A Novel Approach to Pathogen Recognition for Diarrhoeagenic Enteroaggregative *Escherichia coli*

By

Marie Anne Chattaway

University College London

Submitted for examination in the degree of

Doctor of Philosophy

June 2014

This PhD is dedicated to

Nadia Ahmod

1980-2012

A great scientist and friend, you were the first person to encourage me to do a PhD when I was a medical technical officer doing PCR on the bench!

Statement of authorship

I, Marie Anne Chattaway, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Support in different aspects of my thesis is described below.

Data Analysis:
Marie Anne Chattaway, Public Health England (PHE)

Statistical Analysis:
Ross Harris, Statistician, PHE
Marie Anne Chattaway, PHE

Multi-locus Sequence Typing
PCR, Sequencing PCR: Marie Anne Chattaway, PHE
Sequencing PCR product on ABI: Oxford University

ClonalFrame Analysis
Marie Anne Chattaway, University of Haverford, USA & PHE

Hep-2 Cell Assay
Marie Anne Chattaway, University of Haverford, USA

Worm Model
Marie Anne Chattaway, University of Haverford, USA

Biolog
Marie Anne Chattaway, PHE

Serotyping
Dawn Hedges, Martha Valencia, PHE

Virulence gene PCR
Julia Mtwale, UCL MSc student, PHE
Emma White, Royal Free Hospital trainee clinical scientist, PHE
Danielle Hall, PHE
Marie Anne Chattaway

Bioinformatics support
Development of BioNumerics MLST data base, Steve Platt, Bioinformatician, PHE
MLST scripts, Steve Platt, Bioinformatician, PHE
ClonalFrame scripts – Anthony Underwood, Bioinformatician, PHE
Biolog/ *E .coli* MLST data – Tim Dallman, Bioinformatician, PHE

Marie Anne Chattaway
May 2014

Abstract

Enteroaggregative *E. coli* is known to cause diarrhoeal disease in developing and developed countries, but is also found in asymptomatic carriage and so the causal link between EAEC and disease is unknown. This study assessed bacterial load, the presence of co-infections and demographic data to assess if EAEC was independently associated with intestinal infectious disease in the United Kingdom. This study concluded that EAEC was independently capable of causing disease and accounts for ~1% of intestinal disease and therefore an important burden.

A case control approach of analysis by multi locus sequence typing of 564 EAEC isolates from cases and controls in Bangladesh, Nigeria and the UK spanning the past 29 years, revealed multiple successful lineages of EAEC. The population structure of EAEC indicates some clusters are statistically associated with disease or carriage, further highlighting the heterogeneous nature of this group of organisms. Different clusters have evolved independently as a result of both mutational and recombination events; the EAEC phenotype is distributed throughout the population of *E. coli*.

In vivo models looking at EAEC infection and virulence gene content of EAEC show that different complexes varied in their ability to cause disease further concluding that these complexes may have come from different ecological niches but that the EAEC phenotype enhanced survival and so these defined EAEC complexes have converged to stably retain the plasmid and phenotype.

This study has identified several successful EAEC complexes associated with sporadic disease (ST 10, 38, 40, 295, 278, 394, 678 and 746 Cplx) with subset of complexes showing evidence of patho-adaptation to cause extra-intestinal infections (ST38 Cplx) and outbreaks (ST40, 278 and 678 Cplx) or form hybrid strains with Shiga toxin producing *E. coli* causing severe disease (ST40 and 678 Cplx).

Table of Contents

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
µl	microlitre
AA	amino acid
AA	aggregative adherence
aaf5A/AAF/5	EAEC fimbrial type V
aafA/AAF/2	EAEC fimbrial type II
aaiC	aaiC gene from 042 pheU island
aap	Antiaggregative protein (surface coat protein - dispersin) - adhesion
aat	anti-aggregation protein transporter gene
adk	adenylate kinase gene
AE	attaching and effacing
agg3A/AAF/3	EAEC fimbrial type III
agg4A/AAF/4	EAEC fimbrial type IV
aggA/AAF/1	EAEC fimbrial type I
*agg*R	AAF/I & AAF/II transcriptional activator, Regulator of plasmid and chromosomal genes
*ast*A	enteroaggregative heat stable toxin 1 (EAST-1)
astA	enteroaggregative heat stable toxin 1 (EAST-1)
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	base pairs
CDEC	cell detachment *E. coli*
CDTEC	cytolethal distending toxin-producing *E. coli*
CI	confidence interval
CO_2	carbon dioxide
Cplx	complex
Ct	cyclic threshold
DA	diffusely adherence
DAEC	diffusely adherent *E. coli*
DE	Dorset Egg
DEC	diarrhoeagenic *E. coli*
dH_2O	distilled water
DLV	double locus variant
DMEM	Dulbecco modified Eagle medium
DNA	deoxyribonucleic acid
eae	effacement and attachment gene
EAEC	enteroaggregative *E. coli*
EAF	EPEC adherence factor
EHEC	enterohaemorrahgic *E. coli*
EIEC	enteroinvasive *E. coli*
EM	electrophoretic motilities

EPEC	enteropathogenic *E. coli*
ET	electrophoretic types
ETEC	enterotoxigenic *E. coli*
EtOH	ethanol
ExPEC	Extra-intestinal pathogenic *E. coli*
FAS	fluorescence actin staining
*fum*C	fumarate hydratase gene
g	g force
GBRU	Gastrointestinal Bacteria Reference Unit
GEMS	Global Enteric Multi-center Study
GI	Gastrointestinal infection
GIRU	Gastrointestinal Infection Reference Unit
*gyr*B	DNA gyrase subunit B gene
GTP	guanosine triphosphate
H-	non motile
H?	flagella antigen unknown
Haz. Ratio	hazard ratio
HEp-2	human epithelial cell line
HPA	Health Protection Agency
hra1	chromosomal gene encoding an accessory colonization factor
HUS	haemolytic uraemia syndrome
icd	isocitrate dehydrogenase gene
ICDDR,B	International Centre for Diarrhoeal Disease Research, Bangladesh
IID	Infectious Intestinal Disease
ipaH	invasin gene
irp2	Yersiniabactin biosynthetic gene
K	Capsular antigen
L1-4	larvae stages
LA	Local adherence
LEE	locus of enterocyte effacement
LPS	lipopolysaccharide
mdh	malate dehydrogenase gene
ml	millilitre
MLEE	multi-locus enzyme electrophoresis
MLEE	multi-locus enzyme electrophoresis
MLSA	multi-locus Sequence Analysis
MLST	multi-locus sequence typing
MSTree	minimal spanning tree
mTSB	modified tryptone soya broth
NaOAc	sodium acetate
ng	nanograms
NGS	next generation sequencing
O?	somatic antigen unknown
O^2	oxygen
^oC	degrees centigrade

OR	odds ratio
p	probability
p/0	rho over theta
PAF	population attributable fraction
PBS	phosphate buffered saline
PCR	polymerase chain reaction
P_e	proportion of cases with the exposure
pet	plasmid-encoded toxin, heat-liable Cytotoxin (104kDA)/ EAEC enterotoxin
pet	plasmid encoded toxin
PHE	Public Health England
pic	Secreted protease gene
*pur*A	adenylosuccinate synthetase gene
R	recombination rate
r/m	recombination over mutation
rDNA	ribosomal DNA
*rec*A	ATP/GTP binding motif gene
ROC	receiver-operating characteristic
rpm	rounds per minute
rpm	rounds per minute
RR	risk ratio
RR_e	relative risk of disease
rRNA	ribosomal ribonucleic acid
SD	standard deviation
set1A	*Shigella* enterotoxin gene
SLV	single locus varient
sp.	species
ST	sequence type
Std. Err	standard error
STEC	shiga toxin producing *E. coli*
t	t-test
theta (0)	mutation rate
tia	putative invasion determinant
TLV	triple locus varient
TSB	Tryptone Soya Broth
UCL	University College London
UPEC	uropathogenic *E. coli*
VTEC	verocytoxic *E. coli*
WGS	Whole Genome Sequencing

Statistical Analysis Used in this Study

Chapter 2 Burden of EAEC in the UK

Receiver-operating characteristic (ROC) analysis: Used to define a cut-off in the Ct values to estimate the number of true EAEC IID positive cases in the population. When a measurement is used to make a diagnosis , the choice of the 'best cut-off' is not simple and so a graphical approach is carried out to plot the sensitivity versus the specificity for each possible cut-off. The youden index is then used.

Youden index (sensitivity + specificity-1): Used to calculate the maximum value of sensitivity and specificity to identify the optimum cut-off (Bewick *et al.* 2004;Fluss *et al.* 2005). This is a single statistic that captures the performance of a diagnostic test. In this case, could the presence of EAEC below a certain ct value be used to diagnostically.

Students t –test: This test is used for comparing the actual difference between two means in relation to the variation in the data (expressed as the standard deviation of the difference between two means). This parametric test was used to compare the distribution of Ct values from cases and controls for EAEC positive individuals to ascertain if the bacterial load play a role in determining whether the presence of EAEC would be related to carriage or disease.

Chi-square test: This was used to test if the distribution of the pathogen between cases and controls was as expected by chance (Figure 2.3). It was used to test the association of co-infections with EAEC in cases versus control (Figure 2.4). It was used to test the association of co-infection with EAEC in comparison to norovirus (Figure 2.5). If there is a variable which has a normal distribution then X2 will have a chi squared distribution. This test was used to ascertain if the distributions of the presence of EAEC were normal or played a role in disease, the presence of co-infections were also taken into account.

Odd Ratio (OR): A measure of association between an exposure and an outcome. For example what are the chances of someone having disease if EAEC is present in their stool?

Relative risk (RR): This is the ratio of the probability of an event occurring in an exposed group to the probability of the event occurring in a comparison non-exposed group. For example, the chances of case participants having EAEC present in their stool in comparison to control participants having EAEC present in their stool.

Confidence Interval (CI): This is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate and is calculated from the observations different from sample to sample. It is a range of values which can be confident includes the true value.

Logistic Regression: A logistic regression of univariate and multivariate analysis was carried out using case or control as outcome and infecting agent and age as independent variables. In this way the independent association between EAEC and disease whilst controlling for other pathogens was assessed.

Population attributable fraction (PAF): Model results from logistic regression were used to calculate the population attributable fraction (PAF):

$$PAR = P_e (RR_e - 1) / RR_e$$

Where Pe is the proportion of cases with the exposure (EAEC) and RRe the relative risk of disease. This form allows for confounding of the exposure if an adjusted RR is used, as recommended in Rockhill *et al* (Rockhill *et al.* 1998). In this case, adjusted odds ratios are substituted into this equation to give an approximate, adjusted PAF.

Chapter 3 – Definition of pathogenic EAEC groups by a case control approach

Fisher's exact test: This is a statistical significance test used in the analysis of contingency tables (displays the frequency of distribution of the variables). It is a class of exact tests because the significance of the deviation from a null hypothesis (e.g. probability (p) value) can be calculated exactly. It was used to look out the probability that EAEC complexes were associated with cases or controls and whether EAEC complexes were associated with being EAEC or other *E. coli* pathotypes.

Chapter 4 – Inference of Bacterial Microevolution

ClonalFrame: ClonalFrame is a software package and is a Bayesian method of constructing evolutionary histories that takes both mutation and recombination into account (Didelot and Falush 2007). It segregates the sequences into frames and looks across sequences to ascertain if mutations are evenly distributed across the frames.

Bayesian statistics style model: The inference of probability is updated as additional evidence is learned and in this analysis is based on the Markov chain Monte Carlo method.

Markov Chain Monte Carlo: A class of algorithms for sampling from probability distributions (i.e. distributions of mutations or recombination) based on constructing a Markov chain (i.e. random process on which the next state depends only on the current state and not on the sequence of events that preceded it) that has the desired distribution as its equilibrium distribution (Note that there is no assumption on the starting distribution; the chain converges to the stationary distribution regardless of where it begins). For example, analysis is based on the data/evidence that is processed and no assumptions are made as to how this data occurred. It is used to estimate the distribution and compute a mean for example how often mutation and recombination occurs.

Posterior probability: In Bayesian statistics, the posterior probability of a random event or an uncertain proposition is the conditional probability that is assigned after the relevant evidence is taken into account

Inference of genetic events

Visualisation of parameters and statistics were used to infer microevolution and the phylogeny of the EAEC data (Didelot and Falush 2006;Didelot and Falush 2007;Vos and Didelot 2009). Multiple analyses were carried out including:

Mutation rate (theta) – rate at which change is due to mutation

rho over theta (ρ/θ) – a measure of the frequency at which recombination occurs relative to mutation.

Recombination ratio (r/ᴍ) – ratio of rates at which nucleotides become substituted as a result of recombination and mutation.

External to internal branch length ratio test – computes the distribution of the sum of the lengths of the external branches (i.e. ones that connect a leaf of the tree) divided by the sum of the lengths of the internal branches (i.e. the ones that connect two internal nodes of the tree) (shown as red). Also computes the expected distribution of the external to internal branch length ratio under the coalescent model (shown in blue).

Chapter 5 – Genotypic and Phenotypic models to assess EAEC pathogenicity

Kaplan-Meier method: Data from the assays were analysed using the Kaplan-Meier method (Bland and Altman 1998) which is an estimator for estimating the survival function from lifetime data for example the survival of a worm when fed EAEC.

Log Rank test: This was carried out for equality of survivor functions (Bland and Altman 2004) in which a comparison of all complexes was tested against each other to ascertain if the survival function of the worm was reduced by ingesting EAEC from the different complexes.

Cox regression : The groups were further analysed by cox regression analysis (Altman and Andersen 1989), against the complex strains against the control strains. This study was limited by the small numbers so there would need to be extreme differences to show a significance determined by a hazard ratio of <0.5 or >2 and a probability <0.05.

Hazard ratio: In survival analysis the hazard ratio is the ratio of hazard rates corresponding to the conditions described by two levels of an explanatory variable. For example in the EAEC study, the worms fed EAEC strains may die at twice the rate per unit time as the control non-pathogenic *E. coli* strain. The hazard ratio would be 2, indicating higher hazard of death

Pearson Chi-square test: This statistical hypothesis test was used to test the hypothesis of virulence genes being associated with cases or complexes. It is used when the sampling distribution of the test statistic is a chi-squared distribution when the null hypothesis is true resulting in a probability. The Pearson method is a chi-squared test for independence (mentioned without any modifiers) and is meant for an exact testing place of a Fishers exact test (which is appropriate for small sizes under 5). This test was appropriate as the sample size was over 5.

Linear regression: Linear regression is an approach for modelling the relationship between scalar dependent variable y and one or more variables denoted x. This method was used to analyse if these complexes have a higher virulence than a reference group. ST38 complex was chosen as the reference groups as it contained the largest sample size and best representation of the data.

Confidence Interval (CI): this is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate and is calculated from the observations different from sample to sample. It is a range of values which can be confident includes the true value.

Coefficient: This was used to assess the virulence score of the complexes in comparison to a reference group. The coefficient is the change in outcome per unit increase - this just being one, as they are indicators for that complex vs. the reference group, and therefore can be interpreted as a differences in means compared to the control group.

t-test: The model coefficients have a t distribution, so their significance is tested via a t-statistic, which is given, and its p-value - which is the probability that you would achieve results as extreme or more than that observed, given the null hypothesis of there being no difference between the complex and the reference group.

Standard deviation: This shows how much variation or dispersion from the average exists, A low standard deviation indicates that the data points tend to be very close to the mean (or expected value) whereas a high standard deviation indicates that the data points are spread out over a large range of values.

Statistical analysis was carried out in STATA v 13.1 for Chapters 2, 3 and 5. ClonalFrame software was used for statistical analysis for Chapter 4.

Acknowledgements

My PhD supervisors Claire Jenkins, John Wain and Tim McHugh for their guidance throughout my thesis.

Professor Iruka Okeke, who gave me the opportunity to come to Haverford University and guided me in my thesis throughout. The experience I gained in America led me to go to Sierra Leone in response to the Cholera outbreak and develop an enteric laboratory.

My boss Claire Jenkins for her amazing support and giving me time to complete my part-time PhD without making me feel guilty for not being in the laboratory.

Jason, for his on going support and ridiculous amount of patience with my constant 13 hour days and weekends lost, though I did enjoy all of the dinners you made for me because I was studying.

My mum for taking the time to proof read my PhD, a feat for many scientists.

My Dad for provision of laptops and technical supplies, I have enough USB sticks to start a shop, many of which have more memory than my work computer.

Val Bevan, previous Head of the training committee and mentor who advised me on my career and enabled me to get funding to complete it.

My work colleagues in the laboratory at Public Health England for their understanding and support during this period of study. Though when I come back to the laboratory full time and start quality improvement, you will wish I was still doing my PhD.

Chapter One

Introduction

Chapter 1 – Introduction

1.0 What are *Escherichia coli?*

1.1.1 The species *E. coli*

Strains of *Escherichia coli* belong to a diverse bacterial species that reside in the human gut. Some strains can be protective and form part of the commensal gut flora while others are pathogenic and cause us harm. *E. coli* was first isolated during a study of faecal flora of neonates. It was originally described as *Bacterium coli commune* by Theodor Escherich in 1885 and renamed *Escherichia coli* in 1919 (Castellani and Chalmers 2005). *E. coli* colonises our gastrointestinal (GI) tract shortly after birth (Escherich 1885;Robins-Browne 1987), competing with other organisms in a hostile environment, until our deaths. *E. coli* is the type species of the Enterobacteriaceae family, are ubiquitous in nature and capable of adapting to environmental pressure, including exposure to antibiotics. (Bailey *et al.* 2010).

1.1.2 Pathotypes of *E. coli*

E. coli are the 'Jekyll and Hyde' of the bacterial community, and are both protective and pathogenic. Studies have revealed how *E. coli* has evolved into multiple lineages causing varied types of infection via multiple and diverse pathogenic mechanisms (Kaper *et al.* 2004;Wirth *et al.* 2006). Certain *E. coli* pathotypes are defined by the clinical presentation and site of infection. Extra-intestinal *E. coli* (ExPEC) are isolated from extra-intestinal sources such as blood cultures and wounds and also includes Uropathogenic *E. coli* (UPEC) where infection is in the urethra or bladder. Diarrhoeagenic *E. coli* (DEC) cause infection in the gut, are divided into additional pathotypes and are generally defined by the way they they adhere and/or interact with human epithelial cells (HEp-2 cells) or by the presence of certain pathogenicity factors as summarised in table 1.1.

1.1.3 Pathogenicity of *E. coli*

There are multiple mechanisms of pathogenicity for different DEC including adherence mechanisms, production of toxins and iron acquisition mechanisms and the majority of strains can be placed in one or more of the groups described in detail below:

1.1.3.1 - EPEC

Enteropathogenic *E. coli* (EPEC) belong to a group of bacteria collectively known as attaching and effacing (A/E) pathogens with their ability to form distinctive lesions on the surfaces of intestinal epithelial cells. The A/E phenotype is related to the presence of a 35-kb cluster of virulence genes called the locus of enterocyte effacement (LEE) and encoded on a chromosomal pathogenicity island (PAI) (Croxen *et al.* 2013). Further classification of EPEC depends on the presence or absence of the *E. coli* adherence factor plasmid (pEAF) and is termed 'typical' or 'atypical' respectively (Trabulsi *et al.* 2002).The attaching and effacing (*eae*) gene (encoded on the LEE) produces an intimin 94-kDa protein which forms pedestal-like structures via cytoskeletal changes enabling the bacteria to intimately attach to the intestinal epithelial cells (McDaniel *et al.* 1995). Though EPEC are not known for producing toxins, they do produce a protein called lymphostatin which inhibits lymphocyte activation (Klapproth *et al.* 2000). Diarrhoea is likely to result from multiple mechanisms such as the loss of absorptive surfaces resulting from the microvillus effacement, increased intestinal permeability and intestinal inflammation (Kaper *et al.* 2004). EPEC symptoms include diarrhoea that is often accompanied by fever, vomiting, dehydration and onset of diarrhoea in human volunteers can be as early as 3 hours with wild-type bacteria (Shariff *et al.* 1993). EPEC infection is usually associated in children and occurrence of EPEC infections decreases with age, infection can be acute or persistent for more than 2 weeks and generally treated with rehydration therapy (Croxen *et al.* 2013). EPEC is found across the globe and morbidity and mortality is higher in developing countries due to malnutrition and poor healthcare. Transmission is faecal oral and the reservoir for EPEC can

vary from a human source for typical EPEC (Levine 1978) to both humans and animals for atypical EPEC (Croxen *et al.* 2013).

1.1.3.2 - STEC

Shiga Toxin-producing *E. coli* (STEC) is classified by the presence of the Shiga toxin 1 or 2 gene (*stx*1 or *stx*2) which is typically acquired by a lambdoid bacteriophage. STEC typical strains also have the *eae* gene enabling intimate adherence and A/E phenotype. STEC alternative nomenclature is vero-cytotoxic producing *E. coli* (VTEC) where the *stx* genes are referred to as *vtx* genes. This thesis will refer to the nomenclature as STEC. The key virulence factor is the shiga-toxin encoded by the *stx* gene. Stx is produced in the colon and travels to the kidneys via the bloodstream, where it damages renal endothelial cells and obstructs the microvasculature via direct toxicity and induction of local cytokine and chemokine production (Andreoli *et al.* 2002). This damage can lead to haemolytic-uremic syndrome (HUS) which is characterised by haemolytic anaemia (anaemia caused by destruction of red blood cells), acute kidney failure (uraemia) and a low platelet count (thrombocytopenia). Enterohemorrhagic *E. coli* (EHEC) is a subset of STEC and described by its association with haemorrhagic colitis and HUS. STEC symptoms can range from mild watery diarrhoea to bloody diarrhoea and incubation period before onset of symptoms for STEC O157:H7 is about three days (Bell *et al.* 1994). STEC infections are found in all ages but the immunocompromised have a higher risk of developing severe symptoms. STEC tend to have higher rates of infection in developed countries with a variable incident rate across Europe, though the European Center for Disease Prevention and Control recommends not comparing incidences between countries due to the difference in detection methods (European Centre for Disease Prevention and Control and European Food Safety Authority 2011). Current treatment recommendations are to maintain hydration to prevent thrombotic complications (Serna and Boedeker 2008). Transmission is faecal oral and the reservoir for STEC is mainly associated with

animals, particularly ruminants such as cattle and exposure to petting farms (Chase-Topping *et al.* 2007).

1.1.3.3 - ETEC

Enterotoxigenic *E. coli* (ETEC) is classified by the presence of enterotoxins (heat-labile , LT and/or heat stable, ST). The ST gene encodes for small ST peptides which cause watery diarrhoea by mimicking the hormone guanylin and ultimately resulting in the secretion of chloride into the intestinal lumen and preventing absorption (Weiglmeier *et al.* 2010). The LT has multiple variants and is found encoded in the chromosome and may have been acquired by a phage (Jobling and Holmes 2012). The LT gene encodes for a large toxin which is assembled as an AB_5 toxin and about 80% identical to cholera toxin. The pentameric B subunits bind to GM1 gangliosides at lipid rafts to deliver the catalytic A subunit inside the cell. This ultimately results in increasing cAMP levels resulting in electrolytes and fluid loss into the intestinal lumen causing profuse watery diarrhoea (Kreisberg *et al.* 2011). Adherence to the small bowel mucosa is facilitated by over 20 antigenically diverse colonisation factors (Wolf 1997). ETEC is a major cause of travellers' diarrhoea and highly prevalent in developing countries particularly in regions such as Latin America, Africa and certain regions of Asia in which ETEC has shown to be the aetiological agent in approximately 30% of cases (Shah *et al.* 2009). Oral rehydration fluid should be sufficient to treat this self-limiting disease but treatment is not always easily accessible in developing countries and can lead to mortality particularly in children (Nataro and Kaper 1998). Transmission is faecal oral and exposure to ETEC is usually from contaminated foods and water. ETEC has been shown to survive in freshwater for three months (Lothigius *et al.* 2010).

1.1.3.4 - EIEC

Enteroinvasive *E. coli* (EIEC) are facultative intracellular pathogens, shares biochemical, genetic and pathogenic properties with *Shigella* and both are the etiological agents of bacillary dysentery or shigellosis (Nataro & Kaper 1998). This highly invasive pathogen has multiple virulence determinants

that are encoded by both chromosomal and plasmid loci and include a type 3 secretion system (Mxi-Spa proteins), chaperones (Ipg sub-units and Spa15), transcriptional regulators (Vir F, VirB and MxiE) and translocators (ipAB, IpaC, and IpaD) (Schroeder and Hilbi 2008). EIEC are recognised as containing the *Ipa* invasive effector genes and other genetic elements present on a large invasive mosaic plasmid (Buchrieser *et al.* 2000). The detection of the *ipaH* gene by is used by the UK reference laboratories for characterisation of this pathotype. These virulence factors facilitate EIEC to invade the colonic epithelial cell, lyse the phagosome and move through the cell by nucleating actin microfilaments (Kaper *et al.* 2004). EIEC can vary in its severity and symptoms will range from invasive inflammatory colitis and occasionally dysentery to watery diarrhoea that is indistinguishable from other *E. coli* pathogens (Nataro & Kaper 1998). The epidemiology of EIEC is not well known as the focus is to detect *Shigella,* which tends to have more severe clinical manifestations (Croxen *et al.* 2013). EIEC hasn't been found in surveillance studies in the UK over the past 15 years (Tam *et al.* 2012a;Wilson *et al.* 2001) and is not thought to play a major role in diarrhoeal disease. EIEC is generally self-limiting and treatment would involve oral rehydration therapy. Transmission is faecal oral and exposure to EIEC is via contamination of food, it is likely that EIEC is host restricted like *Shigella,* as no particular reservoir has been described.

1.1.3.5 - DAEC

Diffusely adherent *E.coli* (DAEC) has been classified by its diffuse adherence (DA) to cultured epithelial HEp-2 cells (Scaletsky *et al.* 1984). A well described pathogenicity factor is the presence of *Afa* genes that encode Afa/Dr adhesins. These adhesins bind to brush border-associated decay-accelerating factor (DAF) which induces cytoskeleton rearrangement destroying the microvilli (Le and Servin 2006;Servin 2005). It is uncertain if this mechanism causes diarrhoea but destruction of the microvilli can result in increased permeability. DAEC is associated with watery diarrhoea which can become persistent and increase in severity of disease in young children (Servin 2005). Detection methods are still being developed for diarrhoeagenic strains of DAEC and there is no current universal method, thus the epidemiology of diarrhoeagenic DAEC remains unclear. Rehydration

therapy is currently the only treatment recommended for watery diarrhoea caused by DAEC. Transmission is faecal oral but is it unknown how DAEC is transmitted or its reservoir (Croxen *et al.* 2013).

1.1.3.6 - EAEC

Enteroaggregative *E. coli* (EAEC) is classified by its aggregative adherence pattern on HEp-2 cells (Nataro *et al.* 1987) (1.10.1). Colonisation of the intestinal mucosa and colon is carried out by a combination of aggregative adherence fimbriae (Boisen *et al.* 2008;Czeczulin *et al.* 1997;Nataro *et al.* 1993) and a dispersin protein (Baudry *et al.* 1990). This is encoded by the anti-aggregator transporter *aat* gene, which forms a loosely associated layer on the surface of EAEC strains, perhaps facilitating the spread across the mucosal layer as it counters the strong aggregating affects (Kaper *et al.* 2004). There are several toxins that EAEC can produce including a plasmid encoded toxin which has enterotoxic activity and can lead to changes in the epithelial-cell and cytoskeletal protein spectrin (Navarro-Garcia *et al.* 1999), an enteroaggregative *E. coli* heat-stable toxin which could contribute to watery diarrhoea (Savarino *et al.* 1993) and a *Shigella* enterotoxin 1 thought to contribute to secretory diarrhoea (Noriega *et al.* 1995). However, not all EAEC strains produce all, if any of these toxins and the encoding genes for numerous adhesions and proteins associated with virulence are highly variable among strains (Croxen *et al.* 2013). Even the site of infection is not uniform, where some strains infect the small bowel while others infect the small bowel and colon (Okhuysen and DuPont 2010). EAEC can cause acute and persistent diarrhoea and EAEC strains associated with persistent diarrhoea in children were shown to have significant elevations in fecal lactoferrin, interleukin (IL)-8 and IL-1 beta. It is hypothesised that stimulation of IL-8 from intestinal epithelial cells may be caused by EAEC flagellin (Steiner *et al.* 1998;Steiner *et al.* 2000). EAEC is a global pathogen (Okeke and Nataro 2001) but is particularly severe in developing countries where EAEC has been shown to impair growth while malnutrition worsens with persistent EAEC infection (Roche *et al.* 2010). Though EAEC pathogenicity factors are heterogeneous in nature, a general three-part model of EAEC pathogenesis has emerged: (i) adherence to the intestinal mucosa, (ii)

production of enterotoxins and cytotoxins, and (iii) mucosal inflammation (Estrada-Garcia and Navarro-Garcia 2012). EAEC is usually self-limiting and rehydration therapy is recommended but antibiotics will be considered particularly in persistent cases. Transmission is faecal oral and though atypical EAEC (strains lacking *aggR*) have been found in animals (Uber *et al.* 2006), the reservoir of clinically relevant EAEC is generally thought to be human. Recent screening of EAEC from environmental and animal sources confirmed that typical EAEC are not found in animals (Chattaway *et al.* 2014a). The characterisation of EAEC is further described in section 1.9.

The genes associated with the different diarrhoeagenic pathotypes (Table 1.1) enable interactions with eukaryotic cells in very specific and individual ways (Figure 1.1). Though, different pathotypes have been defined, the question of what defines *E. coli* is still being investigated today. This introduction describes how the definition of this organism has evolved over the last century with the continued introduction of novel scientific techniques.

Table 1 1 Description of the different pathotypes of *Escherichia coli*

Abbreviation	Meaning	Description	Disease
DEC	Diarrhoeagenic *E. coli*	Any defined group of *E. coli* which has been associated with the ability to cause diarrhoea	Intestinal/ Diarrheal
DAEC	Diffusely-adherent E. coli	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by a specific pattern of diffuse adherence on HEp-2 cells.	Intestinal/ Diarrheal
EAEC	Enteroaggregative *E. coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by a specific pattern of aggregative aggregation on HEp-2 cells or the presence of the plasmid borne anti-aggregative transporter (*aa*t) or EAEC regulatory (*Agg*R) gene.	Intestinal/ Diarrheal
EIEC	Enteroinvasive *E. coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by the presence of the ipaH invasion gene also found in *Shigella*.	Intestinal/ Diarrheal
EPEC	Enteropathogenic *E. coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea originally defined as specific serotypes and by localised adherence on HEp-2 cells but now by the presence of certain virulence factors including the locus of enterocyte effacement (*eae*) gene.	Intestinal/ Diarrheal
ETEC	Enterotoxigenic *E. coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by the presence of heat stable or heat labile toxins	Intestinal/ Diarrheal
VTEC	Verocytotoxin-producing *E. coli* Also referred to as Shiga toxin-producing *E. coli*	A group of E. coli which been associated with the ability to cause diarrhoea defined by the presence of a verocyto toxin gene, *vtx*, which has activity against cultured vero cells	Intestinal/ Diarrheal
EHEC	Enterohemorrhagic *E. coli*	VTEC/STEC causing bloody diarrhoea, haemorrhagic colitis and haemolytic uremic syndrome (HUS) characterised by acute renal failure, haemolytic anaemia (anaemia due to haemolysis) and thrombocytopenia (low number of platelets)	Intestinal/ Diarrheal
ExPEC	Extraintestinal Pathogenic *E. coli*	Any defined group of E. coli which has been associated with the ability to cause extra-intestinal disease.	Extra-intestinal/ Septecemia/ Menigitidis/ Urinary
UPEC	Uropathogenic *E. coli*	Any defined group of *E. coli* which has been associated with the ability to cause urinary tract disease.	Extra-intestinal/ Urinary
CDEC	Cell detaching *E. coli*	A group of *E. coli* which has the capacity to detach tissue culture cells from solid supports in adherence assays or in a cell-detaching assay	Intestinal/ Diarrheal/ Urinary
CDTEC	cytolethal distending toxin-producing *E. coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by the presence of a toxin called cytolethal distending toxin. This causes morphological changes on HEp-2 cell lines including elongation of cells at 24 h; this is followed by progressive cellular distension and cytotoxicity	Intestinal/ Diarrheal

Figure 1. 1 Attachment and interaction of *E. coli* pathotypes with Eukaryotic Cells

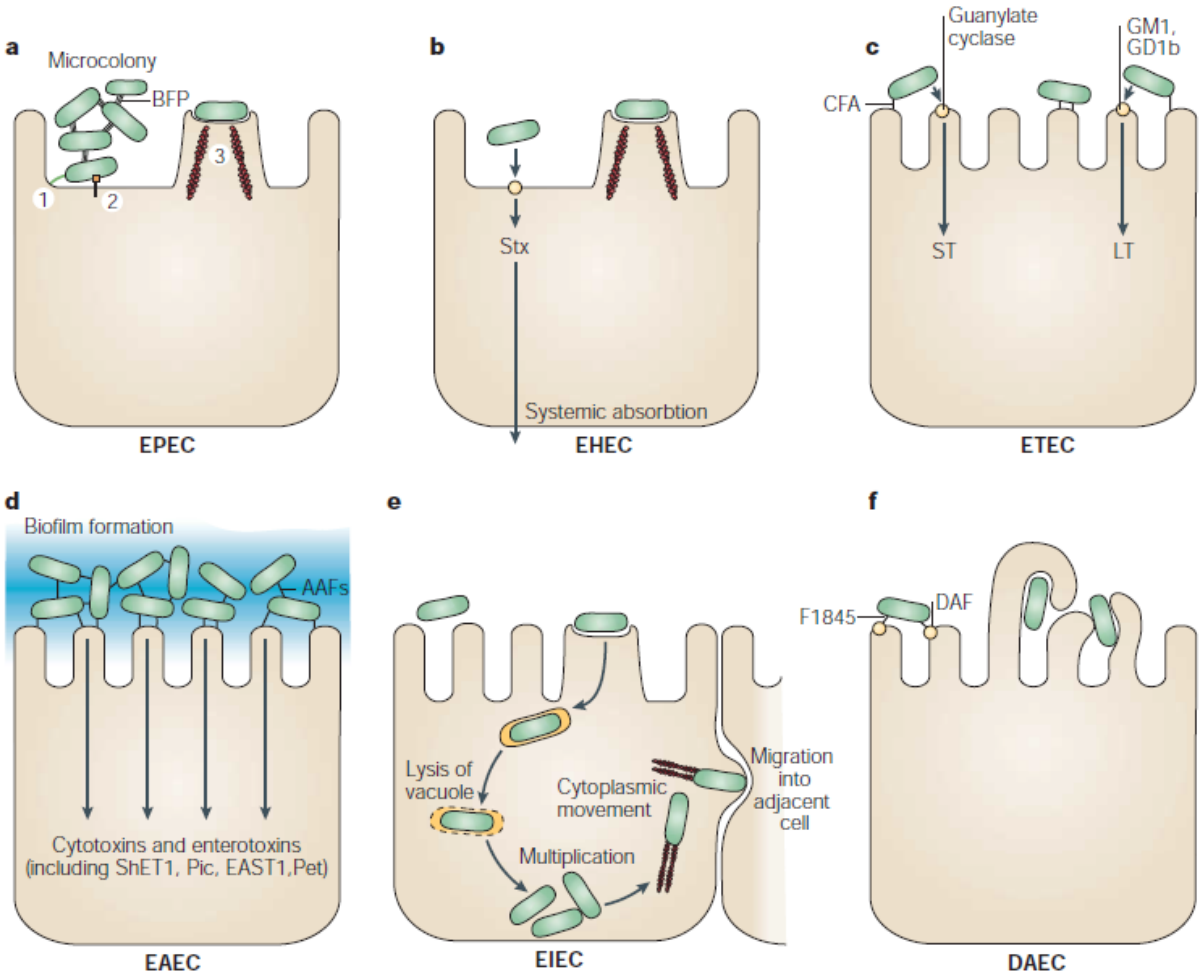


Figure 1 | **Pathogenic schema of diarrhoeagenic *E. coli*.** The six recognized categories of diarrhoeagenic *E. coli* each have unique features in their interaction with eukaryotic cells. Here, the interaction of each category with a typical target cell is schematically represented. These descriptions are largely the result of *in vitro* studies and might not completely reflect the phenomena that occurs in infected humans. **a** | EPEC adhere to small bowel enterocytes, but destroy the normal microvillar architecture, inducing the characteristic attaching and effacing lesion. Cytoskeletal derangements are accompanied by an inflammatory response and diarrhoea. 1. Initial adhesion, 2. Protein translocation by type III secretion, 3. Pedestal formation. **b** | EHEC also induce the attaching and effacing lesion, but in the colon. The distinguishing feature of EHEC is the elaboration of Shiga toxin (Stx), systemic absorption of which leads to potentially life-threatening complications. **c** | Similarly, ETEC adhere to small bowel enterocytes and induce watery diarrhoea by the secretion of heat-labile (LT) and/or heat-stable (ST) enterotoxins. **d** | EAEC adheres to small and large bowel epithelia in a thick biofilm and elaborates secretory enterotoxins and cytotoxins. **e** | EIEC invades the colonic epithelial cell, lyses the phagosome and moves through the cell by nucleating actin microfilaments. The bacteria might move laterally through the epithelium by direct cell-to-cell spread or might exit and re-enter the baso-lateral plasma membrane. **f** | DAEC elicits a characteristic signal transduction effect in small bowel enterocytes that manifests as the growth of long finger-like cellular projections, which wrap around the bacteria. AAF, aggregative adherence fimbriae; BFP, bundle-forming pilus; CFA, colonization factor antigen; DAF, decay-accelerating factor; EAST1, enteroaggregative *E. coli* ST1; LT, heat-labile enterotoxin; ShET1, *Shigella* enterotoxin 1; ST, heat-stable enterotoxin.

Figure 1.1: Figure taken from (Kaper *et al.* 2004) describing the different interactions of the six recognised categories of diarrhoeagenic *E. coli* with eukaryotic cells.

1.2 In the Beginning – Microscopy

The first practical microscope was built by Anton Van Leeuwenhoek in Holland and led to the first description of bacteria in 1676 (Gest 2004). As technology improved, microscopy was pivotal to the study of bacteria in the environment and their association with disease. Theodor Escherich was probably the first paediatric infectious diseases physician who studied the intestinal flora of neonates. Using basic culture techniques, Gram stain and development of anaerobic conditions, he was able to describe *Bacterium coli commune* as a facultative Gram negative coli form that produced gas during fermentation (Escherich 1885;Shulman *et al.* 2007). In 1919, Castellani and Chalmers proposed the designation *Escherichia coli* (Castellani A 1919). The family name Enterobacteriaceae, was first proposed in 1937 by Rahn compromising multiple taxa including *E. coli* (Rahn 1937) still within this family today. Although the proposal of the Enterobacteriaceae group had several nomenclature flaws and was made illegitimate by the International Journal of Systematic Bacteriology (Farmer, III *et al.* 1980), the name was widely used and the Judical Commission voted it to be conserved in 1958. The official name of *E. coli* was also sanctioned as the name of the common colon bacillus discovered by Escherich (Judicial Commission of the International committee on Bacteriological Nomenclature 1958).

1.3 Definition of *E. coli* by biochemistry

As biochemical properties of *E. coli* and other organisms were investigated, multiple taxa were described and classification systems proposed. The validity of the taxonomic groups within this family have been much debated, with the genus *Shigella* being labelled as a 'dubious group' and suggestions that any attempt to separate lactose fermenters as individual groups would be superficial (Borman *et al.* 1944). In 1944, *E. coli* was grouped simply as genus II colobactrum by its ability to ferment lactose and produce gas with other genus separated out by distinct properties (Borman *et al.* 1944), this was an attempt to simplify the growing numbers of species that are

currently called coliforms in clinical laboratories. These biochemical properties are the basis of many of the groups we see in the Enterobacteriaceae today (Edwards and Ewing 1962).

Bergy's Systematic Manual continually revises speciation of all bacteria and with each edition, new species are added or reassigned. Due to the complex and interconnected relationship of the biochemical properties of this family there is a great reluctance to re-define this group and reclassify mis-assigned subgroups.

1.4 Linking pathogenic *E. coli* via serology

It is now well known that *E. coli* can be found in healthy and diseased individuals but the first indication that there could be harmful diarrhoeagenic and harmless commensal strains was described by Lesage in 1897 on serological grounds. The study showed that serum (the plasma from blood that contains antibodies) from a diarrhoeal patient agglutinated with *E. coli* from other patients during an epidemic but did not react with *E. coli* obtained from healthy children or other enteric pathogens (Lesage 1897;Robins-Browne 1987). Between 1927-1953, a large number of outbreaks of infantile diarrhoea were recorded. Strains of *E. coli* associated with the earlier outbreaks were designated dyspepsiekoli, α, β or D433. By the 1950's serogroups of *E. coli* O55 and O111 were being frequently isolated from cases (Robins-Browne 1987). These early studies led onto the serological typing schemes that we have today. The *E. coli* serotyping scheme is currently organised according to 185 somatic (O) antigens, 49 flagellar (H) and 103 capsular (K) antigens (Kauffmann 1947;Parija 2009).

The characterisation of *E. coli* in relation to serology is an organism that agglutinates against *E. coli* sera. *E. coli* serum is made by using plasma from blood which has been exposed to a known serogroup of *E. coli* and therefore has specific antibodies in the sera. Biochemically defined *E. coli* strains that do not react with the known sera are deemed O antigen unidentifiable. Interestingly, Kauffmann states that **'In the coli group, cultural tests play but a minor role, so that the type division has to rest on a serological basis. The prevailing classification of the coli strains, based on cultural criteria (especially fermentation tests) should therefore be abandoned'**. (Kauffmann 1947). However, in reference microbiology the biochemical confirmation of *E. coli* is vital due to cross reactions with other Enterobacteriaceae giving non-specific false positive agglutination, and both biochemistry and serology should be used for confirmation of identification.

1.5 How the study of genetics complicated the definition of *E. coli*

1.5.1 DNA-DNA re-association

E. coli are defined using a combination of biochemical and serological profiles. This approach clearly distinguishes *E. coli* from *Shigella sp*. Further studies attempted to define the species genetically.

DNA-DNA re-association (where a labelled reference strain nucleotide sequence binds against a test strain) was used to assess genetic relatedness of *E. coli* O112 against different *E. coli* and *Shigella sp.* (Brenner *et al.* 1972). At the optimal re-association temperature of 60^{o}C to form stable re-associations, *E. coli* strains relatedness varied between 85-100%. *Shigella sp*. re-associated with *E. coli* K12 between 80-89%, it was suggested that *Shigella sp.* should be considered a different species within *E. coli*, rather than a separate genus, as no other strains from any other genus of enteric bacteria exhibit greater than 50% relatedness to *E. coli* at 60^{o}C (Brenner *et al.* 1969). If *E. coli* was solely defined as a group of organisms that re-associate at 60^{o}C at 85-100% this would include closely related *Shigella* sp (as defined by biochemistry and serology). However, even today *Shigella sp*. is still considered to be a separate genus.

The Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics acknowledges the importance of applying chemotaxonomic approaches, both structural (phenetic) and phylogenetic (evolutionary), when inferring or proposing hierarchical levels. They define a species containing strains of approximately 70% or greater DNA-DNA relatedness with phenotypic characteristics agreeing with the definition (Wayne *et al.* 1987). Even with the advancement of genetic analysis, phenotypic characteristics still play an underlying role in identification.

1.5.2 16S rRNA Sequencing Analysis

The search of an alternative to DNA-DNA re-association methods led to the sequencing of 16S ribosomal RNA, a component of the 30S small subunit of prokaryotic ribosomes. The gene encoding

this region is 16S rDNA and is found in all bacterial species and can be used to assess phylogenetic relationships (Woese and Fox 1977). The 16S rDNA gene has multiple helices and heterogeneous positions (Figure 1.2) enabling differentiation of bacterial species and is useful in other studies such as microbial ecology (Case *et al.* 2007).

Figure 1. 1 Structure of the 16S rDNA gene

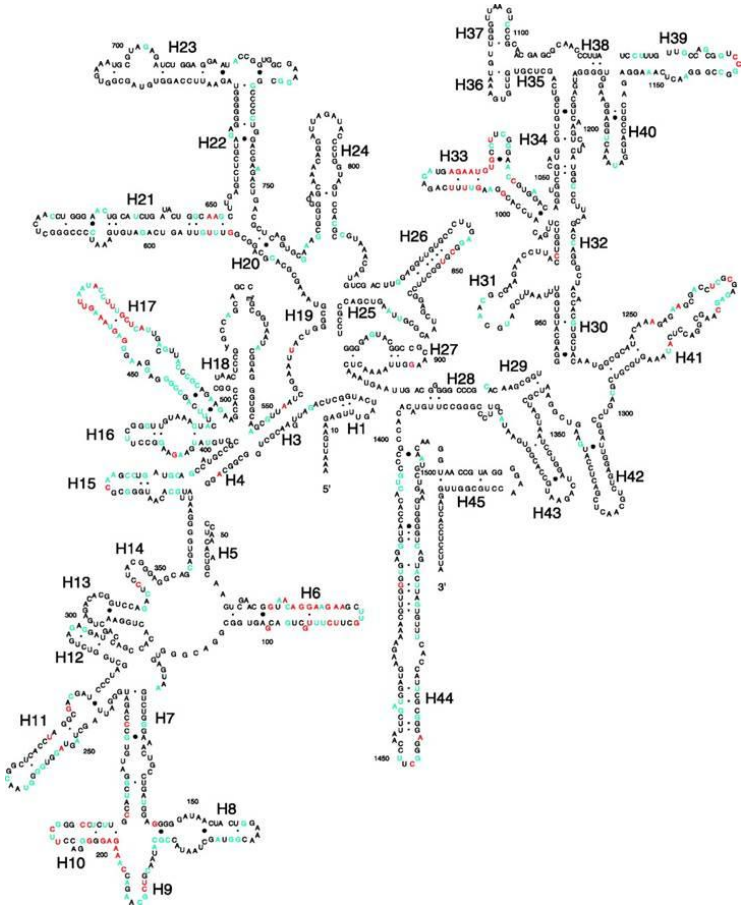


Figure 1.2: Helices (H) are numbered as described by Cannone *et al*.(Cannone *et al.* 2002). Positions indicated in blue are heterogeneous in one species, and those in red are heterogeneous in two or more species (Case *et al.* 2007). The black positions are homogenous in different species and are conserved regions, it is these regions that are used to design primers so that the gene can be amplified and sequenced. The differences can be used to speciate organisms but are not useful when different species are very similar such as *E. coli* and *Shigella*.

Open access of a large bacterial database to search 16S rDNA gene sequences is a useful tool in genus identification but the inability to distinguish closely related groups suggest that 16S rRNA sequencing is not an appropriate method to replace DNA-DNA re-association for the delineation of species and measurement of intraspecies relationships (Stackebrandt and Goebel 1994). This is accurately reflected when comparing *E. coli* and *Shigella sp.* (Figure 1.3), which are not differentiated by this method and confirms that 16S rRNA sequencing is not suitable to analyse inter-strain relationships (Lukjancenko *et al.* 2010).

Figure 1 2 **Phylogenetic tree based on 16S rRNA sequences**

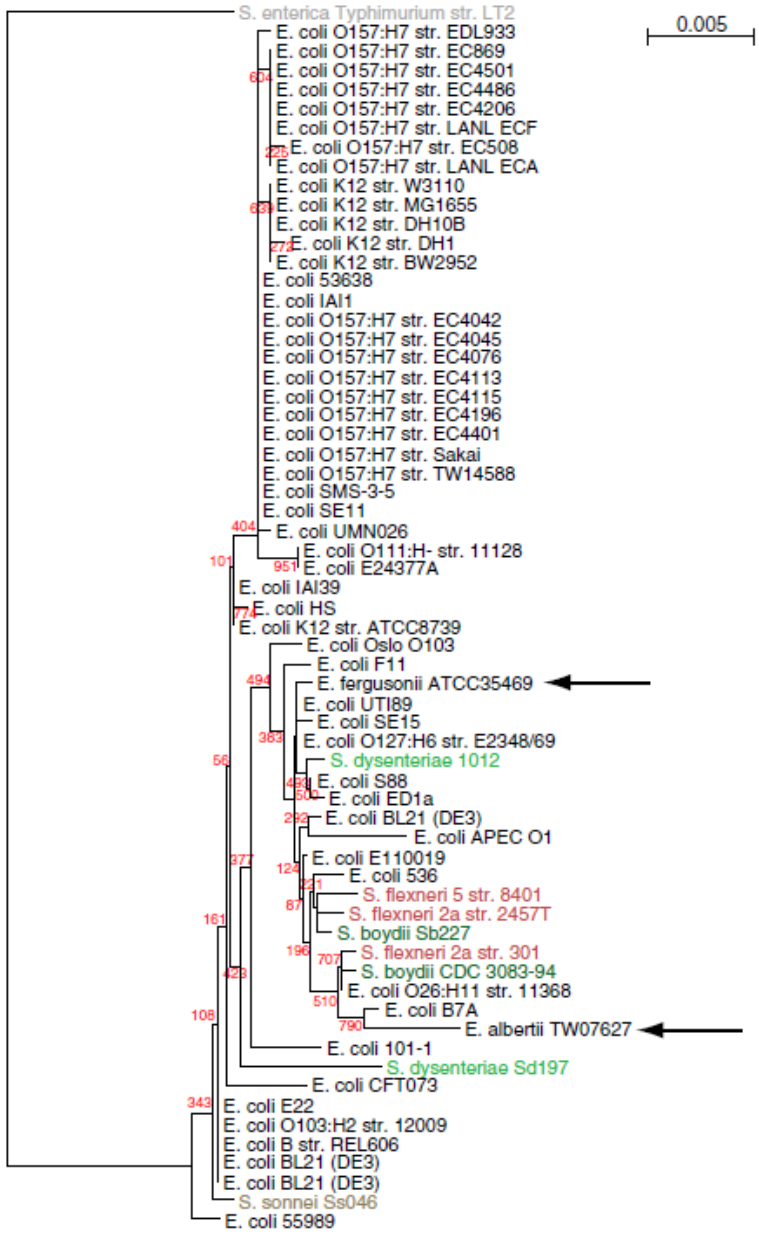


Figure 1.3: (Lukjancenko *et al.* 2010). Tree of 61 sequenced *E. coli* (black) and related species (colored), based on the alignment of the 16S rRNA gene sequence. Apart from Shigella spp., the genes from *E. albertii* and *E. fergusonii* are also included (arrows). The 16S rRNA gene of *S. enterica* Typhimurium LT2 was used as the root. Bootstrap values, indicated in red, show that most nodes are predicted with uncertainty; nevertheless, the genera *Escherichia* spp. and *Shigella* spp. are not separated in this tree, and the three *Escherichia* species are also mixed

1.6 Looking at the Population Structure of *E. coli*

1.6.1 Multi-locus Enzyme Electrophoresis (MLEE)

Multi-locus enzyme electrophoresis (MLEE) had long been a standard method in eukaryotic population genetics (Ayala 1976;Lewontin 1974;Nei 1975). The observation that there were high levels of electrophoretic variation in enzymes essential to the normal functions of metabolism in eurkaryotes, sparked numerous efforts to explain the existence of this variation in prokaryotes. Essentially MLEE is based on the different electrophoretic motilities (EM) of chosen multiple core metabolic enzymes. The enzyme is electrophoresed on an agar gel and the alleles at each locus will define the EM of their products for the different amino acid sequences between each enzyme. This can be shown by how far the band (product) travels in a gel, a difference in how far the band has travelled can show if a mutation has occurred and an amino acid has been substituted affecting the net charge of the enzyme (Figure 1.4) (Selander *et al.* 1986).

Figure 1.3 **The electric mobility of different enzymes in MLEE**

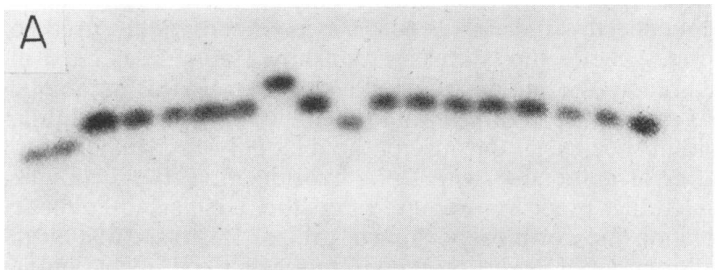


Figure 1.4: (Selander *et al.* 1986). Gel illustrating electrophoretic variation in the enzyme Mannitol 1-phosphate dehydrogenase in *E. coli* in 18 isolates. The different alleles can be clearly seen in the begining and the middle of the gel, as the enzyme has travelled to different points on the gel and therefore has a different electrophoretic motility.

