Innocent, G., McKendrick, I.J., Pearce, M.C., Mellor, D., Porphyre, T., Locking, M., Allison, L., Hanson, M., Matthews, L., Gunn, G.J., Woolhouse, M.E. & Chase-Topping, M.E. (2014). E. coli O157 on Scottish cattle farms: evidence of local spread and persistence using repeat cross-sectional data. *BMC Vet Res*, 10, p. 95.

Herold, S., Karch, H. & Schmidt, H. (2004). Shiga toxin-encoding bacteriophages--genomes in motion. *Int J Med Microbiol*, 294(2-3), pp. 115-21.

Heywood, W., Henderson, B. & Nair, S.P. (2005). Cytolethal distending toxin: creating a gap in the cell cycle. *J Med Microbiol*, 54(Pt 3), pp. 207-16.

Hopkins, K.L. & Hilton, A.C. (2000). Methods available for the sub-typing of Escherichia coli O157. *World Journal of Microbiology and Biotechnology*, 16(8), pp. 741-48.

Howie, H., Mukerjee, A., Cowden, J., Leith, J. & Reid, T. (2003). Investigation of an outbreak of Escherichia coli O157 infection caused by environmental exposure at a scout camp. *Epidemiol Infect*, 131(3), pp. 1063-9.

Hyytia-Trees, E., Lafon, P., Vauterin, P. & Ribot, E.M. (2010). Multilaboratory validation study of standardized multiple-locus variable-number tandem repeat analysis protocol for shiga toxin-producing Escherichia coli O157: a novel approach to normalize fragment size data between capillary electrophoresis platforms. *Foodborne Pathog Dis*, 7(2), pp. 129-36.

Islam, M.Z., Musekiwa, A., Islam, K., Ahmed, S., Chowdhury, S., Ahad, A. & Biswas, P.K. (2014). Regional variation in the prevalence of E. coli O157 in cattle: a meta-analysis and meta-regression. *PLoS One*, 9(4), p. e93299.

Ivarsson, S., Riera Montes, M., Andersson, M. & Boqvist, S. (2012). Epidemiological characteristics and risk factors for sporadic EHEC infection in Sweden – results from a nationwide prospective case control study. *VTEC 2012*. Amsterdam, The Netherlands.

Iyoda, S., Manning, S.D., Seto, K., Kimata, K., Isobe, J., Etoh, Y., Ichihara, S., Migita, Y., Ogata, K., Honda, M., Kubota, T., Kawano, K., Matsumoto, K., Kudaka, J., Asai, N., Yabata, J., Tominaga, K., Terajima, J., Morita-Ishihara, T., Izumiya, H., Ogura, Y., Saitoh, T., Iguchi, A., Kobayashi, H., Hara-Kudo, Y., Ohnishi, M., Arai, R., Kawase, M., Asano, Y., Asoshima, N., Chiba, K., Furukawa, I., Kuroki, T., Hamada, M., Harada, S., Hatakeyama, T., Hirochi, T., Sakamoto, Y., Hiroi, M., Takashi, K., Horikawa, K., Iwabuchi, K., Kameyama, M., Kasahara, H., Kawanishi, S., Kikuchi, K., Ueno, H., Kitahashi, T., Kojima, Y., Konishi, N., Obata, H., Kai, A., Kono, T., Kurazono, T., Matsumoto, M., Matsumoto, Y., Nagai, Y., Naitoh, H., Nakajima, H., Nakamura, H., Nakane, K., Nishi, K., Saitoh, E., Satoh, H., Takamura, M., Shiraki, Y., Tanabe, J., Tanaka, K., Tokoi, Y. & Yatsuyanagi, J. (2014). Phylogenetic Clades 6 and 8 of Enterohemorrhagic Escherichia coli O157:H7 With Particular stx Subtypes are More Frequently Found in Isolates From Hemolytic Uremic Syndrome Patients Than From Asymptomatic Carriers. *Open Forum Infect Dis*, 1(2), p. ofu061.