A dendogram can then be generated from the matrix of pairwise differences between the electrophoretic types (ET) to show the relatedness between isolates. *E. coli* was the first bacterium for which population genetic techniques were introduced (Wirth *et al.* 2006). MLEE inferred evolutionary phylogenetic relationships and indicated that certain combinations of alleles occurred multiple times which was interpreted as indicating a clonal population structure with sporadic recombination amongst many bacteria including *E. coli* populations (Ochman and Selander

1984;Selander and Levin 1980). MLEE laid the foundation of population structure analysis for the bacterial kingdom but was not without its problems. The phenotype of the enzyme could be altered in response to environmental conditions such as phosphorylation, cofactor binding and cleavage of transport sequences, which could badly affect the reproducibility of MLEE results and limit the comparability between different laboratories. Also enzymes with different amino acid (AA) sequences may have a similar EM and any silent mutations where the DNA sequence varies but the AA code is unaltered will also give the same EM. *E. coli* definition via MLEE methodology is a group of organisms that fit within a defined set of EM profiles. The development of an equivalent method, multi-locus sequence typing (Wirth *et al.* 2006) enabled a portable and comparable DNA sequence database that could be easily compared to between different laboratories.

1.6.2 Multi-locus Sequence Typing (MLST)

Multi-locus sequence typing (MLST) is the sequencing of a defined set of multiple housekeeping genes. Each gene will have a defined set number of base pairs that is called an allele, each variant of that allele (i.e. any difference in the sequence of nucleotides) has an allele number assigned. The combination of allele numbers for the seven different genes leads to a given sequence type (ST). For example, allelic profile 10-11-4-8-8-8-2 is ST10. If the allele number differs by one number then this is designated a single locus variant (SLV), if it differs by two numbers then it is a double locus variant (DLV) and if it differs by three it is a triple locus variant (TLV). A complex consists of any strains that are the same ST or a SLV of that group. For example for ST10, an allelic profile of 10-11-4-1-8-8-2 would be ST34 and is an SLV (as the *icd* gene is different) and therefore part of the ST10 complex (Cplx).

There are several MLST schemes available; in this study the University of College Cork (http://mlst.ucc.ie/mlst/) MLST scheme for *E. coli* was used.

The seven housekeeping genes (Table 1.2) chosen are located throughout the chromosome with eight to 20% of the nucleotides being polymorphic (Figure 1.5), which represents the natural genetic

diversity of the genome (Wirth *et al.* 2006). Sequence typing enables us to take a snapshot of the core genetic relatedness of strains of *E. coli* and *Shigella sp*. Minimal spanning trees show how the two groups have clonally expanded into multiple complexes. In this analysis *Shigella sp*. does mainly form distinct clusters from *E. coli* indicating that although they may have a common ancestor, their core genetic content can be resolved (Figure 1.6). Population structure analysis indicates *E. coli* to be a distinct group of organisms forming multiple clonal complexes.

Table 1.2 Housekeeping genes used in MLST

Name	Gene	Function
***Adk* (536bp)**	Adenylate kinase	Catalyzes the reversible transfer of the terminal phosphate group between adenosine triphosphate (ATP)and adenosine monophosphate (AMP). This small ubiquitous enzyme involved in the energy metabolism and nucleotide synthesis, is essential for maintenance and cell growth.
***fumC* (469bp)**	fumarate hydratase	The FH gene provides instructions for making an enzyme called fumarase (also known as fumarate hydratase). Fumarase participates in an important series of reactions known as the citric acid cycle or Krebs cycle, which allows cells to use oxygen and generate energy. Specifically, fumarase helps convert a molecule called fumarate to a molecule called malate.
***gyrB* (460bp)**	DNA gyrase subunit B	DNA gyrase is an essential enzyme that regulates the DNA topology in bacteria. It belongs to the type II DNA topoisomerase family and is responsible for the introduction of negative supercoils into DNA at the expense of hydrolysis of ATP molecules
***Icd* (518bp)**	isocitrate/isopropylmalate dehydrogenase	Isocitrate dehydrogenase is an enzyme that participates in the citric acid cycle. It catalyzes the third step of the cycle: the oxidative decarboxylation of isocitrate, producing alpha-ketoglutarate (α-ketoglutarate) and CO_2 while converting NAD+ to NADH. In aerobic organisms, the citric acid cycle is part of a metabolic pathway involved in the chemical conversion of carbohydrates, fats and proteins into carbon dioxide and water to generate a form of usable energy.
***Mdh* (452bp)**	malate dehydrogenase	Malate dehydrogenase is an enzyme in the citric acid cycle that catalyzes the conversion of malate into oxaloacetate (using NAD+) and vice versa (this is a reversible reaction). Malate dehydrogenase is also involved in gluconeogenesis, the synthesis of glucose from smaller molecules.
***purA* (478bp)**	adenylosuccinate dehydrogenase	Adenylosuccinate synthase (or adenylosuccinate synthetase) is an enzyme that converts IMP to adenylosuccinate. It is involved in purine synthesis,.Purines are Biologically synthesized as nucleotides (bases attached to ribose 5-phosphate).
***recA* (510bp)**	ATP/GTP binding motif	RecA is a 38 kilodalton *Escherichia coli* protein essential for the repair and maintenance of DNA. RecA has a structural and functional homolog in every species in which it has been seriously sought and serves as an archetype for this class of homologous DNA repair proteins. Since it is a DNA-dependent ATPase, RecA contains an additional site for binding and hydrolyzing ATP.. It is also a binding motif for guanosine triphosphate (GTP)

Table 1.2: Table listing the seven MLST housekeeping genes and their functions

Figure 1.4 Genomic location and polymorphisms of MLST genes

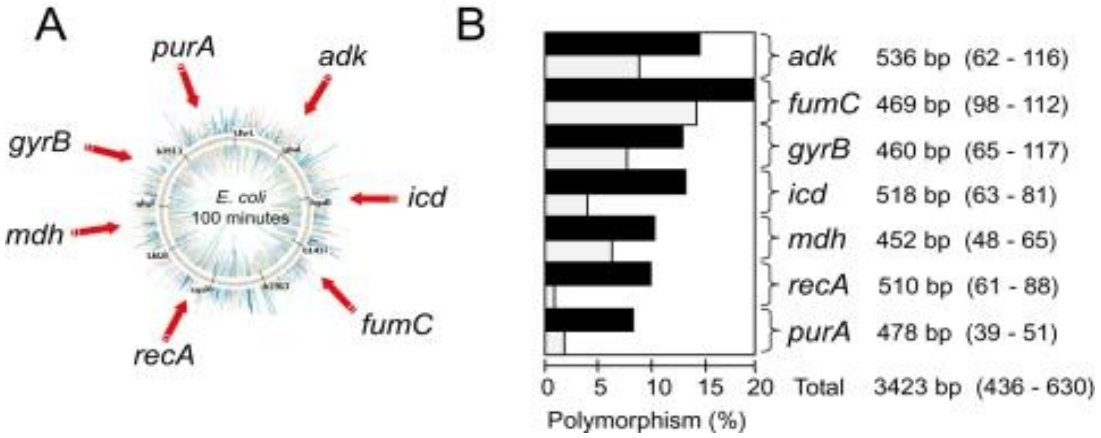


Figure 1.5: (A) Genomic locations and (B) genetic diversity of seven housekeeping genes. (B) Polymorphism levels for each gene are indicated in the histogram in which black bars reflect nucleotide polymorphisms and grey bars indicate amino-acid polymorphisms. Each gene symbol is followed by the length of the sequenced gene fragment (informative sites – polymorphic sites) (Wirth *et al.* 2006).

Figure 1.5 Minimal spanning tree of *E. coli* and *Shigella* sp

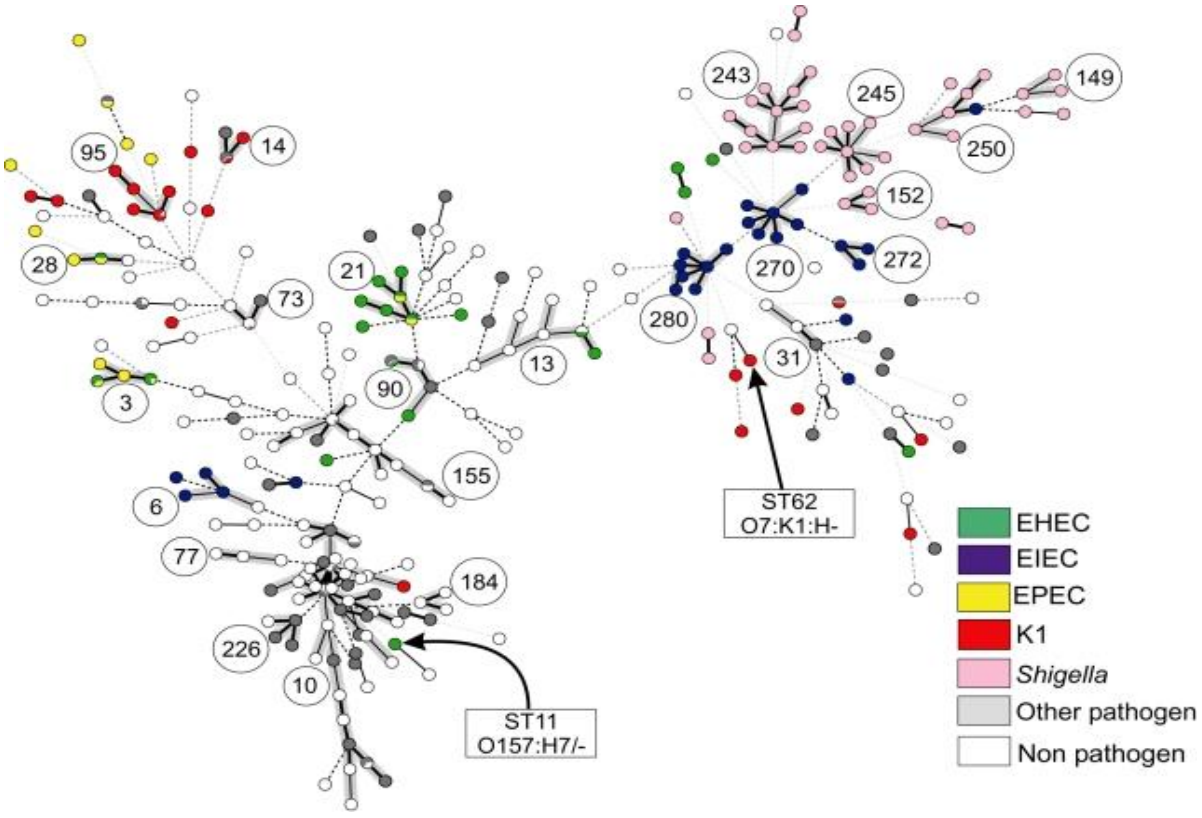


Figure 1.6: Pathogenic types within an MS_{TREE}. Each ST is represented by a dot. Dots with uniform colours indicate that all isolates were of the same pathogen type (see legend) while the small pie charts indicate the fraction of isolates belonging to each pathogen type. Circled numbers indicate ST complexes, whereas arrows indicate STs 11 and 62. Black lines connecting pairs of STs indicate that they share six (thick lines), five (thin) or four alleles (dotted). Grey, dotted lines connecting pairs of STs of increasing line length indicate that they share three to one alleles respectively. In addition, the lines connecting the STs within an ST complex are shaded in grey (Wirth *et al.* 2006).

1.7 Evolving genetic analysis with high throughput sequencing

Whole Genome Sequencing (WGS) involves the sequencing of the entire bacterial DNA content, including the core (essential housekeeping genes which relate to function vital to the survival of the organism and shared by all members of the group) and accessory (dispensable genes, not present in all members of the group, that enhance survival capabilities of the organism but not essential for survival) genomes.

The first sequence of *E. coli* K12 consisting of 4,639,221 base pairs (bp) was published in 1997 (Blattner *et al.* 1997) after six years of sequencing using multiple techniques including traditional Sanger sequencing methodology and long range PCR to sequence any gaps. This project resulted in a fully annotated genome and highlighted at the time that 38 % of genomic data had no attributed function.

Having an accurate fully annotated genome laid an important foundation to our understanding of the genetic content of *E. coli* and the use of future reference platforms for comparative analysis of other strains. Studies comparing full genome sequencing of *E. coli* show that the genetic content of different stains does vary but that unrelated strains can share common mobile elements (Figure 1.7), suggesting that different lineages of *E. coli* are capable of adapting to different environments. *E. coli* are a group that has a complex ancestral history and its constituent strains are promiscuous, picking up genes from a variety of sources. WGS enabled the scientific community to recognise a core genetic content unique to *E. coli*.

Figure 1. 6 **Circular representation of the *E. coli* 042 chromosome in comparison to other *E. coli* strains**

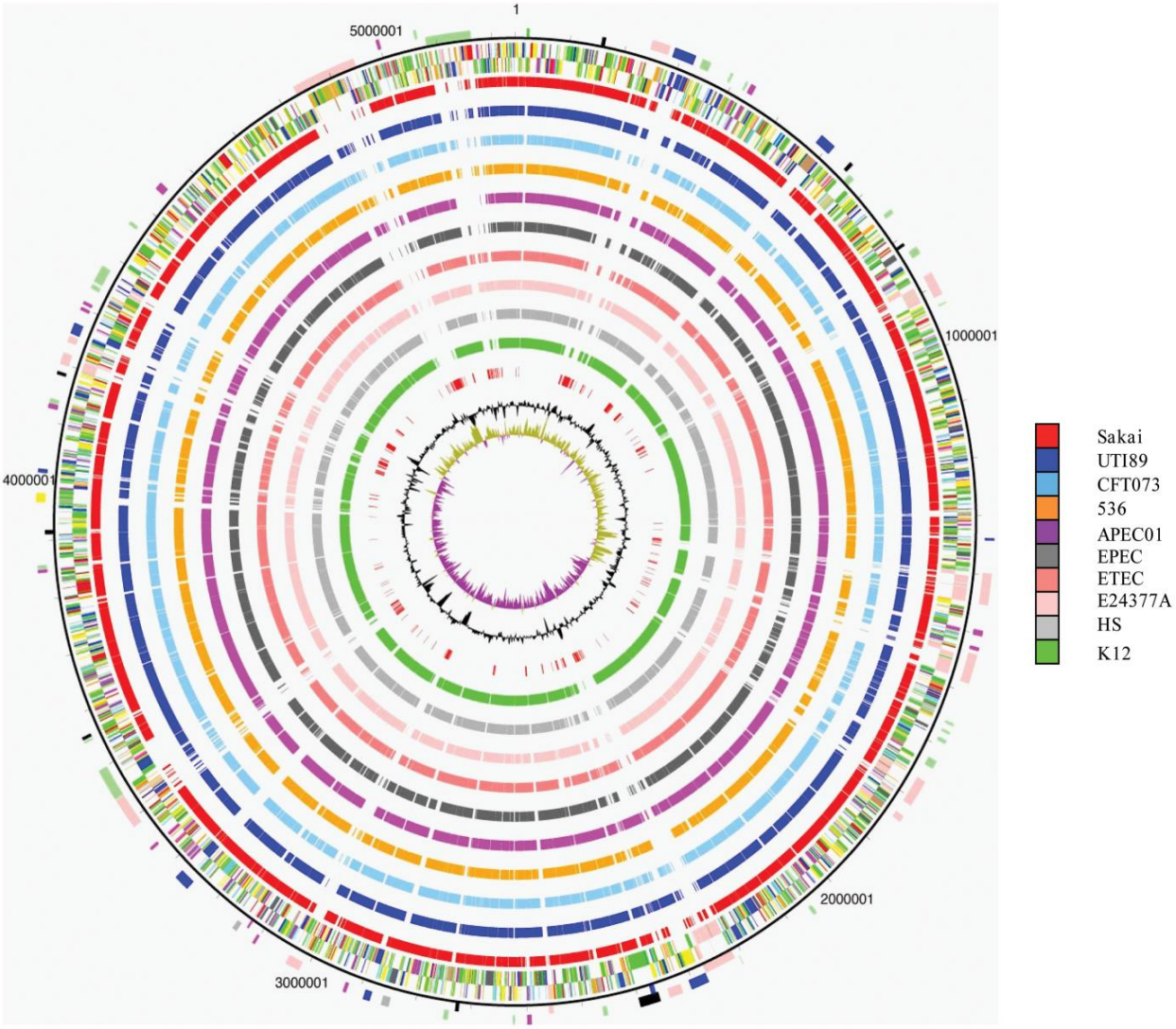


Figure 1.7: (Chaudhuri *et al.* 2010). From the outside in, the outer circle 1 marks the position of regions of difference (including prophage (light pink) fimbrial operons (Dark green) as well as regions differentially present in other *E. coli* strains: blue (Present in 0157:H7 & absent/divergent in UPEC CFT073) Light Green (Present in 0157:H7 absent/divergent in UPEC CFT073). Circle 2 shows the size in bps. Circles 3 and 4 show the position of CDSs transcribed in a clockwise and anticlockwise direction, respectively (for colour codes see below); circle 4 to 13 show the position of *E. coli* 042 genes which have orthologues (by reciprocal FASTA analysis) in other *E. coli* strains (see methods): Sakai (0157:H7; red), UT189 (UPEC; dark blue), CFT073 (UPEC; light blue), 536 (UPEC; orange), APEC 01 (APEC; dark pink), E2348/69 (EPEC; black), H10407 (ETEC; salmon pink), E24377A (ETEC; pale pink), HS (grey), and K-12 MG1655 (green). Circle 14 sows the position of genes unique to *E. coli* 042 unique (red). Circle 15 shows a plot of G+C content (in a 10 Kb window). Circle 16 shows a plot of GC skew ([G−C]/[G+C]; in a 10 Kb window). Genes in circles 3 and 4 are colour coded according to the function of their gene products: dark green = membrane or surface structures, yellow = central or intermediary metabolism, cyan = degradation of macromolecules, red = information transfer/cell division, cerise = degradation of small molecules, pale blue = regulators, Salmon pink = pathogenicity or adaptation, black = energy metabolism, orange = conserved hypothetical, pale green = unknown, brown = pseudogenes

Next generation sequencing technology has reduced the practical limitations of how much comparative analyses we can do with genome sequencing, and reduced the cost substantially. enabling many institutions to carry out comparable genome research simultaneously. Core genome analysis shows *E. coli* strains and *Shigella sp.* have large core content similar to each other and that *E. coli* O157 has additional genes to both genera (Figure 1.8). Sequencing of multiple strains of *E. coli* enables us to consider all gene content across the strains and help build up a 'pan-genome' picture to determine the most accurate definition of the species *E. coli*. As the pan-genome of *E. coli* expands with addition of multiple strains, the core genetic content decreases and any *E. coli* strain will only have 20 % of its core found in all other *E. coli* (Lukjancenko *et al.* 2010).

Accessory genome analysis (i.e. looking at mobile genetic elements or content that is not conserved in a genus or species) such as the pan-genome variable gene content (i.e. looking at the variable content of multiple strains within a genus or species) is also a useful typing tool (Figure 1.9). It is one of the few genetic methodologies that groups *E. coli* into separate clusters from *Shigella sp*. This indicates that perhaps it is the varied and mobile parts of the genome that differentiate between *E. coli* and *Shigella sp*.

Figure 1.7 BLAST atlas of Enterobacteriaceae against *E. coli* O157

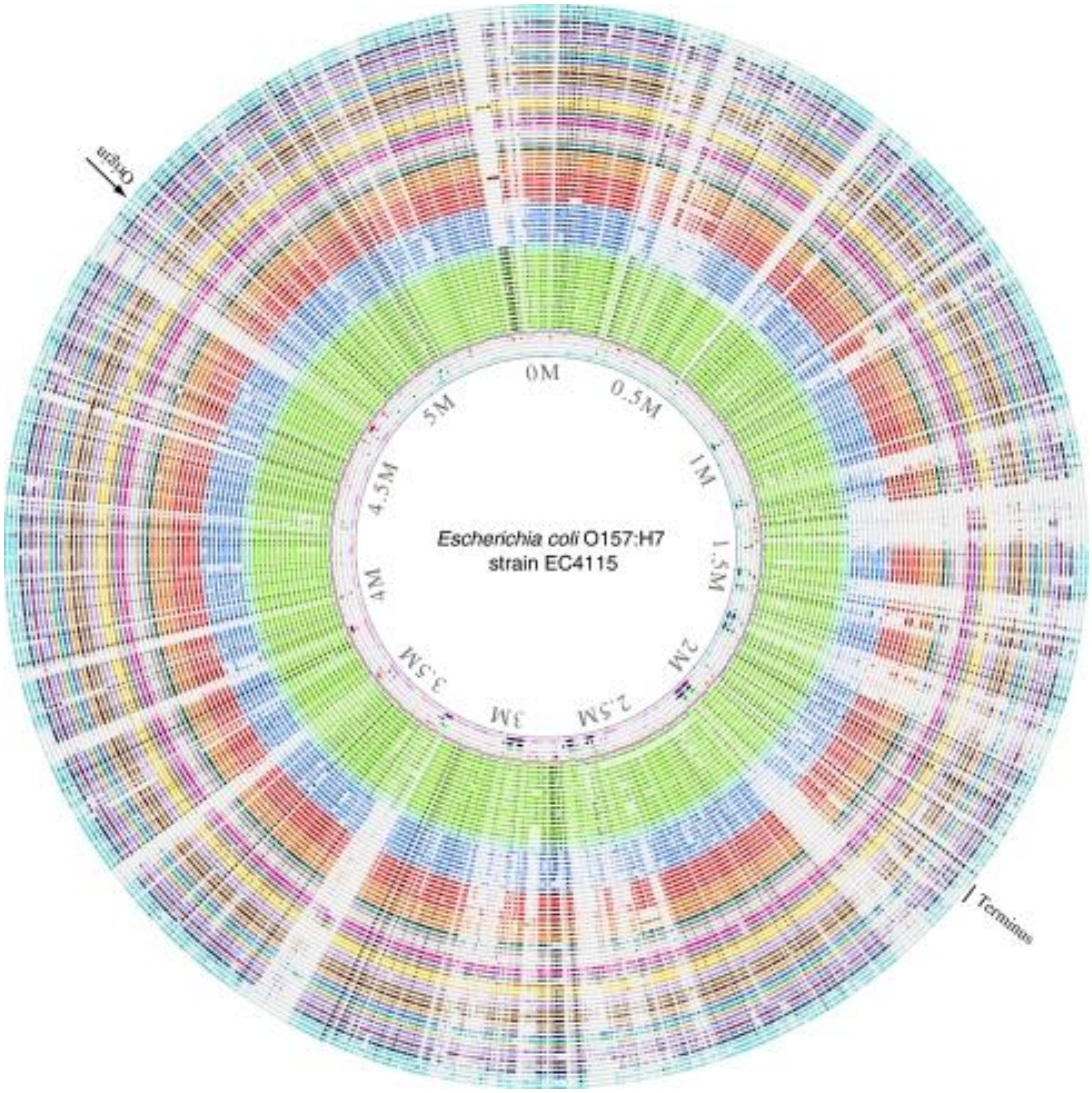


Figure 1.8: (Lukjancenko *et al.* 2010). BLAST atlas. In the middle, a genome atlas of *E. coli* O157:H7 strain EC4115 is shown, around which BLAST lanes are shown. Every lane corresponds to a genome, with the following colors (going outwards): *green E. coli* O157:H7 (15 lanes); *light blue E. coli* LANL strains (two lanes); *dark blue Shigella* spp. (eight lanes); *red E. coli* K12 and derivatives (six lanes); *orange E. coli* strain B phylogroup (four lanes); followed by all other *E. coli* genomes in different colors. The outermost three lanes represent *E. fergusonii*, *E. albertii*, and *S. enterica* Typhimurium LT2. Lack of color indicates that the genes at that position in strain EC4115 were not found in the genome of that lane. The position of replication origin and terminus is indicated

Figure 1.8 **Pan-genome clustering of *E. coli* and related species.**

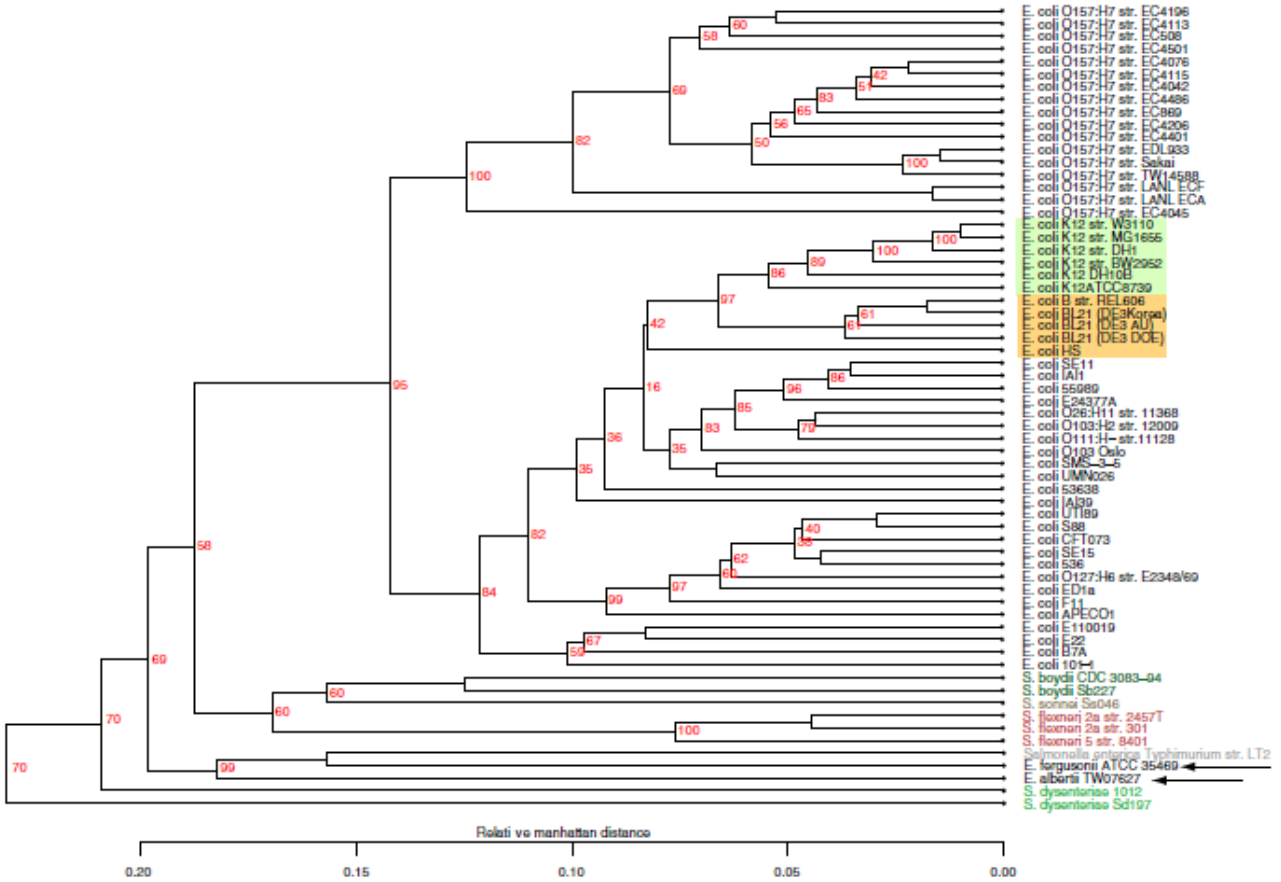


Figure 1.9: (Lukjancenko *et al.* 2010). Pan-genome clustering of *E. coli* (*black*) and related species (*colored*), based on the alignment of their variable gene content. The genomes now cluster according to species and a relatedness between *E. coli* K12 derivatives (*green block*) and group B isolates (*orange block*) is visible

1.8 Definition of *E. coli*

The conventional definition of *E. coli* is the expression of distinct biochemical profiles in which the species is defined as a sucrose, salicin and lactose fermentor and late maltose fermenter. These characteristics were used to define an outbreak on a neonatal ward and showed that this was a heterogeneous group (Bray 1945). Biochemical characterisation could not be used epidemiologically and serotyping schemes were developed to differentiate strains by their serological profiles (somatic, flagella and capsular antigens) (Gross and Rowe 1985;Kauffmann 1947). Eventually sequencing technology led to *E. coli* being defined by a core set of genes which has shown that the majority of different *E. coli* pathotypes can be discriminated from each other genetically.

These technologies show that the more information we receive, the more complicated the picture of *E. coli* becomes and that a polyphasic approach should be used when trying to understand an *E. coli* strain, such as is the practice in the Gastrointestinal Bacteria Reference Unit (Figure 1.10)

Figure 1.9 Polyphasic approach to the detection and identification of E. coli in the reference laboratory, Colindale

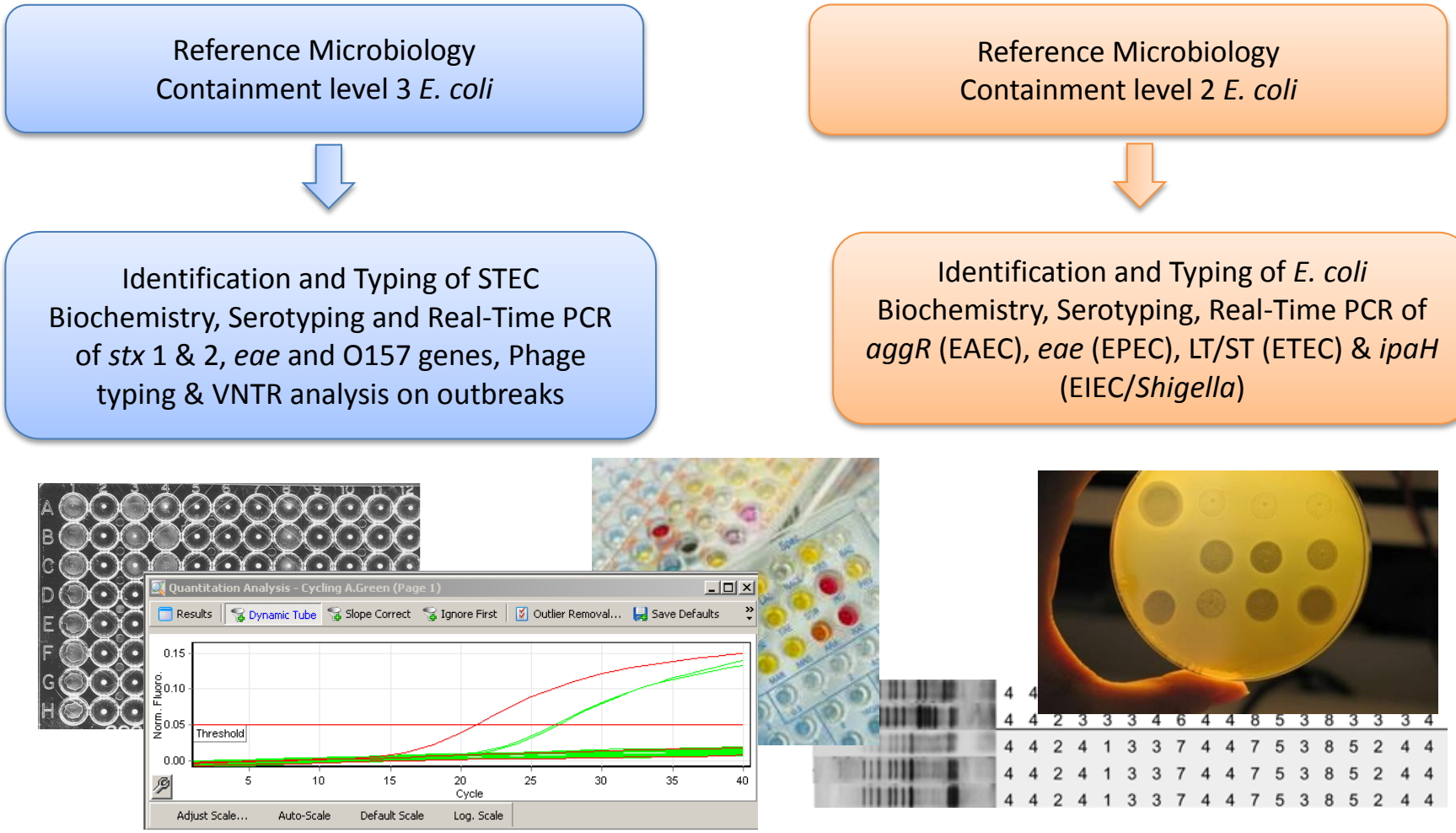


Figure 1.10: Flowchart depicting the polyphasic approach of biochemical, serological and molecular techniques to detect, identify and characterise *E. coli* in the reference laboratory. Key: genes include *stx* (shiga-toxin), *eae* (effacement and attachment), O157 (somatic antigen), *aggR* (regulatory gene), LT/ST (heat liaible and heat stable toxin), *ipaH* (invasion). Pathotypes include VTEC (verocytoxic *E. coli*), EAEC (enteroaggreagative *E. coli*), EPEC (enteropathogenic *E. coli*), EIEC (enteroinvaisive *E. coli*)

1.9 What are Enteroaggregative *E. coli*?

1.10 Defining and Characterising Enteroaggregative *E. coli*

EAEC was originally classified by its aggregative adherence pattern on HEp-2 cells (Nataro *et al.* 1987). Different scientists in the field have defined enteroaggregative EAEC in various ways (Figure 1.11), however, there is no one genotypic characteristic that can be used to define all EAEC.

Figure 1. 10 **What are Enteroaggregative *E. coli* ?**

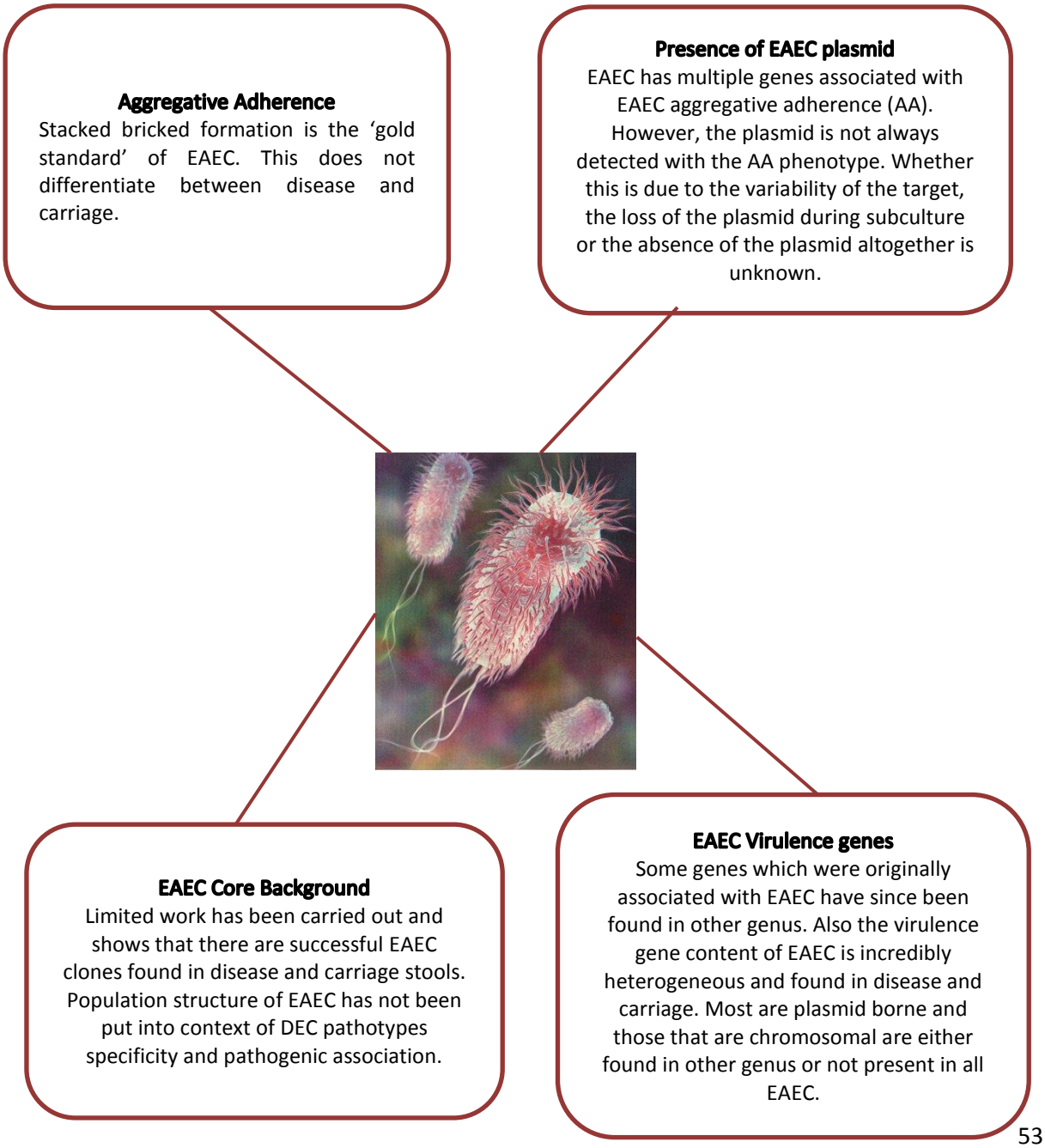


Figure 1.11: Figure descripting the complexity of defining enteroaggregative *E. coli* by previous studies

1.10.1 Identifying Enteroaggregative *E. coli* by adherence on human cells

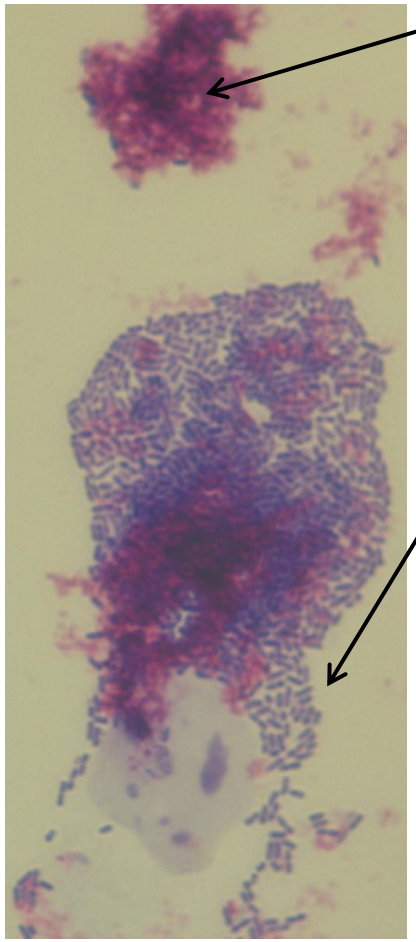
In 1987, whilst looking at aggregation of *E. coli* on human epithelial (HEp-2) cells, it was observed that a sub-set of strains associated with persistent diarrhoeagenic children in Peru had a unique stacked bricked aggregative adherence (AA) pattern (Nataro *et al.* 1987). This was different to the traditional localised adherence patterns found in EPEC and more characteristic than the pattern associated with diffusely adherent *E. coli* (DAEC) (Figure 1.12).

1.10.2 EAEC characterisation by virulence gene detection and profiling

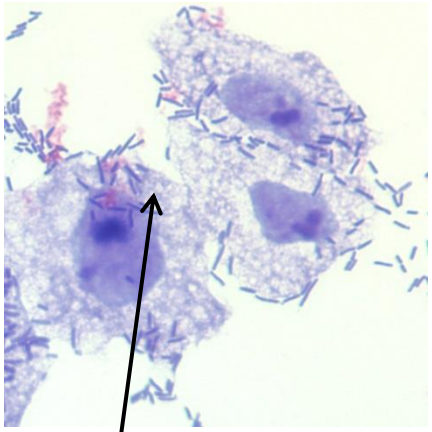
Studies have focused on attempting to characterise EAEC genetically with varied results. The design of a CVD432 probe (Baudry *et al.* 1990) that bound specifically to a subset of AA phenotype strains was later found to be specific for the anti-aggregative transporter (*aat*) gene (responsible for transporting a protein called dispersin across the membrane) (Nishi *et al.* 2003). The *aat* gene was encoded on a plasmid, designated pAA. Initially, the *aat* PCR was used as a PCR target for EAEC plasmid detection but current PCR assays for EAEC, target a gene responsible for regulating many of the EAEC genes (*agg*R) (Nataro *et al.* 1994). The alternative target *agg*R is a better target than *aat* as there are more conserved regions of the gene and the *aat* gene is not always found in the plasmid. However, a small minority of EAEC plasmids do not carry the *aggR* regulatory gene (Huang *et al.* 2007;Jenkins *et al.* 2006a;Jenkins *et al.* 2006b), indicating that the regulatory mechanisms encoding the EAEC phenotype are complicated.

.

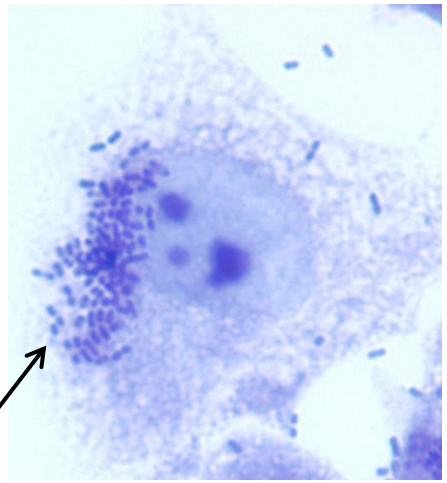
Figure 1.11 Different adherence phenotypes of *E. coli* pathotypes on HEp-2 cells



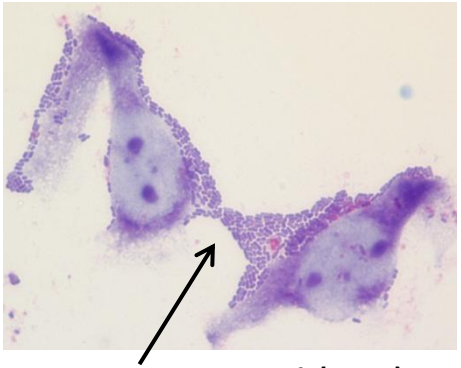
Cell Detachment *E. coli* (CDEC)
Cytotoxicity and destruction to HEp-2 cells. Adherence is lost when the HEp-2 cells are destroyed though some strains of EAEC are also known to bind to the glass slide and can be seen.



Diffusely adherent *E. coli* (DAEC)
Diffuse adherence throughout the HEp-2 cells



Enteropathogenic *E. coli* (EPEC)
Localised adherence along the side of the HEp-2 cells



Enteroaggregative *E. coli* (EAEC)
Classic stacked-bricked adherence along the side of the HEp-2 cells

Figure 1.12: Adherence characteristics of EAEC, EPEC and DAEC on HEp-2 cell. Photos taken by Marie Anne Chattaway, 2011, University of Haverford, USA.

Detection of multiple EAEC plasmid and chromosomal virulence genes also termed as genotypical profiling has been extensively studied (Huang *et al.* 2007;Jenkins *et al.* 2006a;Jenkins *et al.* 2006b;Okeke *et al.* 2000a;Pereira *et al.* 2007;Regua-Mangia *et al.* 2009) and the presence of a combination of key virulence genes is the current favoured method of characterising EAEC (Table 1.3). Studies attempting to characterise the EAEC plasmid have found multiple compatibility types harbouring a variable combination of genes putatively linked to background (Okeke *et al.* 2010). Therefore, characterising a pathogenic EAEC group by detection of plasmid borne genes is not ideal due to the mobility of this genetic element. The key set of characterising EAEC genes has yet to be agreed by experts in the field. To date there is no single gene, or combination of genes, that are EAEC specific and found in all EAEC strains.

Table 1.3 Common EAEC virulence genes

Target gene	Function	Reference	Target gene	Function	Reference
aat	Anti-aggregation protein transporter gene, Part of protein transporter system	(Jenkins *et al.* 2006b)	*Set1A*	*Shigella* enterotoxin, anti-sense strand of pic	(Huang *et al.* 2007)
aaiC	aaiC from 042 pheU island	(Jenkins *et al.* 2006b)	*Irp2*	Yersiniabactin biosynthetic gene	(Schubert *et al.* 1998)
astA	Enteroaggregative heat stable toxin 1 (EAST-1)	(Jenkins *et al.* 2006b)	*tia*	putative invasion determinant	(Jenkins *et al.* 2006a)
aggR	Transcriptional activator of AAFs	(Czeczulin *et al.* 1999)	*aap*	Dispersin gene	(Piva *et al.* 2003)
aggA	AAF/1 fimbrial type I	(Piva *et al.* 2003)	*pet*	Plasmid encoded toxin	(Sheikh *et al.* 2002)
aafA	AAF/II fimbrial type II	(Piva *et al.* 2003)	*aaiC*	Part of the aai gene cluster encoding a type Vi secretion system	(EU Reference Laboratory for E.coli Department of Veterinary Public Health and Food Safety 2013)
agg3A	AAF/III fimbrial type III	(Bernier *et al.* 2002)	*pic*	Secreted protease (146kDA), 116kDa after cleavage, multifunctional protein involved in enteric pathogenesis.	(Piva *et al.* 2003)
Agg4A	AAF/IV fimbrial type IV	(Boisen *et al.* 2009)			

Table 1.3 lists the most commonly targeted virulence targets found in EAEC.

1.10.3 Characterisation of EAEC by population structure

An alternative method to characterising EAEC would be to look at the core background of EAEC to understand the population structure and clonality of this complex pathotype. The MLST *E. coli* scheme devised for population structure analysis focused initially on the well-established pathotypes such as STEC, EIEC, EPEC and *Shigella* spp. (Figure 1.6) (Wirth *et al.* 2006). Although small numbers of EAEC sequence types have been deposited in the public MLST database (http://mlst.ucc.ie/mlst/dbs/Ecoli), the only comprehensive study to date is by Okeke *et al* who investigated association of EAEC with ST groups and disease in children under five in Nigeria, and highlighted links to virulence genes, resistance and plasmid groups. Results indicated that ST10 was linked to cases but there was no other association between other STs and disease or association with certain virulence and resistance profiles or the plasmid compatibility groups (Okeke *et al.* 2010). Until a more precise definition of a pathogenic EAEC is determined, using PCR or MLST in isolation will not define this pathotype or improve the diagnosis of clinically relevant strains.

1.11 Understanding the Burden and Importance of EAEC

It is difficult to determine the clinical and public health significance of EAEC because we are unable to define the pathogenic types or be clear whether EAEC are capable of independently causing disease.

1.11.1 Burden of EAEC

Measuring the burden of infectious disease is essential for the rational design of public health intervention strategies and for the allocation of resources. For intestinal infectious diseases (IID) there is a massive global burden; WHO estimates around two billion cases every year (WHO Media Centre 2009). In England and Wales, detailed surveillance studies have shown that there are up to 17 million sporadic, community cases of IID and one million GP consultations annually in the UK (Tam *et al.* 2012b). Despite routine investigations of IID in the UK for salmonellosis, shigellosis, campylobacteriosis, Cholera, infection by *Escherichia coli* O157, rotavirus, norovirus and parasitic

infections, well over half of the laboratory investigated diarrhoeal episodes are not diagnosed (Tam *et al.* 2012a). One, often under diagnosed but potential aetiological agent is EAEC; this pathotype of *E. coli*, has been associated with cases of gastrointestinal infection (Nataro *et al.* 2006;Tam *et al.* 2012b;Tompkins *et al.* 1999) at a level comparable to *Salmonella* in England and Wales (Amar *et al.* 2007;Tompkins *et al.* 1999).

1.11.2 Co-infections in intestinal infectious disease

Traditionally, gastrointestinal disease has been associated with single pathogens but recent studies using a pan-pathogen PCR approach have shown that cases reporting symptoms of GI disease often have multiple pathogens present in their faecal specimens. Multiple studies have specifically detected and characterised EAEC from faeces, but this is usually limited to research studies (Huang *et al.* 2007;Jenkins *et al.* 2006a;Jenkins *et al.* 2006b;Okeke *et al.* 2000a;Pereira *et al.* 2007;Regua-Mangia *et al.* 2009) where detection of co-infections have not been undertaken. Alternatively, cohorts studies have detected multiple pathogens but further analysis into the true aetiology among co-infections or even analysing the amount of co-infection has not been considered (Amar *et al.* 2007;Tam *et al.* 2012a;Tompkins *et al.* 1999). As diagnostic tools improve, more studies are recognising gut infections of mixed aetiology (Lindsay *et al.* 2011), but when multiple pathogens are present in a diarrhoeic stool, determining the causative agent can be problematic. This is especially true in studies investigating EAEC infection, for example, in Peru multiple pathogens were found in 40% of infants with diarrhoea and with EAEC in their stool (Ochoa *et al.* 2009).

1.11.3 EAEC and outbreaks

Enteroaggregative *E. coli* has been associated with disease globally (Antikainen *et al.* 2009;Cennimo *et al.* 2009;Jenkins *et al.* 2006b;Okeke & Nataro 2001;Okhuysen & DuPont 2010;Toma *et al.* 2003;Usein *et al.* 2009). A number of EAEC outbreaks have been described in the literature, most notably a large outbreak in Japan involving 2697 children (Itoh *et al.* 1997). The EAEC group has

gained notoriety during a recent outbreak in Germany and France (Rubino *et al.* 2011). The outbreak strain *E. coli* ST678 (O104:H4) had both phage encoding *stx* genes and the EAEC plasmid. The acquisition of the EAEC plasmid may have played a role in adherence to the hosts gut mucosa, as the strain lacked the attachment and effacement (*eae*) gene known for intimate adherence in classical STEC strains (Jenkins *et al.* 2003). The emergence of this hybrid pathogen was described previously in 1996 when an O111:H2 strain had caused an outbreak of HUS in France (Mellmann *et al.* 2011), in 1999 when a strains of *E. coli* O86:H- associated with HUS was isolated in Japan (Iyoda *et al.* 2000) and most recently in 2011 when a strain of *E. coli* O111:H21 strain was associated with a family outbreak in Ireland (Dallman *et al.* 2012). All of these cases were associated with severe disease.

The reason EAEC is not widely accepted as a pathogen, like Salmonella for example, is that a significant portion of healthy controls also harbour this pathotype (Huang *et al.* 2007;Okeke *et al.* 2000a;Regua-Mangia *et al.* 2009). Furthermore, research data describing the association of genetic factors with virulence are contradictory (Boisen *et al.* 2012;Huang *et al.* 2007;Jenkins *et al.* 2005). The reliability of virulence factors to identify EAEC for diagnostic purposes is unclear (Okeke & Nataro 2001). Studies have shown EAEC to be a potentially important pathogen in diarrhoeal disease but a causal link between disease and the presence of EAEC in the stool has not been defined.

1.12 How has enteroaggregative *E. coli* evolved?

Population structure analysis of *E. coli* has facilitated the understanding of how the different pathotypes have evolved and the emergence of successful clonal populations such as STEC O26 and O157. The number of strains of EAEC in the MLST public database is relatively low compared to other pathogenic *E. coli* groups. As of 17th January 2013 there were 5425 entries, with the exclusion of the Nigeria dataset which is used in this study, only 2.4% (132 entries) were designated EAEC. Despite the low numbers, analysing the publically available MLST database including the Nigerian dataset (251, 4.6%) to get an understanding of how EAEC are distributed shows that EAEC are dispersed throughout the population of *E. coli* (Figure 1.14).

Overlaying EAEC in the context of both intestinal and extra-intestinal pathotypes is difficult to interpret. Comparing EAEC with the sub-population structure of other diarrhoeagenic *E. coli* pathotypes (1143 entries), shows a clearer picture of potential EAEC clonal populations (mainly based on the Okeke study (Okeke *et al.* 2010)) (Figure 1.15).

Figure 1.12 Distribution of EAEC in the population structure of *E. coli*

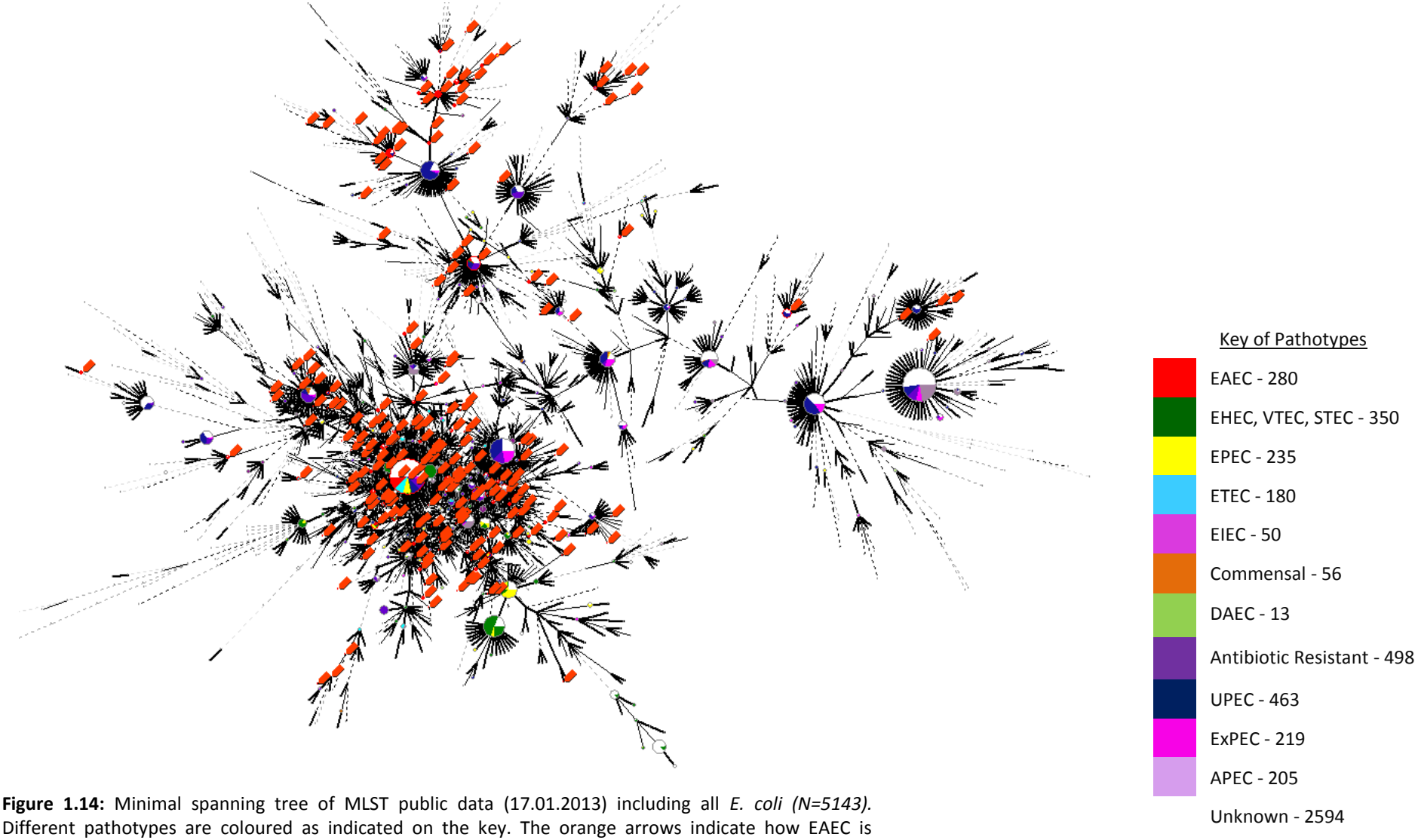


Figure 1.14: Minimal spanning tree of MLST public data (17.01.2013) including all *E. coli (N=5143).* Different pathotypes are coloured as indicated on the key. The orange arrows indicate how EAEC is dispersed throughout the population structure.

Figure 1.13 EAEC groups in the diarrhoeagenic pathotypes

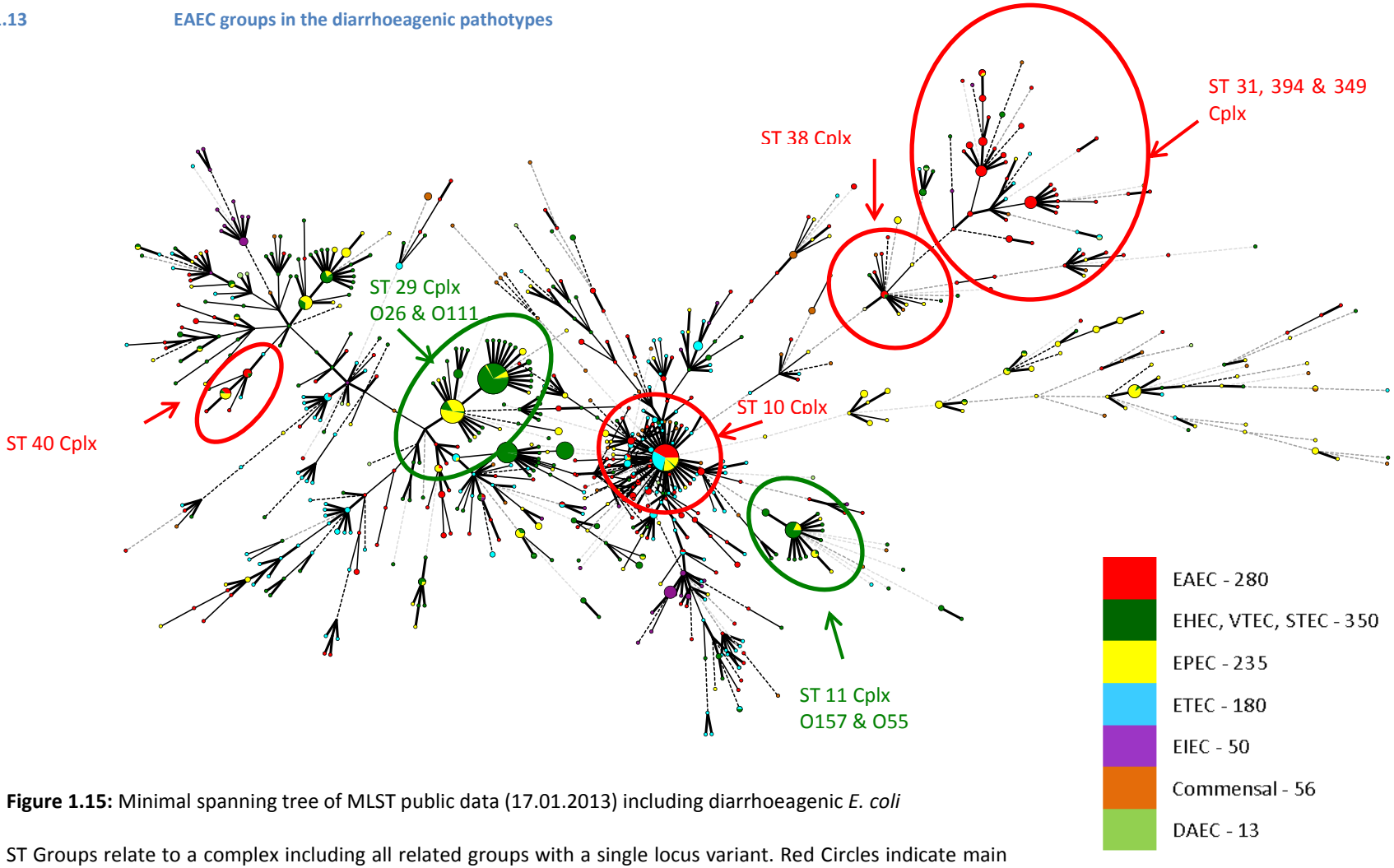


Figure 1.15: Minimal spanning tree of MLST public data (17.01.2013) including diarrhoeagenic *E. coli*

ST Groups relate to a complex including all related groups with a single locus variant. Red Circles indicate main locations of EAEC, Green Circles indicate well known VTEC/EPEC groups.

This snap shot of EAEC background raises questions regarding the complexity of evolution of these groups. Whereas some groups appear to be EAEC specific, such as the ST38, 31, 394 and 349 complex, others share a background with other pathotypes including the ST40 complex with EPEC and the ST10 complex with EPEC, ETEC and to a small number of STEC. This indicates that all EAEC have not evolved from the same ancestral lineage and that perhaps different groups of EAEC have acquired the EAEC phenotype as an advantage to their survival. Okeke *et al* showed that the ST10 complex had a higher mutation rate than the ST31, 38 & 394 complexes, which had a higher recombination rate (Okeke *et al.* 2010). However, the focus was on understanding the EAEC groups from a single geographical location (diarrhoeagenic and asymptomatic children in Nigeria) and association with other important genetic markers. An understanding if there are specific successful EAEC groups that are expanding globally and associated with disease warrants further investigation.

1.13 How is EAEC phenotype linked to genotype and disease?

Understanding the presence of genetic markers and background is an important step in the characterisation of EAEC. Relating this with phenotype and disease may help scientifically validate the definitions of pathogenic EAEC. This may facilitate the design of clinically meaningful diagnostics tests. There are several phenotypic factors to be considered with EAEC, which if linked with successful pathogenic groups, may help us understand the key factors involved in virulence.

1.13.1 Intensity of adherence

First, there is the ability of an organism to adhere to the hosts' mucosa; although we know that EAEC adhere to HEp-2 cells, it is not known whether the intensity of adherence plays a role in the severity of symptoms. The survival and success of all pathogens require that they colonise the host and adhere to cells ,which enables the organism to utilise nutrients and activate metabolic pathways, all of which are used for survival.

1.13.2 Metabolic Utilisation

Characterisation of biochemical reactions required for survival in the gut is a useful tool diagnostically. Studies have shown that metabolic profiles can be linked to a group being more pathogenic (Hwang *et al.* 2010). The traditional view of an evolutionary route towards increased pathogenicity is the acquisition of virulence genes. The loss of genetic material however, can result in a biochemical deficiency linked to pathoadaption and may also be associated with increasing virulence (Sokurenko *et al.* 1999). This has been shown in *Shigella sp.* and enteroinvasive *E. coli* where 'Black holes', i.e. large genomic deletions, have enhanced virulence (Maurelli *et al.* 1998). *Shigella sp.* were shown to contain pathoadaptive mutations via re-organisation of the *cadA* regions that were associated with enhanced virulence, this results in the inability to utilise lysine. The EAEC 101-1 Japan outbreak strain was the largest EAEC outbreak to date affecting over 2,600 children (Itoh *et al.* 1997). The *cadAB* genes were deleted in the outbreak strain and it was suggested that deletion of cad genes may produce hypervirulent EAEC lineages (Hwang *et al.* 2010). The question remains, If pathogenic lineages of EAEC do exist, could metabolism of these groups be slightly different to non-pathogenic EAEC? Looking at end points of utilisation may not answer this question as typical *E. coli* utilise similar substrates, However, assessing real-time values of the utilisation of multiple substrates using a phenotypic array such as the Ominlog (Biolog), and using dendograms to assess the metabolic relationships of strains may identify subtle differences between pathogenic and non-pathogenic groups.

1.13.3 In vivo models for pathogenicity assessment

The EAEC group are potential gut pathogens, but how can we assess if lineages of EAEC are responsible for causing disease? Studies have shown the presence of EAEC can result in the production of interleukin as a response of the mucosal immune system (Goyal *et al.* 2010;Khan *et al.* 2010;Steiner *et al.* 2000) indicating an interaction with the gut. However, the gut is a complex microbiotic environment and symptoms vary between patients, even the ability of EAEC to cause disease varies depending on immune markers in healthy volunteers (Nataro *et al.* 1995). Seeing if

there is a direct link between an EAEC strain from a successful pathogenic group or from a case without confounders (i.e. the complexity of the gut mucosal immune system and co-infections) will enable us understand if the presence of defined diarrhogenic EAEC alone is enough to cause disease. The impractical healthy volunteer challenge model is therefore not suitable for this as it will be unknown if the response is specifically to the presence of EAEC and is ethically challenging. Alternative models for EAEC have been investigated such as T84 cells and human intestinal explants. These have been used to model cytocoxic effect and adhesion of EAEC and showed marked toxic effects most prominently in areas where bacteria were adhering (Nataro *et al.* 1996). Virulence gene expression and diarrhoeagenic effects have been modelled with an abiotic intestinal simulator (in vitro anaerobic continuous culture system) ,a gnotobiotic piglet model and mouse models (Hicks *et al.* 1996;Nataro *et al.* 1996;Tzipori *et al.* 1992). These models show that EAEC produce distinctive intestinal lesions and in some cases were able to differentiate the lesions from those caused by other major categories of diarrhoeagenic *E. coli*. All of these systems are highly specialised, and are only available in one or two laboratories

A simple, economical, practicable and high through-put alternative model which provides an innate response using invertebrates has been investigated. The well-studied *Caenorhabditis elegans* model (Aballay and Ausubel 2002) has been used with multiple bacterial pathogens (Aballay *et al.* 2000;Fuhrman *et al.* 2008;Mellies *et al.* 2006;Mylonakis and Aballay 2005;Tan and Ausubel 2000;Tenor *et al.* 2004) where worms fed bacteria grown on standard nematode growth media (NGM) die over a short period of time (2-3 days). This is the 'slow killing' method and is a consequence of the worms intestinal lumen being colonised by the bacterial pathogen (Mahajan-Miklos *et al.* 1999;Tan *et al.* 1999), that has escaped the mouth grinding action of the worm that should kill the bacteria . Hwang *et al* showed that some EAEC strains killed *C. elegans* when used as a food source, most likely by means of an infection like process via the colonisation of the distal *C. elegans* intestine. This was in comparison to non-pathogenic *E. coli* which did not colonise and kill

the worm. This study was carried out on a few strains focusing on mutation of the pCADA plasmid (Hwang *et al.* 2010). This model could be used to look at strains from different lineages and assess the host-pathogen interactions.

1.14 Definition of enteroaggregative *E. coli*

The original definition of EAEC of stack bricked formation of the bacteria on HEp-2 cells (Nataro *et al.* 1987) has not been successfully replaced by any one alternative test. Due to the specialised and laborious nature of this method, the most popular method for detecting EAEC is by the detection of the *aggR* gene (Chattaway *et al.* 2014a) and currently used in multiple diagnostic laboratories in the UK and the Gastrointestinal Bacteria Reference Unit, UK.

Due to the fact that there is no specific method to define EAEC and to ensure that there is no bias in the data output in this thesis towards a subset of EAEC. All EAEC that had been characterised using multiple methods were included such as the HEp-2 cell assay, the CVD432 probe (Baudry *et al.* 1990), the *aat* gene (Nishi *et al.* 2003) and the *aggR* gene (Nataro *et al.* 1994).

Aims and Objectives

Chapter-1 Introduction.

The introduction is split into two parts. The first part explains the different pathotypes and pathogenicity of *E. coli* and tells the story of how *E. coli* as a species has changed in its definition over the past century with evolving technology and microbiological practices. The second part focuses on the definition and characterisation of EAEC, burden, evolution and phenotype of EAEC in relation to the context of the theories exploited in this thesis.

Chapter 2-Is the burden and aetiology of enteroaggregative *E. coli* in intestinal infectious disease in the UK important?

Chapter 2 gives an insight to how the burden of EAEC in the United Kingdom compares to other important pathogens. It is hypothesized that EAEC is independently capable of causing disease irrespective of co-infections and has played an important and current role in IID over the past 15 years. This has been assessed by statistically analyzing both infectious intestinal disease (IID) databases with emphasis on the pathogenicity of EAEC. Methods were developed to assess disease burden using results of a semi-quantitative real-time PCR assay to diagnose enteroaggregative *E. coli* aetiology in episodes of IID in the UK (Chapter 2).

Chapter 3-Can you differentiate pathogenic EAEC via population structure analysis and use this approach in public health studies?

The aim of this part of the study was to use a case control approach using strains from Bangladesh, Nigeria and the UK based on the hypothesis that certain *E. coli* lineages by Multi-locus sequence typing (MLST) are exclusively EAEC and that these lineages vary in their ability to cause disease. It is hypothesized that the case control isolates will identify pathogenic lineages not currently represented in the public database. This will be proven by analysis of the population structure of

EAEC to place the case control isolates into natural groupings and evaluating the lineages by disease causing EAEC cases and carriage EAEC controls. The EAEC population structure will then be used in two public health studies to ascertain if this approach can be used in other clinical situations.

Chapter 4-What is the Evolutionary history of EAEC?

Here, the study hypothesis that EAEC have a complex and diverse evolutionary history, that has led to the heterogeneous and complex nature of this pathogen is investigated. This will be shown by phylogenetic analysis of the strains of the concatenated sequence of MLST via ClonalFrame analysis.

Chapter 5-Can the Phenotype of EAEC be linked to EAEC Genotype?

The hypothesis that there are pathogenic lineages of EAEC and that these lineages will be phenotypically or genotypically distinct from the carriage EAEC groups is examined in this part of the study. This will be investigated by selecting representative groups and looking at the metabolic profile using the Biolog and their pathogenic potential, using the HEp-2 cell assay and the *Caenorhabditis elegans* worm model. Additional analysis of looking at EAEC strains and linking ST background to serotype and virulence gene profiling was carried out. An assessment of using serotyping to detect pathogenic lineages was also carried out.

Chapter 6-Discussion

Each of the results chapters (chapters 2-5) include a brief background to the work and the methods used for that part of the study. There is no separate methods chapter. The discussion summarises the findings from this thesis.

Chapter Two

Burden of Enteroaggregative *Escherichia coli* in the United Kingdom

Chapter 2 - Burden of EAEC in the United Kingdom

2.1 Background

The aetiological agent is not identified in over half of laboratory investigated diarrhoeal episodes in the UK (Tam *et al.* 2012a). It is possible that a significant number of these cases of gastroenteritis can be attributed to diarrhoeagenic *E. coli* (DEC), including enteroaggregative *Escherichia coli* (EAEC) (1.1.3). Several authors have described high levels of EAEC cases in gastrointestinal disease infection (Nataro *et al.* 2006;Tompkins *et al.* 1999). However, in these studies, EAEC were also found in asymptomatic cases leading to doubt surrounding the role of EAEC in the disease process.

The aim of this chapter was to reanalyse data from two intestinal infection disease (IID) burden studies carried out in the UK over the past 15 years (Tam *et al.* 2012b;Tompkins *et al.* 1999). The first IID study included cases and controls whereas the second IID study included cases only. Real-time PCR methodology, used in both studies for detection of target genes for potential GI pathogens, was semi-quantitative (Barletta *et al.* 2011;Nadkarni *et al.* 2002). It was used to determine bacterial load of pathogens, including the *aat* gene for EAEC in faecal specimens from symptomatic, and asymptomatic patients, to look in detail for a causal link between disease and the presence of EAEC in the stool and whether this methodology could be used to assess causative burden in the second IID study.

Bacterial load and evidence of co-infection (i.e. the presence of other pathogens that may be attributable to disease) were assessed to determine whether or not these factors were associated with increased likelihood of disease.

2.2. Methods

2.2.1 Subjects and datasets used

The first intestinal infectious disease study (IID1) was originally undertaken to identify the microorganisms and toxins in the UK population associated with disease and carriage. Stool samples were collected between August 1993 and January 1996. The study included both cases and controls among the community (based on a population based cohort) and cases presented to their GPs (with age/sex matched controls - from GP centres). During this time, 6743 were tested, 3654 of which were cases. The case definition included people of all ages with loose stools or vomiting (more than once in 24 hours) lasting less than two weeks. Any people that had a known non-infectious cause were excluded. A control definition included people free of loose stools or vomiting for three weeks before the matched case became ill. Controls were matched by age and those over 5 years of age were also matched by sex (Sethi *et al.* 1999;Tompkins *et al.* 1999).

Samples were later re-tested using real-time PCR for detection of the *aat* gene found in EAEC (aatTMF GGGCAGTATATAAACAACAATCAATGG, aatTMR GTAGTTGTTCCTCTCACTAAGCATTTCAAT, AATP VIC-TCTCATCTATTACAGACAGCC-MGB) (Amar *et al.* 2005;Amar *et al.* 2007). The data generated included qPCR cycle threshold (Ct) values inversely proportional to the number of copies of the *aat* gene in 4664 stool specimens (2443 cases, 2221 controls) including 113 cases and 38 controls specimens that were culture positive for EAEC. Real time cycle threshold (Ct) values were obtained for 102 EAEC positive cases and 31 EAEC positive controls and this data was used to assess bacterial load in this study.

Laboratory data from the second IID study (IID2 case study where stool samples were collected from cases only with the same case definition as IID1) (O'Brien *et al.* 2010) , in which a prospective Cohort Study and GP presentation study was carried out during the period 24th April 2008 – 29th March 2009, were also analysed as part of this study. The dataset contained 83 EAEC positive stool specimens (from a total of 3966 stools) with real time PCR Ct Values.

2.2.2 Burden Assessment - Defining diagnostic cut off values for Ct value in EAEC infection

Real- time PCR has been shown to be an accurate tool for detection of infectious disease in faeces (Phillips *et al.* 2009;Phillips *et al.* 2010) and as previously described, the Ct can be used as a measure of aetiological agent in the faeces and is inversely proportional to the amount of organism present in the specimen (Barletta *et al.* 2011;Nadkarni *et al.* 2002). Therefore the lower the Ct value the higher the bacterial load. The number of cycles of PCR replication required to raise the number of copies of the target sequence in the reaction mixture above a pre-determined threshold is represented by the Ct value (Grove 1999). The real-time-PCR assay for the *aat* gene was run for 40 cycles and a specimen was considered positive if the Ct value was <40.

Assessment of the burden of EAEC disease using the most recent (IID2 case only) study was not possible directly. The lack of controls in IID2 and the lack of an absolute association of EAEC presence with diarrhoeal infection meant that the data from the case control study in IID1 was used to redefine the Ct cut-off value of the *aat* gene.

A receiver-operating characteristic (ROC) analysis was used to define a cut-off in the Ct values to estimate the number of true EAEC IID positive cases in the population. This cut-off was used for the estimation of burden of disease caused by EAEC. The ROC analysis Ct value cut-off was not used for the other analyses in this study as a 'case' value, an example being investigating linkage with other pathogens, where a 'case' is the definition used by the original IID studies.

For the ROC analysis, reference groups were selected from the IID1 study using microbiological and clinical characteristics. The positive reference group was defined as in the IID1 study case (people with loose stools or clinically significant vomiting lasting less than 2 weeks in the absence of a known non-infectious cause and preceded by a symptom free period of 3 weeks), detection of EAEC

by real-time PCR for the *aat* gene and culture positive EAEC in the patients' faeces. The negative reference group was also defined as in the IID study control (no IID symptoms in the past 3 weeks), detection of EAEC by real-time PCR for the *aat* gene and culture positive EAEC from faeces. In the ROC analysis, the sensitivity and specificity were calculated for each potential cut-off Ct value and an empirical ROC plot created using StataSE 12.0 (a statistical programme for analysing data) (StataCorp 2011). The Youden index (sensitivity + specificity-1) was calculated and the maximum value used to identify the optimum cut-off (Bewick *et al.* 2004;Fluss *et al.* 2005). This value was then used to predict the number of true positive cases of those with a Ct value in the case only study, IID2. We compared the distribution of Ct values from cases and controls for EAEC positive individuals using the Student's t-test.

2.2.3 Causal link between EAEC and disease - statistical methods

It became clear following the initial analysis, that the relationship between EAEC presence and disease was not absolute and so several methods were used to further investigate the association of EAEC with disease:

2.2.3.1 Carriage rates of EAEC in healthy controls, compared to other pathogens

For each infection the chi squared test was used to test if the distribution of the pathogen between cases and controls was as expected by chance.

2.2.3.2 Association of disease with individual pathogens where multiple pathogens were detected in stool

All EAEC positive individuals with multiple pathogens (both cases and controls) were tested to determine whether individual pathogens were equally distributed between cases and controls using Chi squared tests for independence. Because norovirus was the most common pathogen, a comparison using the Chi squared test for co-infection in all individuals positive for EAEC and all individuals positive for norovirus was carried out to determine if the

presence of other individual pathogens were dependent on infection with EAEC or norovirus.

2.2.3.3 Independent association of EAEC presence with disease

A logistic regression of univariate and multivariate analysis was carried out using case or control as outcome and infecting agent and age as independent variables. In this way the independent association between EAEC and disease whilst controlling for other pathogens was assessed. Model results were then used to calculate the population attributable fraction (PAF):

$$PAR = P_e (RR_e - 1) / RR_e$$

Where P_e is the proportion of cases with the exposure (EAEC) and RR_e the relative risk of disease. This form allows for confounding of the exposure if an adjusted RR is used, as recommended in Rockhill *et al* (Rockhill *et al.* 1998). In this case, adjusted odds ratios are substituted into this equation to give an approximate, adjusted PAF.

2.3 Results

2.3.1 Descriptive statistics

To test if the analysis of data from the IID1 case control study remained relevant in 2009, we compared the demographic data from the two periods. There was no significant difference between the rate of EAEC in the IID1 case – control study (1993 – 96) and IID2 case-only study (2008 – 09), with 1.4% and 1.9%, respectively; individuals with EAEC present in their stool were distributed evenly across all age groups in both IID1 and IID2 (chi-squared p value for non-independence:0.253). For EAEC-positive individuals, there was no significant difference in age between cases and controls (p=0.237). We therefore believe that the epidemiology did not change significantly for EAEC infection between the two periods. Cases tended to be slightly older than controls in IID1 (mean age of cases: 30.1 years, standard deviation (SD): 24.7 years; mean age of controls 28.7 years, SD: 23.9 years; p value for difference:0.051).

2.3.2 Defining diagnostic cut off values for Ct value in EAEC infection

In order to investigate the link between Ct value and disease, the sensitivity and specificity of the Ct value was assessed in EAEC-positive specimens from the case control study (dataset IID1); Ct values were obtained and included 102 cases and 31 controls. Figure 2.1 shows the resulting ROC curve, and Figure 2.2 the distribution of Ct values in cases and controls. The cut-off was chosen to balance sensitivity and specificity and was set at a Ct value of 31 (Figure 2.1). The ratio of false positives versus false negatives with this cut-off point was 1.09 (95% confidence interval (CI): 0.79 – 1.53) (Figure 2.2). The total number of test-positives, although not a good diagnostic for the individual (due to poor sensitivity and specificity), was a reasonable estimate of the total number of cases. Importantly however, in the population studied, there was a significant association between bacterial load and disease state (p=0.039), and further investigations were carried out using the point of <40 to indicate presence of EAEC.

Figure 2.1 **Receiver-operating characteristic (ROC) Analysis of Ct values for enteroaggregative *E. coli* from gastrointestinal disease cases (n=102) and controls (n=31). August 1993-January 1996**

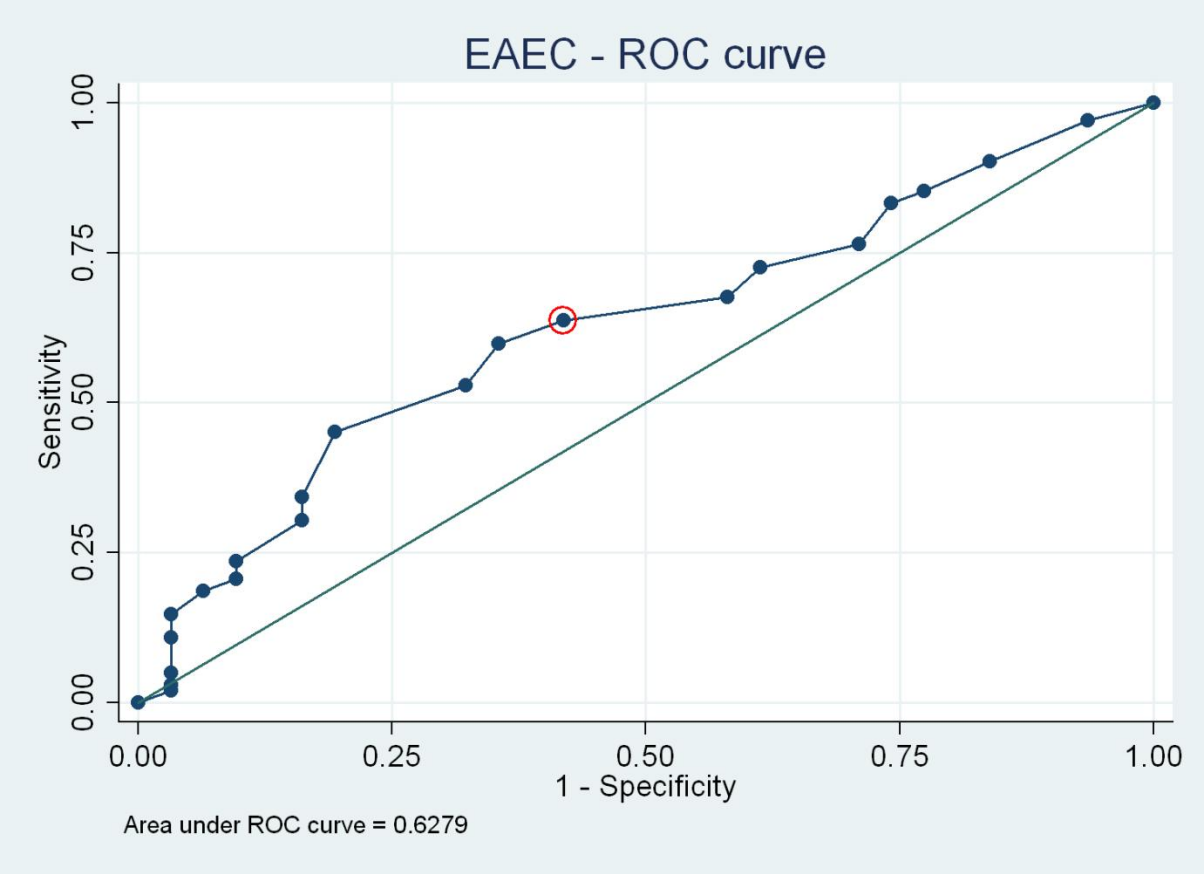


Figure 2.1: The red circle at Ct value 31 indicated the cut-off value which was chosen at the point where sensitivity and specific were equivalent. Figure used in Eurosurveillance publication (Chattaway *et al.* 2013).

Figure 2.2 **Distribution of Ct values for curve analysis of enteroaggregative *E. coli* in gastrointestinal disease cases (n=102) and controls (n=31). August 1993-January 1996**

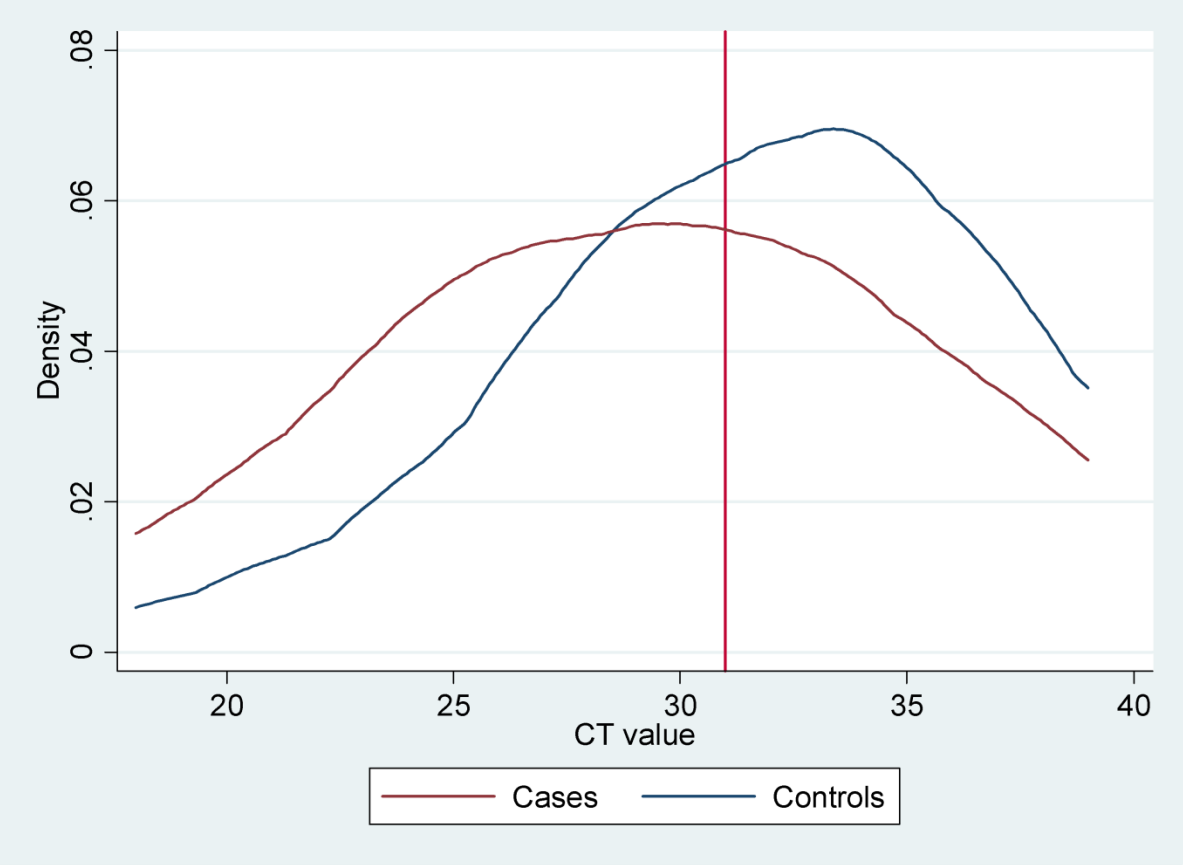


Figure 2.2: Fitted curve distribution of Ct values. The red line indicates the cut-off point where the ratio of false positives versus negatives with this cut-off point was closest to equivalent 1.09; 95% confidence interval: 0.79-1.53. Figure used in Eurosurveillance publication (Chattaway *et al.* 2013).

2.3.3 Investigation of the association of EAEC presence with disease.

2.3.3.1 Carriage rates of EAEC, compared to other pathogens, in healthy controls:

Submitting a stool specimen that was positive for EAEC was positively associated with having disease (Figure 2.3). However, one quarter of all EAEC positive individuals were asymptomatic (38/151).

Figure 2.3 **Organisms present in stool samples from gastrointestinal disease cases (n=2,221) and controls (n=2,213) in the IID1 study, August 1993- January 1996**

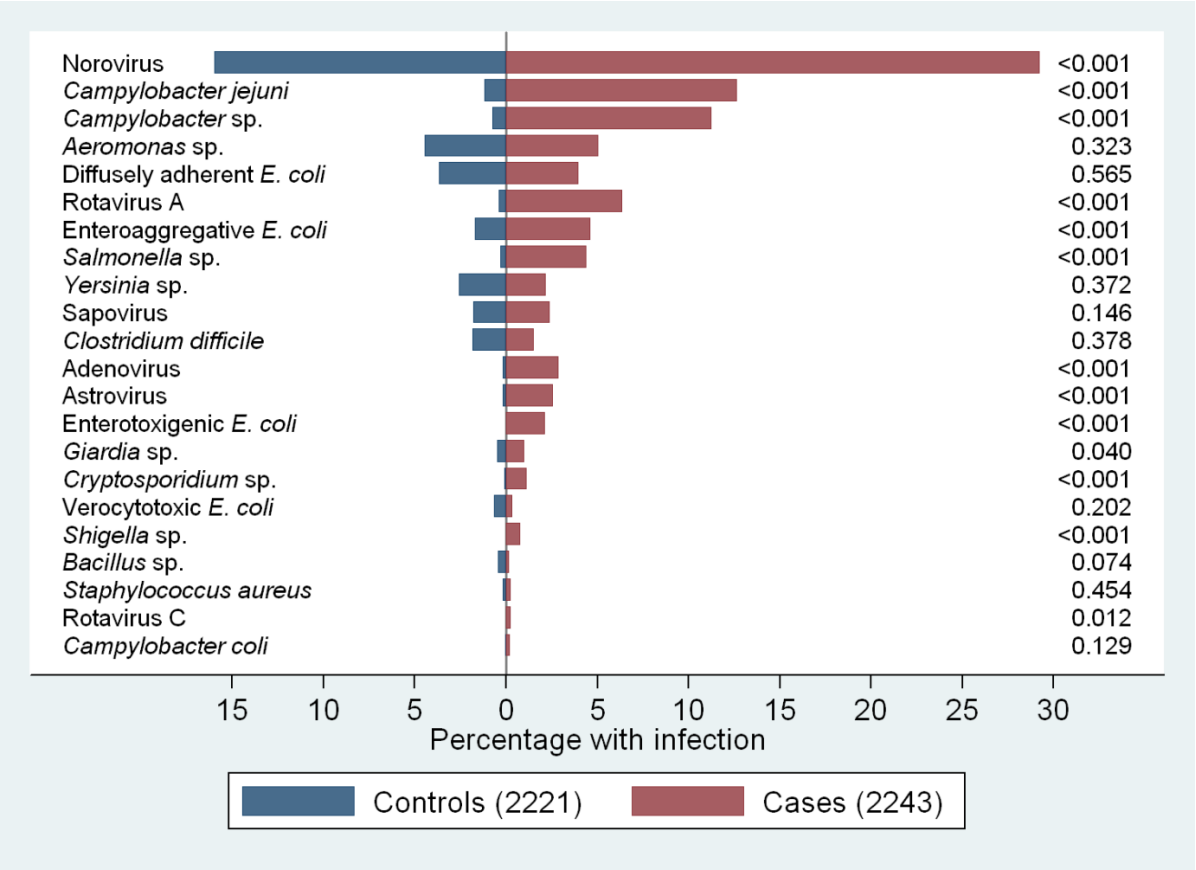


Figure 2.3: Submitting a stool specimen that was positive for enteroaggregative *E. coli* (EAEC) was positively associated with having disease. EAEC was found in <2% of controls, indicating that EAEC is not a ubiquitous commensal organisms. The p values are indicated on the right (chi-square). Figure used in Eurosurveillance publication (Chattaway *et al.* 2013).

2.3.3.2 Association of disease with individual pathogens in persons with multiple pathogens in their stool

The presence of co-infection was almost three times higher in EAEC-positive cases (74/113, 66%) than in EAEC-positive asymptomatic controls (9/38, 24% Figure 2.4) with more multiple co-infections in cases (38/113, 34%) than controls (1/38, 3%) (chi-square test, P<0.001). *Norovirus* and *C. jejuni* were statistically associated with being present in EAEC cases (Figure 2.4). Concomitant presence of other micro-organisms has decreased over the years; in the IID1 EAEC cases data subset came to a total of 42 (41.2%) whereas in the IID2 study, the EAEC cases co-infection total was 30 (35.7%) specimens. This analysis is not statistically relevant (chi-square test, P=>0.05) but the IID1 data showed that there was only one EAEC control that had a co-infection (3%) and so there is a higher association with co-infection and cases then with controls.

2.3.3.3 Investigation of the independent association of EAEC presence with disease.

The logistic regression of EAEC status (but not Ct value) in univariate analysis gave an OR of 2.55, 95% CI: 1.91 - 3.39, P<=0.001; and in multivaritate analysis the OR was 2.41, 95% CI: 1.78 - 3.26, P<=0.001. This means that among IID cases, the odds of EAEC infection were 2.5 times higher compared with asymptomatic controls. The resulting adjusted PAF was 0.033% (95% CI 0.024-0.039%), suggesting that around 3.3% of cases of IID in the UK are attributable to EAEC. This confirms that EAEC is an independent cause of IID.

A comparison of co-infections with norovirus, the most common cause of IID, is presented in Figure 2.5. and shows enterotoxigenic *E. coli* and *Shigella sp.*to be statistically associated with EAEC as a co-infection (Figure 2.5).

Figure 2.4 **Co-infection of Enteroaggregative *E. coli* (EAEC) cases and controls**

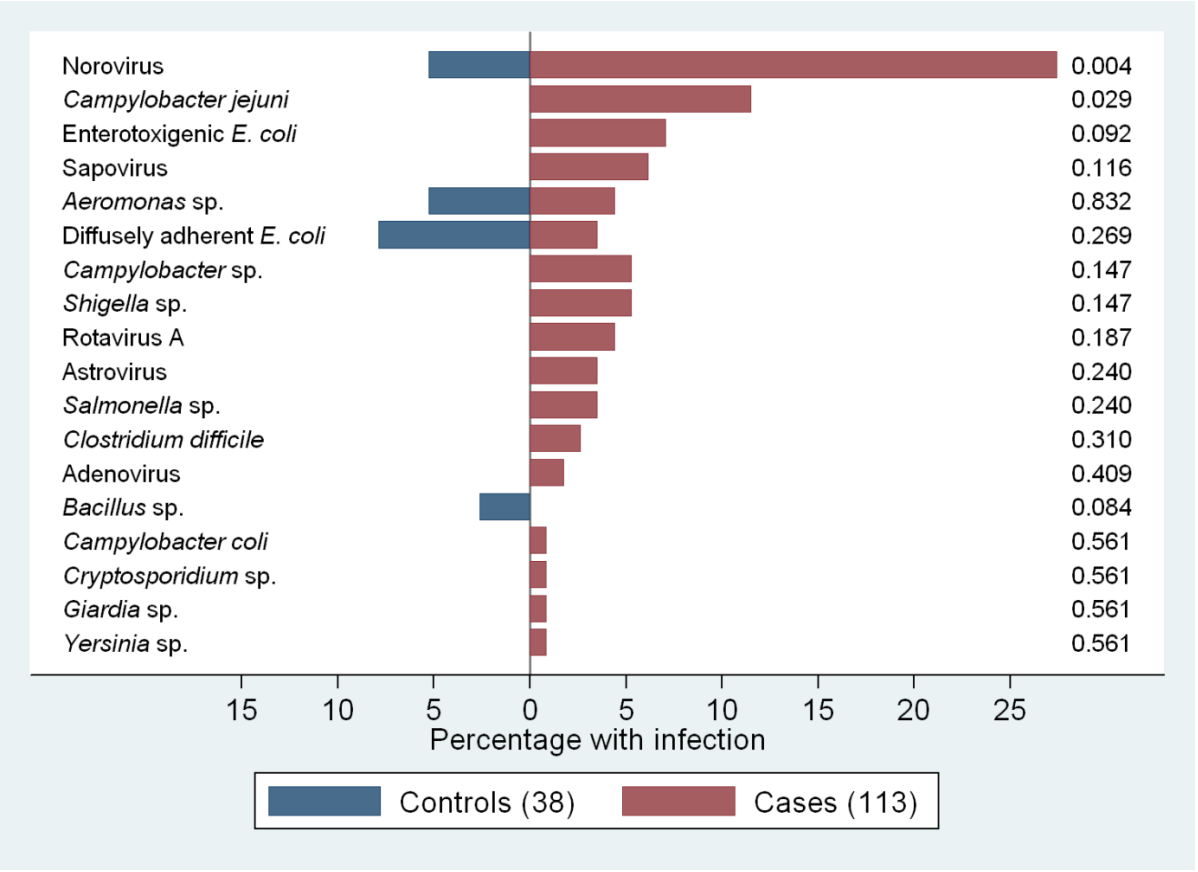


Figure 2.4: There were a higher variety of co-infection types, a higher percentage of co-infections and more multiple co-infections in EAEC positive cases than in EAEC positive controls.
Note: organisms designated sp. Include all species of that genus (except *Campylobacter sp*. *C. jejuni* and *C. coli* are listed separately). The p values are indicated on the right (chi-square). Figure used in Eurosurveillance publication (Chattaway *et al.* 2013).

Figure 2.5 **Comparison of co-infections with Enteroaggregative *E .coli* (EAEC) and Norovirus**

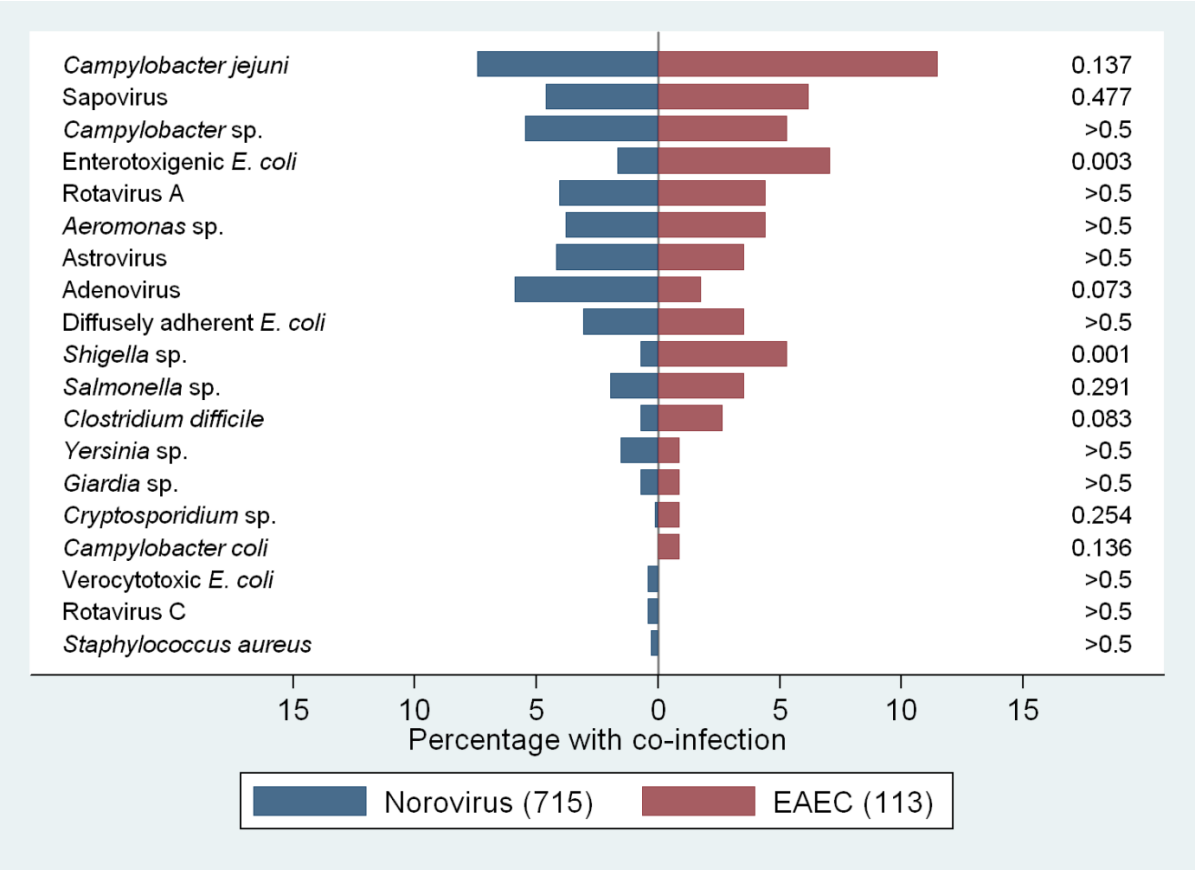


Figure 2.5: Co-infection with EAEC was more common than with norovirus (66% versus 43% respectively). *Staphylococcus aureus* refers to all *S. aureus* >106/g. The p values for individual agents are indicated on the right (chi-square). Figure used in Eurosurveillance publication (Chattaway *et al.* 2013).

2.4. Discussion

2.4.1 Burden of EAEC disease

Although described as a pathogenic group of *E. coli* it is well documented that EAEC may be associated with asymptomatic infection (Huang *et al.* 2007;Okeke *et al.* 2000a;Regua-Mangia *et al.* 2009). In this chapter the question was asked – how much gastrointestinal disease is caused by EAEC? In an attempt to remove healthy carriers from the case definition (a lower bacterial load might be expected in carriers than in cases) data was analysed from a PCR based case control study (IID1). Using data from the IID1 (1993-1996) case control study and the Ct value as an indicator of bacterial load, a defined diagnostic cut-off with 60% sensitivity and specificity (Ct <31) was established. Consideration of the bacterial load has enabled us to design a Ct cut-off values that attributes positive EAEC cases in the population. The cut-off value of Ct 31 represents a value where the number of false positives is closely equal to the number of false negatives (Figure 2.1) and therefore can be used as a value representative to EAEC burden within the population. It is not a cut-off value for diagnosing EAEC infection.

The application of this Ct value indicated that 47% (39/83 EAEC positive faeces below Ct 31) of EAEC positive cases in IID2 (2008-2009) were associated with disease which would represent about 1% of IID cases; this is equivalent to the burden of GI disease called by *Salmonella* (Tam *et al.* 2012b). However, the low sensitivity and specificity values would suggest that estimation of bacterial load by the Ct value of a quantitative PCR for virulence factors is not a useful diagnostic test for EAEC infection in an individual.

There was a strong association between a higher load of EAEC (low Ct) and being a case and so an attempt was made to define more accurately the EAEC positive individuals where EAEC was the causal agent of diarrhoea. The bacterial load data revealed the presence of two overlapping

normally distributed data sets for EAEC: one representing the load in the asymptomatic group (controls) and one in the symptomatic group (cases) (Figure 2.2).

The consideration of any possible confounding effects of age (and so acquired immunity) and co-infection was addressed using logistic regression confirmed by univariate analysis; the results showed that an individual was 2.5 times more likely to be a case then a control if they had EAEC. Therefore it was concluded that EAEC was independently associated with disease and an important burden in intestinal disease.

The results from both IID studies and consideration of factors such as co-infections and patients not reporting mild GI disease for investigation, show a possible under-representation of EAEC infection rate such as 1% of the population (Chattaway *et al.* 2013). This highlights EAEC and as an important aetiological agent of GI disease. EAEC should be regarded as a significant burden of enteric disease in the UK.

2.4.2 Causal link between EAEC presence in stool and disease

2.4.2.1 Carriage rates of EAEC in healthy controls, compared to other pathogens

Studies show that EAEC is detected in the stools of asymptomatic individuals (Figure 2.3). A similar situation is gastrointestinal viral infection where the most likely explanation is that post-infection levels of virus particles, although reduced, persist up to 56 days after symptoms have cleared (Atmar *et al.* 2008;Partridge *et al.* 2012). It is likely that the adherence mechanisms associated with EAEC also enables persistent carriage post-infection.

Another explanation of the presence of a pathogen in the absence of disease is pre-existing immunity to the infection at the time of exposure, which could result in reduced viral or bacterial

replication and a failure to develop symptoms. If pre-existing immunity was the cause of symptomless EAEC carriage we would expect to find an age distribution where adults are less frequently infected (older individuals have a higher chance of exposure and therefore a higher chance of immunity). However, the age distribution was even across the age groups. Having ruled out the significance of bacterial load and pre-existing immunity as explanations of asymptomatic infection, .the role co-infecting pathogens with the association of EAEC with GI disease was explored.

2.4.2.2 Association of disease with individual pathogens in persons with multiple pathogens in their stool

Previous studies have shown that organism load was related to disease in norovirus infection (Phillips *et al.* 2010). This study has shown that this concept cannot be applied to the presence of EAEC. A high EAEC bacterial load does not always directly link to disease and so it was suggested that distribution of co-infections warranted investigation.

The presence of increased co-infection in cases raises two possibilities: (i) - that the co-infecting pathogen rather than the EAEC is the aetiological agent and (ii) the interaction between the two organisms, is causing disease. To test this hypothesis we used norovirus, an infectious agent known to be present in both symptomatic and asymptomatic infection, as a comparator. As norovirus was a very common infection, we removed cases infected simultaneously with both norovirus and EAEC from the calculation: there were more co-infections in EAEC-positive cases than in norovirus positive cases (66% versus 43%). For EAEC co-infection, 12.6% were explained by enterotoxigenic *E. coli* (ETEC) and *Shigella* co-infections (Figure 2.5). This suggests that a proportion of EAEC cases can be explained by the presence or interaction of other pathogens (ETEC and *Shigella* are associated almost exclusively with symptomatic infection).

This high proportion of co-infection with EAEC cases compared to controls from the IID1 study (66% compared with 24%), may be an indicator that infection of the gut, such as with EAEC, may change bowel surfaces (for example the gut mucosa may become inflamed) and lead to favourable conditions for multiple types of pathogen colonisation (Strauman *et al.* 2010). Alternatively, it may be the co-infecting pathogen that enables EAEC to colonise the host more effectively, possibly by removing competition and exposing attachment sites.

Another consideration is that in certain cases the co-infection is causing disease and EAEC is a gut commensal. This hypothesis is supported by statistical analysis showing that an individual is highly likely to be an EAEC case if they have a co-infection (Chi-square, P=0.000). However, the logistic regression univariate data, and also the multivariate analysis, that accounts for co-infections, show that there is an association with EAEC and disease and you are 2.5 times more likely to be a case if you have EAEC. Therefore EAEC is capable of causing disease in certain people. It is clear that there is a relationship between EAEC and co-infecting microbiological agents but direct association with a specific organism in relation to disease could not be elucidated from this analysis. Non-specific changes or damage to the gut by either co-infections or EAEC carriage could be the disease mechanism for EAEC cases rather than interaction with specific co-infection organisms.

2.4.2.3 Independent association of EAEC presence with disease

The logistic regression of co-infection univariate and multivaritate was statistically significant and again confirmed that EAEC was independently associated with disease; the odds of disease were 2.4 times higher if EAEC was present than if not and were still highly significant after controlling for co-infections. The PAF adjustments indicated that EAEC would be responsible for disease in 3.3% of cases, a significant proportion in gastrointestinal disease and higher than for Salmonella (Tam *et al.* 2012b). Although age was an independent predictor for disease overall, controlling for age did not change the association of disease with EAEC, and there was no interaction between EAEC and age.

Therefore EAEC was independently associated with disease and it was concluded that most EAEC strains are capable of causing infection but not all.

The work in this chapter does not directly address causality over association but suggests that bacterial strain variation best explains the results for the following reasons.

There are two common arguments for EAEC being found in high levels in healthy individuals:

1) Low levels of EAEC are present in a commensal relationship in the human gut and only increase to detectable levels after infection with a true pathogen but an independent association of EAEC with disease argues against this for at least half of the infections in this study.

2) Post infection immunity leading to carriage in apparently healthy individuals; a lack of any detectable trends in age distribution and no clear distinction between pathogen load and disease, as seen in norovirus infection (Phillips *et al.* 2009), suggests that acquired immunity is not occurring in the UK population. Transient passage, is also unlikely as there is no known long term reservoir for exposure to EAEC from outside the human gut.