Jay, M.T., Cooley, M., Carychao, D., Wiscomb, G.W., Sweitzer, R.A., Crawford-Miksza, L., Farrar, J.A., Lau, D.K., O'Connell, J., Millington, A., Asmundson, R.V., Atwill, E.R. & Mandrell, R.E. (2007). Escherichia coli O157:H7 in feral swine near spinach fields and cattle, central California coast. *Emerg Infect Dis,* 13(12), pp. 1908-11.

Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M. & Aarestrup, F.M. (2014). Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol,* 52(5), pp. 1501-10.

Johnson, T.J. & Nolan, L.K. (2009). Pathogenomics of the virulence plasmids of Escherichia coli. *Microbiol Mol Biol Rev,* 73(4), pp. 750-74.

Jores, J., Rumer, L. & Wieler, L.H. (2004). Impact of the locus of enterocyte effacement pathogenicity island on the evolution of pathogenic Escherichia coli. *Int J Med Microbiol,* 294(2-3), pp. 103-13.

Kaper, J.B., Nataro, J.P. & Mobley, H.L. (2004). Pathogenic Escherichia coli. *Nat Rev Microbiol,* 2(2), pp. 123-40.

Karama, M. & Gyles, C.L. (2010). Methods for genotyping verotoxin-producing Escherichia coli. *Zoonoses Public Health,* 57(7-8), pp. 447-62.

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(6), pp. 2043-9.

Karch, H., Denamur, E., Dobrindt, U., Finlay, B.B., Hengge, R., Johannes, L., Ron, E.Z., Tonjum, T., Sansonetti, P.J. & Vicente, M. (2012). The enemy within us: lessons from the 2011 European Escherichia coli O104:H4 outbreak. *EMBO Mol Med,* 4(9), pp. 841-8.

Karch, H., Tarr, P.I. & Bielaszewska, M. (2005). Enterohaemorrhagic Escherichia coli in human medicine. *Int J Med Microbiol,* 295(6-7), pp. 405-18.

Karmali, M.A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., Isaac-Renton, J., Clark, C., Rahn, K. & Kaper, J.B. (2003). Association of genomic O island 122 of Escherichia coli EDL 933 with verocytotoxin-producing Escherichia coli seropathotypes that are linked to epidemic and/or serious disease. *J Clin Microbiol,* 41(11), pp. 4930-40.

Khakhria, R., Duck, D. & Lior, H. (1990). Extended phage-typing scheme for Escherichia coli O157:H7. *Epidemiol Infect,* 105(3), pp. 511-20.

Kilonzo, C., Li, X., Vivas, E.J., Jay-Russell, M.T., Fernandez, K.L. & Atwill, E.R. (2013). Fecal shedding of zoonotic food-borne pathogens by wild rodents in a major agricultural region of the central California coast. *Appl Environ Microbiol,* 79(20), pp. 6337-44.

Kim, J., Nietfeldt, J. & Benson, A.K. (1999). Octamer-based genome scanning distinguishes a unique subpopulation of Escherichia coli O157:H7 strains in cattle. *Proc Natl Acad Sci U S A,* 96(23), pp. 13288-93.

Kistemann, T., Zimmer, S., Vagsholm, I. & Andersson, Y. (2004). GIS-supported investigation of human EHEC and cattle VTEC O157 infections in Sweden: geographical distribution, spatial variation and possible risk factors. *Epidemiol Infect,* 132(3), pp. 495-505.

Konowalchuk, J., Speirs, J.I. & Stavric, S. (1977). Vero response to a cytotoxin of Escherichia coli. *Infect Immun,* 18(3), pp. 775-9.

Kudva, I.T., Evans, P.S., Perna, N.T., Barrett, T.J., Ausubel, F.M., Blattner, F.R. & Calderwood, S.B. (2002). Strains of Escherichia coli O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. *J Bacteriol,* 184(7), pp. 1873-9.

Kulldorf, M. (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods,* 26(6), pp. 1481-96.