It is suggested that some, but not all, EAEC cause infection. The explanation for this may lie within defining EAEC by *in vitro* phenotype and a more detailed analysis of the phylogeny and putative pathogenicity genes is required. It is suggested that non-pathogenic EAEC that are able to adhere in an aggregative pattern to HEp-2 cells in the laboratory but unable to cause disease in the human host, are found in controls and in co-infections with true pathogens but pathogenic variants are found as the sole pathogen detected in diarrhoeic stools.

The alternative definition of EAEC by genetic markers to relate to *in vitro* phenotype has been attempted using alternative probes but this has also failed to define those EAEC capable of causing disease. For example, neither the presence of the target genes most commonly used in diagnostic assays, *aat* (anti-aggregative transporter) (Denno *et al.* 2012) and/or *agg*R (a transcriptional

activator) (Antikainen *et al.* 2009;Denno *et al.* 2012;Gomez-Duarte *et al.* 2010;Toma *et al.* 2003) correlate precisely with disease.

It may be that the diagnostic genetic factors used for EAEC are not true virulence factors – rather they encode the ability to adhere to human intestinal cells and facilitate colonisation (especially during infection with a true pathogen). It is likely that a combination of the EAEC associated adherence factors, other virulence factors and genetic background enables EAEC to cause primary infection. This was demonstrated with the strain of ST678 (serotype O104:H4) associated with the outbreak in Germany in 2011 (Chattaway *et al.* 2011) where the EAEC adherence genes were present in the same bacterial host as the shiga-like toxin gene (*stx)* in a stable genetic background. A robust and reliable diagnostic test for pathogenic EAEC is required to elucidate the true burden of GI disease caused by EAEC. Ideally, a suitable diagnostic assay would detect a combination of the EAEC plasmid genes in addition to chromosomally encoded virulence factors. The work carried out during this thesis explores "other" genomic factors associated with diarrhoeagenic EAEC.

The main limitation of this study is the absence of controls in the IID2 study. Although there were 15 years between the IID1 and IID2 studies, the demographic data for cases suggest that the epidemiology has not changed during that period (Chattaway *et al.* 2013). Although there was a slight decrease in co-infection rates, it was not statistically different. It is believed the burden data is still relevant in 2014. Another interesting observation of the study is the range of co-infectious agents identified. Small numbers of cases with co-infections (six cases or less for EAEC co-infections with *C. difficile, Yersinia, Giardia, Cryptosporidium, rotavirus,* VTEC and *Staphylococcus)* meant that the ability to detect statistical differences between cases and controls was limited. However, the study did allow, for the first time, the explanation of the association between EAEC and all potential co-infecting agents as well as the more common pathogens norovirus (n=29) and Campylobacter (n=12).

2.4.3 Application of *the aat* gene Ct value as a mean of diagnosing EAEC infection

The assessment of Ct value in this study has shown a potential method of assessing burden in the population studied, although the EAEC Ct values does not equate to an idealistic ROC curve (Figure 2.1) or a typical normal distribution curve (Figure 2.2) like you would find in norovirus (Phillips *et al.* 2010) and so it is not suitable method to use a cut-off value for diagnosing EAEC infection.

Intestinal infection is multi-factorial (infectious dose, immunity status, underlying physiological conditions) and though the burden of EAEC is generally higher in cases, other factors have not been taken in to account such as co-infection. Microbiological protocol to diagnose infection is to look for a dominant (i.e. most easily isolated and present in high numbers) or recognised known pathogen (i.e. *Salmonella*, *Shigella* e.t.c.) and EAEC is not currently part of these protocols.

If microbiological protocols were to change to include EAEC, the method still needs to be agreed as cell adhesion assays are not practical at the frontline laboratories. Although the *aat* (anti-aggregative transporter encoding gene) probe has become an alternative method for EAEC detection over the gold standard of the HEp-2 assay (Vial *et al.* 1988), and was the probe of choice in both IID studies, studies have shown that not all EAEC carry the plasmid and targeting the *aat* gene doesn't always detect the organism (Jenkins *et al.* 2006a). There are other EAEC specific genes being investigated for detection methods such as the *aggR* gene (a transcriptional activator gene which regulates multiple chromosomal and plasmid virulence factors) (Antikainen *et al.* 2009;Gomez-Duarte *et al.* 2010;Toma *et al.* 2003), but with EAEC being such a heterogeneous organism, until a properly defined sub-set of diarrhoeagenic EAEC is described an improved detection system will be difficult to develop

2.5. Summary

This chapter has provided a unique insight into the burden of IID attributed to EAEC, which was estimated at 1% of the UK population in 2008-2009. The presence of EAEC was not always associated with disease, and the use of a Ct value cut-off to accurately diagnose EAEC as a cause of infection was not robust. This methodology alone cannot be used to diagnose EAEC infection and alternative approaches are needed. It is hypothesised that EAEC is a mixture of pathotypes of which only some groups are capable of causing disease (Chattaway *et al.* 2013). The following chapter describes the population structure of EAEC and analysis of the data in order to define sub-groups within the EAEC population and find associations with disease or carriage.

Chapter Three

Definition of Pathogenic Enteroaggregative Escherichia coli

Chapter 3 - Definition of pathogenic EAEC groups by a case control approach

3.1 Background

Chapter 2 showed that EAEC plays an important role in gastrointestinal infection and hypothesised that certain sub-set populations of EAEC are pathogenic and capable of causing symptoms of GI disease, while other are not pathogenic. The work described in this chapter investigates this hypothesis further.

Attempts to define enteroaggregative *E. coli* (EAEC) have been based on traditional methods, for characterising and typing *E. coli* developed over the past century. Serotyping has been a useful method for identifying the pathotype Shiga toxin producing *E. coli* (STEC) serotype O157:H7, as the serotype is a robust marker for this successful and stable clonal group. In this context, a successful clonal group is defined as a group of closely related strains that have continued to proliferate over time and are still present in the population. STEC O157:H7 has a recognisable genetic background (ST11), and a defined set of virulence factors (*stx* toxin and *eae* intimin gene). However, serotyping doesn't help define the pathogenic profile of all *E. coli* . A more useful method is to detect the presence of specific pathogenicity genes. The approach of genotyping and virulence gene profiling over serotyping has been used to identify *E. coli* pathotypes where the genetic background is stable and key genes are found (Table 1.1).

However, this methodology and approach has been unsuccessful in determining groups or profiles for EAEC. Serotyping has revealed hundreds of combinations of somatic 'O' antigens and flagella 'H' antigens found in both EAEC cases and controls, and so few serotypes have been linked specifically to the EAEC pathotype (Jenkins *et al.* 2006a;Okeke & Nataro 2001). The most common group defining genes used to detect EAEC are plasmid borne (*aat* transporter gene and *agg*R regulator gene). Virulence gene profiling in various case control studies has resulted in the growing number of

putative pathogenicity factors emphasising the heterogeneity of this complex group (Huang *et al.* 2007;Jenkins *et al.* 2006a;Jenkins *et al.* 2006b;Okeke *et al.* 2000a;Pereira *et al.* 2007;Regua-Mangia *et al.* 2009) . Essentially, the traditional methods of characterising EAEC by HEp-2 adherence and EAEC virulence gene content have not shown a direct link to the pathogenicity of EAEC and these are not suitable methods to determine detection of pathogenic EAEC. While most studies focus on the plasmid encoded genes, this study adopted a novel approach based on the use of MLST to determine the population structure of EAEC to investigate and understand the background of isolates from multiple case control studies around the globe, including the UK, Bangladesh and Nigeria. It is hypothesised that this approach can elucidate a clear definition of pathogenic EAEC by assessing the association with disease against the core genetic background which is more stable in comparison to mobile genetic elements.

The applications of novel approaches in research are not always tested in public health situations (to ascertain the validity and impact of new methodologies) in order to facilitate public health investigation. The data obtained from this study were tested in two different public health settings to assess the utility of the methods described in this chapter.

1) The European SAFEFOODERA-ESBL project was initiated to assess the prevalence of Extended spectrum beta-lactamase (ESBL) producing *E. coli* in humans (n=274), food producing animals (n=295) and animal food products (n=59) across Europe to ascertain if there are any zoonotic groups of ESBL in the food chain and characterise the strains. Isolates that were ESBL positive underwent further characterisation and are the isolates used in this study (Wu *et al.* 2013). The population structure of this dataset was assessed using the methods described in this thesis to determine whether there were any ESBL complexes prevailing across Europe. Isolates were investigated from Germany (n=84), The Netherlands (n=254) and the UK (n=291) (Wu *et al.* 2013).

2) The second study was to use MLST in the investigation of EAEC outbreaks to ascertain if there were certain complexes responsible for outbreaks. Known EAEC outbreaks over the past two decades investigated at the Gastrointestinal Bacteria Reference Unit were mapped onto the EAEC population structure developed as part of the work described in this chapter and analysed.

3.2. Methods

3.2.1 Strains Used in the study

Three case control studies, sporadic and outbreak cases of 564 EAEC spanning over 29 years (1985 – 2013) were used in this study (Table 3.1). All of these strains were included to encompass a representation of EAEC in the global community (including UK travellers) over the past three decades. EAEC were defined as having the *aat* gene/CVD432 probe reaction (Baudry *et al.* 1990;Nishi *et al.* 2003), and/or the *aggR* regulatory gene (Jenkins *et al.* 2006a) and/or the aggregative adherence (AA) phenotype (Nataro *et al.* 1987) where the phenotypic test was available (Table 3.1).

The three case control studies included both IID studies from the UK described in Chapter 2 and a Bangladesh and Nigerian study. The Nigerian study defined a case as a child aged 5 years or younger attending one of four primary health care centres in Osun State (Southwest Nigeria) who had frequent stools (usually more than 3 daily), lasting less than 2 weeks. Control subjects were healthy children of the same age range. This resulted in obtaining specimens from 187 cases and 144 controls for testing (Okeke *et al.* 2000a) from which 66 cases and 55 controls EAEC isolates were available for analysis as part of this thesis . The Bangladesh study were strains taken from the Global Enteric Multicenter Study (GEMS) and cases were defined as children of 0-59 months who fulfilled the WHO definition of diarrhoea (3 or more loose stools in 24 hours), controls were matched from the community by age, sex and near by by village who had no diarrhoea for seven days (Kotloff *et al.* 2012). This study resulted in 550 cases and 878 control specimens being tested (Kotloff *et al.* 2013) from which 97 cases and 61 EAEC isolates were available for this study.

In summary, Isolates included strains from multiple case control studies including the UK (273), Bangladesh (169), Nigeria (121) and the prototypical 042 EAEC reference strain from Peru (1) (Table 3.1). In addition to the case control study, EAEC identified at the Gastrointestinal Bacteria Reference Unit over the past 3 decades from clinical specimens or outbreaks were included. Due to the varying

definition of EAEC, all strains were included irrespective of phenotypic and genotypic definition to prevent any bias that may affect the analysis. Where an EAEC outbreak was related to one ST and serotype, only one representative strain was included.

All EAEC strains described above were held in the archive at GBRU, except the strains from the Nigerian study where only the MLST data was made available. All strains, except for the strains from Nigeria, were plated onto Columbia blood agar plates (5% sheep blood) [PHE, Media] to test for purity and archived onto Dorset Egg slopes (egg white and yolk) [PHE, Media] and stored at room temperature and also archived on cryobeads [Prolab] and stored at -80^{o}C.

Table 3.1 List of 564 strains used in this study

Country	Source	Year Range	Case	Control	Reference
Peru	∞042 prototypical strain	1985	1	0	(Nataro *et al.* 1985)
UK	#GBRU Archive Clinical strains	1985-1995	17	0	This Study∞
UK	∞IID1 Case/Control Study	1993-1996	121	36	(Wilson *et al.* 2001)
UK	∞GBRU Outbreak A	1994	2	0	(Spencer *et al.* 1999)
UK	∞GBRU Outbreak B	1994	8	0	(Spencer *et al.* 1999)
UK	∞GBRU Outbreak C	1994	1	0	(Spencer *et al.* 1999)
UK	∞GBRU Outbreak D	1995	3	0	(Spencer *et al.* 1999)
Bangladesh	∞GBRU Outbreak E	1998	12	0	This Study∞
Nigeria	∞Nigeria Case/Control Study	1999	66	55	(Okeke *et al.* 2010)
UK	#IID2 case study	2008-2009	25	0	(Chattaway *et al.* 2013)
Bangladesh	∞GEMS Case/Control Study	2007-2011	97	61	(Kotloff *et al.* 2012;Panchalingam *et al.* 2012)
Germany	#O104:H4 VTEC Outbreak	2011	1	0	(Chattaway *et al.* 2011)
UK	#O111:H2 Household Outbreak	2012	1	0	(Dallman *et al.* 2012)
UK	#GBRU Clinical Strains	2009-2013	38	0	This Study∞
UK	#GBRU Spice Outbreak	2013	19	0	(Dallman *et al.* 2014)

Table 3.1 ∞Strains from this study not previously described include archived clinical strains received by GBRU for typing between 1985-1995, Outbreak E of enteroaggregative *E. coli* that occurred in Bangladesh in 1998, recent clinical strains received by GBRU for typing between 2009-2013. #EAEC were defined as having the *aat* and/or *aggR* gene. ∞Other EAEC strains were defined as having the *aat* gene /CVD432 probe reaction and/or the aggregative adherence (AA) phenotype.

3.2.2 Identification and serotyping of isolates from Bangladesh and UK

Biochemical confirmation of the identification of 443 Bangladesh, UK and Peru isolates as *Escherichia coli* was performed (Castellani and Chalmers 2005). Typical metabolic profiles of *E. coli* included gas production following growth in glucose, positive reactions for glucose, lactose, mannitol, lysine, ornithine, mucate, sodium acetate and indole. Serotyping of the somatic and flagella antigen (Gross and Rowe 1985) was carried out on the heat stable lipopolysaccharide (somatic or O) antigens and the flagellar (H) antigens. Strains which agglutinated with all antigens failed to express the O antigen were termed "rough" and those that did not agglutinate with any of the established serogroups or flagella antigens were termed 'O' or 'H' unidentifiable (O? or H?). Nigerian strains had previously been identified and published (Okeke *et al.* 2000b), strains were not available for serological identification.

3.2.3 Multi-locus Sequence Typing (MLST)

3.2.3.1 DNA Extraction

DNA extraction of the strains was undertaken using the Wizard kit [Promega, UK]. Pure cells were grown overnight in 1.2ml nutrient broth [PHE Media] in a shaking incubator [New Brunswick Scientific] at 37^{o}C in oxygen. The broth was centrifuged [International Equipment Company] at 13,000rpm for 4 min and the supernatant was discarded.

Cells were lysed by gently adding and mixing 630 µl of nucleic lysis solution, incubating for 10 min at 80^{o}C, cooling at room temperature and adding 3 µl of RNase solution. This was then mixed by inverting and incubated at 37^{o}C for 15 min and cooled at room temperature.

Proteins were then precipitated by adding 230 µl of protein precipitation solution and inverting immediately to mixed and then vortexed. They were then incubated on ice for 10 min, inverted and centrifuged at 13,000 rpm.

DNA was precipitated by transferring the clear supernatant to a clean tube containing 650 µl of isopropanol and mixed. The mixture was centrifuged at 13,000rpm for 15 min, the supernatant was

then discarded. One ml of 70% ethanol was added and mixed and the DNA was precipitated as white string. The tube was then centrifuged at 13,000 rpm for 1 min. The supernatant was carefully aspirated and the tube was left to air dry.

DNA was rehydrated in 50 µl of nuclease free water and left to solubilise overnight at 4^{o}C.

DNA was measured using a spectrophotometer [Qubit, Invitrogen] to ensure at least 50 ng was present for the PCR reaction.

3.2.3.2 Polymerase Chain Reaction

Gene fragments from housekeeping *E. coli* genes were amplified as described by Wirth *et al* (Wirth *et al.* 2006) to obtain allele data including *adk* (536 bp), *fum*C (469 bp), *gyr*B (460 bp), *icd* (518 bp), *mdh* (452 bp), *pur*A (478 bp) and *rec*A (510 bp). Each 25 µl reactions contained 1 µl DNA, 1 µl of each primer (10 pmol/µl) [MWG Eurofins] , 12.5 µl PCR master mix [Sigma] and 9.5 µl nuclease free water [Sigma]. Amplification was carried out on a PCR ABI 3700 thermocycler [Applied Biosystems] and included an initial denaturation step at 95^{o}C for 2 min followed by 30 cycles of the following conditions: denaturation at 95^{o}C for 1 min, annealing temperature for each primer set at 56^{o}C for 1 min and extension at 72^{o}C for 2 min, with a final extension step at 72^{o}C for 5 min.

Amplified DNA was prepared for sequencing using the ‘‘ExoSAP’’ method (Amersham Biosciences UK Ltd). Essentially, 1 µl of ExoSAP was added to 10 µl of amplified DNA, the two hydrolytic enzymes (Exonuclease I and Shrimp Alkaline Phosphatase) were activated and unwanted deoxynucleotides and primers were removed by heating the mixture in a thermocycler at 37^{o}C for 30 min and then 80^{o}C for 10 min.

3.2.3.3 Amplification of Sequencing Reaction

Cleaned fragments were sequenced from both ends using the di-deoxy chain terminator method (Sanger *et al.* 1992), with V3.1 Bigdye terminator chemistry (West *et al.* 2005) [Applied Biosystems]. Briefly, a dye PCR mastermix was made consisting of 1 µl of 5 pmol/µl forward primer, 4 µl of terminator ready reaction, 4 µl of nuclease free water and 1 µl of cleaned amplified DNA template

(5-20ng). This was also repeated for the reverse primer as both strands of each fragment were sequenced at least once. The reaction conditions carried out on a PCR ABI 3700 thermocycler [Applied Biosystems] and included an initial denaturation step at 96^{o}C for 1 min followed by 25 cycles of the following conditions: denaturation at 96^{o}C for 10 seconds, annealing temperature for at 50^{o}C for 5 seconds and extension at 60^{o}C for 4 min. The final reaction was stored at -20^{o}C.

3.2.3.4 Cleaning of Amplified Sequencing Reaction

Sequencing reaction was cleaned by adding 15 µl of nuclease free water into each well of amplified sequencing reaction, 52 µl of EtOH/ NaOAc (from stock of 7 ml ethanol and 280 µl sodium acetate pH 5.2 3M [Sigma]) was added, vortexed and incubated at room temperature for 45 min. The plate was then centrifuged for 1 hour at 2800 x g at 4^{o}C.

After spinning, the plate was inverted to decant the supernatant and then placed inverted onto whatman filter paper and spun at 500 x g for 1 min.

Fragments were then washed by adding 150 µl of 70% EtOH and centrifuged at 2800 x g at 4^{o}C for 10 min. After spinning, the plate was inverted to decant the supernatant and then placed inverted onto whatman filter paper and spun at 500 x g for 1 min. The plate was stored at -20^{o}C until sequenced.

3.2.3.5 Sequencing of Cleaned Sequencing Reaction

Both ends were sequenced at least once using the di-deoxy chain terminator method with v 302 Bigdye terminator chemistry The resulting sequencing reactions were analyzed on 3700 ABI sequencing machines [Applied Biosystems, USA].

3.2.3.6 Analysis of Sequenced Data

Sequence data was imported into BioNumerics V 6.5 and fragments were aligned and assessed for quality. Consensus trimming was carried out for each allele and fasta files were exported. Allele and sequence type (ST) assignments were made at the publicly accessible *E. coli* MLST database at http://mlst.ucc.ie/mlst/dbs/Ecoli/.

3.2.4 Population structure analysis

To assess the population structure of the EAEC isolates in this study, Minimal Spanning Trees (MSTree) (where the two STs with the greatest number of single locus and then double locus variants are linked first, preferably using intermediate STs) were constructed using BioNumerics V 6.5 software.

3.2.4.1 Assessment of EAEC Disease and Carriage Groups

As of 18th December 2013, the data available in public database indicated there were 155 EAEC (excluding the 121 Nigerian strains described by *Okeke et al.* 2010*,* out of 6110 *E. coli* entries representing 2.4% of the database. There were 1164 entries of defined diarrhoeagenic pathotypes of *E. coli* of which EAEC accounts for 13 % (155/1164). Diarrhoeagenic pathotypes included ETEC, STEC, EPEC, EIEC and DAEC (Table 1.2)

From the 564 strains used in this study, a complex (defined by a ST and any single locus variants related to that ST) was considered successful if it contained 4 or more strains accounting for a minimum of 2.5% (4/155) of the known EAEC deposited in the public database.

From the EAEC MSTree dataset used in this study, there were 17 complexes (Cplxs) containing four or more EAEC that were deemed representative of successful strains (i.e. strains which have continued to proliferate over time in the population) . The assigned Cplxs were then tested using a fishers exact test (Fisher 1922) for the significance of the groups being associated with disease or carriage in relation to the entire dataset (564 strains). Statistical tests of significance were conducted using the Fisher's exact test on Epi-Info version 2.3.1 (http://www.openepi.com) (Appendix 7.5).

In order to understand if these successful groups were exclusive to EAEC or had evolved with other *E. coli* pathotypes, the public database was compared against each of the 17 Cplxs to rule out groups with a high association with other pathotypes. Extra-intestinal sites of infection for ExPEC (1.1.2) included wounds, meningitis, external sources (ExPEC_Vag) and urinary pathogenic *E. coli* (UPEC). Antibiotic resistance *E. coli* including extended beta-lactamase producing *E. coli* (e.g. presence of

CTX-M-15, AmpC CYM-2, c CMY-2, NDM-1, CTX-M-32 & OXA-48 genes). Other pathotypes included avian pathogenic *E. coli* (APEC) ,non-pathogenic commensal strains and *E. coli* with no defined pathotype.

The 17 Cplxs described in this study were compared against all *E.coli* complexes on the public database and data from this study was tested using a fisher exact test (Fisher 1922) on Epi-Info version 2.3.1 (http://www.openepi.com) (Appendix 7.5), for significance of the 17 Cplxs being associated with EAEC over other pathotypes of *E.coli*.

3.2.5 Statistics

Fishers Exact test (Fisher 1922) was used to test the significance of the groups being associated with cases and controls. Fisher's exact test is a statistical significance test used in the analysis of contingency tables. Although in practice it is employed when sample sizes are small (<5 samples), it is valid for all sample sizes. The test is useful for categorical data and was used in this study to investigate disease versus carriage groups; it is used to examine the significance of the association (contingency) between the two kinds of classification (Altman 1991). In this case, the aim was to determine whether a named disease group is associated with cases or whether a carriage group was associated with control.

3.2.6 Impact of novel approach on public health

Application of MLST to investigate EAEC population structure in a public health setting was assessed in two ways:

1. MLST was performed on the ESBL-producing strains from the SAFEFOODERA study to determine if EAEC is associated with an animal reservoir and if there are any lineages of EAEC that are particularly resistant to antibiotics.
2. Using MLST in EAEC outbreak investigations as a typing method and to determine if there are certain lineages associated with outbreaks. (Wirth *et al.* 2006) (Methods 3.2.3).

3.2.6.1 Detection of EAEC in other E. coli populations

The 359 ESBL-producing *E. coli* isolates used in this study were from the SAFEFOODERA collection (Wu *et al.* 2013) (www.safefoodera.net) and isolated in Germany (73), The Netherlands (158) and the UK (128). Isolates were obtained from multiple sources, including cases of humans infections (140), poultry (137), cattle (63), pigs (16) and dog/cats (3). Isolation sites included urine (97), faeces (92), cattle and poultry meat (64), caecum (51), organs (19), blood (8) and other multiple single sources (28). Isolates had been stored in Luria Bertoli broth and stored at -20^{o}C, having been collected between 2005 and 2009. MLST was carried out as described in method section 3.2.3 and 3.2.4. Isolates were grown aerobically overnight on Columbia blood agar plates (5% sheep blood) [PHE,Media] aerobically at 37^{o}C. DNA was extracted using the Wizard kit [Promega, Mannheim, Germany] (3.2.3.1), stored at 4^{o}C, and used as template for real-time PCR assays. DNA was screened for the presence of the EAEC transport regulator gene (*aggR*), located on the EAEC plasmid (Nataro *et al.* 1994). *AggR* primers and probes were designed for this study (AggR_F 5'-CCATTTATCGCAATCAGATTAA-3' AggR_R 5'-CAAGCATCTACTTTTGATATTCC-3', AggR_P Cy5-CAGCGATACATTAAGACGCCTAAAGGA-BHQ) (Chattaway *et al.* 2014a). Positive control strain used included E60725 (O92:H33).The amplification conditions included an initial denaturation of 95 ^{o}C for 5 min, then 94 ^{o}C for 1 min, 55 ^{o}C for 1 min and 72 ^{o}C for 1 min for 30 cycles, and a final extension of 72 ^{o}C for 10 min on a Rotagene [Qiagen, Manchester, UK].

3.2.6.2 MLST in the investigation of outbreaks

Outbreak strains included (Table 3.1) were colour coded in the population structure from this study to ascertain if there were particular lineages associated with outbreaks or if EAEC outbreaks are easily resolved via MLST.

3.3. Results

3.3.1 Population Structure Results and EAEC Group Assignments

3.3.1.1 Sequence Type distribution and complex structure

Of the 564 EAEC strains studied, there were 126 different sequence types , with additional new sequence types including 57 single locus variants (SLV), 20 double locus variants (DLV) and two triple locus variants (TLV) (Table 3.2). There were 17 main complexes containing 4 or more strains of EAEC totalling 358 strains with the top five complexes including ST10 Cplx (39%, 141/358), ST31 Cplx and ST40 Cplx (12%, 42/358), ST394 Cplx (7%, 26/358) and ST295 Cplx and ST38 Cplx (6%,21/358) (Figure 3.1). There were 35 isolates (6.2%, 35/564) that contained one or more new alleles (40 new alleles in total) not previously found. All new alleles were deposited to the public database (http://mlst.ucc.ie/mlst/dbs/Ecoli) for a new allele and/or ST assignment.

The population of EAEC strains from this study showed a complex clonal structure where multiple lineages have arisen. Some of these lineages have been successful and expanded into large groups and adapted with multiple mutations. Examples include the clonal complex ST10, 200,130, 394 & 38. Other lineages have expanded, apparently in isolation, and are not linked to other groups (ST678, 720 and 1891) and many singletons (i.e. a ST that is not linked to any SLV) exist and are unlinked to any complexes. Generally, the diversity of these strains was observed on a global scale, however, some STs were associated with a specific geographical location such as Bangladesh (ST1891 & 720) and UK (ST40, 1380 and 165). There were no representative groups (i.e.four or more strains) that were only found in the Nigeria collection but there were multiple exclusive ST consisting of a maximum of two strains (Figure 3.2). Some STs were found in two countries, such as Bangladesh and the UK (ST200, 295 and 678), Bangladesh and Nigeria (ST 484 and 155), and others were dispersed throughout all three countries (ST10, 130, 31, 34, 38, 278 and 394) (Figure 3.2, Appendix 7.1).

Table 3.2 Summary of Sequence Types found in EAEC dataset

Sequence Type	Complex	No.	%
10	10	91	16.1
34	10	38	6.7
200	40	31	5.5
31	31	25	4.4
38	38	21	3.7
394	394	18	3.2
295	295	16	2.8
40	40	12	2.1
130	31	12	2.1
278	278	10	1.8
678	None	10	1.8
484	168	8	1.4
1380	394	8	1.4
3748	295	6	1.1
48	10	6	1.1
43	10	6	1.1
449	31	5	0.9
720	None	5	0.9
1891	None	5	0.9
155	155	5	0.9
159	746	4	0.7
30	30	4	0.7
349	349	4	0.7
165	165	4	0.7
226	226	4	0.7
SLV	Various	57	10.1
DLV	Various	10	1.8
TLV	Various	2	0.4
414,746,841,501,206,167,315	Various	3	0.5
433,362,455,499,480,1114,2186, 1295,435,495,456,3570,467,481, 157,459,474,1326,218,58	Various	2	0.4
2166,219,448,223,3107,3051,504,120,520,436,475,510, 488,556,460,483,461,23,466,423,513,489,511,473,940, 424,438,476,512,426,471, 557,464,52,477,515,468,437, 507,506,454,450,434,485,478,46,500,502,469,453,444, 144,728,496,491,425,422,486,2517,73,937,117,1657, 111,152,329,227,164,1136, 2067,101,3670,1490	Various	1	0.2

Table 3.2 Sequence types found in EAEC dataset (N=564), shared complexes are colour coded, ST10 Cplx (red), ST40 Cplx (green), ST394 Cplx (blue), ST295 Cplx (purple), ST31 Cplx (orange)

Figure 3.1 Distribution of Complexes in EAEC dataset (N=358)

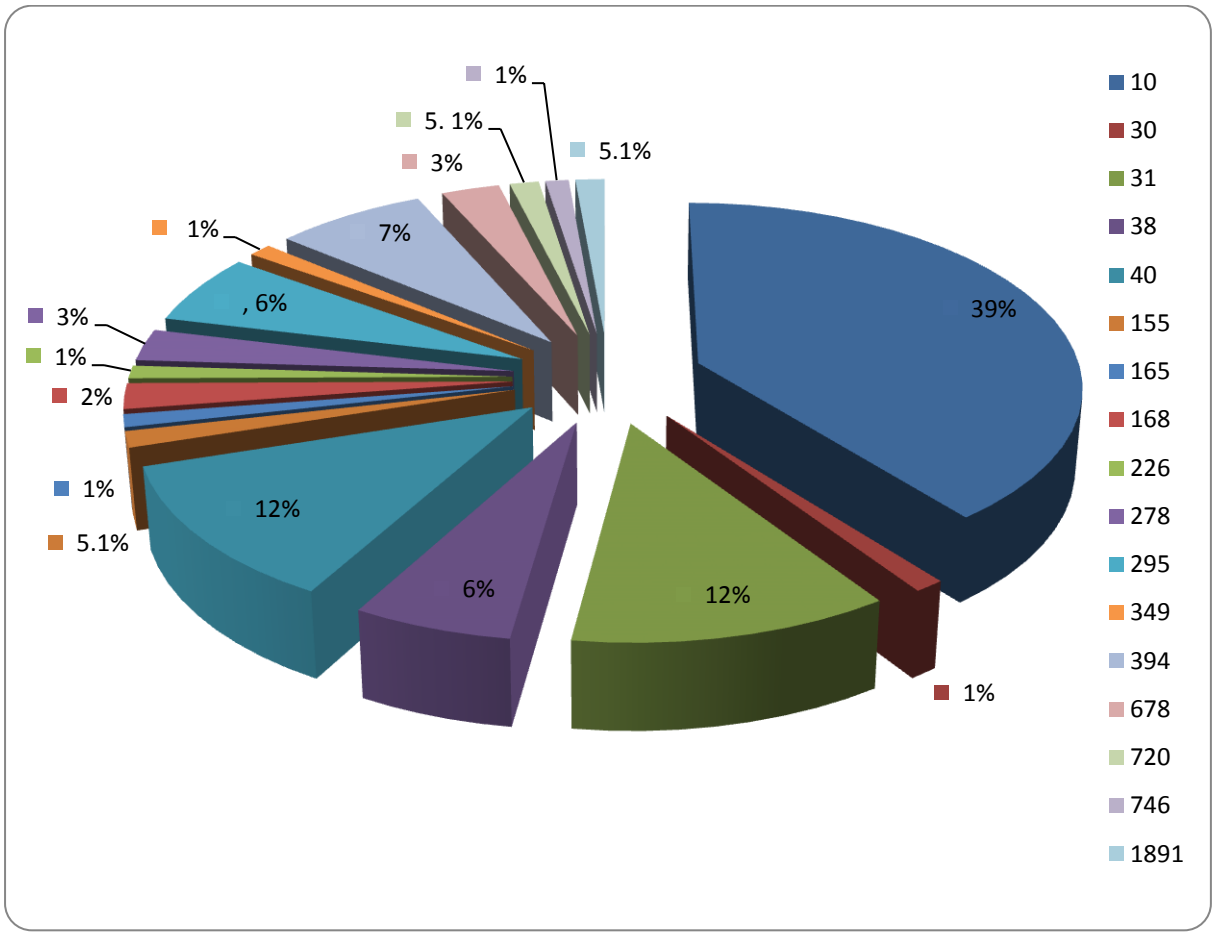


Figure 3.1: Pie Chart showing the main complexes containing 4 or more EAEC strains comprising of 63.5% (358/564) of the dataset.

Figure 3.2 Minimal spanning Tree of EAEC pathogenic disease groups in relation to geographical location

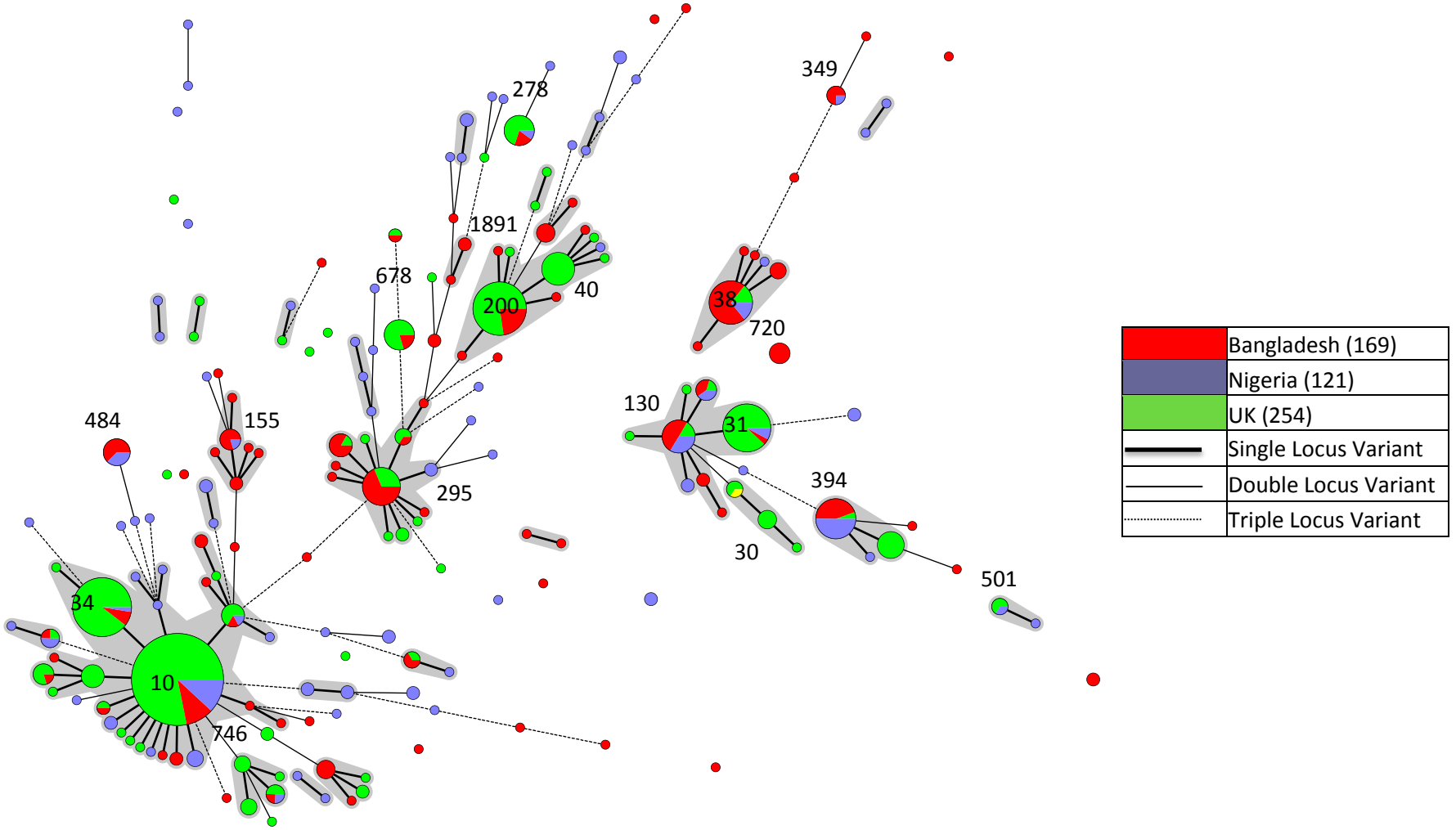


Figure 3.2: Minimal spanning tree of the 564 EAEC used in this study colour coded by isolates from Bangladesh (red), Nigeria (purple) and UK (green) and the prototypical 042 strain from Peru (yellow). Complexes shaded in grey consist of single locus variants (SLV). MSTree shows that complexes are mainly distrusted in at least two countries with only a few small groups geographically specific.

3.3.1.2 EAEC groups associated with disease and carriage

The population structure of EAEC is heterogeneous (Figure 3.1, 3.2 and 3.3) comprising of 17 successful Cplxs of four or more EAEC isolates (Table 3.3). EAEC isolates from cases and controls are dispersed throughout the population structure, ST720 is the only successful group that solely contains isolates from cases (Figure 3.3).

There was a 2.7:1 ratio of case isolates to controls in this study, groups with a higher ratio in cases were deemed associated with cases and groups with a higher ratio in controls were deemed associated with controls. Groups that were below this ratio were deemed to not be associated with cases or controls. This resulted in eleven groups being associated with disease (ST10, 40, 746, 155, 678, 278, 30, 165, 1891, 720 & 501 Cplx), two groups associated with carriage (ST31 & 349 Cplx) and four groups not associated with disease or carriage (ST295, 38, 394 & 168 Cplx).

The disease complexes and carriage complexes were combined and statistical analysis showed both of the disease and carriage complexes were statistically significant (P = <0.001 and P = 0.001 respectively).

Individual complexes were then tested for statistical association with disease or carriage which showed ST10 Cplx and ST40 Cplx were independently statistically significantly (P = 0.01 & 0.03 respectively) associated with disease. ST31 Cplx was independently statistically significantly (Fishers chi-square, p=0.005) associated with carriage (Table 3.3).

Situating the 17 successful EAEC complexes identified in this study within the global *E. coli* phylogeny as represented in the public database (Table 3.4) showed that with the exception of ST155 Cplx, all complexes were significantly associated with being EAEC pathotype (P≤0.01).

Figure 3.3 Minimal spanning Tree of EAEC pathogenic disease groups in relation to case or control

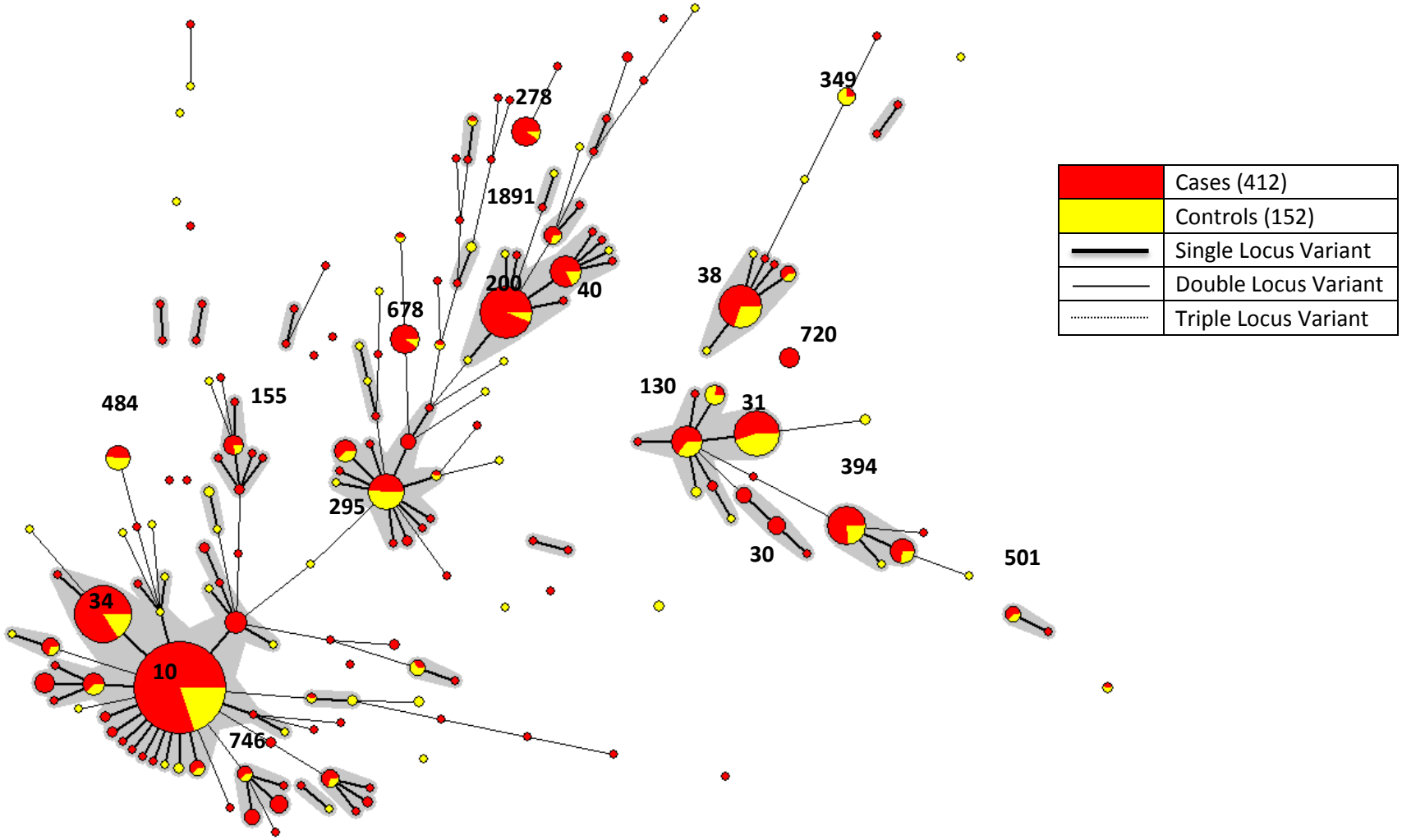


Figure 3.3: Minimal spanning tree of the 564 EAEC used in this study colour coded by isolates from cases (red) and controls (yellow). Complexes shaded in grey consist of single locus variants (SLV). Trees shows that complexes usually contain a mixture of cases and controls with few STs being exclusively associated with cases (ST720, ST30).

Table 3.3 Assessment of EAEC groups associated with cases or controls

EAEC Group	ST complex	UK	Nigeria	Bangladesh	Case	Control	Total	Total % of EAEC	CASE: CONTROL %	P Value
Group 1	10	128	24	21	138	35	173	30.7	80:20	**0.01**
Group 2	40	39	1	12	44	8	52	9.2	85:15	**0.03**
Group 3	31	27	11	12	28	22	50	8.9	56:44	**0.005**
Group 4	295	13	2	21	24	12	36	6.4	67:33	0.24
Group 5	38	3	4	21	19	9	28	5.0	68:32	0.33
Group 6	394	9	10	8	20	7	27	4.8	74:26	0.56
Group 7	746	9	1	1	10	1	11	2.0	90:10	0.16
Group 8	155	0	1	9	9	1	10	1.8	90:10	0.2
Group 9	678 (ST484)	8	0	2	9	1	10	1.8	90:10	0.2
Group 10	278	7	1	2	9	1	10	1.8	90:10	0.2
Group 11	168	0	4	5	5	4	9	1.6	56:44	0.2
Group 12	30	7	0	0	8	0	8	1.4	100:0	0.08
Group 13	165	3	0	5	7	1	8	1.4	83:17	0.32
Group 14	1891	0	0	5	4	1	5	0.9	80:20	0.59
Group 15	720	0	0	5	5	0	5	0.9	100:0	0.21
Group 16	501	2	2	0	3	1	4	0.7	75:25	0.71
Group 17	349	0	1	3	1	3	4	0.7	25:75:25	0.06
Totals	-	248	62	132	343	107	442	-	-	-
Whole Data Set	-	273	121	169	412	152	564	-	-	-

Table 3.3: Table showing the data of EAEC numbers according to complex, country and association with case or control. Groups are in order of complex size from the largest to smallest. Probability (Fishers exact test) of the group being significantly associated with case or control is tabulated at the end. There are two groups statistically associated with cases (ST10 Cplx and ST40 Cplx) and one group statistically associated with controls (ST31 Cplx).

Table 3.4 **Assessment of EAEC associated with other pathotypes**

EAEC Group	ST complex	EAEC PhD	EAEC Public	EPEC	ETEC	STEC	EIEC	DAEC	Commensal	No Pathotype	Other pathotypes	Total DEC	Total *E.coli* inc. EAEC	Other *E.coli* total	%EAEC: DEC	%EAEC: *E.coli*	Total EAEC	P value
Group 1	10	149	42	17	22	4	0	0	5	141	83	234	463	272	81.6	41.3	191	**<0.001**
Group 2	40	51	8	4	0	3	0	0	0	2	0	66	68	9	89.4	86.8	59	**<0.001**
Group 3	31	39	19	0	0	0	0	0	0	6	8	58	72	14	100.0	80.6	58	**<0.001**
Group 4	295	34	1	3	0	0	0	0	1	0	2	38	41	6	92.1	85.4	35	**<0.001**
Group 5	38	24	4	0	0	1	0	0	0	10	27	29	66	38	96.6	42.4	28	**<0.001**
Group 6	394	17	11	0	0	0	0	0	0	3	2	28	33	5	100.0	84.8	28	**<0.001**
Group 7	746	10	0	0	4	0	0	0	0	0	1	14	15	5	71.4	66.7	10	**<0.001**
Group 8	155	9	2	1	3	1	1	0	2	27	22	17	68	57	64.7	16.2	11	0.11
Group 9	678	10	0	0	0	0	0	0	0	1	0	10	11	1	100.0	90.9	10	**<0.001**
Group 10	278	9	0	0	0	0	1	0	0	0	0	10	10	1	90.0	90.0	9	**<0.001**
Group 11	168	5	4	0	0	0	1	2	0	10	8	12	30	21	75.0	30.0	9	**0.003**
Group 12	30	8	2	1	0	0	0	0	0	0	0	11	11	1	90.9	90.9	10	**<0.001**
Group 13	165	8	0	1	7	3	0	0	0	6	1	19	26	18	42.1	30.8	8	**0.005**
Group 14	1891	5	0	0	0	1	0	0	0	0	0	6	6	1	83.3	83.3	5	**<0.001**
Group 15	720	5	0	0	0	1	0	0	0	0	1	6	7	2	83.3	71.4	5	**<0.001**
Group 16	501	2	1	0	0	0	0	0	0	0	0	3	3	0	100.0	100.0	3	**<0.001**
Group 17	349	3	1	2	0	0	0	0	1	0	2	6	9	5	66.7	44.4	4	**0.01**

Table 3.4: Table showing EAEC groups in association with other *E. coli* pathotypes in the public database (all data from 18.12.2013). Nigerian dataset is included under the public database, UK and Bangladesh dataset is included under EAEC PhD. See section 3.2.4 for description of pathotypes included. Total EAEC included is 719 strains (564 from PhD plus 155 EAEC from public database strains), other *E. coli* total is 5955 strains (6674 minus 719 EAEC and minus 141 *Shigella* isolates included in the public database). Probability (Fishers exact test) of the group being significantly associated with EAEC or other pathotypes is tabulated

3.3.2 Impact of novel approach on public health

3.3.2.1 Detection of EAEC in other E. coli populations

From the 359 ESBL-producing *E. coli* isolates screened, eleven isolates contained the *aggR* gene, ten of which were isolated from extra-intestinal sources from human cases. There were no EAEC isolated from animals (Table 3.5, Figure 3.4). Six isolates were ST38 and were statistically associated with causing extra-intestinal infections (P=<0.001, Fisher exact test).

Table 3.5 **Characteristics of EAEC isolated from ESBL producing *E. coli***

Isolate	Pathotype	Serotype	ST	ST Complex	Country	Origin	Detailed source
ESBL-723	EAEC	OR:H30	38	38	UK	Human	Urine
ESBL-746	EAEC	O125ac:H30	38	38	UK	Human	Urine
ESBL-884	EAEC	O19a:H30	38	38	UK	Human	Urine
ESBL-831	EAEC	O19a:H30	38	38	UK	Human	Urine
ESBL-815	EAEC	O19a:H30	38	38	UK	Human	Blood
ESBL-26	EAEC	O153:H30	38	38	Netherlands	Human	Urine
ESBL-221	EAEC	O92:H33	34	10	Germany	Human	Faeces
ESBL-45	EAEC	O?:H26	58	155	Netherlands	Human	Urine
ESBL-46	EAEC	O?:HH-	694	None	Netherlands	Human	Urine
ESBL-48	EAEC	O15:H1	545	None	Netherlands	Human	Urine
ESBL-64	EAEC	O?:H23	224	None	Netherlands	Human	Urine

Table 3.5: Characteristics of EAEC isolated from the 359 ESBL-producing *E. coli* isolates screened for *aggR*. Out of eleven EAEC, six were from ST38 taken from urine and blood samples and were statistically with causing extra-intestinal infections.

Key: **Pathotypes:** EAEC – Enteroaggregative *E. coli* **Serotyping:** R – rough reaction, O? – O unidentifiable, H- not motile, **Genotyping:** ST – Sequence type, Complex – ST complex comprising of single locus variants (SLVs).

Figure 3.4 Venn diagram of diarrhoeagenic *E. coli* Source Attribution of 359 ESBL screened

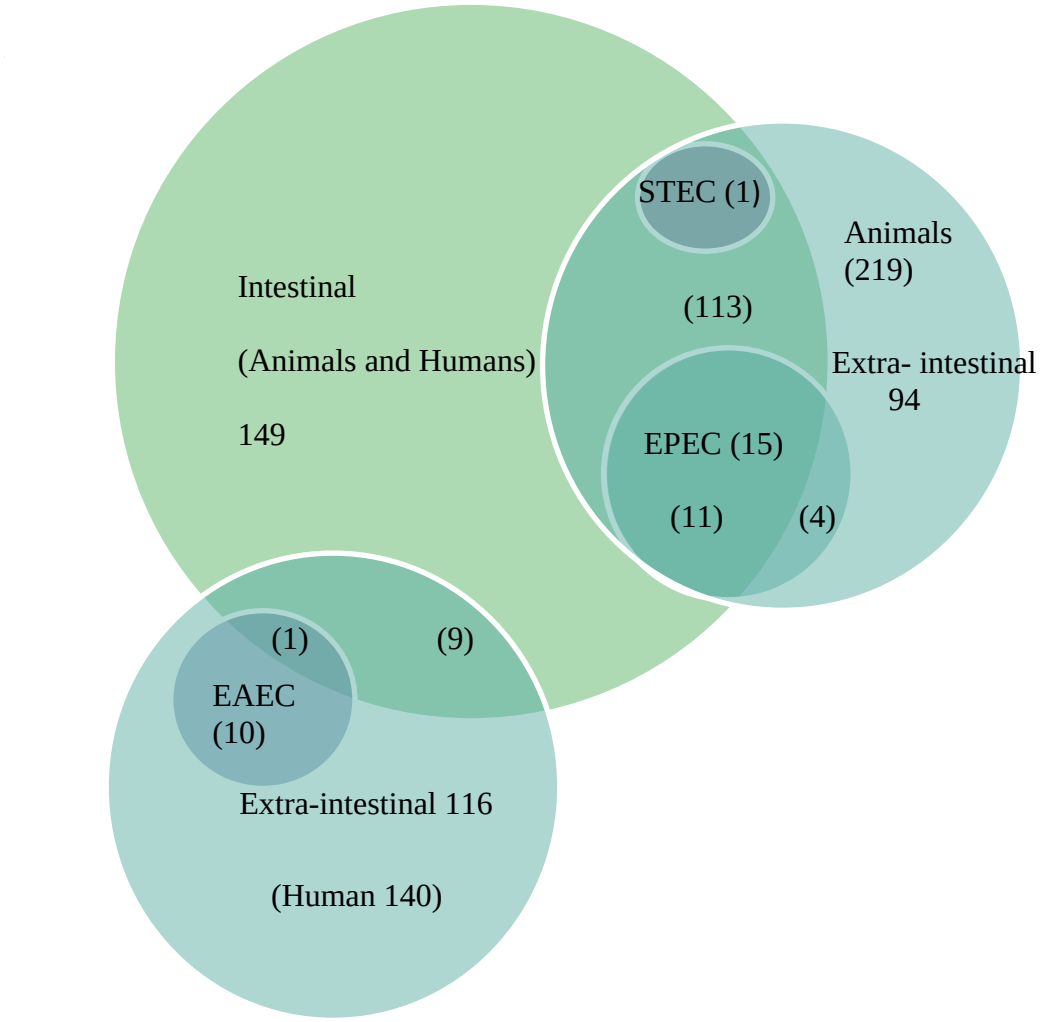


Figure 3.4: Venn diagram showing on overview of where EAEC was isolated in comparison to other diarrhoeagenic *E. coli* including EPEC and STEC. ESBL producing EAEC was predominately isolated from extra-intestinal sources rather than diarrhoeagenic sources indicated that some EAEC such as ST38 have pathoadapted to extra-intestinal niches.

3.3.2.2 MLST in the investigation of outbreaks

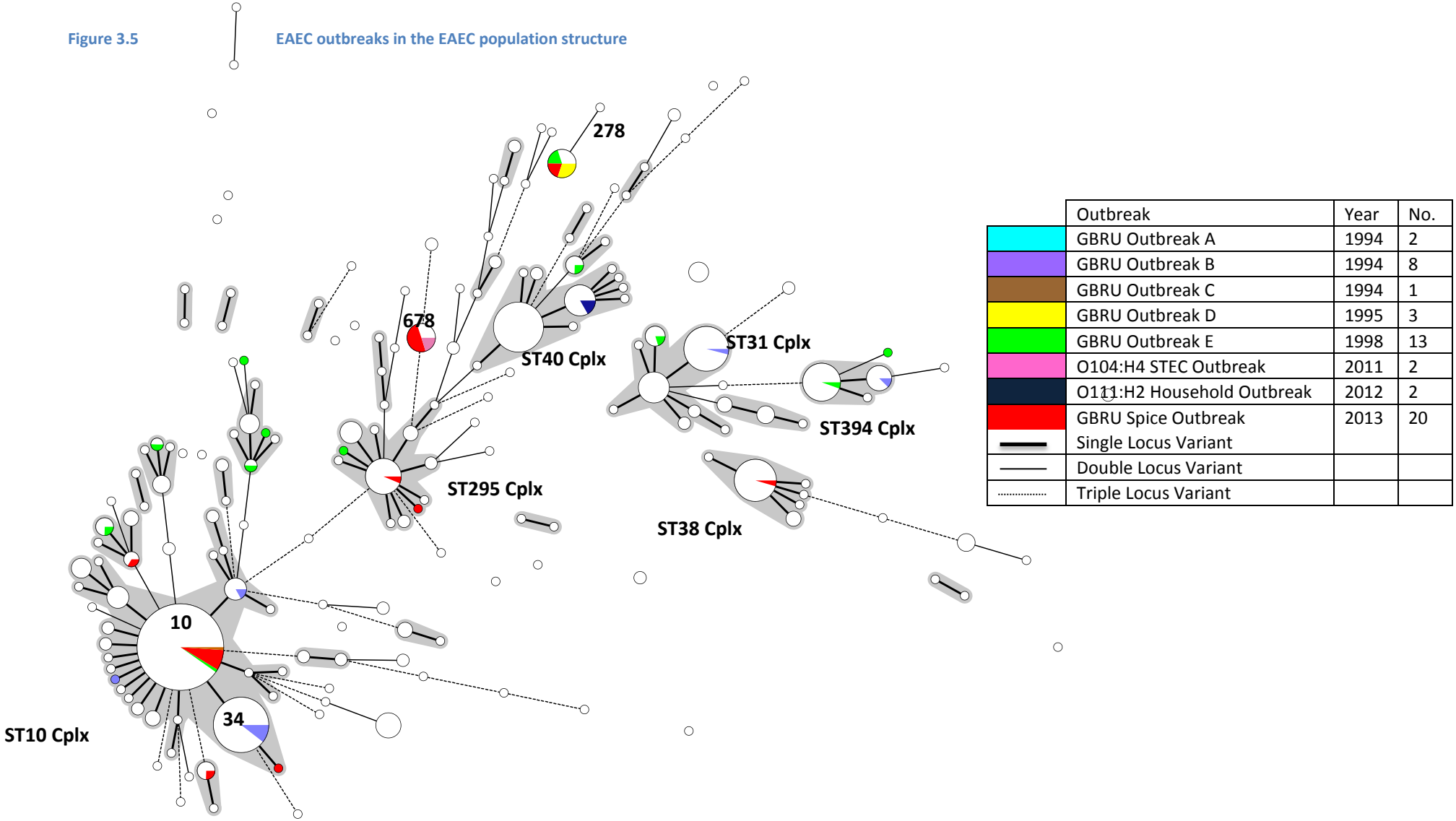
Overlaying outbreaks investigated at the Gastrointestinal Bacteria Reference Unit indicates that strains of EAEC within a single outbreak are rarely resolved into one background including serotype or ST and are distributed across multiple complexes. The exception being the O111 and O104 outbreaks associated with ST40 and ST678 Cplx respectively which was shown to be a point source outbreak (Chattaway *et al.* 2011;Dallman *et al.* 2012). ST278 was associated with three outbreaks from 1994-2013 (Figure 3.5, Table 3.6).

Table 3.6 Serotype and ST outbreak strains

Isolate	Source	Somatic	Flagella	ST	ST Complex	Year	Country
E98527	Outbreak A	19	H-	1114	ST165 Cplx	1994	UK
E98529	Outbreak A	O?	18	31	ST31 Cplx	1994	UK
E96386	Outbreak B	73	18	1380	ST394 Cplx	1994	UK
E96390	Outbreak B	O?	H-	SLV	ST10 Cplx	1994	UK
E96483	Outbreak B	O?	33	34	ST10 Cplx	1994	UK
E96485	Outbreak B	134	27	31	ST31 Cplx	1994	UK
E96487	Outbreak B	O?	H-	34	ST10 Cplx	1994	UK
E97590	Outbreak B	73	13	48	ST10 Cplx	1994	UK
E97820	Outbreak B	62	30	34	ST10 Cplx	1994	UK
E97900	Outbreak B	O?	H-	34	ST10 Cplx	1994	UK
E97470	Outbreak C	86	34	10	ST10 Cplx	1994	UK
E101396	Outbreak D	98	H-	278	ST278 Cplx	1995	UK
E101406	Outbreak D	98	H-	278	ST278 Cplx	1995	UK
E101621	Outbreak D	98	H-	278	ST278 Cplx	1995	UK
E89095	Outbreak E	80	27	SLV	ST155 Cplx	1998	Bangladesh
E89096	Outbreak E	113	H-	159	ST746 Cplx	1998	Bangladesh
E89099	Outbreak E	28ab	18	1657	None	1998	Bangladesh
E89102	Outbreak E	44	18	394	ST394 Cplx	1998	Bangladesh
E89104	Outbreak E	141	49	111	None	1998	Bangladesh
E89105	Outbreak E	80	27	58	ST155 Cplx	1998	Bangladesh
E89106	Outbreak E	R	7	1891	None	1998	Bangladesh
E89107	Outbreak E	O?	27	278	ST278 Cplx	1998	Bangladesh
E89111	Outbreak E	89	18	10	ST10 Cplx	1998	Bangladesh
E89112	Outbreak E	69	11	SLV	ST295 Cplx	1998	Bangladesh
E89114	Outbreak E	44	18	449	ST31 Cplx	1998	Bangladesh
E89115	Outbreak E	162	H-	278	ST278 Cplx	1998	Bangladesh
H125280572/573*	STEC Outbreak	O104	4	678	ST678 Cplx	2011	Germany
H120680226 H120720504*	Household Outbreak	O111	2	40	ST40 Cplx	2012	Ireland
H131920214	Spice Outbreak	O?	19	746	ST746 Cplx	2013	UK
H131920215	Spice Outbreak	20		278	ST278 Cplx	2013	UK
H131920216	Spice Outbreak	104	4	678	None	2013	UK
H131920217	Spice Outbreak	33	16	295	ST295 Cplx	2013	UK
H131920218	Spice Outbreak	104	4	678	None	2013	UK
H131920219	Spice Outbreak	131	27	10	ST10 Cplx	2013	UK
H131920220	Spice Outbreak	131	27	10	ST10 Cplx	2013	UK
H131920221	Spice Outbreak	20	19	278	ST278 Cplx	2013	UK
H131920222*	Spice Outbreak	19a	30	38	ST38 Cplx	2013	UK
H131941060	Spice Outbreak	111ac	4	226	ST226 Cplx	2013	UK
H131941061	Spice Outbreak	55	19	10	ST10 Cplx	2013	UK
H131941062	Spice Outbreak	104	4	678	None	2013	UK
H131941063	Spice Outbreak	104	4	678	None	2013	UK
H131941064	Spice Outbreak	O?	21	227	ST10 Cplx	2013	UK
H131941065	Spice Outbreak	63	12	1664	ST295 Cplx	2013	UK
H131941070	Spice Outbreak	104	4	678	None	2013	UK
H131941071	Spice Outbreak	131	27	10	ST10 Cplx	2013	UK
H131941072	Spice Outbreak	131	27	10	ST10 Cplx	2013	UK
H131941073	Spice Outbreak	131	27	10	ST10 Cplx	2013	UK
H131941074	Spice Outbreak	131	27	10	ST10 Cplx	2013	UK

Table 3.6: Serotype and ST results of strain in outbreaks.
* denotes additional outbreak strains not included in the original dataset

Figure 3.5 EAEC outbreaks in the EAEC population structure



	Outbreak	Year	No.
	GBRU Outbreak A	1994	2
	GBRU Outbreak B	1994	8
	GBRU Outbreak C	1994	1
	GBRU Outbreak D	1995	3
	GBRU Outbreak E	1998	13
	O104:H4 STEC Outbreak	2011	2
	O111:H2 Household Outbreak	2012	2
	GBRU Spice Outbreak	2013	20
	Single Locus Variant		
	Double Locus Variant		
	Triple Locus Variant		

Figure 3.5: Minimal spanning tree of the 564 EAEC used in this study colour coded by outbreaks investigated by GBRU. Complexes shaded in grey consist of single locus variants (SLV). Outbreaks O104 and O111 were shown to be from one ST whereas the other outbreaks are distributed among multiple complexes. ST38 was the only complex not associated with any outbreaks whereas ST278 was associated with three separate outbreaks.

3.4. Discussion

3.4.1 There are successful multiple lineages of EAEC groups that are globally distributed

This study has shown there are successful, statistically significant EAEC clusters associated with disease or carriage. These groups included the EAEC published groups in the public database and comprised ST10, 40, 38, 394 and 349 (Okeke *et al.* 2010). This study also identified MLST complexes that were not, represented as successful (i.e. more than four strains as defined in this thesis) in the public database including ST130, 295, 720, 484 and 678. This indicated that this data, represents a snapshot of the population structure of EAEC from three different countries. The addition of strains across the globe would expand the number of successful established EAEC groupings. It should be noted that the public database is biased towards *E. coli* of interest to the scientific community, such as pathogens and antibiotic resistant strains, with less representation of commensal strains and it seems likely that not all isolates were tested for the aggregative phenotype. A larger, better defined population may show "non-EAEC" (i.e. other *E. coli* pathotypes) present in more of the MLST complexes.

Population structure analysis of EAEC in this study using MLST showed the presence of independent multiple lineages of *E. coli* (i.e. groups that did not have common ancestor within the population). Although there are some STs restricted to one country, the majority of complexes contain isolates from at least two countries indicating global distribution of clusters (Figure 3.2). The explanation for the small exclusive geographical groups could be adaptation to a specific, local ecological niche, sampling bias or small sample size. The global distribution of the MLST complexes is most likely due to human travel and migration. The independent appearance of the EAEC phenotype in discrete complexes across the *E. coli* phylogeny would appear to represent homoplasy - or convergent evolution - suggesting that having the EAEC phenotype represents a biological advantage in certain bacterial genetic backgrounds.

3.4.2 The definition of 'case' or 'control' is ambiguous

There were 17 successful or prevailing EAEC Cplxs defined in this study (i.e. more than four strains and representing 2.5% of the known EAEC population) but there were only two Cplxs statistically associated with disease (ST10 Cplx and ST40 Cplx) and one group associated with carriage (ST31 Cplx), with the majority of groups having a mixture of 'cases' or 'controls'. EAEC belonging to the same ST complex may have been isolated form a case or a control and therefore the presence of the pathogen is not always associated with GI symptoms and the presence of EAEC in the stool may be associated with asymptomatic presentation. There are multiple possible reasons for this including general host susceptibility (often associated with age or physiological conditions), host acquired immunity, virulence gene content, co-infections, post infection carriage and, as shown in this chapter, genetic background. Another consideration is due to sample size, for example complexes containing less than 50 strains were not statistically significant. This does not mean that they are not clinically relevant as the sample size in these groups maybe too small for significance and this may change as the Cplxs increase in size.

The ability for EAEC to retain the plasmid may also play a role in establishing a successful Cplx and the population structure indicates that certain plasmids are stable in certain backgrounds while others (such as singletons) are not and therefore do not expand into a successful EAEC Cplx. The multiple plasmid compatibility types (Okeke *et al.* 2010) harboured by different strains of EAEC indicate that the plasmids of EAEC also have a complex ancestry, possibly suggesting independent but convergent co-evolution of these plasmids with the EAEC strains, that has resulted in a mosaic profile of similar gene sets on different plasmids (Dallman *et al.* 2014), this could explain why the plasmids and gene content are so heterogeneous. It is suggested that strains with different phylogenetic backgrounds (determined by ST) may have an affinity for these plasmids with specific compatibility types.

Any analysis based on the concept of an EAEC case or control being defined as isolating EAEC from symptomatic or asymptomatic subjects will produce a complicated picture as shown in the case/control population structure (Figure 3.3). With the complex nature of EAEC, it is likely, that some of the 'case' or 'control' definitions are misleading and that a strain from a 'control' subject may be a "disease" strain but was misidentified in the original study, due to host acquired immunity (Nataro *et al.* 1995). An EAEC strain from a 'case' may have characteristics of a "carriage" strain (in terms of the combination of virulence factors or the lineages it belongs to) but is classed as a "case" strain because the individuals has diarrhoea caused by a co-infection (Chattaway *et al.* 2013). During the IID study, over 25% of EAEC cases were co-infected with norovirus and over 5 % with *Shigella* (Figure 2.4).

The discovery of the multiple compatibility EAEC plasmids containing the genes responsible for the EAEC phenotype introduces another layer of heterogeneous complexity with this pathotypes. The plasmid can be acquired by any *E. coli* strain irrespective of whether it is pathogenic or commensal. Despite the issues associated with the definition of 'case' or 'control', the analysis in this study describes the population structure that can be used to explore EAEC causing clinical disease and to build an understanding on EAEC Cplxs that are public health threats.

3.4.3 Application of EAEC MLST population structure analysis has had an important impact on public health

3.4.3.1 EAEC have pathoadapted to extra-intestinal niches

Using the applications from this study, an ESBL producing *E. coli* population structure was created (Wu *et al.* 2013) and screening for *aggR*, highlighted that ST38 ESBL producing EAEC were associated with extra-intestinal infections (Table 3.5, Figure 3.4) (Chattaway *et al.* 2014a). ST38 is a successful EAEC diarrhoeagenic group as shown in this study and others (Okeke *et al.* 2010). According the

public database, the only other diarrhoeagenic pathotype in this ST is one STEC (O1:H15) (from a pigeon in Germany, the method of testing for pathotype assignment is unknown). The other pathotypes are urinary tract infections (UTI) and antibiotic resistance *E. coli* including the ESBLs producers (ESBL producers in the MLST database are usually isolated from the urinary tract or other extra-intestinal sites). It seems likely that EAEC strains within ST38 Cplx have adapted to cause disease in both the gut and the urethra/bladder. These strains in the ST38 complex were further investigated and contained multiple EAEC and ExPEC virulence factors: AAFI – EAEC fimbriae type, a*ggR* - EAEC regulatory gene, - *aap,* EAEC dispersin gene, *traT*, -serum resistance, *fimH,* - fimbriae in *E. coli* , *fyuA* and *irp2* –iron acquisition receptors (Chattaway *et al.* 2014a). Thus the ST38 strain described here is likely to have acquired the two phenotypes (UPEC and EAEC) independently suggesting the emergence of a UPEC/EAEC hybrid strain. It is possible that the genetic background of *E. coli* ST38 was stable enough to host the EAEC plasmid. The presence of the plasmid facilitated survival in the gut through increased adherence and has acquired UPEC virulence factors facilitating the exploitation of an extra-intestinal niche, the urinary tract.

Despite the historical characterisation of numerous virulence factors, there is no single genetic feature that currently defines EAEC or UPEC isolates (Wiles *et al.* 2008). As the EAEC ST38 strain had between four and seven ExPEC-associated virulence factors, it is suggested that by epidemiological, microbiological and molecular content, these EAEC ST38 strains should be considered an ExPEC associated with uropathogenic infections (Chattaway *et al.* 2014a;Wiles *et al.* 2008).

ExPEC EAEC is an important emerging group for several reasons, (i) the multiple drug resistance associated with this group can impact on patient care, (ii) the potential of this pathogen to cause multiple infections, such as gastrointestinal disease and UTIs, and (iii) the potential of this group to evolve into a hypervirulent strain causing outbreaks. EAEC has already been shown to acquire additional virulence genes evolve into hypervirulent strains, for example the EAEC/STEC hybrid

which has led to multiple outbreaks of HUS (Buchholz *et al.* 2011;Dallman *et al.* 2012). Although STEC strains have not been associated with extra-intestinal infections, there is the potential for EAEC ST38, which can cause disease in both the gastrointestinal tract and the urinary tract, to acquire *stx1* and/or *stx2* genes and thus cause severe disease.

3.4.3.2 There are successful complexes associated with EAEC outbreaks

Unlike Salmonella, in which the presence of this organisms in a faecal specimen is almost always associated with GI symptoms, EAEC is frequently isolated from both cases and controls. EAEC cannot be definitively defined as a pathogen. The increased understanding of the EAEC npopulation structure gained during this study, has facilitated population structure approach being used in outbreak investigations of EAEC. For example, seeing where adaptation and integration of EAEC with other *E. coli* pathotypes (such as STEC) and causing severe disease fits within the structure . This has been shown in recent outbreaks including a household outbreak of STEC/EAEC hybrid ST40 (O111:H21) (Dallman *et al.* 2012) and the infamous ST678 STEC/EAEC hybrid (O104:H4) outbreak in Germany (Chattaway *et al.* 2011).

In these outbreaks, an EAEC strain acquired the *stx* phage and increased the pathogenic potential and disease severity of the strain. The outbreaks were clonal and caused by one strain. In the cases of investigating other EAEC outbreaks, tracing the source and causative EAEC strains is more complicated. Several outbreaks of EAEC have been investigated in the UK over the last 20 years: Outbreaks A-E (Table 3.1) (small EAEC outbreaks that occurred between 1994-1998) and a recent EAEC outbreak from 2013 with an epidemiological link to contaminated curry leaves, known as the street spice outbreak (Dallman *et al.* 2014). Serological typing and MLST showed the strains were variable (Table 3.6).

Characterisation of EAEC by MLST in isolation is not a useful technique, for example knowing the ST of an EAEC without understanding the heterogeneous nature of EAEC could lead to false

assumptions. In the case of STEC O157 in which the genetic background is consistently ST11, the detection of this ST Cplx via MLST give a good indication that that strains was STEC O157. In the case of EAEC, there can be multiple complexes and they may relate to a different pathotypes and so a polyphasic approach of using MLST and EAEC virulence genes, such as the *aggR* gene would be required. Understanding the population structure of EAEC can help ascertain which strains may have caused disease. What the population structure shows, is that you get dominant ST complexes associated with outbreaks and other sporadic STs. Considering the complex nature of EAEC as described in this study, it is likely that the dominant STs may have particular traits associated with causing outbreak situations and are likely to be more pathogenic. Strains associated with sporadic cases and those that do not prevail in the population could be "carriage" or low pathogenicity EAEC strains or those that have transiently acquired the EAEC plasmid. Alternatively, they may have caused the diarrhoea due to factors such as low immunity, such as HIV cases (Mayer and Wanke 1995;Medina *et al.* 2010) but are not truly pathogenic (like in the volunteer studies (Nataro *et al.* 1995)).

Analysis of these outbreaks and symptomatic cases indicated there were three main complexes successful in outbreaks and likely to be important EAEC ST complexes (Figure 6.2): ST10 complex (Outbreak B, street spice outbreak), shown to be statistically associated with disease in under 5 years old (Okeke *et al.* 2010), and statistically associated with disease in this study (Table 3.3). ST678 (street spice outbreak, German outbreak) notoriously associated with the German outbreak (Chattaway *et al.* 2011) and a dominant ST in the street spice outbreak (Dallman *et al.* 2014). ST278 (outbreak D, E and street spice outbreak) has recently been associated with the street spice outbreak (Dallman *et al.* 2014) and is a dominant ST associated with three outbreaks. Isolating ST278 EAEC strains from the IID1 and Nigerian studies indicate that these strains continue to circulate in the background but also may have specific characteristics that increase its potential to cause outbreaks.

The EAEC street spice outbreak had a similar source (dried curry leaves) to the ST678 outbreak which was linked to fenagreek sprouting seeds in that they were both associated with contaminated seeds and spices (Dallman *et al.* 2014). Studies have shown that EAEC adhere to vegetables, such as salad leaves (Berger *et al.* 2009) and this might explain why they cause outbreaks associated with this type of foodstuff.

3.5 Summary

The novel approach described in this chapter of using a case control approach to assess the population structure of potentially pathogenic EAEC in multiple countries, has shown EAEC to have a complex population structure and diverse phylogeny. Statistical analysis of the complexes indicated certain complexes to be statistically associated with GI symptoms such as ST 10 and ST40 Cplx. Understanding the EAEC population from this study has had a positive impact on public health in terms of understanding EAEC Cplxs associated with outbreaks (Chattaway *et al.* 2011;Dallman *et al.* 2012;Dallman *et al.* 2014) and discovering a pathoadaptive EAEC ST38 group that causes extra-intestinal infections (Chattaway *et al.* 2014a). It is likely that further studies may reveal more regarding EAEC high pathogenic complexes. MLST as an analytical tool alone may not define pathogenic strains of EAEC and a more robust and polyphasic approach is recommended. A suggestion would be to combine the presence of putative EAEC virulence factors with ST complex to get best indicator of pathogenic potential as explored in Chapter 5.

The complex ancestry of EAEC indicates that there is not one ancestral complex that has expanded, but multiple ancestral complexes indicating possible convergent evolution of this successful phenotype (Chattaway *et al.* 2014b). However, the assessment of MLST alleles described in this chapter cannot indicate how these have occurred or whether this is predominantly due to mutation or recombination. To understand the evolutionary events that have driven the success of EAEC, a look into the exact sequence mutations of the different genes are required. Chapter 4 uses an evidence based Bayesian model, the genetic evolution of the dataset was assessed to look at the mutation types within the genes, understand how the EAEC groups have evolved and attempt to explain phylogenetic diversity observed during this study.

Chapter Four

Inference of Bacterial Microevolution

Chapter 4 - Inference of Bacterial Microevolution

4.1 Background

Chapter 3 laid a foundation for understanding the population structure of EAEC demonstrating the diversity of the group but also highlighting certain prevailing or successful lineages and suggesting that certain ST complexes were more likely to be associated with cases with GI symptoms than others. The evolution of these successful groups warrant further investigation.

Defining the sequence variation within the sequenced housekeeping genes is a well-established and simple method which requires input of sequence data into the *E. coli* MLST database (http://mlst.ucc.ie/mlst/dbs/Ecoli/). Interpreting that variation in terms of genetic evolution requires sophisticated mathematical models. ClonalFrame is a computer package for the inference of bacterial microevolution using multilocus sequence data and looks at the sequence data from each gene as opposed to allele analysis. It is a Bayesian statistics style model in which the inference of probability is updated as additional evidence is learned. The model constructs evolutionary histories by taking both mutation and recombination into account (Didelot & Falush 2007). The programme identifies clonal relationships between the members of a sample, and homologous recombination events that may have disrupted the clonal inheritance; it takes both point mutation and recombination into account. (Didelot & Falush 2007). It enables us to see close relationships between strains that may be obscured by recombination. This approach was used to understand how the genetic events in core genomic background, (represented by MLST) of the strains analysed, have influenced the evolution of the defined EAEC disease and carriage groups.

4.2. Methods

4.2.1 Clonal Frame Analysis

The same dataset described in Chapter 3 was used (Table 3.1). Clonal Frame analysis was carried out (http://www.xavierdidelot.xtreemhost.com/clonalframe.htm) to investigate the relationships of the different sequence type complexes. The Graphic User Interface in the ClonalFrame programme was used to construct 75% majority-rule consensus trees, mutational (theta) and recombination rates. Other analysis included the measure of the frequency at which recombination occurs relative to mutation (ρ/θ). The relative effect of recombination on the genetic diversification of populations, ratio r/m in which the ratio of rates at which nucleotides become substituted as a result of recombination and mutation (Vos & Didelot 2009) was also used. Finally, the external to internal branch length ratio was computed which gave the inferred expected values against the coalescent and actual ratios. Analysis was split into assessing the Bangladesh and Nigeria case control studies and UK clinical data set for comparison against the entire dataset.

4.2.2 Placing EAEC in the *E. coli* phylogeny

Multi-locus sequence analysis (MLSA) was performed by concatenating MLST sequence alleles of the EAEC from this dataset and all sequence types representative of the *E. coli* phylogeny. These were aligned and clustered (MEGA V 5.1) and the genetic relationship of isolates designated as assessed in the context of all *E. coli* using a neighbour joining tree phylogeny (MEGA V 5.1 and FigTree V 1.4). Phylogrouping PCR was carried out on the 17 main groups of EAEC (Doumith *et al.* 2012) and labelled on the phylogeny.

4.3. Results

4.3.1 Evolutionary Events leading to successful EAEC disease groups

ClonalFrame analysis showed that EAEC mutation and recombination rates varied between countries (Table 4.1) and complexes (Table 4.2). ST10 Cplx (statistically associated with cases, Table 3.3) had the highest mutation rate (4.05) and recombination rate (1.2) whereas ST295 Cplx (predominantly associated with cases, Table 3.3) the lowest mutation rate (0.02) and lowest recombination rate (0.002). However, both of these groups had a similar mutation to recombination ratio. Recombination had the greatest impact on the evolution of ST40 Cplx (statistically associated with cases, Table 3.3) (12) and ST394 Cplx (predominantly associated with cases, Table 3.3) (10). Recombination occurred 1.7 times more often than the mutation rate in the strains from Bangladesh and Nigeria whereas in EAEC strains from the UK, recombination and mutation rate was almost equal. In the entire dataset recombination events occurred 1.3 times more often than mutation. The geographical location of where an EAEC strain was isolated does not influence phylogeny (with the exception of small geographical specific STs possibly due to sampling bias) and successful EAEC STs were distributed globally (Figure 3.2). The impact of recombination in the diversification of the sample set relative to mutation showed the greatest impact in the Bangladesh strain set, and the least impact in the strains from the UK. These data suggest that recombination may play an important role in the evolution of EAEC (Table 4.1).

External to Internal Branch Length Ratio gave coalescent expectations indicating that all EAEC irrespective of location and including the entire dataset were significantly different (P= <0.001) from the inferred value (Figure 4.1). Coalescent expectation is the expectation of values or output ofa dataset inputted into ClonalFrame of mutation over time, the inferred values is the actual values produced from the analysis or output of the same dataset. The difference between the two is significantly different (P=<0.001) suggesting that a particular event or series of events (such as recombination events) have occurred leading to the extant population of a dataset inputted into

ClonalFrame rather than natural events (such as sporadic mutation) over time. This data suggests that the overall population structure of EAEC has not occurred naturally over time such as would be expected from sporadic mutation. This data indicates that a particular event or more likely a series of events such as acquiring chunks of foreign DNA (recombination) has been pivotal in the evolution of EAEC lineages.

Table 4 1 Mutation and Recombination rates of dataset by geographical source and all Sequence types found in dataset

Parameters	Bangladesh N=169 (108 Cases, 61 Controls)	Nigeria N= 121 (66 cases, 55 controls)	UK N= 254 (228 cases, 36 controls)	All ST N=199 (138 cases, 61 controls)
Mutation Rate (theta 0) Mutational rate & assumed to be constant on the branches of topology	mean:15.033101, credibility_region: 6.951621-26.141660	mean:120.794748, credibility_region: 69.290083-33.005260	mean:70.132709 credibility_region: 49.35070 - 94.014220	mean:16.019035 credibility_region: 8.644529-23.717401
Recombination rate (R) recombination rate & assumed constant on branches of topology	mean:22.5888284, credibility_region: 14.055470-33.461680	mean:31.381483, credibility_region: 19.687600-43.376330	mean:15.664218, credibility_region: 9.840895-22.310770	mean:89.532805 credibility_region: 64.216890-121.960200
view rho over theta (p/0) How often recombination occurs relative to mutations	mean:1.650017, credibility_region: 0.775438-3.148886	mean:1.689762, credibility_region: 0.785667-3.808293	mean:1.048907, credibility_region: 0.505756-1.982207	mean:1.317856, credibility_region: 0.767609-2.072695
view r over m (r/m) The impact of how important the effect of recombination was in the diversification of the sample relative to mutation	mean:4.384019, credibility_region: 2.382602-8.059029	mean:4.103293, credibility_region: 2.138517-8.091730	mean:2.605728, credibility_region: 1.444657-4.395547	mean:2.876673, credibility_region: 1.946228-4.241088
External to Internal Branch Length Ratio Gives the inferred expected values against the coalescent and actual rations. It they are significantly apart then it shows there was a genetic event such as recombination that led to these values.	mean:0.732716, interval:0.545600-0.942675 **Significance:0.00070**	mean:0.569694, interval:0.408030-0.760060 **Significance:0.01583**	mean:0.672798, interval:0.507496-0.885945 **Significance:0.00049**	mean:0.902324, interval:0.722079-1.068309 **Significance:0.00001**

Table 4.1: ClonalFrame mutation and recombination rates as well the impact of recombination over mutation in the diversification of the data and the significance of the expected over the inferred value as to whether the data evolved over a period of time (not significant) or due to a large genetic event (significant). This analysis was applied to the different geographical locations and all STs found in the 564 EAEC dataset.

Table 4.2 Mutation and Recombination rates of dataset by ST complex.

Parameters	ST10 Cplx & DLV	ST38 Cplx & DLV	ST40 Cplx & DLV	ST295Cplx & DLV	ST394Cplx & DLV	ST31 & ST 130Cplx & DLV
Mutation Rate (theta 0)	mean:4.047662 credibility_region: 2.096737-6.310707	mean:0.281472 credibility_region: 0.023902-1003483	mean:0.948535 credibility_region: 0.029708-2.621763	mean:0.0277171 credibility_region: 0.000899-1.872790	mean:0.231795 credibility_region: 0.008852-1.006419	mean:0.657076 credibility_region:0.134456-1.484436
Recombination rate (R)	mean:1.248849 credibility_region: 0.419549-2.841862	mean:0.083915 credibility_region: 0.000868-0.388625	mean:0.612102 credibility_region: 0.001200-1.909099	mean:0.002833 credibility_region: 0.000732-0.011280	mean:0.107088 credibility_region: 0.000807-0.464062	mean:0.378763 credibility_region: 0.035706-0.977376
view rho over theta (p/0)	mean:0.334218, credibility_region: 0.092813-0.825982	mean:0.683871, credibility_region: 0.003663-3.603088	mean:5.557545, credibility_region: 0.000534-46.863189	mean:0.572612, credibility_region: 0.000759-4.492473	mean:4.075097, credibility_region: 0.001951-33.129625	mean:1.074699, credibility_region: 0.044360-5.632848
view r over m (r/m)	mean:1.200017, credibility_region: 0.394262-2.665440	mean:3.558562, credibility_region: 0.019087-19.63949	mean:12.004059, credibility_region: 0.002926-102.352654	mean:0.914214, credibility_region: 0.002019-7.049239	mean:10.392526, credibility_region: 0.007888-74.564922	mean:4.274411, credibility_region: 0.245787-20.061001
External to Internal Branch Length Ratio.	mean:0.481798, interval:0.283358-0.728356 **Significance:0.02534**	mean:0.778292, interval:0.302418-1.510262 **Significance:0.15335**	mean:0.641577**,** interval:0.300055-1.20614 **Significance:0.09755**	mean:0.646918, interval:0.299785-1.250929 **Significance:0.15941**	mean:0.644643 interval:0.237609-1.323429 **Significance:0.24016**	mean:0.567944 interval:0.276826-1.143014 **Significance:0.12012**

Table 4.2: ClonalFrame mutation and recombination rates as well the impact of recombination over mutation in the diversification of the data and the significance of the expected over the inferred value as to whether the data evolved over a period of time (not significant) or due to a large genetic event (significant). This analysis was applied to the different complexes including single locus variants (SLV) and double locus variants (DLV)

Figure 4.1 External to internal branch length ratio of enteroaggregative *E. coli* dataset

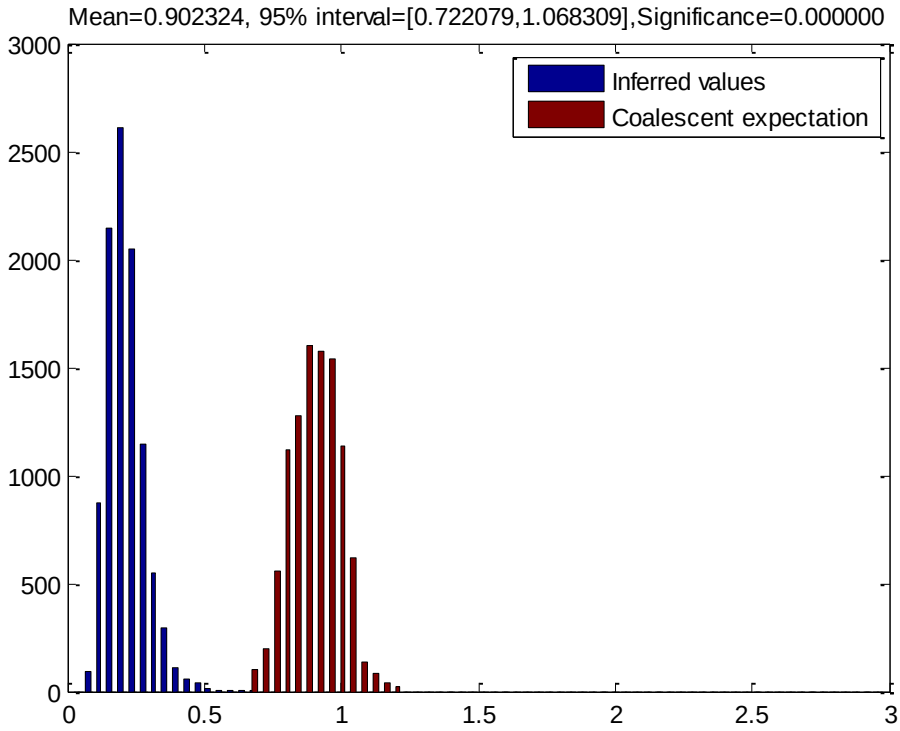


Figure 4.1: ClonalFrame external to internal branch length ratio of EAEC STs found in this study. The difference between the two is significantly different (P=<0.001) indicating that this extant population is not due to natural mutation alone.

4.3.2 Evolution of EAEC in the context of the *E. coli* population

EAEC in the background of the four main branches of *E. coli* phylogeny show a dominant evolution in branch 1 and 2 (Figure 4.2) consisting of phylogroups D, A and B1 respectively. ST30, 38, 394 and 31 Cplx which are grouped together by MLST population structure (Figure 3.3) and cluster tightly on branch 1 (phylogroup D) of the *E. coli* phylogeny. Possibly indicating they evolved during a similar timeframe and share a similar ancestral source. The other large successful groups are dispersed throughout branch 2 and more diverse (larger distance on the tree), so a similar ancestral source and timeframe are more unlikely.

ST10 Cplx shows that some SLVs on the MLST structure are dispersed in the context of the *E. coli* phylogeny (Figure 4.2). This highlights the limitation of using MLST as the EAEC lineages in the context of all *E. coli* are not as closely related as they seem. Using ClonalFrame, some strains that appear to be closely related (ie different by a SLV) are shown to have a larger genetic difference and that recombination is the likely cause.

ST295 Cplx which is linked to ST10 Cplx by ST48 with a triple locus variant via MLST population structure analysis (Figure 3.3) is on the opposite end of branch 2, and therefore evolutionary distant. MLST would show these groups to differ by 3 loci and would be assumed to be different. This is confirmed by ClonalFrame supporting the hypothesis that strains associated with ST 95 and ST10 are evolutionary distinct.

The smaller successful groups with only 4 EAEC strains were found at the end of branch 4 which contained a mixture of phylogroups A and D, indicating that these backgrounds have not enabled EAEC to expand into prevailing lineages. None of the main EAEC complexes (Table 3.3) were found in branch 5 of the *E coli* phylogeny which is generally associated with extra-intestinal infections such as ST131 belonging to phylogroup B2.

Figure 4.2 Neighbour joining tree of all *E. coli* sequence types and EAEC in this study

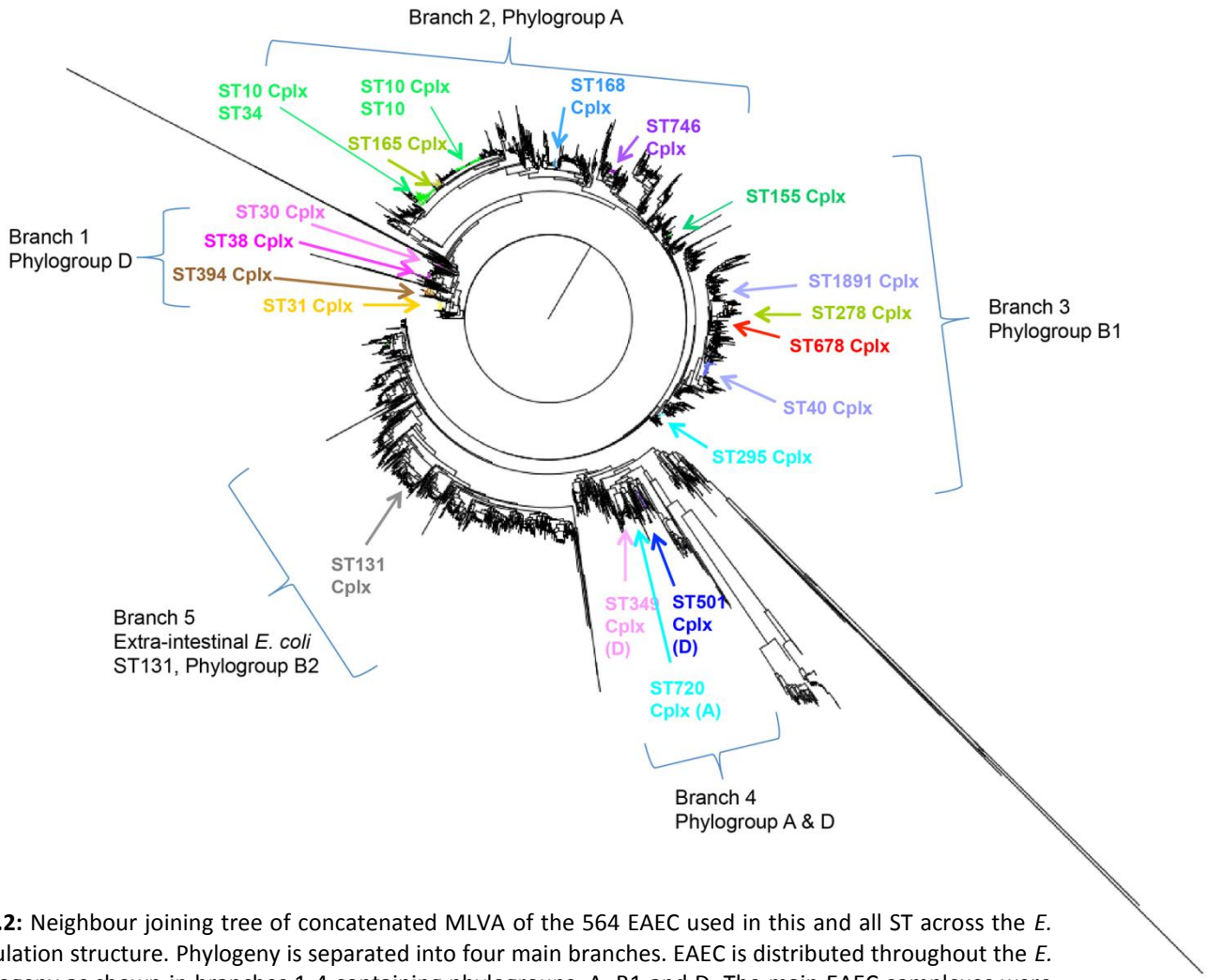


Figure 4.2: Neighbour joining tree of concatenated MLVA of the 564 EAEC used in this and all ST across the *E. coli* population structure. Phylogeny is separated into four main branches. EAEC is distributed throughout the *E. coli* phylogeny as shown in branches 1-4 containing phylogroups, A, B1 and D. The main EAEC complexes were not found in branch 5, phylogroup B2 associated with extra-intestinal infections.

4.4 Discussion

4.4.1 Multiple genetic events have led to the independent evolution of EAEC

In order to understand the genetic events which led to the formation of different EAEC associated MLST complexes, Clonal Frame analysis of the branching events for each node was carried out. Variation in the frequency of recombination or mutation which occurred in all of the seven loci at different time points was seen (data not shown) indicating continuous multiple genetic events. The relative frequency of recombination as compared to mutation (ρ/θ) for the entire data set was 1.31 and is comparable to the rates for *E. coli* proposed by Wirth *et al* (Wirth *et al.* 2006) and Touchon *et al* (Touchon *et al.* 2009) but higher than computed rates for the *E. coli* species via MLST who estimated recombination at essentially zero (Perez-Losada *et al.* 2006).

The parameters of rates and impact are based on the Markov model (Didelot & Falush 2006) which assumes such that horizontal gene transfer events are equally probable between any pair of lineages, irrespective of phylogenetic and ecological proximity (Galtier 2007). The analysis clearly shows that this is not the case and that recombination rates vary within the EAEC pathotype between different lineages, the most ancestral being ST10 Cplx with the least impact of recombination in comparison to the other lineages (Table 4.2).

Multiple successful complexes vary in mutation and recombination rates (Table 4.2) and are distributed throughout the *E. coli* population (Figure 4.2). These complexes have clearly evolved independently through multiple genetic events that have led to the phenotypic congruency of this pathotype. These EAEC strains in the population structure show different, apparent, mutation/ recombination rates and suggests that certain bacterial backgrounds allow the biological advantage (such as strong adherence to the gut with the AA plasmid) to be expressed - possibly influenced by the ability to retain the EAEC plasmid. Fast radiation (expansion of successful strains) of the complexes after population bottlenecks (a sharp reduction in the size of a population due to

environmental events and assumed to play a key role in the maintenance of social traits inmicrobes) and frequent recombination seems a likely explanation for this population structure - and this has been observed in other *E. coli* populations (Wirth *et al.* 2006). Gastrointestinal EAEC complexes were not found in the extra-intestinal *E. coli* phylogeny branch as maintaining the pAA, may not be a biological advantage to strains of ExPEC as their pathogenicity is not based on the need to adhere to the gut.

4.2.2 Factors responsible for EAEC evolution

Due to the heterogeneous nature of EAEC, it is likely that several factors are responsible for these convergent groups. Distinct populations within a species may emerge because of differential local adaptation or genetic drift (Vos & Didelot 2009). This concept may be applied to successful EAEC groups which represent clusters of closely related genotypes and can be termed ecotypes (Cohan 2002). These groups differ in their homologous recombination events because of adaptive evolution or environmental constraints (Vos & Didelot 2009). This is shown by the variable recombination rate in different complexes which may have evolved from different environments. The variable recombination rate from each country will depend on the complexes found from the sample size tested. Although recombination had the lowest impact on the UK EAEC isolates, it is recognised that a portion of these had recently travelled outside the UK and would therefore include EAEC from many different countries. In summary, there are several factors such as competition with other microbes and adaptation to environmental niches responsible for the evolution of these groups which has occurred from different ancestral sources, over different time periods and from different environments.

Virulent pathotypes have been shown to recombine more than non-pathogens pointing towards the theory that that virulence is the driving force for more frequent recombination (Wirth *et al.* 2006) or

that frequent recombination increases pathogenic potential. This is demonstrated with the ST40 Cplx which was statistically significantly associated with disease (P=0.03) and recombination which had the highest impact of on its diversification. ST10 Cplx, which was also statistically significantly associated with disease (P=0.01) in this study and in a previous EAEC study (Okeke *et al.* 2010) had the highest rate of mutation among the groups and the impact of recombination was almost equal to mutation (1.2). This indicates that both types of genetic events are important in the pathogenic evolution of EAEC.

The EAEC data set showed that the external to internal branch length ratio is significantly higher than expected (Table 4.1, Figure 4.1). This means that the inferred genealogy is consistent with either an expansion of the population size or acquisition of a fitness advantage early in the history of the sample (Didelot & Falush 2006). Therefore, although mutation and recombination are important in the evolution of the most pathogenic EAEC groups, other factors have contributed to the evolution of the successful established groups.

One fitness advantage could be the acquisition of one (or more) of the plasmids associated with the EAEC phenotype harbouring genes that have increase its binding ability giving it the characteristic 'stacked bricked' HEp-2 assay phenotype and more importantly biologically and pathologically enabling the EAEC pathotypes to adhere to the host gut mucosa. The ability to adhere to and colonise the gut for an extended period of time may confer a selective advantage on this pathotype. Independent acquisition of an EAEC plasmid could account for the different rates of mutation and recombination between the groups. Whereas the ancestral ST10 Cplx already had the background mutations to be able to acquire and retain the EAEC plasmid, other ST Cplxs such as ST40 might not have and therefore needed large recombination events to stably retain the EAEC plasmid.

One could argue that this adherence plasmid would not give the organism a particular advantage of success as *E. coli* already have the ability to adhere in the gut via multiple mechanisms (Kaper *et al.* 2004), though the efficient binding of EAEC is reflected in persistent diarrhoea. A recent example, however, of how acquisition of this EAEC plasmid can increase pathogenicity is the recent ST678 (0104:H4) STEC German outbreak (Chattaway *et al.* 2011). This is a VTEC strain that did not have the *eae* gene (attachment and effacement loci for intimate adherence and typically associated with severe disease) but did have the *aat* plasmid and its factors associated with adherence. This strain was particularly virulent with high HUS rates despite harbouring the same toxin type as many other STEC strains, the difference perhaps being its increased ability to adhere to the host gut mucosa and hence expose the host to the toxin for longer. This mechanism of attachment could provide a fitness advantage that will enable this relatively new pathotype to become established.

This chapter clearly shows the complexity of the evolution of EAEC, while it is evident that the same lineages prevail in multiple global locations indicative of either clonal expansion or convergence, there are also lineages ecologically adapted that account for the inconsistent impact rates of recombination between different geographical locations.

4.5. Summary

The work described in this chapter further analyses the MLST data by specifically looking at the sequences of the house-keeping genes in the MLST scheme and gives a description of the phylogeny of the largest multi-sourced EAEC dataset to date and a unique insight into the evolution of specific EAEC designated groups. These groups have evolved independently of each other multiple times in multiple locations via a combination of mutation and recombination and converged to the EAEC phenotype. With the selection of important EAEC complexes and the understanding how these have

evolved, investigation to the assessment of pathogenicity of these groups using phenotypic and genotypic models was assessed as discussed in Chapter 5.

Chapter Five

Phenotypic and Genotypic Models

Chapter 5 - Linking Phenotype and Genotype Models

5.1 Background

The work described in Chapter 4 suggested there are lineages of EAEC associated with either asymptomatic carriage or GI disease. In this chapter the virulence gene profiles and the phenotypic properties of these lineages are investigated phenotypically to look for differences between the two groups. Specifically the phenotype of strains within a complex was examined by selecting representative isolates and investigating the following: the adherence capability by HEp-2 cell assay; their pathogenicity using the *Caenorhabditis elegans* worm model; their metabolic profile using the Biolog; and their serotype. In addition, genotypic virulence profiling was assessed for association with EAEC complexes.

5.2 Methods

5.2.1 Strain Details

The HEp-2 assay and *Caenorhabditis elegans* worm model was carried out in the laboratory of Prof. Iruka Okeke, Haverford, USA. A representation of prominent and successful ST lineages from strains available in Prof. Okeke's collection was selected. These represented a selection of different ST complexes defined in this thesis as being associated with disease or carriage. These included EAEC disease associated group (ST394 Cplx), EAEC carriage associated group (ST31 Cplx), one group exclusive to EAEC (ST 168 Cplx) and not associated with either disease or carriage and one group associated with EAEC and specifically uropathogens (ST38 Cplx).

For the Biolog metabolic profiling, carried out in the GBRU, a random selection of 97 strains from the UK and Bangladesh from cases and controls and including the successful complexes ST10, ST295, ST38, ST168 and ST40 Cplxs (all associated with cases, Table 3.3) were used. All UK and Bangladesh strains (443 strains) were serotyped. Bangladesh strains were genotyped.

5.2.2 Human epithelial (HEp-2) cell assay

The method originally described by Cravioto *et al* (Cravioto 1979) was used with modifications necessary for delineating aggregative adherence stipulated by Vial *et al* (Vial *et al.* 1990).

5.2.2.1 Preparation of HEp-2 cells

Gloves were always used and changed regularly throughout this process especially whenever the cabinet was entered in order to reduce the high risk of contaminating the tissue cell culture. The tissue culture water bath was heated to just below 37^{o}C (over 40^{o}C and the media will become inactive) and was monitored throughout. All sterile manipulation work was carried out inside the Purifier Class II Biosafety Cabinet [Labconco]. A clean and dirty incubator was used depending on whether the flasks were sterile or inoculated with viable material. The area was wiped down with 70% ethanol regularly. Sterile incubating media was prepared with 50ml of Fetal Bovine Serum [Gibco] (to support cell growth), 5ml of 100 µg/ml penicillin/streptomycin [Gibco] (to prevent contamination) and 500ml of Dulbecco modified Eagle medium (DMEM) media [Gibco]. A small tissue culture flask was filled with 8ml of incubating media and inoculated with HEp-2 cells and incubated at 37^{o}C for 48 hours in CO_2 (5%) incubator [VWR].

The flasks were ready to use when the HEp-2 cells formed a 50% confluent growth monolayer. Flasks were divided to keep the cells fresh to maintain optimum conditions for testing. After 48 hours incubation, an inverted microscope [Zeiss Invertoskop] was used to check if there was a minimum of 50% confluent growth and whether the cells appeared healthy. Cells were not allowed to grow over 90% confluent as their physiology would start to change which could affect any testing. If 50% confluent growth had not been achieved, the incubating media was replaced with fresh media and the flask was incubated and checked every 24 hours. When the HEp-2 monolayers were ready for dividing they were washed three times with Dulbecco's phosphate-buffered saline [Gibco], this washed away the antibiotics and the serum that would inhibit the trypsin. Trypsin [Gibco] (0.8mls) was added to the monolayer and left for 30-180 seconds. The flask was gently banged to dislodge

the cells and the trypsin started to appear cloudy. The cells were continually checked under the inverted microscope, when the cells were trypsinized (cell disassociation, round and free floating), 8 ml of incubating media was added. This inhibited the trypsin to stop the process. If the cells were over trypsinized, the physiology of the cells would change or they would die affecting the test. Loose cell media was diluted by adding 4 ml of trypsinized cells and 4 ml of fresh incubating media to an 8 ml tissue culture flask. An 8 chamber well glass slide [Thermo Scientific] was seeded with 0.4 ml of diluted cell loose media and was incubated overnight at 37^{o}C with 5% CO_2.

5.2.2.2 Preparation of strains

Pure strains were grown statically in 2ml Luria broth overnight at 37^{o}C aerobically in a non-shaking incubator. It was important to not shake the broths as this would affect the adherence capability. A positive control (pathogenic EAEC strain 042) and a negative control (laboratory *E. coli* control strain pir116) were used in each experiment and all strains were tested in duplicate.

5.2.2.3 Preparation of adherence media

Sterile adherence media was made by adding 50mls of 10% mannose (0.5g mannose: 50ml milli q water) to 500ml of DMEM which made a final concentration of 1% mannose. Mannose is added as *E. coli* have type 1 fimbriae that will bind to mannose receptors on the HEp-2 cells, the additional of mannose will prevent type 1 fimbriae binding to the cells so the adherence you see will relate specific *E. coli* pathotype type binding (Figure 5.1) and not non-specific mannose binding.