Kusumoto, M., Okitsu, T., Nishiya, Y., Suzuki, R., Yamai, S. & Kawamura, Y. (2001). Spontaneous reactivation of Shiga toxins in Escherichia coli O157:H7 cells caused by transposon excision. *J Biosci Bioeng,* 92(2), pp. 114-20.

La Ragione, R.M., Best, A., Woodward, M.J. & Wales, A.D. (2009). Escherichia coli O157:H7 colonization in small domestic ruminants. *FEMS Microbiol Rev,* 33(2), pp. 394-410.

Lahti, E., Keskimaki, M., Rantala, L., Hyvonen, P., Siitonen, A. & Honkanen-Buzalski, T. (2001). Occurrence of Escherichia coli O157 in Finnish cattle. *Vet Microbiol,* 79(3), pp. 239-51.

Laidler, M.R., Tourdjman, M., Buser, G.L., Hostetler, T., Repp, K.K., Leman, R., Samadpour, M. & Keene, W.E. (2013). Escherichia coli O157:H7 infections associated with consumption of locally grown strawberries contaminated by deer. *Clin Infect Dis,* 57(8), pp. 1129-34.

Laing, C.R., Buchanan, C., Taboada, E.N., Zhang, Y., Karmali, M.A., Thomas, J.E. & Gannon, V.P. (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, p. 287.

Larsson, J.T., Torpdahl, M., group, M.w. & Moller Nielsen, E. (2013). Proof-of-concept study for successful inter-laboratory comparison of MLVA results. *Euro Surveill,* 18(35), p. 20566.

Lejeune, J.T., Abedon, S.T., Takemura, K., Christie, N.P. & Sreevatsan, S. (2004). Human Escherichia coli O157:H7 genetic marker in isolates of bovine origin. *Emerg Infect Dis,* 10(8), pp. 1482-5.

Leopold, S.R., Magrini, V., Holt, N.J., Shaikh, N., Mardis, E.R., Cagno, J., Ogura, Y., Iguchi, A., Hayashi, T., Mellmann, A., Karch, H., Besser, T.E., Sawyer, S.A., Whittam, T.S. & Tarr, P.I. (2009). A precise reconstruction of the emergence and constrained radiations of Escherichia coli O157 portrayed by backbone concatenomic analysis. *Proc Natl Acad Sci U S A,* 106(21), pp. 8713-8.

Leopold, S.R., Shaikh, N. & Tarr, P.I. (2010). Further evidence of constrained radiation in the evolution of pathogenic Escherichia coli O157:H7. *Infect Genet Evol,* 10(8), pp. 1282-5.

Leotta, G.A., Miliwebsky, E.S., Chinen, I., Espinosa, E.M., Azzopardi, K., Tennant, S.M., Robins-Browne, R.M. & Rivas, M. (2008). Characterisation of Shiga toxin-producing Escherichia coli O157 strains isolated from humans in Argentina, Australia and New Zealand. *BMC Microbiol,* 8, p. 46.

Levin, B.R., Lipsitch, M. & Bonhoeffer, S. (1999). Population biology, evolution, and infectious disease: convergence and synthesis. *Science,* 283(5403), pp. 806-9.

Lim, J.Y., Hong, J.B., Sheng, H., Shringi, S., Kaul, R., Besser, T.E. & Hovde, C.J. (2010). Phenotypic diversity of Escherichia coli O157:H7 strains associated with the plasmid O157. *J Microbiol,* 48(3), pp. 347-57.

Lim, J.Y., Li, J., Sheng, H., Besser, T.E., Potter, K. & Hovde, C.J. (2007). Escherichia coli O157:H7 colonization at the rectoanal junction of long-duration culture-positive cattle. *Appl Environ Microbiol,* 73(4), pp. 1380-2.

Lindstedt, B.A. (2005). Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis,* 26(13), pp. 2567-82.

Lindstedt, B.A., Torpdahl, M., Vergnaud, G., Le Hello, S., Weill, F.X., Tietze, E., Malorny, B., Prendergast, D.M., Ni Ghallchoir, E., Lista, R.F., Schouls, L.M., Soderlund, R., Borjesson, S. & Akerstrom, S. (2013). Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. *Euro Surveill,* 18(4), p. 20385.

Liu, K., Knabel, S.J. & Dudley, E.G. (2009). rhs genes are potential markers for multilocus sequence typing of Escherichia coli O157:H7 strains. *Appl Environ Microbiol,* 75(18), pp. 5853-62.

Low, J.C., McKendrick, I.J., McKechnie, C., Fenlon, D., Naylor, S.W., Currie, C., Smith, D.G., Allison, L. & Gally, D.L. (2005). Rectal carriage of enterohemorrhagic Escherichia coli O157 in slaughtered cattle. *Appl Environ Microbiol,* 71(1), pp. 93-7.

MacFarquhar, J.K. & Williams, C. (2013). *Outbreak report: Escherichia coli O157:H7 outbreak associated with a county fair, North Carolina 2012*: North Carolina Department of Health and Human Services, Division of Public Health.

Madic, J., Peytavin de Garam, C., Vingadassalon, N., Oswald, E., Fach, P., Jamet, E. & Auvray, F. (2010). Simplex and multiplex real-time PCR assays for the detection of flagellar (H-antigen) fliC alleles and intimin (eae) variants associated with enterohaemorrhagic Escherichia coli (EHEC) serotypes

O26:H11, O103:H2, O111:H8, O145:H28 and O157:H7. *J Appl Microbiol,* 109(5), pp. 1696-705.

Magnusson, U. (2012). Overview of infectious diseases and the wildlife-livestock interface. In: J.M. Levengood & Norrgren, L. (eds) *Ecology and Animal Health*, pp. 183-185.

Manning, S.D., Motiwala, A.S., Springman, A.C., Qi, W., Lacher, D.W., Ouellette, L.M., Mladonicky, J.M., Somsel, P., Rudrik, J.T., Dietrich, S.E., Zhang, W., Swaminathan, B., Alland, D. & Whittam, T.S. (2008). Variation in virulence among clades of Escherichia coli O157:H7 associated with disease outbreaks. *Proc Natl Acad Sci U S A,* 105(12), pp. 4868-73.

Matthews, L., Reeve, R., Gally, D.L., Low, J.C., Woolhouse, M.E., McAteer, S.P., Locking, M.E., Chase-Topping, M.E., Haydon, D.T., Allison, L.J., Hanson, M.F., Gunn, G.J. & Reid, S.W. (2013). Predicting the public health benefit of vaccinating cattle against Escherichia coli O157. *Proc Natl Acad Sci U S A,* 110(40), pp. 16265-70.

McGee, P., Scott, L., Sheridan, J.J., Earley, B. & Leonard, N. (2004). Horizontal transmission of Escherichia coli O157:H7 during cattle housing. *J Food Prot,* 67(12), pp. 2651-6.

Mead, P.S. & Griffin, P.M. (1998). Escherichia coli O157:H7. *Lancet,* 352(9135), pp. 1207-12.

Mele, C., Remuzzi, G. & Noris, M. (2014). Hemolytic uremic syndrome. *Semin Immunopathol,* 36(4), pp. 399-420.

Mellmann, A., Lu, S., Karch, H., Xu, J.G., Harmsen, D., Schmidt, M.A. & Bielaszewska, M. (2008). Recycling of Shiga toxin 2 genes in sorbitol-fermenting enterohemorrhagic Escherichia coli O157:NM. *Appl Environ Microbiol,* 74(1), pp. 67-72.

Mellor, G.E., Sim, E.M., Barlow, R.S., D'Astek, B.A., Galli, L., Chinen, I., Rivas, M. & Gobius, K.S. (2012). Phylogenetically Related Argentinean and Australian Escherichia coli O157 Isolates Are Distinguished by Virulence Clades and Alternative Shiga Toxin 1 and 2 Prophages. *Appl Environ Microbiol,* 78(13), pp. 4724-31.