Figure 5.1 Different pathotype binding phenotypes of *E. coli*

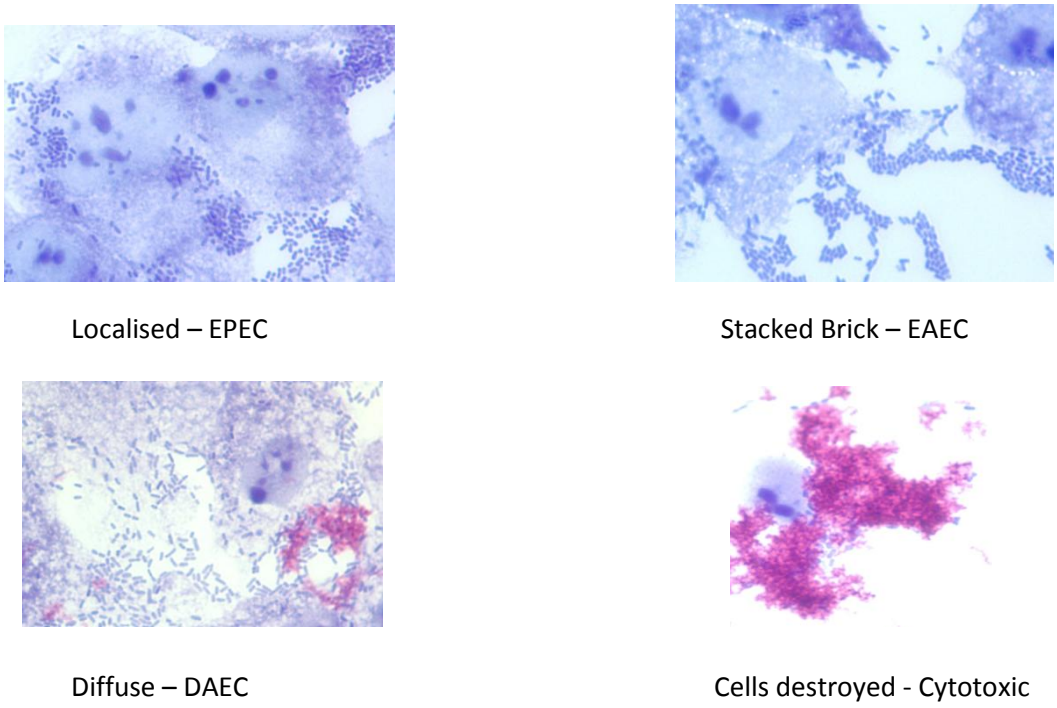


Figure 5.1: Photos of the difference adherence patterns of *E. coli* on HEp-2 cells

5.2.2.4 Adherence assay

Chamber wells were washed three times with PBS and 0.4ml of adherence media (DMEM with mannose) was added to each well in the safety cabinet. On the laboratory bench, 10µl of bacterial suspension was added to each well and incubated for 3 hours at 37^{o}C with 5% CO_2 in a non-sterile incubator. The monolayers were washed three times with PBS and fixed for 20 min with 70% methanol. Each well was stained with 0.4ml Giemsa Stain [Gibco] (1ml concentrate: 5ml PBS), for 20 min. The chamber and rubber seal was removed and the cells washed very gently with water and dried at room temperature.

5.2.2.5 Analysis of HEp-2 Assay

Cells were examined at X1000 magnification under oil immersion. They were checked to ensure they were consistent across the slide, a minimum of five fields were checked and two representative pictures were taken. Semi-quantitative methods were used to assess the Intensity of adherence of each strain and was scored as shown and described in table 5.1

Table 5.1 Hep-2 Scoring Assignment

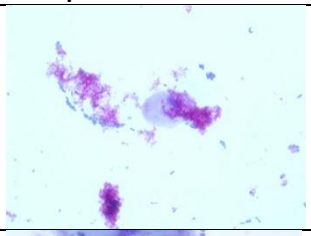
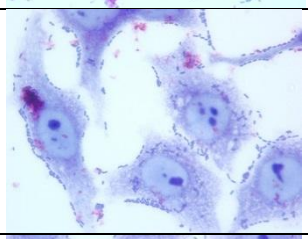
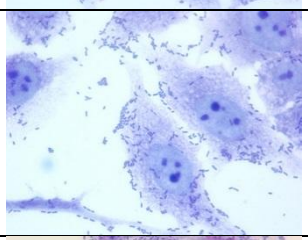
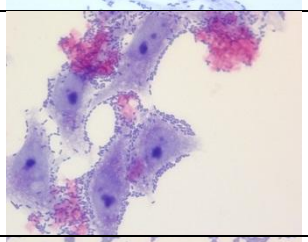
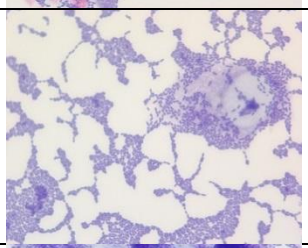
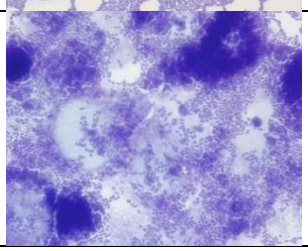
Score	Definition	Example
0 = Cytotoxic	Destruction of HEp-2 cells, auto agglutination of EAEC is usually still present showing stacked bricked adherence.	
1 = Very Light	Very little binding, ~ 20% of cell sides covered by EAEC.	
2 = Light	Light binding of EAEC around the cells, <50% of cell sides covered by EAEC	
3 = Medium	Medium binding of EAEC, ~70% of the side of cells covered by EAEC, ~10 % cell surface binding	
4 = Heavy	Heavy binding of EAEC, ~90% HEp-2 cell sides covered by EAEC, >20% cell surface binding	
5 = Very Heavy	Excessive binding, 100% cells covered by EAEC on sides and >80% cell surface binding, evidence of dense clumping of bacteria	

Table 5.1: Parameters of score given for the intensity of adherence, evidenced by photos.

5.2.3 *Caenorhabditis elegans* worm model

5.2.3.1 Preparation of worms and media

A worm picker was prepared using a glass pipette and thin wire. The wire was inserted into the glass pipette and sealed with a Bunsen burner. Aseptic techniques were used throughout and the wire was flamed between worm picking. Worms were maintained on nematode growth medium (NGM) (Brenner 1974;Lewis and Fleming 1995) on lawns of *E. coli* (strain OP50) at 15^{o}C. All plates were incubated at 37^{o}C for 24 hours and placed at room temperature for 8-12 hours. Test plates were prepared by spreading 5µl of overnight bacterial broth culture on 3.5cm diameter plates containing NGM and incubated for four hours at 37^{o}C. This was repeated for each selected enteroaggregative *E. coli* strain. A positive control of the virulent prototypical 042 strain and a negative control of the laboratory attenuated pir116 strain were used. The standard N2 Bristol worm strain was used for assays and the slow-kill assay protocol was adapted from Tan *et al* (Tan *et al.* 1999). Worms were found in four stages of growth, L1 being the smallest and L4 being the largest (Figure 5.2). A dead carcass (containing multiple L1 larvae) was transferred onto a fresh NGM medium agar plate and grown overnight in aerobically at 37^{o}C. Stage L2 larvae were selected for the experiments the next day.

5.2.3.2 Slow-kill assay

For the slow kill assay, ten N2 hermaphrodites at the L2 stage were seeded into each test plate and incubated at 22^{o}C for 24 hours. Survival, motility rate and reproducibility were assayed every 24 hours. Worms were transferred with the worm picker to new inoculated test plates every 24 hours to allow the distinction of subsequent generations. Worms were considered dead when they no longer responded to touch (Figure 5.2).

Figure 5.2 Flow diagram of slow-kill assay procedure

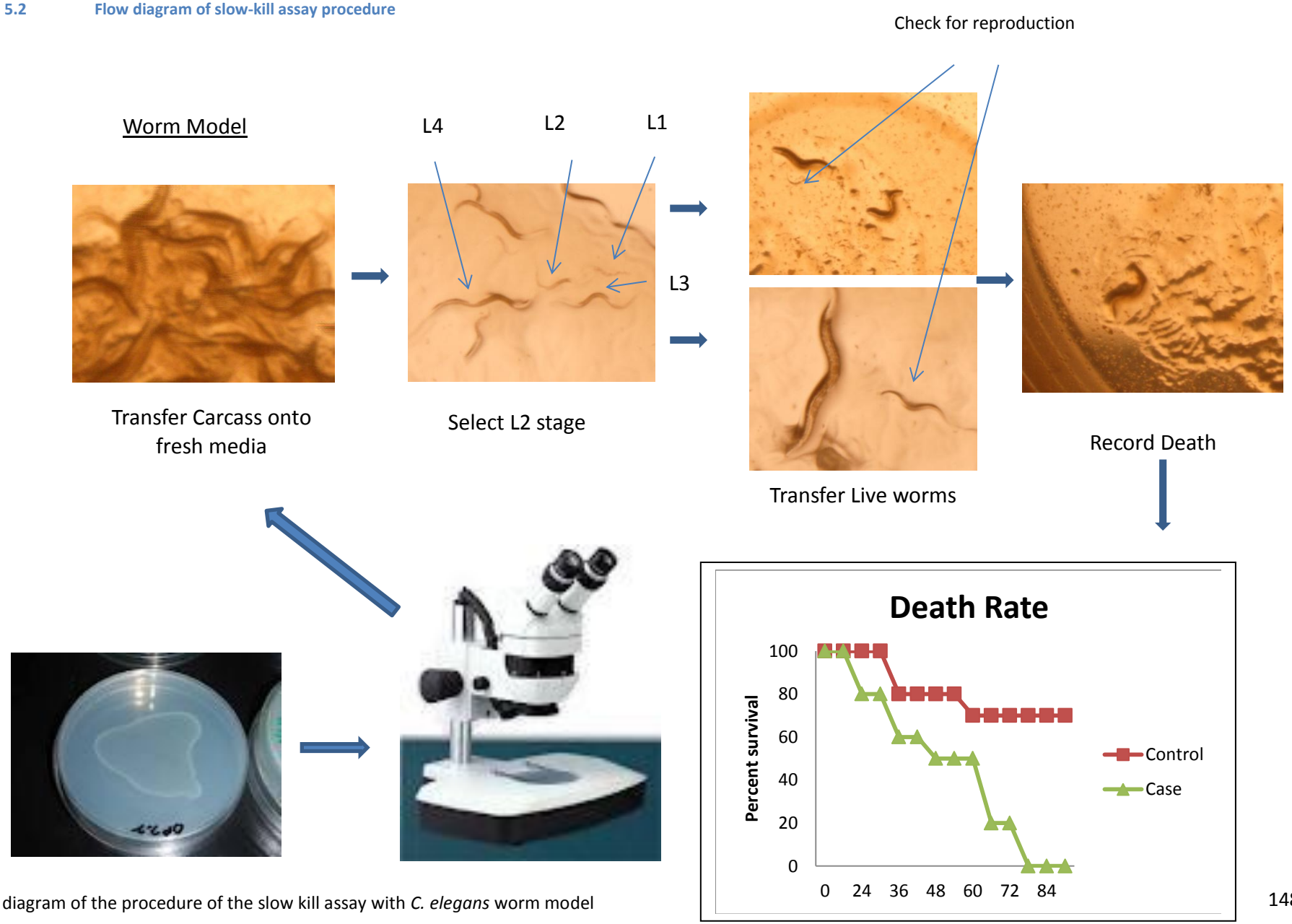


Figure 5.2: Flow diagram of the procedure of the slow kill assay with *C. elegans* worm model

5.2.3.3 Analysis of Caenorhabditis elegans worm model

Data from the assays were analysed using the Kaplan-Meier method (Bland & Altman 1998) and STATA was performed on the results. Log Rank test (Bland & Altman 2004) was carried out for equality of survivor functions which was a comparison of all complexes against each other. A cox regression (Altman & Andersen 1989) analysis was made against the complex strains against the control strains.

5.2.4 Metabolite profiling using the Biolog

Strains were streaked onto Columbia blood agar plates (5% sheep blood) [PHE Media] and incubated overnight at 37^{o}C. A single colony was picked and inoculated into type A broth [Technopath] using a flat end cotton swab. Biolog GEN III well plates which contained a negative control and 95 metabolites (Appendix 7.2) were inoculated immediately with 100µl of type A broth [Technopath] and placed into the Biolog with an incubation temperature of 37^{o}C. Reading of the 96 well Biolog GEN III plates for metabolite utilisation was carried out every 15 min for 22 hours. This is a routine method used within the GBRU

5.2.4.1 Analysis of Biolog Data

A neighbour-joining tree of the Biolog outputs was constructed using Phylogeny Inference Package (PHYLIP) and viewed in FigTree version 1.4. The first two consisted of EAEC strains only and were coloured by case/control or by complex.

As a comparison with the metabolic activity of the clinical *E. coli* population a third neighbour joining tree was then constructed with the 97 EAEC strains overlaid with 1479 clinical *E. coli* strains processed in GBRU from 2010 – 2014.

5.2.5 Serotyping of isolates from Bangladesh and UK

Serotyping of the somatic and flagella antigen (Gross and Rowe 1985) was carried out on the heat stable lipopolysaccharide (Somatic or O) antigens and the flagellar (H) antigens on the Bangladesh and UK strains. Strains that did not express LPS reacted with all antigens and were termed rough and those that did not react with any were designated 'O unidentified' or 'H unidentified'. Nigerian strains had previously been identified and published (Okeke *et al.* 2000b), strains were not accessible and serological identification was not performed.

5.2.6 Virulence typing of isolates from Bangladesh

DNA extraction and PCR was carried on 153 EAEC isolates from Bangladesh (93 cases and 60 controls) out as described in Chapter 3 (3.2.3.1 – 3.2.3.2) with annealing temperatures varying according to the target detected (Table 5.2). Previously published virulence factor genes included *aat*, *aaiC*, *astA*, *aggR*, *aafA*, *aggA*, *agg3A*, agg4A, *pic*, *set1A, irp2, tia and pet.* Virulence factor *aaf5A* primers and probes were provided by Julia Mtwale (University College London). Products were visualised on a 1.5% agarose gel and 10X tris acetate EDTA buffer standing with ethidium bromide [BIORAD Gel Doc 2000]. Multiple controls were used that were positive for one or more virulence factors (Table 5.3).

Pearson Chi-Square test was carried out to ascertain if there was an association with any genes being present in EAEC isolates from cases over controls. Pearson Chi-Square test was repeated to see if the presence of a virulence gene was associated with a complex. For this test a sample size of 5 or more was needed, therefore complex included ST10 (n=18), 155 (8), 165 (5), 168 (5), 295 (20), 31 (11), 38 (21), 394 (7) and 40 (10).

The mean virulence score (total of virulence genes present) was listed for these complexes. A linear regression was used to analyse if these complexes have a higher virulence in comparison to a reference group. ST38 complex was chosen as the reference groups as it contained the largest sample size and best representation of the data. All statistical analysis was carried out on STATA v 13.

Table 5.2 **Table of virulence factor genes used in this study**

Target gene	Primer	Primer Sequence	Amplicon length (bp)	Function	Reference
aat	aat_F aat_R	CTG GCG AAA GAC TGT ATC AT CAA TGT ATA GAA ATC CGC TGT T	630	Anti-aggregation protein transporter gene, Part of protein transporter system	(Jenkins *et al.* 2006b)
aaiC	aaiC_F aaiC_R	CTC TTA GCA GGG AGT TTG TC GCT TTG TTT ACC GAC TGA AC	430	aaiC from 042 pheU island	(Jenkins *et al.* 2006b)
astA	astA_F astA_R	CCA TCA ACA CAG TAT ATC CGA GGT CGC GAG TGA CGG CTT TGT	111	Enteroaggregative heat stable toxin 1 (EAST-1)	(Jenkins *et al.* 2006b)
aggR	aggR_F aggR_R	CTA ATT GTA CAA TCG ATG TA ATG AAG TAA TTC TTG AAT	308	Transcriptional activator of AAFs	(Czeczulin *et al.* 1999)
aggA	aggA_F aggA_R	GCT AAC GCT GCG TTA GAA AGA CC GGA GTA TCA TTC TAT ATT CGC C	421	AAF/1 fimbrial type I	(Piva *et al.* 2003)
aafA	aafA_F aafA_R	GAC AAC CGC AAC GCT GCG CTG GAT AGC CGG TGT AAT TGA GCC	133	AAF/II fimbrial type II	(Piva *et al.* 2003)
agg3A	agg3A_F agg3A_R	GTA TCA TTG CGA GTC TGG TAT TCA G GGGC TGT TAT AGA GTA ACT TCC AG	462	AAF/III fimbrial type III	(Bernier *et al.* 2002)
agg4A	agg4A_F agg4A_R	ATA CTT TAG ATA CCC CTC ACG CAG TCC ATT ATG TCA GGC TGC AA	411	AAF/IV fimbrial type IV	(Boisen *et al.* 2009)
aaf5A	aaf5A_F aaf5A_R aaf5A_P	GACTGGATTCTTCAGCTTAAATTAAG TTCATTTGATGCTGGATTGA GAGCCCGAGCCTGTACATAGATTTGT	250	AAF/V fimbrial type V	Mtwale unpublished
pic	Pic_F Pic_R	TTC AGC GGA AAG ACG AA TCT GCG CAT TCA TAC CA	500	Secreted protease (146kDA), 116kDa after cleavage, multifunctional protein involved in enteric pathogenesis.	(Piva *et al.* 2003)
Set1A	Set1A_F Set1A_R	TCA CGC TAC CAT CAA AGA TAT CCC CCT TTG GTG GTA	309	*Shigella* enterotoxin, anti-sense strand of pic	(Huang *et al.* 2007)

Table 5.2 Table of virulence factor genes used in this study… continued

Target gene	Primer	Primer Sequence	Amplicon length (bp)	Function	Reference
Irp2	Irp2_F Irp2_R	AAG GAT TCG CTG TTA CCG GAC TCG TCG GGC AGC GTT TCT TCT	280	Yersiniabactin biosynthetic gene	(Schubert *et al.* 1998)
tia	tia_F tia_R	AGT GAT AGC GGA GAT GAT TG CTC ACC CCG CTA TTT ATA TT	705	putative invasion determinant	(Jenkins *et al.* 2006a)
aap	aap_F aap_R	CTTTTCTGGCATCTTGGGT GTAACAACCCCTTTGGAAGT	232	Dispersin gene	(Piva *et al.* 2003)
pet	pet_F pet_R	CCGCAAATGGAGCTGCAAC CGAGTTTTCCGCCGTTTTC	1132	Plasmid encoded toxin	(Sheikh *et al.* 2002)
aaiC	aaiC_F aaiC_R aaiC_P	CATTTCACGCTTTTTCAGGAAT CCTGATTTAGTTGATTCCCTACG CACATACAAGACCTTCTGGAGAA	160	Part of the aai gene cluster encoding a type Vi secretion system	(EU Reference Laboratory for E.coli Department of Veterinary Public Health and Food Safety 2013)

Table 5.2: List of primers used for virulence typing

Table 5.3 Table of controls used in this study

Control Strains	Source	*aggA* (AAF/I)	*aafA* (AAF/II)	*agg3A* (AAF/III)	*agg4A* (AAF/IV)	*aaf5A* (AAF/V)	*aat*	*aggR*	*pic*	*astA*	*set1A*	*aaiC*	*irp2*	*tia*	*hra1*	*aap*	*pet*
042 (O44:H18)	Nataro	-	+	-	-	-	+	+	+	+	+	+	+	-	-	+	+
8089 (O7:H4)	ICDDR,B	-	-	-	+	-	+	+	-	-	-	-	+	-	-	+	-
601010 (O15:H23)	ICDDR,B	-	+	+	-	-	+	+	+	+	+	+	+	+	-	+	-
900732 (O?:H7)	ICDDR,B	-	-	-	+	-	-	+	-	-	-	-	+	-	+	+	-
900063(O?:H23)	ICDDR,B	+	-	-	-	+	+	+	-	+	+	-	-	-	-	+	-
E099518 (O104:H4)	LGP	-	-	+	-	-	+	+	NT	NT	NT	NT	NT	NT	NT	NT	NT

Table 5.3: Table of control strains used in this study. E099518 was only used for AAF/III and was not tested (NT) against all virulence factor

5.3 Results

5.3.1 HEp-2 Adherence

Strains showed variation in intensity of adherence to HEp-2 cells (Appendix 7.3) between each ST complex. In general, EAEC disease group ST394 Cplx showed heavy to very heavy binding (ST130 and ST394 Cplx) whereas the EAEC uropathogenic associated ST38 Cplx (Chattaway *et al.* 2014a) showed very light to medium binding. The EAEC exclusive, ST168 Cplx defined as being associated in equal numbers of cases and controls showed the most varied of binding capabilities and the groups associated with controls (ST31 Cplx) show a medium to very heavy binding phenotype with several cytotoxic strains destroying HEp-2 cells altogether (Table 5.4). There was no association between the intensity of individual strain adherence and whether it was associated with case or control.

Table 5.4 HEp-2 Adherence scores

Isolate	ST	ST Complex	HEp2	Case	Pathotype	Country	Year	Source	Origin
042	414	None	5 = Very Heavy	Case	EAEC	Peru	1987	EAEC reference strain	Human
H145-1R	31	ST31 Cplx	4 = Heavy	Case	EAEC	Peru	None	Okeke	Human
44-1R	31	ST31 Cplx	0 = Cytotoxic, autoaggregative	Case	EAEC	Thailand	None	Okeke	Human
309-1R	31	ST31 Cplx	5 = Very Heavy	Case	EAEC	Thailand	None	Okeke	Human
E54H	31	ST31 Cplx	3 = Medium	Control	EAEC	Nigeria	1995	Okeke	Human
E56H	31	ST31 Cplx	4 = Heavy	Control	EAEC	Nigeria	1994	Okeke	Human
G149H	474	ST31 Cplx	0 = Cytotoxic, autoaggregative	Control	EAEC	Nigeria	1994	Okeke	Human
G121aH	474	ST31 Cplx	4 = Heavy	Control	EAEC	Nigeria	1995	Okeke	Human
C16D	512	ST31 Cplx	4 = Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
C20D	130	ST31 Cplx	4 = Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
C60H	130	ST31 Cplx	4 = Heavy	Control	EAEC	Nigeria	1995	Okeke	Human
G116H	130	ST31 Cplx	4 = Heavy	Control	EAEC	Nigeria	1995	Okeke	Human
G23D	38	ST38 Cplx	1 = Very Light	Case	EAEC	Nigeria	1995	Okeke	Human
G29D	38	ST38 Cplx	3 = Medium	Case	EAEC	Nigeria	1995	Okeke	Human
G28D	426	ST38 Cplx	2 = Light	Case	EAEC	Nigeria	1995	Okeke	Human
G59D	394	ST394 Cplx	4 = Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
G10D	394	ST394 Cplx	5 = Very Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
C14D	394	ST394 Cplx	5 = Very Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
G17aD	394	ST394 Cplx	4 = Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
C08D	394	ST394 Cplx	5 = Very Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
E30D	394	ST394 Cplx	4 = Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
G108H	394	ST394 Cplx	5 = Very Heavy	Control	EAEC	Nigeria	1995	Okeke	Human
E64H	394	ST394 Cplx	5 = Very Heavy	Control	EAEC	Nigeria	1995	Okeke	Human
E33D	484	ST168 Cplx	4 = Heavy	Case	EAEC	Nigeria	1995	Okeke	Human
G30D	484	ST168 Cplx	2 = Light	Case	EAEC	Nigeria	1995	Okeke	Human
G110H	484	ST168 Cplx	0 = Cytotoxic, autoaggregative	Control	EAEC	Nigeria	1995	Okeke	Human
D09D	460	ST168 Cplx	3 = Medium	Case	EAEC	Nigeria	1995	Okeke	Human

Table 5.4: Table of results showing HEp-2 scores of EAEC strains representing the main groups defined in this study. See Appendix 7.3 for photos

5.3.2 Worm model

The graphs below show the worms fed on the control non-pathogenic strain pir116 to have the longest survival time whereas worms fed the EAEC strains from each complex were lethal including ST31 (carriage group) and ST394 Cplx (disease group) (Figure 5.3), the EAEC uropathogenic ST38 complex (Figure 5.4) and EAEC exclusive group ST484 (ST168 Cplx) (Figure 5.5). There was no association between the survival time and whether individual EAEC strains were from a case or control within a complex.

Figure 5.3 Survival of *C. elegans* fed with ST31 & ST394 Cplx EAEC strains

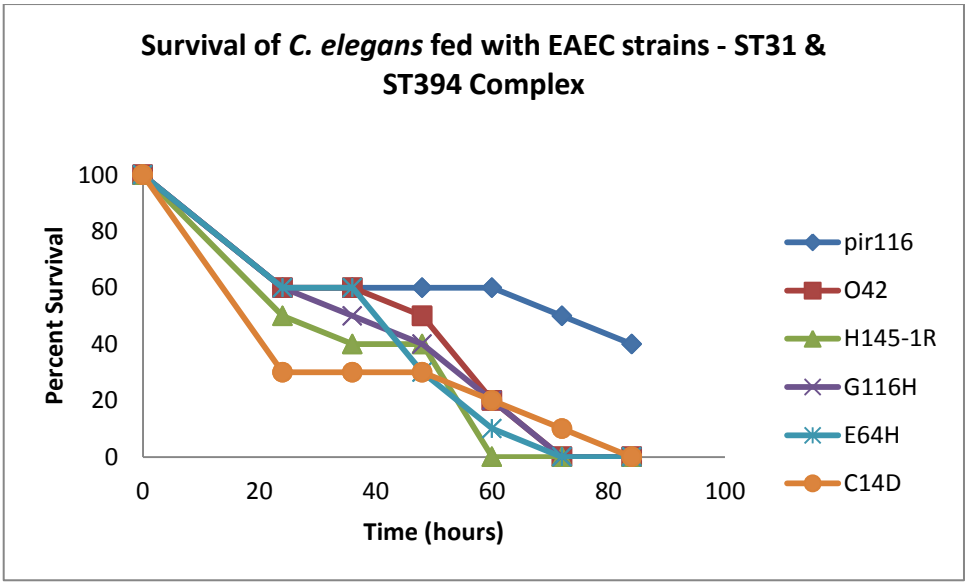


Figure 5.3: Survival of *C. elegans* fed with prototypical 042 EAEC virulent strain, laboratory attenuated pir116 control strain. Test strains include a case and control EAEC strain from EAEC ST31 Cplx (H145-1R & G116H) and ST394 Cplx (C14D & E64H) respectively.

Figure 5.4 Survival of *C. elegans* fed with ST38 EAEC strains

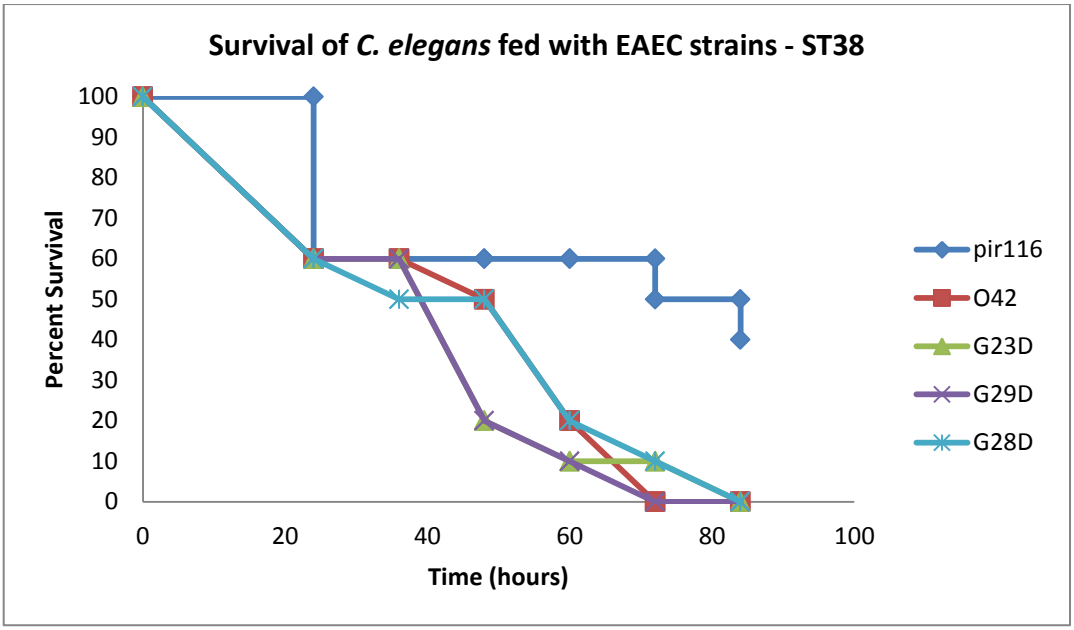


Figure 5.4: Survival of *C. elegans* fed with prototypical 042 EAEC virulent strain, laboratory attenuated pir116 control strain. Test strains include a case strains from the EAEC uropathogenic ST38 group

Figure 5.5 Survival of *C. elegans* fed with ST168 Cplx EAEC strains

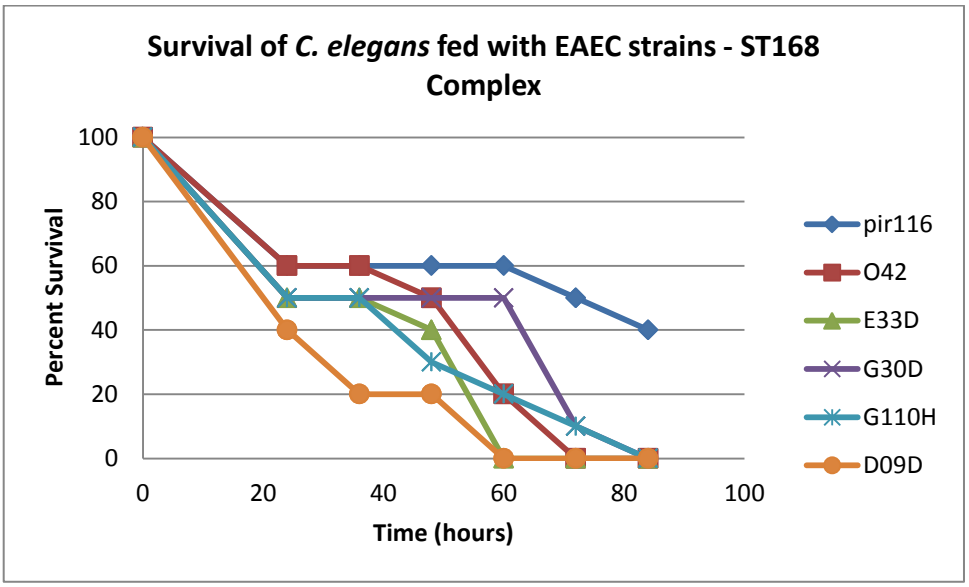


Figure 5.5: Survival of *C. elegans* fed with prototypical 042 EAEC virulent strain, laboratory attenuated pir116 control strain. Test strains include EAEC case strains from the mixed pathotype ST168 complex

The log rank test for equality of survivor functions indicated there was a significant difference between the survival of the worms in the different complexes in relation to each other (Chi square, P = 0.0170).

The groups were further analysed by cox regression, this study was limited by the small numbers so there would need to be extreme differences to show a significance determined by a hazard ratio of <0.5 or >2 and a probability <0.05.

1) The complex groups were comparable with the case reference strain 042 (chi square, P= 0.649).

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
ST168 Cplx (n=4)	1.11025	.3928736	0.30	0.768	.5548997 2.221403
ST31 Cplx (n=11)	.880873	.2916214	-0.38	0.702	.4603784 1.685434
ST38 Cplx (n=3)	1.097805	.4014096	0.26	0.799	.5361493 2.247838
ST394 Cplx (n=8)	.7842502	.2651851	-0.72	0.472	.404232 1.521523

2) The complex groups were statistically significant with the control reference strain pir116 (Chi square, P = 0.276). ST168 Cplx and ST38 Cplx were the most pathogenic with the highest Hazard ratio (3.1)

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
ST168 Cplx (n=4)	3.150686	1.388907	2.60	0.009	1.327922 7.475456
ST31 Cplx (n=11)	2.511154	1.058663	2.18	0.029	1.099057 5.737551
ST38 Cplx (n=3)	3.109995	1.400264	2.52	0.012	1.286794 7.516404
ST394 Cplx (n=8)	2.237626	.9526613	1.89	0.059	.9713851 5.154464

3) Case strains from each complex were comparable with the case reference strain 042 (Chi square, P=0.3422).

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
ST168 Cplx Cases (3)	1.120294	.4094586	0.31	0.756	.5472988 2.293187
ST31 Cplx Cases (5)	.7960973	.276724	-0.66	0.512	.4027985 1.573419
ST38 Cplx Cases (3)	1.085061	.3971413	0.22	0.823	.5295509 2.223315
ST394 Cplx Cases (6)	.7678135	.264989	-0.77	0.444	.3903802 1.510163

4) Control strains from each complex were comparable to the control reference strain pir116 (Chi square, P=0.112), ST31 Cplx control strains did show a difference from pir116. For example, Cplx control strains were still more pathogenic than the non-EAEC negative control strain. ST 168 Cplx was the most pathogenic with the highest hazard ratio, it was 2.7 times more likely to be pathogenic.

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
ST168 Cplx Controls (1)	2.720906	1.416667	1.92	0.055	.9806791 7.549184
ST31 Cplx Controls (6)	2.456715	1.066777	2.07	0.038	1.04891 5.75402
ST394 Cplx Controls (2)	1.987913	.9506842	1.44	0.151	.7786192 5.07539

5) Case strains from each complex were statistically different from the control reference strain pir116 (Chi square, 0.0250). ST168 was 3.7 times more likely to be pathogenic over the non-EAEC negative control strain.

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
ST168 Cplx (n=4)	3.364593	1.52294	2.68	0.007	1.385629 8.169925
ST31 Cplx (n=11)	2.418352	1.052245	2.03	0.042	1.030753 5.673935
ST38 Cplx (n=3)	3.254318	1.470943	2.61	0.009	1.341897 7.892251
ST394 Cplx (n=8)	2.331175	1.004852	1.96	0.050	1.001532 5.426063

6) All case strains (irrespective of complex) were statistically different from the control reference strain pir116 (Chi square, P=0.0078).

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
Cases vs pir116	2.604575	1.089875	2.29	0.022	1.146977 5.914515

7) All control strains (irrespective of complex) were statistically different from the control reference strain pir116 (Chi square, P=0.0229).

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
Controls vs pir116	2.366515	1.011242	2.02	0.044	1.024189 5.468125

8) All case strains (irrespective of complex) were comparable with the case reference strain 042 (Chi square, P= 0.693)

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
Cases vs 042	.878149	.2868496	-0.40	0.691	.4629361 1.665771

9) All control strains (irrespective of complex) were comparable with the case reference strain 042 (Chi square, P= 0.8706)

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
Controls vs 042	.9464675	.3174477	-0.16	0.870	.4904678 1.826421

The slow-kill indicated that EAEC strains were more lethal for worms than control strain pir116 and comparable with virulent 042 EAEC prototypical strain but that there was no significant difference between strains from cases and those from controls within the Cplxs. ST168 Cplx was the most pathogenic complex with this model.

5.3.3 Biolog

Analysis of the metabolite respiration of 97 EAEC (see Appendix 7.1) clustered into three main clades. Strains varied in metabolism irrespective of complex and while some complexes clustered together on the dendrogram, they were integrated with other complexes (Figure 5.6). There was no clustering of EAEC strains from cases or controls (Figure 5.7). This data showed that Biolog analysis of real-time metabolite respiration could not be used to differentiate EAEC complexes or EAEC from cases or controls which were distributed across the three clades.

Overlaying EAEC metabolic respiration in comparison to 1479 *E. coli* clinical strains (2010-2014) showed that EAEC had 5 main clusters (Figure 5.8) indicating that groups of EAEC did resolve together metabolically.

Figure 5.6 Dendrogram of Biolog of 97 EAEC strains coloured by complex

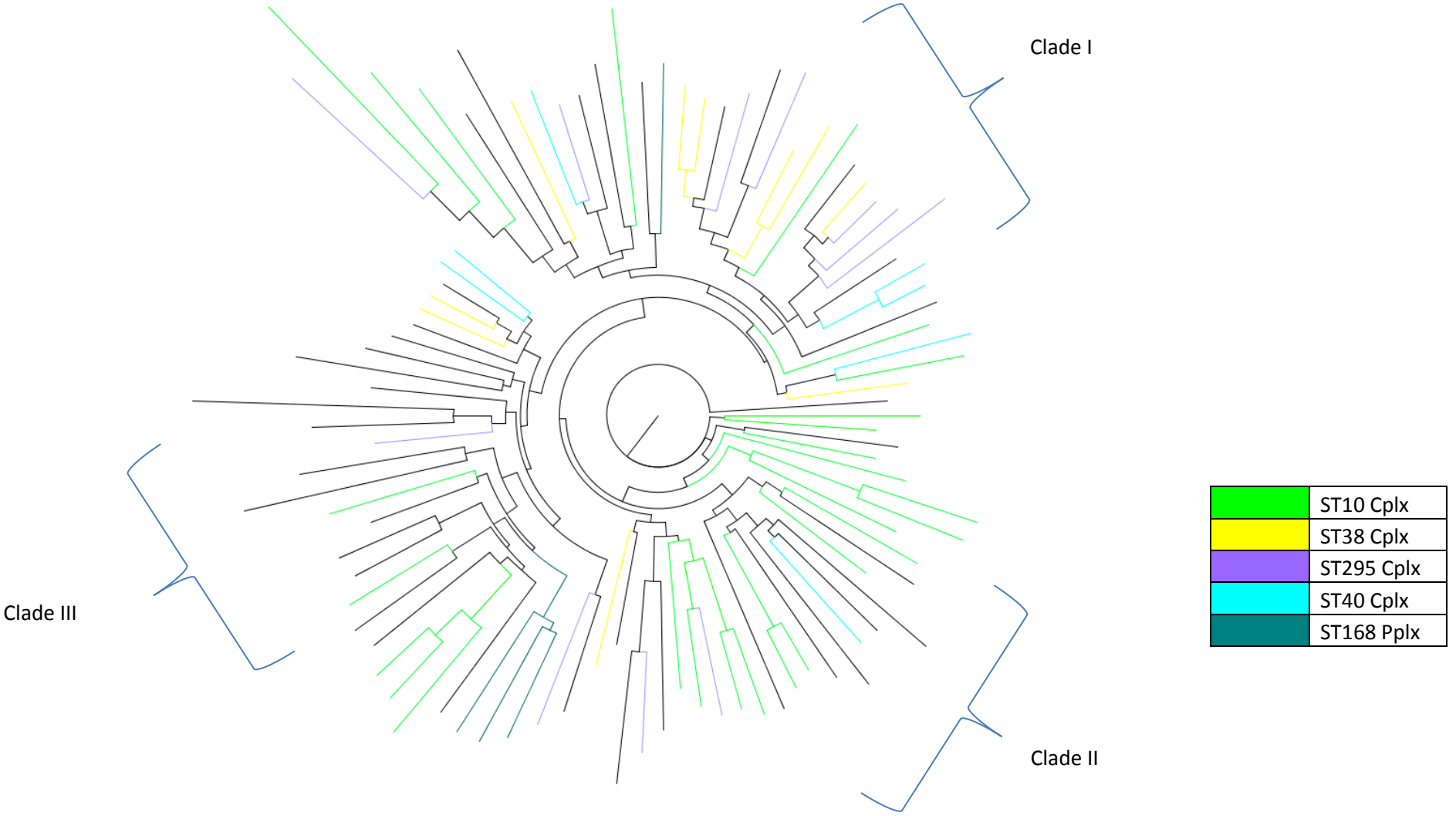


Figure 5.6: Dendogram of EAEC metabolism of 97 EAEC strains separates into three main clades. Strains are coloured by complex and indicated that Biolog cannot be used to resolve EAEC into complexes as they are all metabolically heterogeneous.

Figure 5.7 Dendrogram of Biolog of 97 EAEC strains coloured by case or control strain

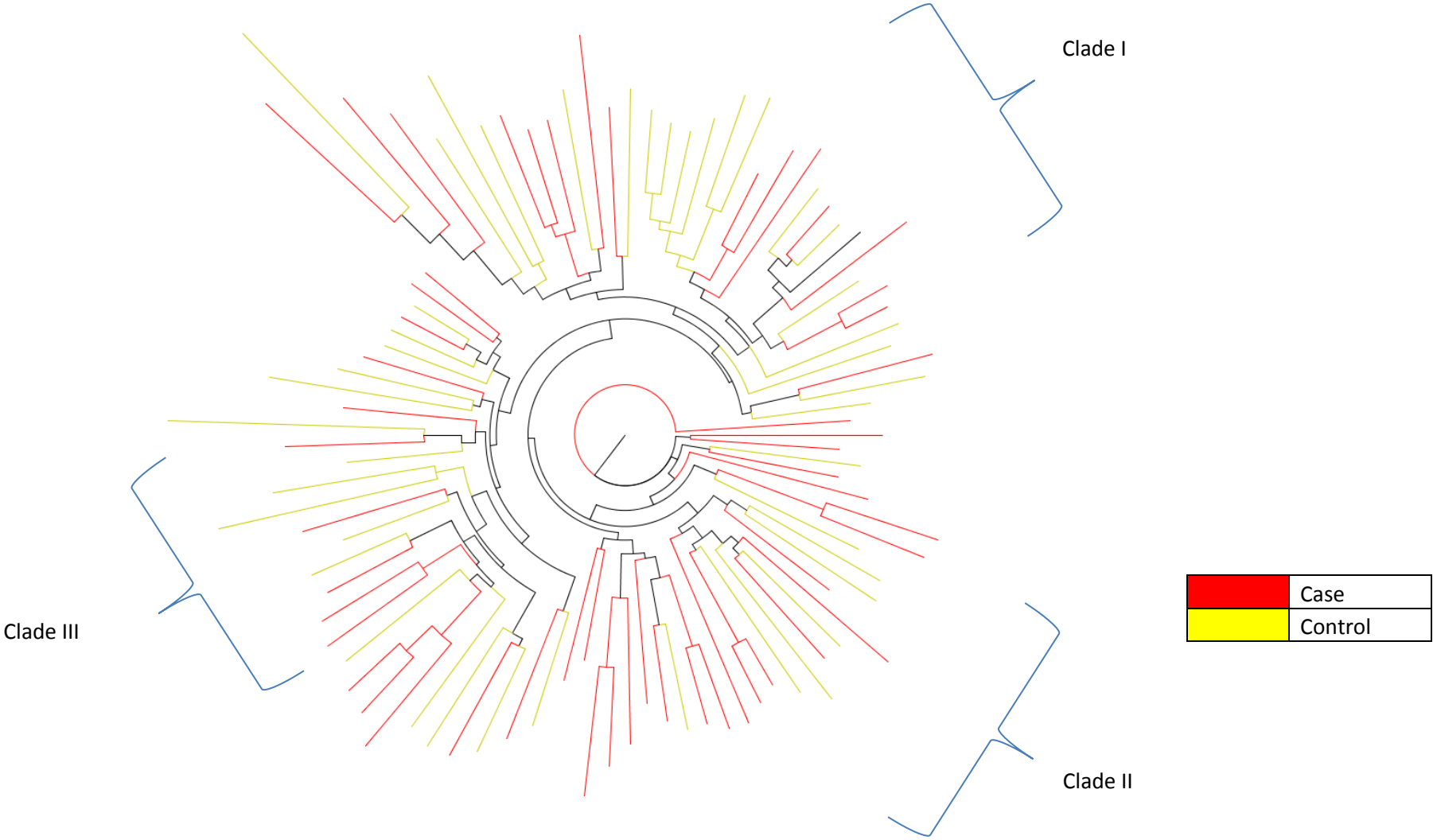


Figure 5.7: Dendogram of EAEC metabolism of 97 EAEC strains separates into three main clades. Strains are coloured by complex and indicated that Biolog cannot be used to resolve EAEC into case and control strains as they are all metabolically heterogeneous.

Figure 5.8 Dendrogram of Biolog of 97 EAEC strains (pink) with 1479 *E. coli*

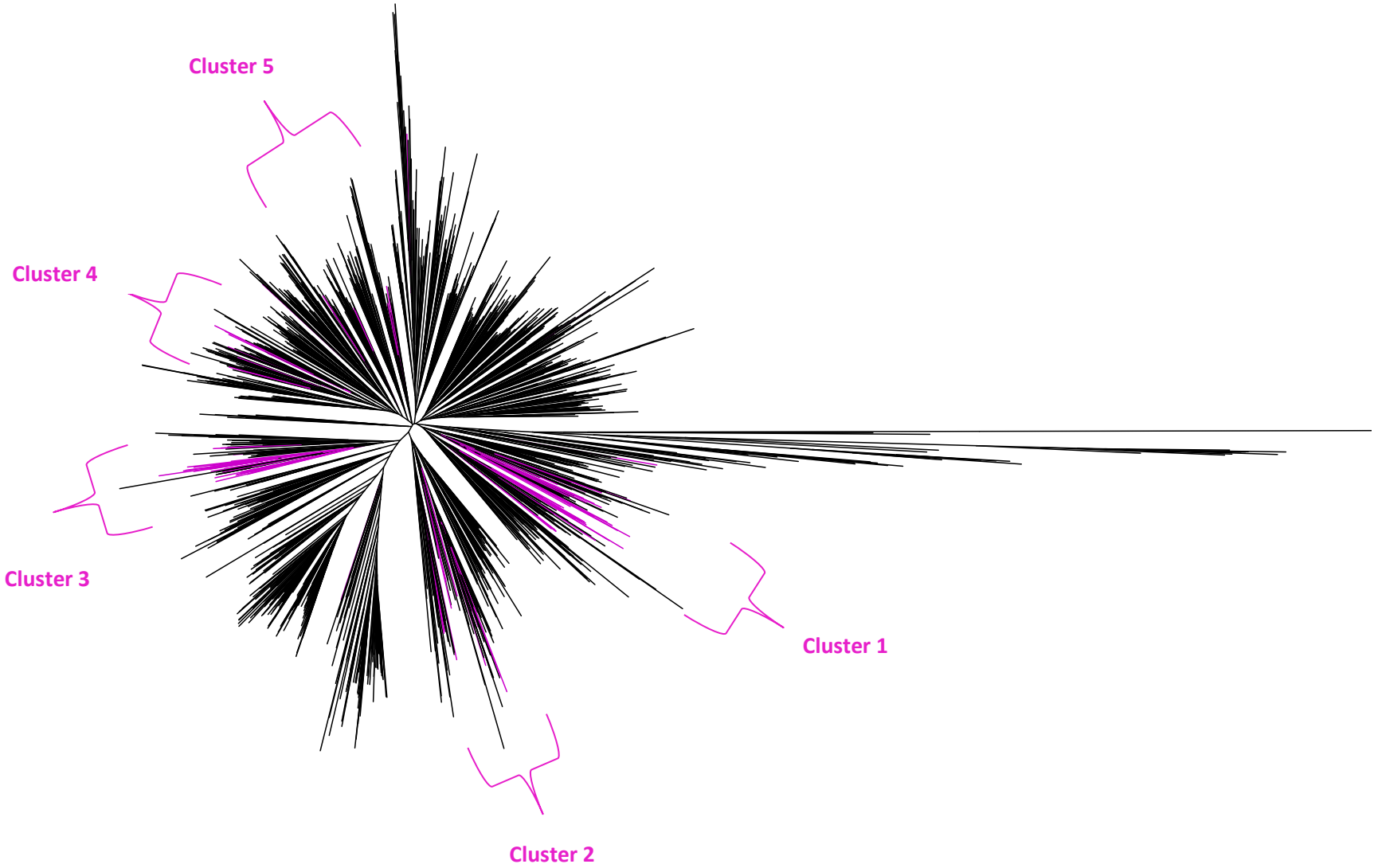


Figure 5.8: Dendogram of EAEC metabolism of 97 EAEC strains with 1479 clinical *E. coli* isolates processed from 2010-2014. Dendogram shows that EAEC fall into 4 main clusters.

5.3.4 Serological Distribution with EAEC clonal structure

Most EAEC serotypes were heterogeneous with respect to ST and dispersed throughout the population structure. Some serotypes were predominantly associated with STs (O104:H4-ST678, O125ac:H9-ST295, O111:H21-ST40, O153:H30-ST38, O7:H4-ST484) while others were found in multiple STs (O126:H27-ST200 & SLV, ST155, O166:H15-ST349 &SLV/DLV, ST130, ST394, O44:H18-ST449, ST414, ST30). There were no mutually exclusive ST and serotypes found in the EAEC population structure (Figure 5.9).

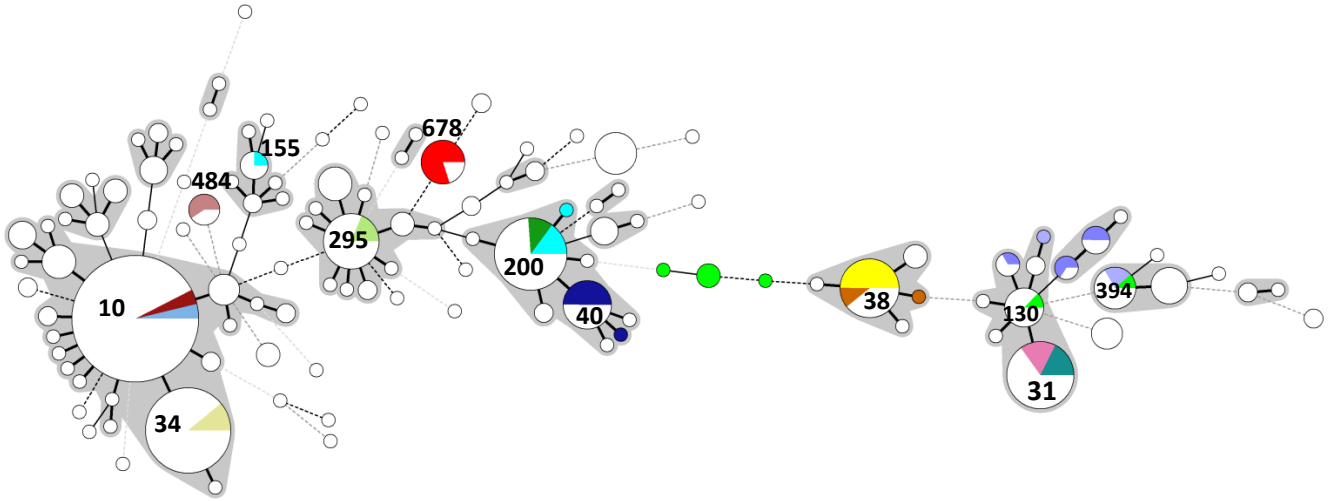
5.3.5 Virulence profiling in EAEC Groups

Virulence profiling was heterogeneous irrespective of complex, serotype and whether the isolate was from a case or control (Appendix 7.4). The average virulence score (total number of virulence genes present) was higher in cases (7.3) than controls (6.2), p-value=0.027. The proportion of specific individual genes present was similar in cases and controls was similar for most genes (5% difference) with p-values for association >0.196, with the exception of the chromosomal gene *ipr2* which was higher in cases (70%) than controls (58%) and borderline significant (p0.107) (Table 5.5).

The majority of virulence EAEC genes (*aggR, aat, aap, AAF1-4, astA, pet,pic, setA, ipr2, aaiC* and *tia*) were associated with the main complexes (ST 10, 155, 165, 168, 295, 31, 38,394 & 40 Cplx). EAEC fimbriae type 5 was not associated with the EAEC main complexes.

Mean of virulence scores per EAEC complex from Bangladesh indicated ST40 and 295 Cplx being the most virulent. A multivariable linear regression model with genes as predictors of virulence score indicated that ST40 and 295Cplx were statistically associated with more virulence genes than ST 38 Cplx (the reference group), whereas ST10 and 155 were the least virulent.

Figure 5.9 Minimal Spanning Tree of 443 EAEC serotypes



Key	Serogroup
	O153:H30 (9)
	O104:H4 (8)
	O166:H15 (7)
	O111:H21 (6)
	O126:H27 (5)
	O134:H27 (4)
	O130:H27 (4)
	O4:H33 (4)
	O44:H18 (4)
	O44:H34 (3)
	O86:H30 (3)
	O125ac:H9 (3)
	O3:H2 (3)
	O18ac:H30 (3)
	O175:H31 (3)
	O7:H4 (3)
	Serotype n=<3
	Single Locus Variant
	Double Locus Variant
	Triple Locus Variant

Figure 5.9: Minimal Spanning Tree of 443 EAEC serotyped in this study. Tree is colour coded by serotypes containing 4 or more isolates. Complexes shaded in grey consist of single locus variants (SLV). MSTree shows that some serotypes are ST specific such whereas others are dispersed across different STs indicated that serotyping cannot be used to look at genetic relatedness

Table 5.5 Virulence gene content in EAEC from cases and controls

	aggR	aat	aap	aggA	aafA	agg3A	agg4A	aaf5A	astA	pet	pic	setA	irp2	tia	aaiC
Case (93)	72% (67)	80% (74)	86% (80)	14% (13)	22% (20)	17% (16)	15% (14)	18% (17)	43% (40)	22.% (21)	44% (41)	45% (42)	70% (66)	37% (34)	45% (42))
Control (60)	67% (40)	83% (50)	83% (50)	17% (10)	22% (13)	13% (8)	17% (10)	13% (8)	47% (28)	23% (14)	48% (29)	40% (24)	58% (35)	37% (22)	56% (33)
Probability	0.479	0.562	0.65	0.65	0.981	0.52	0.811	0.419	0.657	0.914	0.607	0.529	0.107	0.989	0.196

Table 5.5: Table showing the content of each gene in association with cases and controls from 153 EAEC from Bangladesh.

Table 5.6 Virulence gene content in EAEC by association of complex (>5 isolates)

	aggR	aat	aap	aggA	aafA	agg3A	agg4A	aaf5A	astA	pet	pic	setA	irp2	tia	aaiC
Probability associated with complex	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.001**	0.217	**0.002**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.002**	**<0.001**

Table 5.6: Table showing the Complex included sample sizes of 5 or more including ST10 (n=18), 155 (8), 165 (5), 168 (5), 295 (20), 31 (11), 38 (21), 394 (7) and 40 (10) complexes. Majority of virulence EAEC genes were associated with the main complexes, EAEC fimbrie type 5 was not associated of being present with the EAEC main complexes.