Muniesa, M., Blanco, J.E., De Simon, M., Serra-Moreno, R., Blanch, A.R. & Jofre, J. (2004). Diversity of stx2 converting bacteriophages induced from Shiga-toxin-producing Escherichia coli strains isolated from cattle. *Microbiology,* 150(Pt 9), pp. 2959-71.

Nart, P., Naylor, S.W., Huntley, J.F., McKendrick, I.J., Gally, D.L. & Low, J.C. (2008). Responses of cattle to gastrointestinal colonization by Escherichia coli O157:H7. *Infect Immun,* 76(11), pp. 5366-72.

Naylor, S.W., Low, J.C., Besser, T.E., Mahajan, A., Gunn, G.J., Pearce, M.C., McKendrick, I.J., Smith, D.G. & Gally, D.L. (2003). Lymphoid follicle-dense mucosa at the terminal rectum is the principal site of colonization of enterohemorrhagic Escherichia coli O157:H7 in the bovine host. *Infect Immun,* 71(3), pp. 1505-12.

Nielsen, E.M., Tegtmeier, C., Andersen, H.J., Gronbaek, C. & Andersen, J.S. (2002). Influence of age, sex and herd characteristics on the occurrence of Verocytotoxin-producing Escherichia coli O157 in Danish dairy farms. *Vet Microbiol,* 88(3), pp. 245-57.

Noller, A.C., McEllistrem, M.C., Stine, O.C., Morris, J.G., Jr., Boxrud, D.J., Dixon, B. & Harrison, L.H. (2003). Multilocus sequence typing reveals a lack of diversity among Escherichia coli O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J Clin Microbiol,* 41(2), pp. 675-9.

Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., Kodama, T., Abe, H., Nakayama, K., Kurokawa, K., Tobe, T., Hattori, M. & Hayashi, T. (2009). Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic Escherichia coli. *Proc Natl Acad Sci U S A,* 106(42), pp. 17939-44.

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(2), pp. 604-12.

Ooka, T., Ogura, Y., Asadulghani, M., Ohnishi, M., Nakayama, K., Terajima, J., Watanabe, H. & Hayashi, T. (2009). Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in Escherichia coli O157 genomes. *Genome Res,* 19(10), pp. 1809-16.

Paiba, G.A., Wilesmith, J.W., Evans, S.J., Pascoe, S.J., Smith, R.P., Kidd, S.A., Ryan, J.B., McLaren, I.M., Chappell, S.A., Willshaw, G.A., Cheasty, T., French, N.P., Jones, T.W., Buchanan, H.F., Challoner, D.J., Colloff, A.D., Cranwell, M.P., Daniel, R.G., Davies, I.H., Duff, J.P., Hogg, R.A., Kirby, F.D., Millar, M.F., Monies, R.J., Nicholls, M.J. & Payne, J.H. (2003). Prevalence of faecal excretion of verocytotoxigenic Escherichia coli O157 in cattle in England and Wales. *Vet Rec,* 153(12), pp. 347-53.

Paunio, M., Pebody, R., Keskimaki, M., Kokki, M., Ruutu, P., Oinonen, S., Vuotari, V., Siitonen, A., Lahti, E. & Leinikki, P. (1999). Swimming-associated outbreak of Escherichia coli O157:H7. *Epidemiol Infect,* 122(1), pp. 1-5.

Persson, S., Olsen, K.E., Ethelberg, S. & Scheutz, F. (2007). Subtyping method for Escherichia coli shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *J Clin Microbiol,* 45(6), pp. 2020-4.

Pierard, D., Muyldermans, G., Moriau, L., Stevens, D. & Lauwers, S. (1998). Identification of new verocytotoxin type 2 variant B-subunit genes in human and animal Escherichia coli isolates. *J Clin Microbiol,* 36(11), pp. 3317-22.

Rangel, J.M., Sparling, P.H., Crowe, C., Griffin, P.M. & Swerdlow, D.L. (2005). Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002. *Emerg Infect Dis,* 11(4), pp. 603-9.