Table 5.7 Mean virulence score of EAEC complexes

Complex	Sample size	Mean of virulence score
10	18	6.8
155	8	6.8
165	5	4.6
168	5	6.8
295	**20**	**9.2**
31	11	7.7
38	21	6.3
394	7	5.6
40	**10**	**10.4**

Table 5.7: Mean of virulence scores per EAEC complex from Bangladesh. ST40 and 295 Cplx are the most virulent in terms of average virulent gene content.

Table 5.8 Virulence score of complexes in comparison to reference group ST38 complex via linear regression.

Complex	Coefficient	Std. Err.	t	P>/t/	95% conf. Interval
10	0.44	0.85	0.520	0.601	-1.23, 2.12
155	0.29	1.10	0.270	0.790	-1.88, 2.46
165	-1.73	1.31	-1.320	0.189	-4.33, 0.87
168	0.47	1.31	0.360	0.723	-2.13, 3.07
295	**2.87**	**0.82**	**3.480**	**0.001**	**1.23, 4.50**
31	1.39	0.98	1.420	0.158	-0.55, 3.34
394	-0.76	1.15	-0.660	0.509	-3.04, 1.52
40	**4.07**	**1.01**	**4.020**	**0.000**	**2.06, 6.07**

Table 5.8: Comparison of virulence of complexes in relation to the reference group ST38 Cplx, ST 40 and 295 Cplx are statistically associated as being more virulent than ST 38 Cplx.

5.4 Discussion

5.4.1 Different EAEC complex's show variation in intensity of HEp-2 binding

There are multiple studies reporting proteins responsible for the aggregative adherence (AA) phenotype, notably the dispersin protein encoded by *aap*, EAEC fimbriae, *aggR* regulator (Bhargava *et al.* 2009;Boisen *et al.* 2008;Farfan *et al.* 2008;Moreira *et al.* 2003;Nataro *et al.* 1994;Sheikh *et al.* 2002). However, there have been no studies to date to assess if different lineages of EAEC have different binding capabilities. In this study EAEC lineages have been defined and their association with disease or carriage investigated. For reasons already discussed such as carriage, post infection immunity and co-infections, sequence type complexes invariably contain a mixture of EAEC from cases and controls; the pathogenicity of the lineage itself was investigated in this chapter.

Adherence varied in each of the ST complexes studied irrespective of whether the complex was associated with cases or controls. For example adherence was heavy or very heavy in ST394 Cplx (associated with cases) and ST31 Cplx (statistically associated with controls, Table 3.3). Variability was also seen looking at individual strains where a control or case strain could be light or heavy. This study indicated that intensity of adherence was independent of sequence type and not linked to whether the isolate had come from a case or a control.

Adherence capabilities of EAEC lineages based on this small study is not an appropriate method to assess phenotype for several reasons:

1) The case/control definition varies in literature and assumption that EAEC isolated from a diarrhoeal case is the causative agent maybe flawed. For example; co-infections were found in 40% of EAEC cases in some studies (Chattaway *et al.* 2013) demonstrating that EAEC from a case may not be capable of causing disease and incorrectly labelled as a case strain.

Conversely, strains isolated from a control may not have caused symptoms in that host, but maybe capable of causing symptoms in others.

2) The HEp-2 cell assay is laborious with multiple stages and techniques required, it is technically demanding and there are multiple stages where the test can go wrong affecting the end result. The final result is also qualitative and down to interpretation leading to discrepancies between laboratories.

3) There may have been laboratory discrepancies (such as misinterpretation of the slide) historically so it would not be possible to do a fair comparison to with current results. Also, technicians only stated an end point of adherence type, there have been no attempts to quantify the levels of adherence (light to heavy) and so this information was not available.

4) There was no control over strain selection at the time of this study which was based in another laboratory with access to a selection of the Nigerian EAEC strains. The main complexes of EAEC were not yet defined and so the strains selected were based on results of the Nigerian study (Okeke *et al.* 2010) which did not represent all of the important groups identified in this study, such as ST295 and 40 Cplx.

This study shows a complicated picture of adherence in relation to ST lineages and demonstrates further that EAEC pathogenicity is complex and multi-factorial. While ST groups associated with disease show heavy binding capabilities, so do control groups. In strains belonging to other lineages the binding capabilities were more variable (Table 5.4). What this study does clearly show is that strain adherence capability alone is not associated with cases or controls. Other factors such as co-infections (as discussed in Chapter 2) (Chattaway *et al.* 2013) or immunity (Nataro *et al.* 1995) or

virulence factors (Jenkins *et al.* 2007) as discussed in this chapter are all likely to play a role in infection.

5.4.2 EAEC complexes vary in lethality in worms

The EAEC complexes were statistically significant in reducing the survival time of the worms but due to the limitations of the log rank test (which can only assess data as a whole) it was not possible to determine which complex were the most pathogenic. This analysis highlights that the successful complexes of EAEC have evolved independently and in a variety of ways as described in Chapter 4 (Chattaway *et al.* 2014b) and therefore show different phenotypic characterisitics and have different pathogenicity potential. This is unsurprising since EAEC are heterogeneous in terms of genetic background, virulence gene content and plasmid compatibility (Okeke *et al.* 2010).

Data was further analysed using a Cox regression analysis which indicated that all EAEC within the complexes were statistically more lethal to worms than non – EAEC control strain pir116. EAEC have been shown to colonize distal to the worm grinder (and hence avoid being destroyed by the grinder) and throughout the entire length of the worms intestine whereas colonisation is not seen with non-EAEC control strains (OP50) (Hwang *et al.* 2010). It was therefore expected that the EAEC strains would be comparable with the pathogenic control EAEC strain 042 (the prototypical EAEC strain associated with disease) and this was confirmed in this study. The most pathogenic group in terms of colonising and killing the worms was ST168 (P=0.007) followed by ST38 (P=0.009), ST31 (P=0.42) and ST394 Cplxs (p=0.050). The results were not comparable with the adherence intensity of the HEp-2 assay in this study where ST31 and 394 Cplxs had the overall heaviest adherence to HEp-2 cells and ST38 and 168 were more variable and lighter in its adherence. This shows that intensity of adherence on HEp-2 cells is not linked to colonisation in the worm model.

There was no difference in colonisation (as seen in the worm model) as to whether the EAEC isolate was from a ‘case’ or ‘control’. For example a ‘control’ isolate in a complex could have a faster kill

time than a ‘case’ whereas overall ‘case’ isolates within a complex were statistically different from the non-EAEC control strain indicating that phenotypic colonisation varies within a complex. This could be evidence of homoplasy where strains within each ST complex have not come from the same ancestral source but exhibit similar properties and have converged from different sources to the same ST complex and therefore exhibit different phenotypes.

The *C. elegans* worm model has been shown to be a successful model for assessing EAEC pathogenic ability where EAEC colonise the distal *C. elegans* intestine, whereas non-pathogenic *E. coli* strains do not (Hwang *et al.* 2010). It is logical therefore that EAEC strains from control sources would also have a lethal effect on *C. elegans*. These data illustrates that, even on a small subset of EAEC from different successful complexes, there was variability between different complexes with respect to their ability to cause disease, that strains within a complex can vary and that there was a slight difference in the disease EAEC strains versus the carriage EAEC strains within a complex.

The EAEC strains were the same set tested with the HEp-2 and therefore too small a sample set to make statistically significant conclusions. Larger studies of more EAEC isolates within a complex with multiple biological repeats and including additionalcomplexes are required to assess if the difference described here are consistent as the dataset expands. The worm model assay is also laborious and technically demanding and while it has been shown to be a successful model in linking genes to EAEC pathogenicity via mutational studies (Hwang *et al.* 2010) , in this study there was too much variability of the strains within a complex to use this approach to assess the pathogenicity of each lineage.

5.4.3 EAEC metabolism varies with complexes

We know that EAEC strains have not evolved from a recent common ancestor rather there has been selection of distinct lineages to create the EAEC group. Strains exposed to the same ecological niche, the human GI tract, should be under the same selective pressures and should have evolved to

adaptat to their environment in the same way. We therefore hypothesised that metabolic ability would be similar within the EAEC groups. This metabolic similarity is observed in other groups of pathogenic *E. coli* such as enteroinvasive *E. coli* and *Shigella sonnei*. This study clearly shows that metabolism of 95 different carbon sources utilised by EAEC strains do not fully resolve into clusters associated with ST complexes. This suggests that EAEC strains within a complex may have adapted differently from each other. There are clear clusters of complexes, perhaps indicative of ancestral evolution, but there are also mixtures of different complexes clustering together and examples where the same complex spread across the dendogram (Figure 5.6). This variable metabolism further strengthens the theory that groups of EAEC evolved independently and are converging to several distinct successful complexes. There was no resolution of metabolic capabilities according as to whether a strain was from a case or control (Figure 5.7). With additional factors such as host immunity in which a pathogenic EAEC may not cause disease and co-infections where a carriage EAEC strain would be isolated from a case, it is unlikely that the definition of case and control strains of EAEC in this study are robust. However, this analysis does indicate, that EAEC continue to be a highly adaptive organism capable of surviving in a variety of ecological niches and therefore variability of phenotypic ability is vital in its survival.

Comparison of EAEC metabolic ability in the content of 1479 clinical *E .coli* gave distinct clustering of EAEC into five main clusters (Figure 5.8). Though these clusters contained a variety of serotypes and ST complexes (5.3.4 & Appendix 7.1), and therefore it was not possible to differentiate lineages, it was possible to identify EAEC signatures. Further work should involve expanding the EAEC strain collection representing all of the main complexes and linking information such as clinical source (blood, urine faeces, commensal) and symptoms to ascertain if the which metabolic activity is linked to clinical symptoms or specific environmental source.

5.5.4 Can serotyping be used to identify pathogenic lineages of EAEC?

In this study we found no correlation between the EAEC complexes and serotype (Figure 5.9). Since the development of sequence based typing, such as MLST, the use of traditional typing methods, such as serotyping as a means of defining population structure (Hartl and Dykhuizen 1984;Tenaillon *et al.* 2010) have come under close scrutiny. Studies have shown that the same serogroups are found in genetically unrelated strains of *E. coli* indicating possible horizontal gene transfer (Beutin *et al.* 2005) of the cassette encoding the serogroup genes.

This demonstrates that typing bacteria on the basis of a single set of genes clustered at the same locus, with a product under diversifying selection can result in the clustering of strains which are not related ancestrally. On the other hand genes for MLST, which are spread across the whole genome (Figure 1.5) are chosen because they encode housekeeping functions (Table 1.2) and so are under stabilising selection and accumulate mutations through genetic drift (Cooper and Feil 2004;Turner and Feil 2007). This provides a population framework based on ancestry where identical genotypes are closely related. This is important because it enables the identification of high risk clones for epidemiology such as STEC ST 11 (O157:H7), ST 21 and 29 (O26:H11) and ST678 (O104:H4) (Chattaway *et al.* 2011).

Although some serotypes were associated with single clonal complexes (Figure 5.9), we conclude that serotyping is not a suitable method for determining ancestral relatedness of EAEC and cannot be used for the identifying pathogenic lineages within EAEC. This may in part be attributable to the high recombination rates in some lineages.

5.5.5 Some EAEC complexes are more virulent than others

Multiple studies including this one, have shown that EAEC is a heterogeneous group in relation to virulence gene content , and importantly, there is no consistent chromosomal marker (Okeke *et al.* 2010). There are no studies assessing whether EAEC complexes contain more virulence genes than

others. This has not been attempted as EAEC complexes have not been previously defined as described in Chapter 3.

All EAEC genes with the exception of *aaf5A* were statistically associated with being present in the main complexes (Table 5.6). EAEC fimbriae type 5 is a relatively new fimbriae type (Dallman *et al.* 2012) and found in low numbers in this dataset and may explain why this gene is not associated with the complexes in this dataset. The statistical association of the other EAEC virulent genes support the MLST data that these are successful EAEC complexes able to stably retain the plasmid and chromosomal EAEC markers.

The association of genes versus EAEC from cases and control was assessed and compared to previous studies (Jenkins *et al.* 2006a). Results showed that the virulence genes had no statistical association with cases over controls. This includes the *aaiC* gene (a chromosomal marker encoding a secreted protein of the EAEC pathogenicity island AAI, which is co-ordinately regulated by the a*ggR* activator). This *aaiC* gene has been suggested in a recent study to be an important component of regulating the EAEC pathogenicity AAI Island and that AAI operon deficient strains may have reduced pathogenicity irrespective of other virulence gene content (Dallman *et al.* 2014). In this study, the presence of *aaiC* was higher in controls (56%) than cases (45%) from Bangladesh EAEC strains (Table 5.5) conflicting this previous study. However, the Dallman *et al* study was analysing strains relating to a complex multi-strain outbreak of travellers in which the larger numbers of specific ST types did have the *aaiC* gene. The Bangladesh dataset is from an endemic region where 'control' patients (non-travellers) may have acquired immunity. What this study does highlight is that the EAEC strains were not consistently found with the regulator gene *aggR*, indicating that even regulation of genes in EAEC is heterogeneous and that there may be multiple regulation factors involved.

Previous studies have mainly shown that there is no pattern of specific makers being related to disease, as all of the makers are found in either cases or controls (Jenkins *et al.* 2006a;Okeke *et al.* 2010). There is one exception, where a small study in Brazil found a capsular (*kps*) marker only found in cases, but this was in one community and found in only 50% of case strains (Regua-Mangia *et al.* 2009). The dataset in this part of the study comprised the Bangladesh strains only and although it represented a large proportion of EAEC complexes it did not take into consideration the global variation of the EAEC strains. However, results indicate that characterising EAEC from a small population without exposure to travel (and therefore an additional subset of EAEC in the community) still results in a very heterogeneous group. Virulence typing continues to produce heterogeneous results for characterising EAEC and there is no definite pathogenic EAEC marker.

Diagnostically, if the true burden of EAEC is to be determined then the local hospital and regional laboratories need a robust assay targeting an EAEC specific gene or combination of genes. Although this study and recent studies have found *aaiC* in both cases and controls (Boisen *et al.* 2012;Lima *et al.* 2013), it is recognised as an important marker for clinically significant EAEC and is now used alongside *aat* to detect EAEC (Boisen *et al.* 2012;Dutta *et al.* 2013;Taniuchi *et al.* 2013). In the United Kingdom, *aggR* is used as the primary target for screening diagnostically as it has found to be more consistently present in EAEC than *aat* (Public Health England, Colindale). *AaiC* is included in follow-up characteristic studies. Historically previous studies focused on the virulence genes found on the plasmid and genetic background (i.e. EAEC complexes) was not considered as important factor in association with EAEC disease. This study showed that it is essential to consider the contribution of genetic background to the pathogenic potential of strains of EAEC. It is concluded that there are important, successful and prevailing complexes associated with cases that contain multiple EAEC virulence factors.

ST40 and 295 had the highest virulence score (9.6 & 8.3 respectively) (Table 5.7). This was confirmed when comparing the virulence score of the complexes against a reference group ST38 Cplx (Table 5.8). ST38 Cplx was chosen as a reference group as it contained the largest number of strains and therefore was assumed to be representative of strain variation with in complex. The ST31 Cplx was statistically associated with being a control group (P=0.005, table 3.3) but its virulence score was higher than other disease complexes. This may be attributed to the low numbers of Bangladesh ST31 Cplx strains (11/50 strains, 22%) and so the strains included for analysis may not be representative of the virulence content within this complex. This study shows that virulence within complexes does vary and that some complexes are more virulent than others. ST40 Cplx had the highest virulence profile score (Table 5.7) and was shown to be statistically associated with disease (Table 3.3) and was significantly associated with being an EAEC complex (Table 3.4). ST295 Cplx also had a high virulence score and statistically associated as an EAEC complex (Table 3.4) but not statistically associated with disease (Table 3.3). When analysing table 3.3, it was seen that the largest complexes (i.e. containing the most strains) including ST10, 40 and 31 Cplx are statistically relevant and contain over 50 isolates but any groups containing less than 50 isolates are not significant indicating that these groups are potentially clinically relevant but that the dataset is too small for statistical significance.

Although the functions of many of these genes have been described, a full understanding of the functions and how they interact with other genes are not known. While each virulence genes was not independently significantly associated with the ability to cause disease, the number and combination of virulence genes could be significant. This study illustrates how a combination of defined EAEC complexes in association with known virulence gene profiles can contribute to the clinical picture and that the genomic background and virulence gene content may relate to disease. For example it is known that from previous studies ST11 (O157:H7) with the *stx* gene is clinically important, this study highlights that ST40 EAEC is associated with disease and high virulence

potential as shown a recent EAEC/STEC household outbreak (Dallman *et al.* 2012). ST38 Cplx (Table 3.5), shown to be associated with extra-intestinal infection had an average high virulence score of 9 (including extra-intestinal virulence markers) (Chattaway *et al.* 2014a). ST38 Cplx in the Bangladesh set had an average high virulence score of 6 though over 50% of the strains had a higher score (Figure 7.4). Other complexes such as ST678 (Chattaway *et al.* 2011) and 278 Cplx (both of which are not represented by the Bangladesh subset) are associated with outbreaks (Figure 3.5) as described in Chapter 3 and are also likely to have a high virulence score.

5.5 Summary

This study demonstrated that the diverse genetic backgrounds of EAEC also have heterogeneous phenotypes; there is no obvious link between different phenotypic models of EAEC complexes and pathogenesis as defined by belonging to a case or control. There is not one EAEC complex that is directly linked to a particular phenotype or disease capability. This confirms the heterogeneous nature of EAEC and provides further evidence of the likelihood of independent evolution, homoplasy and convergence of the EAEC phenotype from multiple ancestral lineages. In this work, because of the small sample size of the strains selected and the complexity of the the *in vitro* models of pathogenesis, this approach did not help resolve the central question of the thesis. However, the genotypic model of linking virulence genes to defined EAEC complexes showed how certain EAEC complexes are more virulent and therefore clinically relevant in public health.

Chapter Six

Discussion

Chapter 6 - Discussion

Prior to this study in 2009, EAEC was defined as the ability of adhere to HEp-2 cells in a stacked bricked formation and/or contain the AA plasmid and was known to cause diarrhoea in developing and developed countries (Okeke & Nataro 2001). EAEC was not recognised as an important pathogen in the UK as shown by protocols for the detection of gastrointestinal (GI) pathogens at frontline hospital laboratories which did not include EAEC. The second intestinal infectious disease (IID) study in 2008-2009 was reanalysed in this study to ascertain if the association between EAEC carriage and diarrhoea was causal. Taking bacterial load and co-infection of other pathogens into account, EAEC accounted for approximately 1% of diarrhoeal disease annually in the UK, equivalent to *Salmonella,* confirming that EAEC was an important pathogen in the UK (Chattaway *et al.* 2013).

In 2011 awareness of EAEC as a significant pathogen of GI disease was further increased by a large outbreak of HUS in Germany caused by an EAEC strain (ST678 and serotype O104:H4) that acquired the *stx* phage from a STEC (Chattaway *et al.* 2011). In preparation for the London Olympics in 2012, a multiplex PCR was developed for the detection of common gastrointestinal pathogens, including EAEC. This assay was subsequently was rolled out to frontline laboratories. Select laboratories continue to use this assay in the event of an outbreak which has detected EAEC in samples that would have previously been missed. These samples were sent to the Gastrointestinal Bacteria Reference Unit (GBRU) for isolation and typing. GBRU introduced routine detection of EAEC in 2012 and have seen an increase of EAEC for identification in the past two years (Figure 6.1). EAEC have been isolated from routinely screened faeces originally submitted for testing for STEC or *Shigella* indicating that some strains of EAEC can cause severe symptoms and that many clinicians are still not aware of EAEC as a potential aetiological agent for IID.

Figure 6.1 Number of EAEC detected at GBRU

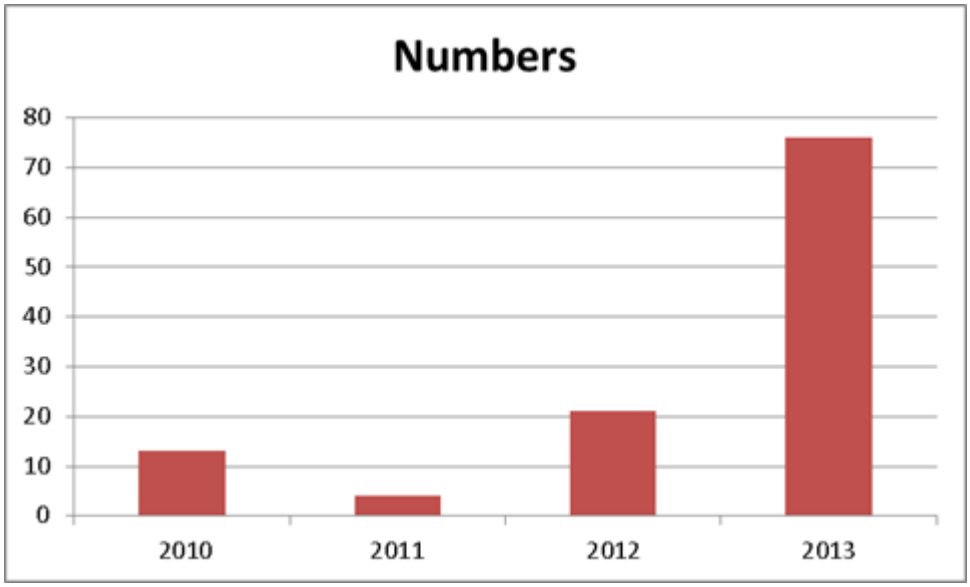


Figure 6.1: Number of EAEC detected over the past four years at Gastrointestinal Bacteria Reference Unit, Public Health England, Colindale.

EAEC is independently capable of causing disease even when accounting for co-infections and people are 2.5 times more likely to have IID if they have EAEC present in their stool (Chattaway *et al.* 2013). Clearly EAEC is an important cause of gastrointestinal disease and accounts for ~ 1% of IID in the UK, but it can also be carried asymptomatically (Chattaway *et al.* 2013) suggesting that there are different strains that vary in their ability to cause disease. An investigation of the population structure of EAEC using isolates from case control studies and routinely isolated strains was carried out to look for lineages associated with disease or carriage.

Within the global EAEC population over the past 30 years, there were multiple successful EAEC lineages characterised by the ability to maintain carriage of the EAEC plasmid (Chapter 3). These lineages have continued to expand over time and develop single, double and triple locus variants particularly the ST10, 31, 40 and 295 Cplxs and were found in multiple countries across the globe (Figure 3.2). This study defined successful EAEC lineages associated with disease, the largest (10 strains or more) being ST 10, 38, 40, 295, 278, 394, 678 and 746 (DLV to ST10) Cplx. When using this

population structure approach in the context of other public health concerns it was discovered that there is an emerging extra-intestinal ST38 EAEC Cplx with multiple drug resistance (Chattaway *et al.* 2014a) and two important groups associated with multiple EAEC outbreaks including ST678 (Chattaway *et al.* 2011) and ST278 Cplx (Chapter 3, Figure 3.5). Now that these clinically relevant EAEC groups have been defined they can be used when interpreting data.

Further analysing this population structure via ClonalFrame indicated multiple genetic events across the MLST loci and that a combination of mutation and recombination events were responsible for the evolution of this pathotype (Chapter 4, Figure 4.2) (Chattaway *et al.* 2014b). Overlaying the EAEC population structure with the entire *E. coli* phylogeny showed that these successful EAEC groups were dispersed throughout the *E. coli* population (Chapter 4) (Chattaway *et al.* 2014b), again indicating independent evolution of EAEC lineages. These lineages were not resolved into distinct case and control lineages indicating that other factors such as the bacterial phenotype (for example intensity of adherence), the definition of a 'case', co-infection or host variation may play a role the ability of a given EAEC strain to cause disease.

When looking at pathogenicity models to link genotype to phenotype, results were variable in terms of strain variability and reproducibility between models (Chapter 5). EAEC was more lethal in the worm model than non-pathogenic *E. coli* but the rate of killing was variable between strains within a complex. The least survival time of the worm model was linked to ST168 Cplx yet the adherence assay showed light adherence. The highest level of adherence was seen with ST31 & ST394 Cplx but these were the least pathogenic groups in the worm model. The phenotypic HEp-2 assay and worm model were inconclusive. A larger more comprehensive study may resolve results, particularly in light of the knowledge gained during the cause of this study.

Although there were serotypes exclusively from cases (O3:H2, O44:H18, O104:H4, O111:H21, O126:H27and O134:H27) we found no link between a sequence type and a single serotype in this study. Although some serotypes were associated with single clonal complexes, they were not mutually exclusive and high recombination rates in some lineages meant that a given serotype could also be distributed in different complexes (Chapter 5, Figure 5.9). Due to the fact that serotyping did not always correlate with genetic relatedness or disease, it cannot be used to infer genetic background or pathogenicity.

The metabolic activity, as measured by the Biolog, was variable with all of the complexes with different complexes falling into the same clusters (Figure 5.6). These data produced further evidence that these organisms did not evolve from one ancestral source expanding into different complexes with the same phenotypic abilities. Instead, it seems likely that there were multiple events in different lineages and that these groups evolved independently and have converged into the successful and established EAEC phenotype. This model of evolution suggests selective pressure is at work.

The genotypic model of linking virulence gene content to the defined EAEC complexes in this study showed that some complexes were more virulent than others. ST40 Cplx had the highest virulence score (Chapter 5, Table 5.34 & 5.3.5), was statistically associated with disease (Table 3.3) and being associated as an EAEC complex (Table 3.4) and has been implicated in a recent EAEC/STEC household outbreak (Dallman *et al.* 2012). The approach used in this study has been unique by defining successful EAEC complexes associated with disease and then testing these groups to assess clinical relevance in the context of public health.

6.1 Further work

This work could be taken further by investigating Biolog as a tool to identify EAEC from other *E. coli*. Biolog is currently used as a tool for identification of *E. coli* at the reference laboratory (Public Health England) but this study has highlighted potential clusters of EAEC that separate from the other *E. coli* (Figure 5.8). These clusters are only based on 97 EAEC and a small subset so a larger dataset would be required to assess if EAEC do differentiate from other *E. coli*. It was shown that the clusters are not by defined 'cases' or complexes but could be for other reasons such as the source of *E. coli* (urine, blood, faeces) and severity of symptoms which could be assessed.

This project has been based on MLST which is small representation on the core genome, ideally, this dataset could be further sequenced by whole genome sequencing and the core and accessory genome used for further analysis.

This would help ascertain:

1) If the MLST population structure is a true representation of the EAEC phylogeny (using core genome)
2) A base line for variation of the complexes for future outbreak investigation (such as single nucleotide polymorphism analysis of the core genome)
3) A true picture of all virulence genes present rather than a specific selection, other genes maybe highlighted as being important.
4) Redefining what is an EAEC case, currently the definition is an EAEC isolated from a case but this does not mean it is pathogenic.

This study has enabled an understanding of the complexes associated with disease but none of these methods such as the HEp-2 assay, virulence typing or population structure define disease. There are clearly sub-populations of EAEC that are pathogenic but can still be in asymptomatic controls. This highlights the complexity and heterogeneous nature of this organism and that more information is needed. A comparison analysis of clinical information and the techniques used throughout this thesis

can be used with whole genome sequence data to determine markers and redefine the group as pathogenic and not whether it is 'EAEC'.

6.2 Conclusion

The analysis of the data generated during this study has enabled the original objectives to be met as described below:

i. Methods to assess disease burden using results of a semi-quantitative real-time PCR assay to diagnose EAEC aetiology in episodes of IID in the UK were developed (Chapter 2). This study has shown that EAEC is capable of causing disease independently of other co-infections and is responsible for 1% of gastrointestinal infection in the UK (Chattaway *et al.* 2013).

ii. The population structure of EAEC *globally* has been defined and divided into phylogenetically relevant groups to allow analysis of the association between defined groups and disease (Chapter 3). This study has defined clinically relevant EAEC Cplxs and the approach has been used in public health investigations (Chattaway *et al.* 2011;Chattaway *et al.* 2014a;Dallman *et al.* 2012)

iii. The ancestry of the defined groups has been analysed (Chapter 4). This study has shown how these groups have independently evolved via mutation and recombination events (Chattaway *et al.* 2014b).

iv. The genotype phenotype associations to define characteristics which may be useful for diagnosis has been investigated (Chapter 5). This study has shown the phenotypic association with disease Cplxs and that genes associated with disease can be useful in public health investigation (Dallman *et al.* 2014).

EAEC is an important aetiological agent of gastrointestinal disease and several clinically relevant complexes have been defined to facilitate public health investigations. This study showed that many groups of *E. coli* known as EAEC have independently evolved and that the heterogeneous nature has enabled EAEC to expand globally and patho-adapt to cause extra-intestinal infections and STEC outbreaks. The novel approach in understanding the population structure of EAEC used in this study has had a positive impact on public health by facilitating outbreak investigations

References

Aballay, A. & Ausubel, F.M. 2002. Caenorhabditis elegans as a host for the study of host-pathogen interactions. *Curr.Opin.Microbiol.*, 5, (1) 97-101 available from: PM:11834377

Aballay, A., Yorgey, P., & Ausubel, F.M. 2000. Salmonella typhimurium proliferates and establishes a persistent infection in the intestine of Caenorhabditis elegans. *Curr.Biol*, 10, (23) 1539-1542 available from: PM:11114525

Altman, D. G. 1991, "Comparing groups - categorical data," *In Practical Statistic for Medical Research*, London: Chapman and Hall, pp. 229-276.

Altman, D.G. & Andersen, P.K. 1989. Bootstrap investigation of the stability of a Cox regression model. *Stat.Med*, 8, (7) 771-783 available from: PM:2672226

Amar, C.F., East, C.L., Grant, K.A., Gray, J., Iturriza-Gomara, M., Maclure, E.A., & McLauchlin, J. 2005. Detection of viral, bacterial, and parasitological RNA or DNA of nine intestinal pathogens in fecal samples archived as part of the english infectious intestinal disease study: assessment of the stability of target nucleic acid. *Diagn.Mol.Pathol.*, 14, (2) 90-96 available from: PM:15905692

Amar, C.F., East, C.L., Gray, J., Iturriza-Gomara, M., Maclure, E.A., & McLauchlin, J. 2007. Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996). *Eur.J.Clin.Microbiol.Infect.Dis.*, 26, (5) 311-323 available from: PM:17447091

Andreoli, S.P., Trachtman, H., Acheson, D.W., Siegler, R.L., & Obrig, T.G. 2002. Hemolytic uremic syndrome: epidemiology, pathophysiology, and therapy. *Pediatr.Nephrol.*, 17, (4) 293-298 available from: PM:11956886

Antikainen, J., Tarkka, E., Haukka, K., Siitonen, A., Vaara, M., & Kirveskari, J. 2009. New 16-plex PCR method for rapid detection of diarrheagenic Escherichia coli directly from stool samples. *Eur.J.Clin.Microbiol.Infect.Dis.*, 28, (8) 899-908 available from: PM:19238467

Atmar, R.L., Opekun, A.R., Gilger, M.A., Estes, M.K., Crawford, S.E., Neill, F.H., & Graham, D.Y. 2008. Norwalk virus shedding after experimental human infection. *Emerg.Infect.Dis.*, 14, (10) 1553-1557 available from: PM:18826818

Ayala, F.J. 1976. *Molecular evolution* Sunderlannd,Mass., Sinauer Associates.

Bailey, J.K., Pinyon, J.L., Anantham, S., & Hall, R.M. 2010. Commensal Escherichia coli of healthy humans: a reservoir for antibiotic-resistance determinants. *J.Med Microbiol.*, 59, (Pt 11) 1331-1339 available from: PM:20671087

Barletta, F., Ochoa, T.J., Mercado, E., Ruiz, J., Ecker, L., Lopez, G., Mispireta, M., Gil, A.I., Lanata, C.F., & Cleary, T.G. 2011. Quantitative real-time polymerase chain reaction for enteropathogenic Escherichia coli: a tool for investigation of asymptomatic versus symptomatic infections. *Clin.Infect.Dis.*, 53, (12) 1223-1229 available from: PM:22028433

Baudry, B., Savarino, S.J., Vial, P., Kaper, J.B., & Levine, M.M. 1990. A sensitive and specific DNA probe to identify enteroaggregative Escherichia coli, a recently discovered diarrheal pathogen. *J.Infect.Dis.*, 161, (6) 1249-1251 available from: PM:2189007

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., & . 1994. A multistate outbreak of Escherichia coli O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. *JAMA*, 272, (17) 1349-1353 available from: PM:7933395

Berger, C.N., Shaw, R.K., Ruiz-Perez, F., Nataro, J.P., Henderson, I.R., Pallen, M.J., & Frankel, G. 2009. Interaction of enteroaggregative Escherichia coli with salad leaves. *Environ.Microbiol.Rep.*, 1, (4) 234-239 available from: PM:23765852

Bernier, C., Gounon, P., & Le, B.C. 2002. Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative Escherichia coli as a sensitive probe for detecting the AAF-encoding operon family. *Infect.Immun.*, 70, (8) 4302-4311 available from: PM:12117939

Beutin, L., Tao, J., Feng, L., Krause, G., Zimmermann, S., Gleier, K., Xia, Q., & Wang, L. 2005. Sequence analysis of the Escherichia coli O15 antigen gene cluster and development of a PCR assay for rapid detection of intestinal and extraintestinal pathogenic E. coli O15 strains. *J.Clin.Microbiol.*, 43, (2) 703-710 available from: PM:15695667

Bewick, V., Cheek, L., & Ball, J. 2004. Statistics review 13: receiver operating characteristic curves. *Crit Care*, 8, (6) 508-512 available from: PM:15566624

Bhargava, S., Johnson, B.B., Hwang, J., Harris, T.A., George, A.S., Muir, A., Dorff, J., & Okeke, I.N. 2009. Heat-resistant agglutinin 1 is an accessory enteroaggregative Escherichia coli colonization factor. *J.Bacteriol*, 191, (15) 4934-4942 available from: PM:19482929

Bland, J.M. & Altman, D.G. 1998. Survival probabilities (the Kaplan-Meier method). *BMJ*, 317, (7172) 1572 available from: PM:9836663

Bland, J.M. & Altman, D.G. 2004. The logrank test. *BMJ*, 328, (7447) 1073 available from: PM:15117797

Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., & Shao, Y. 1997. The complete genome sequence of Escherichia coli K-12. *Science*, 277, (5331) 1453-1462 available from: PM:9278503

Boisen, N., Ruiz-Perez, F., Scheutz, F., Krogfelt, K.A., & Nataro, J.P. 2009. Short report: high prevalence of serine protease autotransporter cytotoxins among strains of enteroaggregative Escherichia coli. *Am.J.Trop.Med.Hyg.*, 80, (2) 294-301 available from: PM:19190229

Boisen, N., Scheutz, F., Rasko, D.A., Redman, J.C., Persson, S., Simon, J., Kotloff, K.L., Levine, M.M., Sow, S., Tamboura, B., Toure, A., Malle, D., Panchalingam, S., Krogfelt, K.A., & Nataro, J.P. 2012. Genomic characterization of enteroaggregative Escherichia coli from children in Mali. *J.Infect.Dis.*, 205, (3) 431-444 available from: PM:22184729

Boisen, N., Struve, C., Scheutz, F., Krogfelt, K.A., & Nataro, J.P. 2008. New adhesin of enteroaggregative Escherichia coli related to the Afa/Dr/AAF family. *Infect.Immun.*, 76, (7) 3281-3292 available from: PM:18443096

Borman, E.K., Stuart, C.A., & Wheeler, K.M. 1944. Taxonomy of the Family Enterobacteriaceae. *J.Bacteriol.*, 48, (3) 351-367 available from: PM:16560842

Bray, J. 1945. Isolation of antigenically homogenous strains of Bact. Coli neapolitanum from summer diarrhoea in infants. *J.Pathol.Bacteriol.*, 57, 239-247

Brenner, D.J., Fanning, G.R., Johnson, K.E., Citarella, R.V., & Falkow, S. 1969. Polynucleotide sequence relationships among members of Enterobacteriaceae. *J.Bacteriol*, 98, (2) 637-650 available from: PM:4891264

Brenner, D.J., Fanning, G.R., Skerman, F.J., & Falkow, S. 1972. Polynucleotide sequence divergence among strains of Escherichia coli and closely related organisms. *J.Bacteriol*, 109, (3) 953-965 available from: PM:4551757

Brenner, S. 1974. The genetics of Caenorhabditis elegans. *Genetics*, 77, (1) 71-94 available from: PM:4366476

Buchholz, U., Bernard, H., Werber, D., Bohmer, M.M., Remschmidt, C., Wilking, H., Delere, Y., an der, H.M., Adlhoch, C., Dreesman, J., Ehlers, J., Ethelberg, S., Faber, M., Frank, C., Fricke, G., Greiner, M., Hohle, M., Ivarsson, S., Jark, U., Kirchner, M., Koch, J., Krause, G., Luber, P., Rosner, B., Stark, K., & Kuhne, M. 2011. German outbreak of Escherichia coli O104:H4 associated with sprouts. *N.Engl.J.Med*, 365, (19) 1763-1770 available from: PM:22029753

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., Sansonetti, P., & Parsot, C. 2000. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri. *Mol.Microbiol.*, 38, (4) 760-771 available from: PM:11115111

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., Pande, N., Shang, Z., Yu, N., & Gutell, R.R. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC.Bioinformatics.*, 3, 2 available from: PM:11869452

Case, R.J., Boucher, Y., Dahllof, I., Holmstrom, C., Doolittle, W.F., & Kjelleberg, S. 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl.Environ.Microbiol.*, 73, (1) 278-288 available from: PM:17071787

Castellani A, C.A. 1919. *Manual of tropical medicine*, 3rd ed. New York, William Wood.

Castellani, A. & Chalmers, AJ. 2005, "Family I. Enterobacteriaceae. Genus I. *Escherichia*," *In Bergy's Manual of Systematic Bacteriology*, 2nd ed. vol. 2 F. Scheutz & N. A. Strockbine, eds., USA: Springer, pp. 607-624.

Cennimo, D., Abbas, A., Huang, D.B., & Chiang, T. 2009. The prevalence and virulence characteristics of enteroaggregative Escherichia coli at an urgent-care clinic in the USA: a case-control study. *J.Med.Microbiol.*, 58, (Pt 4) 403-407 available from: PM:19273633

Chase-Topping, M.E., McKendrick, I.J., Pearce, M.C., MacDonald, P., Matthews, L., Halliday, J., Allison, L., Fenlon, D., Low, J.C., Gunn, G., & Woolhouse, M.E. 2007. Risk factors for the presence of high-level shedders of Escherichia coli O157 on Scottish farms. *J.Clin.Microbiol.*, 45, (5) 1594-1603 available from: PM:17360845

Chattaway, M., Harris, R., Jenkins, C., Tam, C., Coia, J., Gray, J., Iturriza-Gomara, M., & Wain, J. 2013. Investigating the link between the presence of enteroaggregative Escherichia coli and infectious intestinal disease in the United Kingdom, 1993 to 1996 and 2008 to 2009. *Euro.Surveill*, 18, (37) available from: PM:24079400

Chattaway, M.A., Dallman, T., Okeke, I.N., & Wain, J. 2011. Enteroaggregative E. coli O104 from an outbreak of HUS in Germany 2011, could it happen again? *J.Infect.Dev.Ctries.*, 5, (6) 425-436

Chattaway, M.A., Jenkins, C., Ciesielczuk, H., Day, M., DoNascimento, V., Day, M., Rodriguez, I., van Essen-Zandbergen, A., Schink, A.K., Wu, G., Threlfall, J., Woodward, M.J., Coldham, N., Kadlec, K., Schwarz, S., Dierikx, C., Guerra, B., Helmuth, R., Mevius, D., Woodford, N., & Wain, J. 2014a. Evidence of Evolving Extraintestinal Enteroaggregative Escherichia coli ST38 Clone. *Emerg.Infect.Dis.*, 20, (11) 1935-1937 available from: PM:25340736

Chattaway, M.A., Jenkins, C., Rajendram, D., Cravioto, A., Talukder, K.A., Dallman, T., Underwood, A., Platt, S., Okeke, I.N., & Wain, J. 2014b. Enteroaggregative Escherichia coli Have Evolved Independently as Distinct Complexes within the E. coli Population with Varying Ability to Cause Disease. *PLoS.One.*, 9, (11) e112967 available from: PM:25415318

Chaudhuri, R.R., Sebaihia, M., Hobman, J.L., Webber, M.A., Leyton, D.L., Goldberg, M.D., Cunningham, A.F., Scott-Tucker, A., Ferguson, P.R., Thomas, C.M., Frankel, G., Tang, C.M., Dudley, E.G., Roberts, I.S., Rasko, D.A., Pallen, M.J., Parkhill, J., Nataro, J.P., Thomson, N.R., & Henderson, I.R. 2010. Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative Escherichia coli strain 042. *PLoS.One.*, 5, (1) e8801 available from: PM:20098708

Cohan, F.M. 2002. What are bacterial species? *Annu.Rev.Microbiol.*, 56, 457-487 available from: PM:12142474

Cooper, J.E. & Feil, E.J. 2004. Multilocus sequence typing--what is resolved? *Trends Microbiol.*, 12, (8) 373-377 available from: PM:15276613

Cravioto, A. 1979. ADHESIVE FACTOR FOUND IN STRAINS OF ESCHERICHIA-COLI BELONGING TO THE TRADITIONAL INFANTILE ENTEROPATHOGENIC SEROTYPES. *CURRENT MICROBIOLOGY*, 3, (2) 95-99

Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M., & Finlay, B.B. 2013. Recent advances in understanding enteric pathogenic Escherichia coli. *Clin.Microbiol.Rev.*, 26, (4) 822-880 available from: PM:24092857

Czeczulin, J.R., Balepur, S., Hicks, S., Phillips, A., Hall, R., Kothary, M.H., Navarro-Garcia, F., & Nataro, J.P. 1997. Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative Escherichia coli. *Infect.Immun.*, 65, (10) 4135-4145 available from: PM:9317019

Czeczulin, J.R., Whittam, T.S., Henderson, I.R., Navarro-Garcia, F., & Nataro, J.P. 1999. Phylogenetic analysis of enteroaggregative and diffusely adherent Escherichia coli. *Infect.Immun.*, 67, (6) 2692-2699 available from: PM:10338471

Dallman, T., Smith, G.P., O'Brien, B., Chattaway, M.A., Finlay, D., Grant, K.A., & Jenkins, C. 2012. Characterization of a verocytotoxin-producing enteroaggregative Escherichia coli serogroup O111:H21 strain associated with a household outbreak in Northern Ireland. *J.Clin.Microbiol.*, 50, (12) 4116-4119 available from: PM:23035193

Dallman, T.J., Chattaway, M.A., Cowley, L.A., Doumith, M., Tewolde, R., Wooldridge, D.J., Underwood, A., Ready, D., Wain, J., Foster, K., Grant, K.A., & Jenkins, C. 2014. An Investigation of the Diversity of Strains of Enteroaggregative Escherichia coli Isolated from Cases Associated with a Large Multi-Pathogen Foodborne Outbreak in the UK. *PLoS.One.*, 9, (5) e98103 available from: PM:24844597

Denno, D.M., Shaikh, N., Stapp, J.R., Qin, X., Hutter, C.M., Hoffman, V., Mooney, J.C., Wood, K.M., Stevens, H.J., Jones, R., Tarr, P.I., & Klein, E.J. 2012. Diarrhea Etiology in a Pediatric Emergency Department: A Case Control Study. *Clin.Infect.Dis.* available from: PM:22700832

Didelot, X. & Falush, D. ClonalFrame User Guide Version 1.1. 1-33. 2006.
Ref Type: Generic

Didelot, X. & Falush, D. 2007. Inference of bacterial microevolution using multilocus sequence data. *Genetics*, 175, (3) 1251-1266 available from: PM:17151252

Doumith, M., Day, M.J., Hope, R., Wain, J., & Woodford, N. 2012. Improved multiplex PCR strategy for rapid assignment of the four major Escherichia coli phylogenetic groups. *J.Clin.Microbiol.*, 50, (9) 3108-3110 available from: PM:22785193

Dutta, S., Guin, S., Ghosh, S., Pazhani, G.P., Rajendran, K., Bhattacharya, M.K., Takeda, Y., Nair, G.B., & Ramamurthy, T. 2013. Trends in the prevalence of diarrheagenic Escherichia coli among hospitalized diarrheal patients in Kolkata, India. *PLoS.One.*, 8, (2) e56068 available from: PM:23457500

Edwards, P.R. & Ewing, W.H. 1962. *Identification of Enterobacteriaceae*, 2nd ed. Minneapolis, Burgess Publishing Co.

Escherich, T. Die Darmbakterien des Neugeborenen und Sauglings. Fortschr Med 3, 515-554. 1885.
Ref Type: Journal (Full)

Estrada-Garcia, T. & Navarro-Garcia, F. 2012. Enteroaggregative Escherichia coli pathotype: a genetically heterogeneous emerging foodborne enteropathogen. *FEMS Immunol.Med Microbiol.*, 66, (3) 281-298 available from: PM:22775224

EU Reference Laboratory for E.coli Department of Veterinary Public Health and Food Safety. Detection of enteroaggregative *E. coli* in food by real-time PCR amplication of *aagR* and *aaiC* genes. 1-4. 15-4-2013.
Ref Type: Generic

European Centre for Disease Prevention and Control and European Food Safety Authority 2011, *Shiga toxin/verotoxin-producing E. coli in humans, food and animals in the EU/EAA, with special reference to the German outbreak strain STEC)104.*, **European Centre for Disease Prevention and Control and European Food Safety Authority**, Stockhom, Sweden.

Farfan, M.J., Inman, K.G., & Nataro, J.P. 2008. The major pilin subunit of the AAF/II fimbriae from enteroaggregative Escherichia coli mediates binding to extracellular matrix proteins. *Infect.Immun.*, 76, (10) 4378-4384 available from: PM:18591223

Farmer, J. J., III, Brenner, D. J., & Ewing, W. H. Opposition to recent proposals which would reject the family name Enterobacteriaceae and Escherichia as its type genus. Int.J.Sys.Bacteriol 30[4], 660-673. 1980.
Ref Type: Journal (Full)

Fisher, R.A. 1922. On the interpretation of X^2 from contingency tables, and the calculation of P. *J.Roy.Stat.Soc*, 85, 87-94

Fluss, R., Faraggi, D., & Reiser, B. 2005. Estimation of the Youden Index and its associated cutoff point. *Biom.J.*, 47, (4) 458-472 available from: PM:16161804

Fuhrman, L.E., Shianna, K.V., & Aballay, A. 2008. High-throughput isolation and mapping of C. elegans mutants susceptible to pathogen infection. *PLoS.One.*, 3, (8) e2882 available from: PM:18682730

Galtier, N. 2007. A model of horizontal gene transfer and the bacterial phylogeny problem. *Syst.Biol*, 56, (4) 633-642 available from: PM:17661231

Gest, H. 2004. The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. *Notes Rec.R Soc Lond*, 58, (2) 187-201 available from: PM:15209075

Gomez-Duarte, O.G., Arzuza, O., Urbina, D., Bai, J., Guerra, J., Montes, O., Puello, M., Mendoza, K., & Castro, G.Y. 2010. Detection of Escherichia coli enteropathogens by multiplex polymerase chain reaction from children's diarrheal stools in two Caribbean-Colombian cities. *Foodborne.Pathog.Dis.*, 7, (2) 199-206 available from: PM:19839760

Goyal, A., Konar, M., Setia, A., Narang, A., & Ghosh, S. 2010. Galactose specific adhesin of enteroaggregative E. coli induces IL-8 secretion via activation of MAPK and STAT-3 in INT-407 cells. *Biochim.Biophys.Acta*, 1800, (6) 574-579 available from: PM:20304034

Gross, R. J. & Rowe, B. 1985, "Serotyping of Escherichia coli," *In The virulence of Escherichia coli*, M. Sussman, ed., Cambridge: CambridgeUniversity Press, pp. 345-360.

Grove, D.S. 1999. Quantitative real-time polymerase chain reaction for the core facility using TaqMan and the Perkin-Elmer/Applied Biosystems Division 7700 Sequence Detector. *J.Biomol.Tech.*, 10, (1) 11-16 available from: PM:19499001

Hartl, D.L. & Dykhuizen, D.E. 1984. The population genetics of Escherichia coli. *Annu.Rev.Genet.*, 18, 31-68 available from: PM:6099090

Hicks, S., Candy, D.C., & Phillips, A.D. 1996. Adhesion of enteroaggregative Escherichia coli to formalin-fixed intestinal and ureteric epithelia from children. *J.Med.Microbiol.*, 44, (5) 362-371 available from: PM:8636951

Huang, D.B., Mohamed, J.A., Nataro, J.P., DuPont, H.L., Jiang, Z.D., & Okhuysen, P.C. 2007. Virulence characteristics and the molecular epidemiology of enteroaggregative Escherichia coli isolates from travellers to developing countries. *J.Med.Microbiol.*, 56, (Pt 10) 1386-1392 available from: PM:17893178

Hwang, J., Mattei, L.M., VanArendonk, L.G., Meneely, P.M., & Okeke, I.N. 2010. A pathoadaptive deletion in an enteroaggregative Escherichia coli outbreak strain enhances virulence in a Caenorhabditis elegans model. *Infect.Immun.*, 78, (9) 4068-4076 available from: PM:20584976

Itoh, Y., Nagano, I., Kunishima, M., & Ezaki, T. 1997. Laboratory investigation of enteroaggregative Escherichia coli O untypeable:H10 associated with a massive outbreak of gastrointestinal illness. *J.Clin.Microbiol.*, 35, (10) 2546-2550 available from: PM:9316905

Iyoda, S., Tamura, K., Itoh, K., Izumiya, H., Ueno, N., Nagata, K., Togo, M., Terajima, J., & Watanabe, H. 2000. Inducible stx2 phages are lysogenized in the enteroaggregative and other phenotypic Escherichia coli O86:HNM isolated from patients. *FEMS Microbiol.Lett.*, 191, (1) 7-10 available from: PM:11004392

Jenkins, C., Chart, H., Willshaw, G.A., Cheasty, T., & Smith, H.R. 2006a. Genotyping of enteroaggregative Escherichia coli and identification of target genes for the detection of both typical and atypical strains. *Diagn.Microbiol.Infect.Dis.*, 55, (1) 13-19 available from: PM:16500068

Jenkins, C., Chart, H., Willshaw, G.A., Cheasty, T., & Tompkins, D.S. 2007. Association of putative pathogenicity genes with adherence characteristics and fimbrial genotypes in typical enteroaggregative Escherichia coli from patients with and without diarrhoea in the United Kingdom. *Eur.J.Clin.Microbiol.Infect.Dis.*, 26, (12) 901-906 available from: PM:17899229

Jenkins, C., Tembo, M., Chart, H., Cheasty, T., Willshaw, G.A., Phillips, A.D., Tompkins, D., & Smith, H. 2006b. Detection of enteroaggregative Escherichia coli in faecal samples from patients in the community with diarrhoea. *J.Med.Microbiol.*, 55, (Pt 11) 1493-1497 available from: PM:17030907

Jenkins, C., van Ijperen, C., Dudley, E.G., Chart, H., Willshaw, G.A., Cheasty, T., Smith, H.R., & Nataro, J.P. 2005. Use of a microarray to assess the distribution of plasmid and chromosomal virulence genes in strains of enteroaggregative Escherichia coli. *FEMS Microbiol.Lett.*, 253, (1) 119-124 available from: PM:16243450

Jenkins, C., Willshaw, G.A., Evans, J., Cheasty, T., Chart, H., Shaw, D.J., Dougan, G., Frankel, G., & Smith, H.R. 2003. Subtyping of virulence genes in verocytotoxin-producing Escherichia coli (VTEC) other than serogroup O157 associated with disease in the United Kingdom. *J.Med Microbiol.*, 52, (Pt 11) 941-947 available from: PM:14532337

Jobling, M.G. & Holmes, R.K. 2012. Type II heat-labile enterotoxins from 50 diverse Escherichia coli isolates belong almost exclusively to the LT-IIc family and may be prophage encoded. *PLoS.One.*, 7, (1) e29898 available from: PM:22242186

Judicial Commission of the International committee on Bacteriological Nomenclature. Conservation of the family name Enterobacteriaceae, of the name of the type genus, and designation of the type species. INTERNATIONAL BULLETIN OF BACTERIOLCGICAL NOMENCLATURE AND TAXONOMY 8[1], 73-74. 15-1-1958.
Ref Type: Generic

Kaper, J.B., Nataro, J.P., & Mobley, H.L. 2004. Pathogenic Escherichia coli. *Nat.Rev.Microbiol.*, 2, (2) 123-140 available from: PM:15040260

Kauffmann, F. 1947. The serology of the coli group. *J.Immunol.*, 57, 71-100

Khan, K., Konar, M., Goyal, A., & Ghosh, S. 2010. Enteroaggregative Escherichia coli infection induces IL-8 production via activation of mitogen-activated protein kinases and the transcription factors NF-kappaB and AP-1 in INT-407 cells. *Mol.Cell Biochem.*, 337, (1-2) 17-24 available from: PM:19898747

Klapproth, J.M., Scaletsky, I.C., McNamara, B.P., Lai, L.C., Malstrom, C., James, S.P., & Donnenberg, M.S. 2000. A large toxin from pathogenic Escherichia coli strains that inhibits lymphocyte activation. *Infect.Immun.*, 68, (4) 2148-2155 available from: PM:10722613

Kotloff, K.L., Blackwelder, W.C., Nasrin, D., Nataro, J.P., Farag, T.H., van, E.A., Adegbola, R.A., Alonso, P.L., Breiman, R.F., Faruque, A.S., Saha, D., Sow, S.O., Sur, D., Zaidi, A.K., Biswas, K., Panchalingam, S., Clemens, J.D., Cohen, D., Glass, R.I., Mintz, E.D., Sommerfelt, H., & Levine, M.M. 2012. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. *Clin.Infect.Dis.*, 55 Suppl 4, S232-S245 available from: PM:23169936

Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalingam, S., Wu, Y., Sow, S.O., Sur, D., Breiman, R.F., Faruque, A.S., Zaidi, A.K., Saha, D., Alonso, P.L., Tamboura, B., Sanogo, D., Onwuchekwa, U., Manna, B., Ramamurthy, T., Kanungo, S., Ochieng, J.B., Omore, R., Oundo, J.O., Hossain, A., Das, S.K., Ahmed, S., Qureshi, S., Quadri, F., Adegbola, R.A., Antonio, M., Hossain, M.J., Akinsola, A., Mandomando, I., Nhampossa, T., Acacio, S., Biswas, K., O'Reilly, C.E., Mintz, E.D., Berkeley, L.Y., Muhsen, K., Sommerfelt, H., Robins-Browne, R.M., & Levine, M.M. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet*, 382, (9888) 209-222 available from: PM:23680352

Kreisberg, R.B., Harper, J., Strauman, M.C., Marohn, M., Clements, J.D., & Nataro, J.P. 2011. Induction of increased permeability of polarized enterocyte monolayers by enterotoxigenic Escherichia coli heat-labile enterotoxin. *Am.J.Trop.Med Hyg.*, 84, (3) 451-455 available from: PM:21363985

Le, B.C. & Servin, A.L. 2006. Diffusely adherent Escherichia coli strains expressing Afa/Dr adhesins (Afa/Dr DAEC): hitherto unrecognized pathogens. *FEMS Microbiol.Lett.*, 256, (2) 185-194 available from: PM:16499605

Lesage, A. A. Contribution a l'etude des enterites infantiles - sero-diagnostic des races de Bacterium coli. C R Soc Biol 49, 900-9001. 1897.
Ref Type: Journal (Full)

Levine, M.M. 1978. Escherichia coli strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. *Lancet*, 1, (8074) 1119-1122

Lewis, J. & Fleming, J. 1995, "Basic culture methods," *In Caenorhabditis elegans:a modern biological analysis of an organism*, vol. 48 H. Epstein & H. Shakes, eds., San Diego, CA: Academic press, pp. 4-30.