Reinstein, S., Fox, J.T., Shi, X., Alam, M.J. & Nagaraja, T.G. (2007). Prevalence of Escherichia coli O157:H7 in the American bison (Bison bison). *J Food Prot,* 70(11), pp. 2555-60.

Riordan, J.T., Viswanath, S.B., Manning, S.D. & Whittam, T.S. (2008). Genetic differentiation of Escherichia coli O157:H7 clades associated with human disease by real-time PCR. *J Clin Microbiol,* 46(6), pp. 2070-3.

Said, B., Wright, F., Nichols, G.L., Reacher, M. & Rutter, M. (2003). Outbreaks of infectious disease associated with private drinking water supplies in England and Wales 1970-2000. *Epidemiol Infect,* 130(3), pp. 469-79.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239(4839), pp. 487-91.

Sartz, L., De Jong, B., Hjertqvist, M., Plym-Forshell, L., Alsterlund, R., Lofdahl, S., Osterman, B., Stahl, A., Eriksson, E., Hansson, H.B. & Karpman, D. (2008). An outbreak of Escherichia coli O157:H7 infection in southern Sweden associated with consumption of fermented sausage; aspects of sausage production that increase the risk of contamination. *Epidemiol Infect,* 136(3), pp. 370-80.

Schets, F.M., During, M., Italiaander, R., Heijnen, L., Rutjes, S.A., van der Zwaluw, W.K. & de Roda Husman, A.M. (2005). Escherichia coli O157:H7 in drinking water from private water supplies in the Netherlands. *Water Res,* 39(18), pp. 4485-93.

Scheutz, F., Teel, L.D., Beutin, L., Pierard, D., Buvens, G., Karch, H., Mellmann, A., Caprioli, A., Tozzoli, R., Morabito, S., Strockbine, N.A., Melton-Celsa, A.R., Sanchez, M., Persson, S. & O'Brien, A.D. (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J Clin Microbiol,* 50(9), pp. 2951-63.

Schmidt, H. (2001). Shiga-toxin-converting bacteriophages. *Res Microbiol,* 152(8), pp. 687-95.

Schmidt, H., Bielaszewska, M. & Karch, H. (1999). Transduction of enteric Escherichia coli isolates with a derivative of Shiga toxin 2-encoding bacteriophage phi3538 isolated from Escherichia coli O157:H7. *Appl Environ Microbiol,* 65(9), pp. 3855-61.

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(2), pp. 265-72.

Schmidt, M.A. (2010). LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol,* 12(11), pp. 1544-52.

Schoeni, J.L. & Doyle, M.P. (1994). Variable colonization of chickens perorally inoculated with Escherichia coli O157:H7 and subsequent contamination of eggs. *Appl Environ Microbiol,* 60(8), pp. 2958-62.

Schouten, J.M., Graat, E.A., Frankena, K., van de Giessen, A.W., van der Zwaluw, W.K. & de Jong, M.C. (2005). A longitudinal study of Escherichia coli O157 in cattle of a Dutch dairy farm and in the farm environment. *Vet Microbiol,* 107(3-4), pp. 193-204.

Sekse, C., Solheim, H., Urdahl, A.M. & Wasteson, Y. (2008). Is lack of susceptible recipients in the intestinal environment the limiting factor for transduction of Shiga toxin-encoding phages? *J Appl Microbiol,* 105(4), pp. 1114-20.

Serna, A.t. & Boedeker, E.C. (2008). Pathogenesis and treatment of Shiga toxin-producing Escherichia coli infections. *Curr Opin Gastroenterol,* 24(1), pp. 38-47.

Sheng, H., Lim, J.Y., Knecht, H.J., Li, J. & Hovde, C.J. (2006). Role of Escherichia coli O157:H7 virulence factors in colonization at the bovine terminal rectal mucosa. *Infect Immun,* 74(8), pp. 4685-93.

Snedeker, K.G., Shaw, D.J., Locking, M.E. & Prescott, R.J. (2009). Primary and secondary cases in Escherichia coli O157 outbreaks: a statistical analysis. *BMC Infect Dis,* 9, p. 144.

Soderlund, R., Jernberg, C., Ivarsson, S., Hedenstrom, I., Eriksson, E., Bongcam-Rudloff, E. & Aspan, A. (2014). Molecular typing of Escherichia coli O157:H7 isolates from Swedish cattle and human cases: population dynamics and virulence. *J Clin Microbiol,* 52(11), pp. 3906-12.

Soderstrom, A., Osterberg, P., Lindqvist, A., Jonsson, B., Lindberg, A., Blide Ulander, S., Welinder-Olsson, C., Lofdahl, S., Kaijser, B., De Jong, B., Kuhlmann-Berenzon, S., Boqvist, S., Eriksson, E., Szanto, E., Andersson, S., Allestam, G., Hedenstrom, I., Ledet Muller, L. & Andersson, Y. (2008). A large Escherichia coli O157 outbreak in Sweden associated with locally produced lettuce. *Foodborne Pathog Dis,* 5(3), pp. 339-49.

Spinale, J.M., Ruebner, R.L., Copelovitch, L. & Kaplan, B.S. (2013). Long-term outcomes of Shiga toxin hemolytic uremic syndrome. *Pediatr Nephrol,* 28(11), pp. 2097-105.

Strachan, N.J., Dunn, G.M., Locking, M.E., Reid, T.M. & Ogden, I.D. (2006). Escherichia coli O157: burger bug or environmental pathogen? *Int J Food Microbiol,* 112(2), pp. 129-37.

Swirski, A.L., Pearl, D.L., Williams, M.L., Homan, H.J., Linz, G.M., Cernicchiaro, N. & LeJeune, J.T. (2014). Spatial epidemiology of Escherichia coli O157:H7 in dairy cattle in relation to night roosts Of Sturnus vulgaris (European Starling) in Ohio, USA (2007-2009). *Zoonoses Public Health,* 61(6), pp. 427-35.

Talley, J.L., Wayadande, A.C., Wasala, L.P., Gerry, A.C., Fletcher, J., DeSilva, U. & Gilliland, S.E. (2009). Association of Escherichia coli O157:H7 with filth flies (Muscidae and Calliphoridae) captured in leafy greens fields and experimental transmission of E. coli O157:H7 to spinach leaves by house flies (Diptera: Muscidae). *J Food Prot,* 72(7), pp. 1547-52.

Tam, P.J. & Lingwood, C.A. (2007). Membrane cytosolic translocation of verotoxin A1 subunit in target cells. *Microbiology,* 153(Pt 8), pp. 2700-10.

Tesh, V.L. (2010). Induction of apoptosis by Shiga toxins. *Future Microbiol,* 5(3), pp. 431-53.

Tilden, J., Jr., Young, W., McNamara, A.M., Custer, C., Boesel, B., Lambert-Fair, M.A., Majkowski, J., Vugia, D., Werner, S.B., Hollingsworth, J. & Morris, J.G., Jr. (1996). A new route of transmission for Escherichia coli: infection from dry fermented salami. *Am J Public Health,* 86(8), pp. 1142-5.

Tobe, T., Beatson, S.A., Taniguchi, H., Abe, H., Bailey, C.M., Fivian, A., Younis, R., Matthews, S., Marches, O., Frankel, G., Hayashi, T. & Pallen, M.J.

(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(40), pp. 14941-6.

Trachtman, H., Austin, C., Lewinski, M. & Stahl, R.A. (2012). Renal and neurological involvement in typical Shiga toxin-associated HUS. *Nat Rev Nephrol,* 8(11), pp. 658-69.

Waldor, M.K. & Friedman, D.I. (2005). Phage regulatory circuits and virulence gene expression. *Curr Opin Microbiol,* 8(4), pp. 459-65.

Wallace, J.S., Cheasty, T. & Jones, K. (1997). Isolation of vero cytotoxin-producing Escherichia coli O157 from wild birds. *J Appl Microbiol,* 82(3), pp. 399-404.

Walsh, M.J., Dodd, J.E. & Hautbergue, G.M. (2013). Ribosome-inactivating proteins: potent poisons and molecular tools. *Virulence,* 4(8), pp. 774-84.

Wang, G. & Doyle, M.P. (1998). Survival of enterohemorrhagic Escherichia coli O157:H7 in water. *J Food Prot,* 61(6), pp. 662-7.

Whittam, T.S., Wolfe, M.L., Wachsmuth, I.K., Orskov, F., Orskov, I. & Wilson, R.A. (1993). Clonal relationships among Escherichia coli strains that cause hemorrhagic colitis and infantile diarrhea. *Infect Immun,* 61(5), pp. 1619-29.

Wick, L.M., Qi, W., Lacher, D.W. & Whittam, T.S. (2005). Evolution of genomic content in the stepwise emergence of Escherichia coli O157:H7. *J Bacteriol,* 187(5), pp. 1783-91.

Widgren, S., Soderlund, R., Eriksson, E., Fasth, C., Aspan, A., Emanuelson, U., Alenius, S. & Lindberg, A. (2015). Longitudinal observational study over 38 months of verotoxigenic Escherichia coli O157:H7 status in 126 cattle herds. *Prev Vet Med*.

Vogler, A.J., Keys, C., Nemoto, Y., Colman, R.E., Jay, Z. & Keim, P. (2006). Effect of repeat copy number on variable-number tandem repeat mutations in Escherichia coli O157:H7. *J Bacteriol,* 188(12), pp. 4253-63.

Wong, A.R., Pearson, J.S., Bright, M.D., Munera, D., Robinson, K.S., Lee, S.F., Frankel, G. & Hartland, E.L. (2011). Enteropathogenic and enterohaemorrhagic Escherichia coli: even more subversive elements. *Mol Microbiol,* 80(6), pp. 1420-38.

Xu, X., McAteer, S.P., Tree, J.J., Shaw, D.J., Wolfson, E.B., Beatson, S.A., Roe, A.J., Allison, L.J., Chase-Topping, M.E., Mahajan, A., Tozzoli, R., Woolhouse, M.E., Morabito, S. & Gally, D.L. (2012). Lysogeny with Shiga toxin 2-encoding bacteriophages represses type III secretion in enterohemorrhagic Escherichia coli. *PLoS Pathog,* 8(5), p. e1002672.

Yang, Z., Kovar, J., Kim, J., Nietfeldt, J., Smith, D.R., Moxley, R.A., Olson, M.E., Fey, P.D. & Benson, A.K. (2004). Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative Escherichia coli O157:H7 from bovine production environments and human clinical samples. *Appl Environ Microbiol,* 70(11), pp. 6846-54.

Yokoyama, E., Hirai, S., Hashimoto, R. & Uchimura, M. (2012). Clade analysis of enterohemorrhagic Escherichia coli serotype O157:H7/H- strains and hierarchy of their phylogenetic relationships. *Infect Genet Evol,* 12(8), pp. 1724-8.

Yoon, J.W. & Hovde, C.J. (2008). All blood, no stool: enterohemorrhagic Escherichia coli O157:H7 infection. *J Vet Sci*, 9(3), pp. 219-31.

Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A.K., Taboada, E. & Gannon, V.P. (2007). Genome evolution in major Escherichia coli O157:H7 lineages. *BMC Genomics*, 8, p. 121.

Zhang, Y., Laing, C., Zhang, Z., Hallewell, J., You, C., Ziebell, K., Johnson, R.P., Kropinski, A.M., Thomas, J.E., Karmali, M. & Gannon, V.P. (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(2), pp. 474-82.

Zhou, Z., Li, X., Liu, B., Beutin, L., Xu, J., Ren, Y., Feng, L., Lan, R., Reeves, P.R. & Wang, L. (2010). Derivation of Escherichia coli O157:H7 from its O55:H7 precursor. *PLoS One*, 5(1), p. e8700.