Lewontin, R.C. 1974. *The genetic basis of evolutionary change* New York, Columbia University Press.

Lima, I.F., Boisen, N., Quetz, J.S., Havt, A., de Carvalho, E.B., Soares, A.M., Lima, N.L., Mota, R.M., Nataro, J.P., Guerrant, R.L., & Lima, A.A. 2013. Prevalence of enteroaggregative Escherichia coli and its virulence-related genes in a case-control study among children from north-eastern Brazil. *J.Med Microbiol.*, 62, (Pt 5) 683-693 available from: PM:23429698

Lindsay, B., Ramamurthy, T., Sen, G.S., Takeda, Y., Rajendran, K., Nair, G.B., & Stine, O.C. 2011. Diarrheagenic pathogens in polymicrobial infections. *Emerg.Infect.Dis.*, 17, (4) 606-611 available from: PM:21470448

Lothigius, A., Sjoling, A., Svennerholm, A.M., & Bolin, I. 2010. Survival and gene expression of enterotoxigenic Escherichia coli during long-term incubation in sea water and freshwater. *J.Appl.Microbiol.*, 108, (4) 1441-1449 available from: PM:19804537

Lukjancenko, O., Wassenaar, T.M., & Ussery, D.W. 2010. Comparison of 61 sequenced Escherichia coli genomes. *Microb.Ecol.*, 60, (4) 708-720 available from: PM:20623278

Mahajan-Miklos, S., Tan, M.W., Rahme, L.G., & Ausubel, F.M. 1999. Molecular mechanisms of bacterial virulence elucidated using a Pseudomonas aeruginosa-Caenorhabditis elegans pathogenesis model. *Cell*, 96, (1) 47-56 available from: PM:9989496

Maurelli, A.T., Fernandez, R.E., Bloch, C.A., Rode, C.K., & Fasano, A. 1998. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli. *Proc.Natl.Acad.Sci.U.S.A*, 95, (7) 3943-3948 available from: PM:9520472

Mayer, H.B. & Wanke, C.A. 1995. Enteroaggregative Escherichia coli as a possible cause of diarrhea in an HIV-infected patient. *N.Engl.J.Med.*, 332, (4) 273-274 available from: PM:7808505

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S., & Kaper, J.B. 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc.Natl.Acad.Sci.U.S.A*, 92, (5) 1664-1668 available from: PM:7878036

Medina, A.M., Rivera, F.P., Romero, L.M., Kolevic, L.A., Castillo, M.E., Verne, E., Hernandez, R., Mayor, Y.E., Barletta, F., Mercado, E., & Ochoa, T.J. 2010. Diarrheagenic Escherichia coli in human immunodeficiency virus (HIV) pediatric patients in Lima, Peru. *Am.J.Trop.Med Hyg.*, 83, (1) 158-163 available from: PM:20595495

Mellies, J.L., Barron, A.M., Haack, K.R., Korson, A.S., & Oldridge, D.A. 2006. The global regulator Ler is necessary for enteropathogenic Escherichia coli colonization of Caenorhabditis elegans. *Infect.Immun.*, 74, (1) 64-72 available from: PM:16368958

Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S.F., Henkhaus, J.K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P.M., Moore, R.L., Guenther, S., Rothberg, J.M., & Karch, H. 2011. Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. *PLoS.One.*, 6, (7) e22751 available from: PM:21799941

Moreira, C.G., Carneiro, S.M., Nataro, J.P., Trabulsi, L.R., & Elias, W.P. 2003. Role of type I fimbriae in the aggregative adhesion pattern of enteroaggregative Escherichia coli. *FEMS Microbiol.Lett.*, 226, (1) 79-85 available from: PM:13129611

Mylonakis, E. & Aballay, A. 2005. Worms and flies as genetically tractable animal models to study host-pathogen interactions. *Infect.Immun.*, 73, (7) 3833-3841 available from: PM:15972468

Nadkarni, M.A., Martin, F.E., Jacques, N.A., & Hunter, N. 2002. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology*, 148, (Pt 1) 257-266 available from: PM:11782518

Nataro, J.P., Baldini, M.M., Kaper, J.B., Black, R.E., Bravo, N., & Levine, M.M. 1985. Detection of an adherence factor of enteropathogenic Escherichia coli with a DNA probe. *J.Infect.Dis.*, 152, (3) 560-565 available from: PM:2863319

Nataro, J.P., Deng, Y., Cookson, S., Cravioto, A., Savarino, S.J., Guers, L.D., Levine, M.M., & Tacket, C.O. 1995. Heterogeneity of enteroaggregative Escherichia coli virulence demonstrated in volunteers. *J.Infect.Dis.*, 171, (2) 465-468 available from: PM:7844392

Nataro, J.P., Hicks, S., Phillips, A.D., Vial, P.A., & Sears, C.L. 1996. T84 cells in culture as a model for enteroaggregative Escherichia coli pathogenesis. *Infect.Immun.*, 64, (11) 4761-4768 available from: PM:8890237

Nataro, J.P. & Kaper, J.B. 1998. Diarrheagenic Escherichia coli. *Clin.Microbiol.Rev.*, 11, (1) 142-201 available from: PM:9457432

Nataro, J.P., Kaper, J.B., Robins-Browne, R., Prado, V., Vial, P., & Levine, M.M. 1987. Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. *Pediatr.Infect.Dis.J.*, 6, (9) 829-831 available from: PM:3313248

Nataro, J.P., Mai, V., Johnson, J., Blackwelder, W.C., Heimer, R., Tirrell, S., Edberg, S.C., Braden, C.R., Glenn, M.J., Jr., & Hirshon, J.M. 2006. Diarrheagenic Escherichia coli infection in Baltimore, Maryland, and New Haven, Connecticut. *Clin.Infect.Dis.*, 43, (4) 402-407 available from: PM:16838226

Nataro, J.P., Yikang, D., Giron, J.A., Savarino, S.J., Kothary, M.H., & Hall, R. 1993. Aggregative adherence fimbria I expression in enteroaggregative Escherichia coli requires two unlinked plasmid regions. *Infect.Immun.*, 61, (3) 1126-1131 available from: PM:8094379

Nataro, J.P., Yikang, D., Yingkang, D., & Walker, K. 1994. AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative Escherichia coli. *J.Bacteriol.*, 176, (15) 4691-4699 available from: PM:7913930

Navarro-Garcia, F., Sears, C., Eslava, C., Cravioto, A., & Nataro, J.P. 1999. Cytoskeletal effects induced by pet, the serine protease enterotoxin of enteroaggregative Escherichia coli. *Infect.Immun.*, 67, (5) 2184-2192 available from: PM:10225873

Nei, M. 1975. *Molecular polymorphism and evolution* Amsterdam, North Holland Publishing Co.

Nishi, J., Sheikh, J., Mizuguchi, K., Luisi, B., Burland, V., Boutin, A., Rose, D.J., Blattner, F.R., & Nataro, J.P. 2003. The export of coat protein from enteroaggregative Escherichia coli by a specific ATP-binding cassette transporter system. *J.Biol.Chem.*, 278, (46) 45680-45689 available from: PM:12933818

Noriega, F.R., Liao, F.M., Formal, S.B., Fasano, A., & Levine, M.M. 1995. Prevalence of Shigella enterotoxin 1 among Shigella clinical isolates of diverse serotypes. *J.Infect.Dis.*, 172, (5) 1408-1410 available from: PM:7594690

O'Brien, S.J., Rait, G., Hunter, P.R., Gray, J.J., Bolton, F.J., Tompkins, D.S., McLauchlin, J., Letley, L.H., Adak, G.K., Cowden, J.M., Evans, M.R., Neal, K.R., Smith, G.E., Smyth, B., Tam, C.C., & Rodrigues, L.C. 2010. Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). *BMC.Med.Res.Methodol.*, 10, 39 available from: PM:20444246

Ochman, H. & Selander, R.K. 1984. Evidence for clonal population structure in Escherichia coli. *Proc.Natl.Acad.Sci.U.S.A*, 81, (1) 198-201 available from: PM:6364134

Ochoa, T.J., Ecker, L., Barletta, F., Mispireta, M.L., Gil, A.I., Contreras, C., Molina, M., Amemiya, I., Verastegui, H., Hall, E.R., Cleary, T.G., & Lanata, C.F. 2009. Age-related susceptibility to infection with diarrheagenic Escherichia coli among infants from Periurban areas in Lima, Peru. *Clin.Infect.Dis.*, 49, (11) 1694-1702 available from: PM:19857163

Okeke, I.N., Lamikanra, A., Czeczulin, J., Dubovsky, F., Kaper, J.B., & Nataro, J.P. 2000a. Heterogeneous virulence of enteroaggregative Escherichia coli strains isolated from children in Southwest Nigeria. *J.Infect.Dis.*, 181, (1) 252-260 available from: PM:10608774

Okeke, I.N., Lamikanra, A., Steinruck, H., & Kaper, J.B. 2000b. Characterization of Escherichia coli strains from cases of childhood diarrhea in provincial southwestern Nigeria. *J.Clin.Microbiol.*, 38, (1) 7-12 available from: PM:10618054

Okeke, I.N. & Nataro, J.P. 2001. Enteroaggregative Escherichia coli. *Lancet Infect.Dis.*, 1, (5) 304-313 available from: PM:11871803

Okeke, I.N., Wallace-Gadsden, F., Simons, H.R., Matthews, N., Labar, A.S., Hwang, J., & Wain, J. 2010. Multi-locus sequence typing of enteroaggregative Escherichia coli isolates from Nigerian children uncovers multiple lineages. *PLoS.One.*, 5, (11) e14093 available from: PM:21124856

Okhuysen, P.C. & DuPont, H.L. 2010. Enteroaggregative Escherichia coli (EAEC): a cause of acute and persistent diarrhea of worldwide importance. *J.Infect.Dis.*, 202, (4) 503-505 available from: PM:20594108

Panchalingam, S., Antonio, M., Hossain, A., Mandomando, I., Ochieng, B., Oundo, J., Ramamurthy, T., Tamboura, B., Zaidi, A.K., Petri, W., Houpt, E., Murray, P., Prado, V., Vidal, R., Steele, D., Strockbine, N., Sansonetti, P., Glass, R.I., Robins-Browne, R.M., Tauschek, M., Svennerholm, A.M., Kotloff, K., Levine, M.M., & Nataro, J.P. 2012. Diagnostic microbiologic methods in the GEMS-1 case/control study. *Clin.Infect.Dis.*, 55 Suppl 4, S294-S302 available from: PM:23169941

Parija, S.C. 2009. *Textbook of Microbiology and Immunology* India, Elsevier.

Partridge, D.G., Evans, C.M., Raza, M., Kudesia, G., & Parsons, H.K. 2012. Lessons from a large norovirus outbreak: impact of viral load, patient age and ward design on duration of symptoms and shedding and likelihood of transmission. *J.Hosp.Infect.*, 81, (1) 25-30 available from: PM:22445090

Pereira, A.L., Ferraz, L.R., Silva, R.S., & Giugliano, L.G. 2007. Enteroaggregative Escherichia coli virulence markers: positive association with distinct clinical characteristics and segregation into 3 enteropathogenic E. coli serogroups. *J.Infect.Dis.*, 195, (3) 366-374 available from: PM:17205475

Perez-Losada, M., Browne, E.B., Madsen, A., Wirth, T., Viscidi, R.P., & Crandall, K.A. 2006. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect.Genet.Evol.*, 6, (2) 97-112 available from: PM:16503511

Phillips, G., Lopman, B., Tam, C.C., Iturriza-Gomara, M., Brown, D., & Gray, J. 2009. Diagnosing norovirus-associated infectious intestinal disease using viral load. *BMC.Infect.Dis.*, 9, 63 available from: PM:19442278

Phillips, G., Tam, C.C., Conti, S., Rodrigues, L.C., Brown, D., Iturriza-Gomara, M., Gray, J., & Lopman, B. 2010. Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. *Am.J.Epidemiol.*, 171, (9) 1014-1022 available from: PM:20360244

Piva, I.C., Pereira, A.L., Ferraz, L.R., Silva, R.S., Vieira, A.C., Blanco, J.E., Blanco, M., Blanco, J., & Giugliano, L.G. 2003. Virulence markers of enteroaggregative Escherichia coli isolated from children and adults with diarrhea in Brasilia, Brazil. *J.Clin.Microbiol.*, 41, (5) 1827-1832 available from: PM:12734212

Rahn, O. 1937. New principles of the classification of bacteria. *Zentrabl.Bakteriol.Parasitenkd.Infektionskr.Hyg.*, 2, (96) 273-286

Regua-Mangia, A.H., Gomes, T.A., Vieira, M.A., Irino, K., & Teixeira, L.M. 2009. Molecular typing and virulence of enteroaggregative Escherichia coli strains isolated from children with and without diarrhoea in Rio de Janeiro city, Brazil. *J.Med.Microbiol.*, 58, (Pt 4) 414-422 available from: PM:19273635

Robins-Browne, R.M. 1987. Traditional enteropathogenic Escherichia coli of infantile diarrhea. *Rev.Infect.Dis.*, 9, (1) 28-53 available from: PM:3547577

Roche, J.K., Cabel, A., Sevilleja, J., Nataro, J., & Guerrant, R.L. 2010. Enteroaggregative Escherichia coli (EAEC) impairs growth while malnutrition worsens EAEC infection: a novel murine model of the infection malnutrition cycle. *J.Infect.Dis.*, 202, (4) 506-514 available from: PM:20594107

Rockhill, B., Newman, B., & Weinberg, C. 1998. Use and misuse of population attributable fractions. *Am.J.Public Health*, 88, (1) 15-19 available from: PM:9584027

Rubino, S., Cappuccinelli, P., & Kelvin, D.J. 2011. Escherichia coli (STEC) serotype O104 outbreak causing haemolytic syndrome (HUS) in Germany and France. *J.Infect.Dev.Ctries.*, 5, (6) 437-440 available from: PM:21727641

Savarino, S.J., Fasano, A., Watson, J., Martin, B.M., Levine, M.M., Guandalini, S., & Guerry, P. 1993. Enteroaggregative Escherichia coli heat-stable enterotoxin 1 represents another subfamily of E. coli heat-stable toxin. *Proc.Natl.Acad.Sci.U.S.A*, 90, (7) 3093-3097 available from: PM:8385356

Scaletsky, I.C., Silva, M.L., & Trabulsi, L.R. 1984. Distinctive patterns of adherence of enteropathogenic Escherichia coli to HeLa cells. *Infect.Immun.*, 45, (2) 534-536 available from: PM:6146569

Schroeder, G.N. & Hilbi, H. 2008. Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin.Microbiol.Rev.*, 21, (1) 134-156 available from: PM:18202440

Schubert, S., Rakin, A., Karch, H., Carniel, E., & Heesemann, J. 1998. Prevalence of the "high-pathogenicity island" of Yersinia species among Escherichia coli strains that are pathogenic to humans. *Infect.Immun.*, 66, (2) 480-485 available from: PM:9453599

Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N., & Whittam, T.S. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl.Environ.Microbiol.*, 51, (5) 873-884 available from: PM:2425735

Selander, R.K. & Levin, B.R. 1980. Genetic diversity and structure in Escherichia coli populations. *Science*, 210, (4469) 545-547 available from: PM:6999623

Serna, A. & Boedeker, E.C. 2008. Pathogenesis and treatment of Shiga toxin-producing Escherichia coli infections. *Curr.Opin.Gastroenterol.*, 24, (1) 38-47 available from: PM:18043231

Servin, A.L. 2005. Pathogenesis of Afa/Dr diffusely adhering Escherichia coli. *Clin.Microbiol.Rev.*, 18, (2) 264-292 available from: PM:15831825

Shah, N., DuPont, H.L., & Ramsey, D.J. 2009. Global etiology of travelers' diarrhea: systematic review from 1973 to the present. *Am.J.Trop.Med Hyg.*, 80, (4) 609-614 available from: PM:19346386

Shariff, M., Bhan, M.K., Knutton, S., Das, B.K., Saini, S., & Kumar, R. 1993. Evaluation of the fluorescence actin staining test for detection of enteropathogenic Escherichia coli. *J.Clin.Microbiol.*, 31, (2) 386-389 available from: PM:7679403

Sheikh, J., Czeczulin, J.R., Harrington, S., Hicks, S., Henderson, I.R., Le, B.C., Gounon, P., Phillips, A., & Nataro, J.P. 2002. A novel dispersin protein in enteroaggregative Escherichia coli. *J.Clin.Invest*, 110, (9) 1329-1337 available from: PM:12417572

Shulman, S.T., Friedmann, H.C., & Sims, R.H. 2007. Theodor Escherich: the first pediatric infectious diseases physician? *Clin.Infect.Dis.*, 45, (8) 1025-1029 available from: PM:17879920

Sokurenko, E.V., Hasty, D.L., & Dykhuizen, D.E. 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol.*, 7, (5) 191-195 available from: PM:10354593

Spencer, J., Smith, H.R., & Chart, H. 1999. Characterization of enteroaggregative Escherichia coli isolated from outbreaks of diarrhoeal disease in England. *Epidemiol.Infect.*, 123, (3) 413-421 available from: PM:10694151

Stackebrandt, E. & Goebel, B. M. Taxonomic Note: A Place for DNA-DNA Reassociation and 16s rRNA Sequence Analysis in the Present Species Definition in Bacteriology. Int.J.Syst.Bacteriol. 846-849. 1994.
Ref Type: Journal (Full)

StataCorp. Stata Statistical Software. [Release 12]. 2011. College Station, TX, StataCorp LP.
Ref Type: Computer Program

Steiner, T.S., Lima, A.A., Nataro, J.P., & Guerrant, R.L. 1998. Enteroaggregative Escherichia coli produce intestinal inflammation and growth impairment and cause interleukin-8 release from intestinal epithelial cells. *J.Infect.Dis.*, 177, (1) 88-96 available from: PM:9419174

Steiner, T.S., Nataro, J.P., Poteet-Smith, C.E., Smith, J.A., & Guerrant, R.L. 2000. Enteroaggregative Escherichia coli expresses a novel flagellin that causes IL-8 release from intestinal epithelial cells. *J.Clin.Invest*, 105, (12) 1769-1777 available from: PM:10862792

Strauman, M.C., Harper, J.M., Harrington, S.M., Boll, E.J., & Nataro, J.P. 2010. Enteroaggregative Escherichia coli disrupts epithelial cell tight junctions. *Infect.Immun.*, 78, (11) 4958-4964 available from: PM:20823198

Tam, C.C., O'Brien, S.J., Tompkins, D.S., Bolton, F.J., Berry, L., Dodds, J., Choudhury, D., Halstead, F., Iturriza-Gomara, M., Mather, K., Rait, G., Ridge, A., Rodrigues, L.C., Wain, J., Wood, B., & Gray, J.J. 2012a. Changes in Causes of Acute Gastroenteritis in the United Kingdom Over 15 Years: Microbiologic Findings From 2 Prospective, Population-Based Studies of Infectious Intestinal Disease. *Clin.Infect.Dis.*, 54, (9) 1275-1286 available from: PM:22412058

Tam, C.C., Rodrigues, L.C., Viviani, L., Dodds, J.P., Evans, M.R., Hunter, P.R., Gray, J.J., Letley, L.H., Rait, G., Tompkins, D.S., & O'Brien, S.J. 2012b. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut*, 61, (1) 69-77 available from: PM:21708822

Tan, M.W. & Ausubel, F.M. 2000. Caenorhabditis elegans: a model genetic host to study Pseudomonas aeruginosa pathogenesis. *Curr.Opin.Microbiol.*, 3, (1) 29-34 available from: PM:10679415

Tan, M.W., Mahajan-Miklos, S., & Ausubel, F.M. 1999. Killing of Caenorhabditis elegans by Pseudomonas aeruginosa used to model mammalian bacterial pathogenesis. *Proc.Natl.Acad.Sci.U.S.A*, 96, (2) 715-720 available from: PM:9892699

Taniuchi, M., Sobuz, S.U., Begum, S., Platts-Mills, J.A., Liu, J., Yang, Z., Wang, X.Q., Petri, W.A., Jr., Haque, R., & Houpt, E.R. 2013. Etiology of diarrhea in Bangladeshi infants in the first year of life analyzed using molecular methods. *J.Infect.Dis.*, 208, (11) 1794-1802 available from: PM:24041797

Tenaillon, O., Skurnik, D., Picard, B., & Denamur, E. 2010. The population genetics of commensal Escherichia coli. *Nat.Rev.Microbiol.*, 8, (3) 207-217 available from: PM:20157339

Tenor, J.L., McCormick, B.A., Ausubel, F.M., & Aballay, A. 2004. Caenorhabditis elegans-based screen identifies Salmonella virulence factors required for conserved host-pathogen interactions. *Curr.Biol*, 14, (11) 1018-1024 available from: PM:15182677

Toma, C., Lu, Y., Higa, N., Nakasone, N., Chinen, I., Baschkier, A., Rivas, M., & Iwanaga, M. 2003. Multiplex PCR assay for identification of human diarrheagenic Escherichia coli. *J.Clin.Microbiol.*, 41, (6) 2669-2671 available from: PM:12791900

Tompkins, D.S., Hudson, M.J., Smith, H.R., Eglin, R.P., Wheeler, J.G., Brett, M.M., Owen, R.J., Brazier, J.S., Cumberland, P., King, V., & Cook, P.E. 1999. A study of infectious intestinal disease in England: microbiological findings in cases and controls. *Commun.Dis.Public Health*, 2, (2) 108-113 available from: PM:10402744

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M.E., Frapy, E., Garry, L., Ghigo, J.M., Gilles, A.M., Johnson, J., Le, B.C., Lescat, M., Mangenot, S., Martinez-Jehanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M.A., Pichon, C., Rouy, Z., Ruf, C.S., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Medigue, C., Rocha, E.P., & Denamur, E. 2009. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS.Genet.*, 5, (1) e1000344 available from: PM:19165319

Trabulsi, L.R., Keller, R., & Tardelli Gomes, T.A. 2002. Typical and atypical enteropathogenic Escherichia coli. *Emerg.Infect.Dis.*, 8, (5) 508-513 available from: PM:11996687

Turner, K.M. & Feil, E.J. 2007. The secret life of the multilocus sequence type. *Int.J.Antimicrob.Agents*, 29, (2) 129-135 available from: PM:17204401

Tzipori, S., Montanaro, J., Robins-Browne, R.M., Vial, P., Gibson, R., & Levine, M.M. 1992. Studies with enteroaggregative Escherichia coli in the gnotobiotic piglet gastroenteritis model. *Infect.Immun.*, 60, (12) 5302-5306 available from: PM:1452364

Uber, A.P., Trabulsi, L.R., Irino, K., Beutin, L., Ghilardi, A.C., Gomes, T.A., Liberatore, A.M., de Castro, A.F., & Elias, W.P. 2006. Enteroaggregative Escherichia coli from humans and animals differ in major phenotypical traits and virulence genes. *FEMS Microbiol.Lett.*, 256, (2) 251-257 available from: PM:16499614

Usein, C.R., Tatu-Chitoiu, D., Ciontea, S., Condei, M., & Damian, M. 2009. Escherichia coli pathotypes associated with diarrhea in Romanian children younger than 5 years of age. *Jpn.J.Infect.Dis.*, 62, (4) 289-293 available from: PM:19628907

Vial, P.A., Mathewson, J.J., DuPont, H.L., Guers, L., & Levine, M.M. 1990. Comparison of two assay methods for patterns of adherence to HEp-2 cells of Escherichia coli from patients with diarrhea. *J.Clin.Microbiol.*, 28, (5) 882-885 available from: PM:2191011

Vial, P.A., Robins-Browne, R., Lior, H., Prado, V., Kaper, J.B., Nataro, J.P., Maneval, D., Elsayed, A., & Levine, M.M. 1988. Characterization of enteroadherent-aggregative Escherichia coli, a putative agent of diarrheal disease. *J.Infect.Dis.*, 158, (1) 70-79 available from: PM:2899125

Vos, M. & Didelot, X. 2009. A comparison of homologous recombination rates in bacteria and archaea. *ISME.J.*, 3, (2) 199-208 available from: PM:18830278

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrant, E., & Starr, M. P. T. H. G. Report of the Ad Hoc Committee on Reconciliation of Apporaches to Bacterial Systematics. Int.J.Syst.Bacteriol. [37], 463-464. 1987.
Ref Type: Journal (Full)

Weiglmeier, P.R., Rosch, P., & Berkner, H. 2010. Cure and curse: E. coli heat-stable enterotoxin and its receptor guanylyl cyclase C. *Toxins.(Basel)*, 2, (9) 2213-2229 available from: PM:22069681

WHO Media Centre 2009, *Diarrhoeal disease* Fact Sheet 330.

Wiles, T.J., Kulesus, R.R., & Mulvey, M.A. 2008. Origins and virulence mechanisms of uropathogenic Escherichia coli. *Exp.Mol.Pathol.*, 85, (1) 11-19 available from: PM:18482721

Wilson, A., Evans, J., Chart, H., Cheasty, T., Wheeler, J.G., Tompkins, D., & Smith, H.R. 2001. Characterisation of strains of enteroaggregative Escherichia coli isolated during the infectious intestinal disease study in England. *Eur.J.Epidemiol.*, 17, (12) 1125-1130 available from: PM:12530772

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden, M.C., Ochman, H., & Achtman, M. 2006. Sex and virulence in Escherichia coli: an evolutionary perspective. *Mol.Microbiol.*, 60, (5) 1136-1151 available from: PM:16689791

Woese, C.R. & Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc.Natl.Acad.Sci.U.S.A*, 74, (11) 5088-5090 available from: PM:270744

Wolf, M.K. 1997. Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic Escherichia coli. *Clin.Microbiol.Rev.*, 10, (4) 569-584 available from: PM:9336662

Wu, G., Day, M.J., Mafura, M.T., Nunez-Garcia, J., Fenner, J.J., Sharma, M., van Essen-Zandbergen, A., Rodriguez, I., Dierikx, C., Kadlec, K., Schink, A.K., Wain, J., Helmuth, R., Guerra, B., Schwarz, S., Threlfall, J., Woodward, M.J., Woodford, N., Coldham, N., & Mevius, D. 2013. Comparative Analysis of ESBL-Positive Escherichia coli Isolates from Animals and Humans from the UK, The Netherlands and Germany. *PLoS.One.*, 8, (9) e75392 available from: PM:24086522

Chapter Seven

Appendix

Chapter 7 - Appendix

7.1 Characteristics of 564 EAEC strains, isolated from human faeces, used in this study

Isolate	MOLIS No.	Country	Source	Pathogen	Year	Somatic	Flagella	ST	ST Complex	Allelic Profile
3026	H103320332	Bangladesh	GEMS Study	Control	2007-2011	O?	27	TLV	TLV	6 11 4 8 9 2 7
1037	H103320336	Bangladesh	GEMS Study	Case	2007-2011	159	23	1490	None	9 6 204 131 24 8 7
1038	H103320334	Bangladesh	GEMS Study	Case	2007-2011	25	H-	SLV	None	6 11 4 8 0 78 2
3036	H103320339	Bangladesh	GEMS Study	Control	2007-2011	113	H-	DLV	None	0 11 4 8 67 8 2
7064	H102400217	Bangladesh	GEMS Study	Control	2007-2011	153	36	157	None	33 57 58 63 37 5 4
7067	H102340354	Bangladesh	GEMS Study	Control	2007-2011	86	27	2166	None	6 29 5 16 9 8 44
7092	H102340349	Bangladesh	GEMS Study	Control	2007-2011	125ac	8	1295	None	6 4 5 9 9 8 2
7121	H102340363	Bangladesh	GEMS Study	Control	2007-2011	77	18	DLV	None	35 35 12 52 5 0 4
900694	H103820323	Bangladesh	GEMS Study	Case	2007-2011	O?	24	219	None	58 53 53 58 24 1 42
900693	H103820322	Bangladesh	GEMS Study	Case	2007-2011	44	H-	720	None	35 3 58 6 5 16 4
900553	H103800321	Bangladesh	GEMS Study	Case	2007-2011	51	12	DLV	None	8 0 1 8 24 1 42
900512	H103780407	Bangladesh	GEMS Study	Case	2007-2011	O?	23	157	None	33 57 58 63 37 5 4
900063	H103760530	Bangladesh	GEMS Study	Case	2007-2011	O?	23	720	None	35 3 58 6 5 16 4
900851	H104080228	Bangladesh	GEMS Study	Case	2007-2011	125ac	11	SLV	None	76 15 13 15 30 14 2
900794	H104080226	Bangladesh	GEMS Study	Case	2007-2011	O?	10	SLV	None	6 11 4 8 24 8 14
900753	H104120574	Bangladesh	GEMS Study	Case	2007-2011	166	15	SLV	None	34 36 207 87 67 8 4
900732	H113860291	Bangladesh	GEMS Study	Case	2007-2011	O?	7	1891	None	6 4 3 26 7 8 6
900550	H104080224	Bangladesh	GEMS Study	Case	2007-2011	128	12	1326	None	6 220 3 26 9 7 7
900442	H103840253	Bangladesh	GEMS Study	Case	2007-2011	121	H?	SLV	None	8 0 1 8 7 18 6
900245	H104080225	Bangladesh	GEMS Study	Case	2007-2011	161	H-	TLV	None	0 0 4 8 0 8 2
900770	H104120576	Bangladesh	GEMS Study	Case	2007-2011	69	4	678	None	6 6 5 136 9 7 7
900769	H104120575	Bangladesh	GEMS Study	Case	2007-2011	69	4	1891	None	6 4 3 26 7 8 6
600961	H104400252	Bangladesh	GEMS Study	Case	2007-2011	69	38	3107	None	10 11 5 8 7 219 2
600965	H104400253	Bangladesh	GEMS Study	Case	2007-2011	126	7	1891	None	6 4 3 26 7 8 2
600982	H104180063	Bangladesh	GEMS Study	Control	2007-2011	166	15	DLV	None	34 36 2 25 5 16 4
601062	H104560240	Bangladesh	GEMS Study	Case	2007-2011	2	42	SLV	None	92 4 87 96 70 58 11
601087	H104580566 & H113860293	Bangladesh	GEMS Study	Case	2007-2011	O?	45	3051	None	225 274 307 248 24 2 64
601098	H104320306	Bangladesh	GEMS Study	Control	2007-2011	O?	45	SLV	None	52 54 46 48 0 40 38
601101	H104500402 & H113860294	Bangladesh	GEMS Study	Case	2007-2011	O?	23	SLV	None	6 4 5 16 11 8 2
601106	H104320307	Bangladesh	GEMS Study	Control	2007-2011	O?	23	2186	None	6 4 5 16 11 8 6
601155	H113860295	Bangladesh	GEMS Study	Control	2007-2011	O?	30	678	None	6 6 5 136 9 7 7
601158	H104500404	Bangladesh	GEMS Study	Case	2007-2011	2	H-	3738	None	92 4 87 96 70 58 7
601174	H113860297	Bangladesh	GEMS Study	Control	2007-2011	78	7	1891	None	6 4 3 26 7 8 6

601175	H104500405 & H113860298	Bangladesh	GEMS Study	Case	2007-2011	8	9	720	None	35 3 58 6 5 16 4
601176	H104500406	Bangladesh	GEMS Study	Case	2007-2011	O?	H-	720	None	35 3 58 6 5 16 4
601180	H104380396	Bangladesh	GEMS Study	Control	2007-2011	O?	23	2186	None	6 4 5 16 11 8 6
900108	H104060670	Bangladesh	GEMS Study	Case	2007-2011	21	8	1295	None	6 4 5 9 9 8 2
H103060197	H103060197	UK	GBRU	Case	2010	O?	H?	SLV	None	76 43 9 36 0 14 10
H093880675	H093880675	UK	GBRU	Case	2009	O?	H?	504	None	76 43 9 36 17 14 10
900883	H104120572	Bangladesh	GEMS Study	Case	2007-2011	44	23	720	None	35 3 58 6 5 16 4
2506	H101200081	UK	IID2	Case	2008	153	2	120	None	49 4 44 9 11 35 7
G146	-	Nigeria	Okeke Study	Control	1995	-	-	483	None	56 11 4 110 8 83 83
G125	-	Nigeria	Okeke Study	Control	1995	-	-	499	None	64 7 12 8 78 18 6
G150a	-	Nigeria	Okeke Study	Control	1995	-	-	499	None	64 7 12 8 78 18 6
G115	-	Nigeria	Okeke Study	Control	1995	-	-	513	None	56 133 12 1 9 5 7
D05	-	Nigeria	Okeke Study	Case	1995	-	-	489	None	56 4 15 112 9 2 7
G112a	-	Nigeria	Okeke Study	Control	1995	-	-	511	None	108 4 33 114 20 12 7
G80a	-	Nigeria	Okeke Study	Case	1995	-	-	940	None	6 6 22 16 11 1 7
G05b	-	Nigeria	Okeke Study	Case	1995	-	-	455	None	56 6 22 16 11 1 7
C66	-	Nigeria	Okeke Study	Control	1995	-	-	455	None	56 6 22 16 11 1 7
G74	-	Nigeria	Okeke Study	Case	1995	-	-	424	None	6 30 32 16 11 8 7
G30a	-	Nigeria	Okeke Study	Case	1995	-	-	438	None	40 13 9 13 16 0 9
G01b	-	Nigeria	Okeke Study	Case	1995	-	-	557	None	56 4 15 18 9 8 6
D08	-	Nigeria	Okeke Study	Case	1995	-	-	464	None	56 4 3 18 9 8 6
G01	-	Nigeria	Okeke Study	Case	1995	-	-	435	None	56 4 15 102 9 73 6
E10	-	Nigeria	Okeke Study	Case	1995	-	-	435	None	56 4 15 102 9 73 6
D34a	-	Nigeria	Okeke Study	Case	1995	-	-	481	None	56 11 100 1 20 8 2
D33b	-	Nigeria	Okeke Study	Case	1995	-	-	481	None	56 11 100 1 20 8 2
G37a	-	Nigeria	Okeke Study	Case	1995	-	-	52	None	6 11 5 1 20 8 2
G67	-	Nigeria	Okeke Study	Case	1995	-	-	515	None	57 11 1 109 7 8 2
C78a	-	Nigeria	Okeke Study	Control	1995	-	-	468	None	57 11 1 109 7 81 2
C05a	-	Nigeria	Okeke Study	Case	1995	-	-	437	None	9 117 99 18 11 8 6
E66	-	Nigeria	Okeke Study	Control	1995	-	-	470	None	56 4 14 16 24 40 14
G83	-	Nigeria	Okeke Study	Case	1995	-	-	506	None	20 45 0 105 7 50 80
G16	-	Nigeria	Okeke Study	Case	1995	-	-	454	None	20 45 56 105 7 50 80
E59	-	Nigeria	Okeke Study	Control	1995	-	-	456	None	56 11 95 104 8 7 2
G107	-	Nigeria	Okeke Study	Control	1995	-	-	456	None	56 11 95 104 8 7 2
G106	-	Nigeria	Okeke Study	Control	1995	-	-	450	None	6 11 95 104 8 7 2
E16	-	Nigeria	Okeke Study	Case	1995	-	-	434	None	52 116 55 101 35 40 0
E07	-	Nigeria	Okeke Study	Case	1995	-	-	485	None	52 116 55 101 35 40 38
G112	-	Nigeria	Okeke Study	Control	1995	-	-	500	None	6 11 15 9 63 18 7
G122	-	Nigeria	Okeke Study	Control	1995	-	-	502	None	64 7 17 1 8 8 2

E54a	-	Nigeria	Okeke Study	Control	1995	-	-	362	None	62 100 17 31 5 5 4
E72	-	Nigeria	Okeke Study	Control	1995	-	-	362	None	62 100 17 31 5 5 4
C54a	-	Nigeria	Okeke Study	Control	1995	-	-	459	None	46 4 87 29 77 9 22
C54b	-	Nigeria	Okeke Study	Control	1995	-	-	459	None	46 4 87 29 77 9 22
D31b	-	Nigeria	Okeke Study	Case	1995	-	-	144	None	13 43 9 36 30 44 25
E04	-	Nigeria	Okeke Study	Case	1995	-	-	728	None	64 4 96 1 24 8 6
C83	-	Nigeria	Okeke Study	Control	1995	-	-	496	None	56 11 103 113 8 1 2
G110b	-	Nigeria	Okeke Study	Control	1995	-	-	491	None	13 39 19 36 30 14 82
G04	-	Nigeria	Okeke Study	Case	1995	-	-	425	None	64 196 188 83 24 8 6
G52	-	Nigeria	Okeke Study	Case	1995	-	-	422	None	56 95 3 18 11 8 14
C28	-	Nigeria	Okeke Study	Case	1995	-	-	486	None	0 0 101 39 8 8 2
E100869	-	UK	IID1	Case	1994	125ab	27	495	None	10 11 4 10 12 8 2
E105358	-	UK	IID1	Control	1994	O?	H?	SLV	None	6 29 5 26 8 8 6
E105835	-	UK	IID1	Case	1994	125ab	H?	495	None	10 11 4 10 12 8 2
E107532	-	UK	IID1	Case	1994	O?	H?	937	None	6 29 5 26 24 8 6
E093414	-	UK	IID1	Control	1993	128ac	H?	1326	None	6 220 3 26 9 7 7
E099534	-	UK	IID1	Case	1994	86	34	117	None	20 45 41 43 5 32 2
E89099	-	Bangladesh	Outbreak E	Case	1998	28ab	18	1657	None	21 35 2 52 5 5 182
E89104	-	Bangladesh	Outbreak E	Case	1998	141	49	111	None	6 29 14 16 24 8 2
E89106	-	Bangladesh	Outbreak E	Case	1998	R	7	1891	None	6 4 3 26 7 8 6
H123280788	H123280788	UK	GBRU	Case	2012	104	4	678	None	6 6 5 136 9 7 7
H123630425	H123630425	UK	GBRU	Case	2012	136	54	329	None	64 4 5 83 24 8 6
H125280573	H125280573	UK	GBRU	Case	2012	104	4	678	None	6 6 5 136 9 7 7
H131920216	H131920216	UK	Spice Outbreak	Case	2013	104	4	678	None	6 6 5 136 9 7 7
H131920218	H131920218	UK	Spice Outbreak	Case	2013	104	4	678	None	6 6 5 136 9 7 7
H131941062	H131941062	UK	Spice Outbreak	Case	2013	104	4	678	None	6 6 5 136 9 7 7
H131941063	H131941063	UK	Spice Outbreak	Case	2013	104	4	678	None	6 6 5 136 9 7 7
H131941070	H131941070	UK	Spice Outbreak	Case	2013	104	4	678	None	6 6 5 136 9 7 7
H112180280	H112180280	UK	German Outbreak	Case	2011	104	4	678	None	6 6 5 136 9 7 7
767	H100720284	UK	IID2	Case	2008	59	H-	1136	None	6 212 4 1 9 48 7
7004	H102340360	Bangladesh	GEMS Study	Control	2007-2011	O?	26	2067	None	6 95 3 18 11 122 2
1773	H100280372	UK	IID2	Case	2008	O?	30	DLV	None	6 196 1 1 8 8 20
7071	H102420449	Bangladesh	GEMS Study	Case	2007-2011	130	35	3670	None	6 4 4 16 24 2 14
7155	H102320115	Bangladesh	GEMS Study	Case	2007-2011	25	H-	SLV	None	6 11 4 8 0 78 2
1024	H103320333	Bangladesh	GEMS Study	Case	2007-2011	130	27	10	ST10 Cplx	10 11 4 8 8 8 2
1027	H103320341	Bangladesh	GEMS Study	Case	2007-2011	O?	H-	SLV	ST10 Cplx	10 0 4 8 8 8 2
1116	H103320335	Bangladesh	GEMS Study	Case	2007-2011	O?	33	10	ST10 Cplx	10 11 4 8 8 8 2
3074	H102400210	Bangladesh	GEMS Study	Control	2007-2011	56	48	34	ST10 Cplx	10 11 4 1 8 8 2
900616	H103820324	Bangladesh	GEMS Study	Case	2007-2011	113	H-	SLV	ST10 Cplx	0 11 4 8 8 8 2

900008	H103780405	Bangladesh	GEMS Study	Case	2007-2011	O?	10	34	ST10 Cplx	10 11 4 1 8 8 2
3042	H103760527	Bangladesh	GEMS Study	Case	2007-2011	3	2	10	ST10 Cplx	10 11 4 8 8 8 2
900575	H103840255	Bangladesh	GEMS Study	Case	2007-2011	O?	10	SLV	ST10 Cplx	24 11 4 8 8 8 0
600978	H104460624	Bangladesh	GEMS Study	Case	2007-2011	O?	11	48	ST10 Cplx	6 11 4 8 8 8 2
601035	H104300176	Bangladesh	GEMS Study	Control	2007-2011	117	27	SLV	ST10 Cplx	10 11 4 8 8 8 14
601051	H104300178	Bangladesh	GEMS Study	Control	2007-2011	154	19	10	ST10 Cplx	10 11 4 8 8 8 2
601134	H104320310	Bangladesh	GEMS Study	Control	2007-2011	O?	10	34	ST10 Cplx	10 11 4 1 8 8 2
601090	H104480138	Bangladesh	GEMS Study	Case	2007-2011	O?	33	10	ST10 Cplx	10 11 4 8 8 8 2
601091	H104480139	Bangladesh	GEMS Study	Case	2007-2011	O?	H-	10	ST10 Cplx	10 11 4 8 8 8 2
601096	H104480140	Bangladesh	GEMS Study	Case	2007-2011	O?	33	10	ST10 Cplx	10 11 4 8 8 8 2
601197	H104380400	Bangladesh	GEMS Study	Control	2007-2011	3	41	10	ST10 Cplx	10 11 4 8 8 8 2
2278	H101200133	UK	IID2	Case	2008	3	2	10	ST10 Cplx	10 11 4 8 8 8 2
G49	-	Nigeria	Okeke Study	Case	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
C10	-	Nigeria	Okeke Study	Case	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
G04a	-	Nigeria	Okeke Study	Case	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
D25a	-	Nigeria	Okeke Study	Case	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
C14a	-	Nigeria	Okeke Study	Case	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
C17	-	Nigeria	Okeke Study	Case	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
G63	-	Nigeria	Okeke Study	Case	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
C55	-	Nigeria	Okeke Study	Control	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
G133	-	Nigeria	Okeke Study	Control	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
G123	-	Nigeria	Okeke Study	Control	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
G129	-	Nigeria	Okeke Study	Control	1995	-	-	10	ST10 Cplx	10 11 4 8 8 8 2
G41	-	Nigeria	Okeke Study	Case	1995	-	-	48	ST10 Cplx	6 11 4 8 8 8 2
G155a	-	Nigeria	Okeke Study	Control	1995	-	-	520	ST10 Cplx	56 11 4 8 8 8 2
G36	-	Nigeria	Okeke Study	Case	1995	-	-	167	ST10 Cplx	10 11 4 8 8 13 2
G85	-	Nigeria	Okeke Study	Case	1995	-	-	167	ST10 Cplx	10 11 4 8 8 13 2
C53	-	Nigeria	Okeke Study	Control	1995	-	-	167	ST10 Cplx	10 11 4 8 8 13 2
C22	-	Nigeria	Okeke Study	Case	1995	-	-	436	ST10 Cplx	10 11 4 8 8 74 2
C29	-	Nigeria	Okeke Study	Case	1995	-	-	218	ST10 Cplx	10 11 4 12 8 8 2
G50	-	Nigeria	Okeke Study	Case	1995	-	-	218	ST10 Cplx	10 11 4 12 8 8 2
G126	-	Nigeria	Okeke Study	Control	1995	-	-	475	ST10 Cplx	56 11 4 8 8 78 2
G155	-	Nigeria	Okeke Study	Control	1995	-	-	510	ST10 Cplx	6 11 4 8 8 86 2
E12	-	Nigeria	Okeke Study	Case	1995	-	-	488	ST10 Cplx	56 11 102 8 8 8 2
G113	-	Nigeria	Okeke Study	Control	1994	-	-	556	ST10 Cplx	111 11 12 8 8 8 2
G170	-	Nigeria	Okeke Study	Control	1995	-	-	34	ST10 Cplx	10 11 4 1 8 8 2
E099963	-	UK	IID1	Case	1994	18ac	30	10	ST10 Cplx	10 11 4 8 8 8 2
E099964	-	UK	IID1	Case	1994	18ac	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E099966	-	UK	IID1	Case	1994	151	11	48	ST10 Cplx	6 11 4 8 8 8 2
E099969	-	UK	IID1	Control	1994	134	H-	10	ST10 Cplx	10 11 4 8 8 8 2

E099970	-	UK	IID1	Case	1994	53	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E099971	-	UK	IID1	Control	1994	53	2	10	ST10 Cplx	10 11 4 8 8 8 2
E099973	-	UK	IID1	Case	1994	151	11	SLV	ST10 Cplx	6 11 4 8 0 8 2
E099974	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E099977	-	UK	IID1	Case	1994	86	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E100866	-	UK	IID1	Case	1994	7	2	10	ST10 Cplx	10 11 4 8 8 8 2
E100868	-	UK	IID1	Case	1994	4	33	34	ST10 Cplx	10 11 4 1 8 8 2
E100870	-	UK	IID1	Case	1994	R	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E100872	-	UK	IID1	Case	1994	62	30	34	ST10 Cplx	10 11 4 1 8 8 2
E100875	-	UK	IID1	Case	1994	7	2	10	ST10 Cplx	10 11 4 8 8 8 2
E100876	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E101092	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E101093	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E101094	-	UK	IID1	Case	1994	4	33	34	ST10 Cplx	10 11 4 1 8 8 2
E101097	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E103597	-	UK	IID1	Case	1994	R	H?	43	ST10 Cplx	24 11 4 8 8 8 2
E103599	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E103619	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E104927	-	UK	IID1	Control	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E104942	-	UK	IID1	Case	1994	162	10	34	ST10 Cplx	10 11 4 1 8 8 2
E104943	-	UK	IID1	Control	1994	O?	H?	43	ST10 Cplx	24 11 4 8 8 8 2
E104965	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E104966	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E104968	-	UK	IID1	Case	1994	131	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E104972	-	UK	IID1	Case	1994	62	30	34	ST10 Cplx	10 11 4 1 8 8 2
E104973	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E105355	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E105389	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E105837	-	UK	IID1	Control	1994	O?	H?	43	ST10 Cplx	24 11 4 8 8 8 2
E106505	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E106776	-	UK	IID1	Control	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E107241	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E107246	-	UK	IID1	Control	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E107248	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E107249	-	UK	IID1	Case	1994	21	4	10	ST10 Cplx	10 11 4 8 8 8 2
E107525	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E107528	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E107529	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E107530	-	UK	IID1	Case	1994	O?	H?	SLV	ST10 Cplx	10 11 4 8 8 8 133
E107533	-	UK	IID1	Case	1994	O?	H?	43	ST10 Cplx	24 11 4 8 8 8 2

E107536	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E107537	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E107541	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E107743	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E107755	-	UK	IID1	Control	1995	21	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E107756	-	UK	IID1	Case	1995	131	H-	DLV	ST10 Cplx	10 7 4 8 0 0 2
E092843	-	UK	IID1	Case	1993	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E094704	-	UK	IID1	Case	1993	R	H?	SLV	ST10 Cplx	24 11 4 55 8 8 2
E094712	-	UK	IID1	Case	1994	3	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E096612	-	UK	IID1	Case	1994	R	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E096616	-	UK	IID1	Control	1994	5	4	10	ST10 Cplx	10 11 4 8 8 8 2
E097477	-	UK	IID1	Case	1994	81	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E097480	-	UK	IID1	Case	1994	6	10	SLV	ST10 Cplx	24 11 4 8 8 8 0
E097494	-	UK	IID1	Case	1994	81	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E097496	-	UK	IID1	Case	1994	86	2	10	ST10 Cplx	10 11 4 8 8 8 2
E097499	-	UK	IID1	Case	1994	6	10	10	ST10 Cplx	10 11 4 8 8 8 2
E099535	-	UK	IID1	Case	1994	R	27	10	ST10 Cplx	10 11 4 8 8 8 2
E099536	-	UK	IID1	Case	1994	4	33	34	ST10 Cplx	10 11 4 1 8 8 2
E099959	-	UK	IID1	Case	1994	66	25	10	ST10 Cplx	10 11 4 8 8 8 2
E099960	-	UK	IID1	Case	1994	18ac	30	10	ST10 Cplx	10 11 4 8 8 8 2
E099961	-	UK	IID1	Control	1994	18ac	30	10	ST10 Cplx	10 11 4 8 8 8 2
E108398	-	UK	IID1	Case	1994	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E108400	-	UK	IID1	Case	1994	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E108689	-	UK	IID1	Control	1995	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E108690	-	UK	IID1	Case	1995	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E108692	-	UK	IID1	Control	1995	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E108838	-	UK	IID1	Case	1995	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E109634	-	UK	IID1	Control	1995	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E109635	-	UK	IID1	Case	1995	O?	H?	43	ST10 Cplx	24 11 4 8 8 8 2
E109902	-	UK	IID1	Control	1995	4	33	34	ST10 Cplx	10 11 4 1 8 8 2
E109903	-	UK	IID1	Case	1995	21	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E109906	-	UK	IID1	Case	1995	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E110715	-	UK	IID1	Case	1995	6	10	43	ST10 Cplx	24 11 4 8 8 8 2
E110899	-	UK	IID1	Case	1995	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
E110910	-	UK	IID1	Case	1995	118	11	48	ST10 Cplx	6 11 4 8 8 8 2
E111139	-	UK	IID1	Case	1995	4	2	10	ST10 Cplx	10 11 4 8 8 8 2
E111260	-	UK	IID1	Control	1995	4	2	10	ST10 Cplx	10 11 4 8 8 8 2
E111262	-	UK	IID1	Case	1995	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E111268	-	UK	IID1	Case	1995	O?	H?	DLV	ST10 Cplx	10 11 4 0 8 8 2
E111294	-	UK	IID1	Control	1995	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2

E111479	-	UK	IID1	Control	1995	11	H-	10	ST10 Cplx	10 11 4 8 8 8 2
E112215	-	UK	IID1	Case	1995	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
E59905	-	UK	GBRU Archive	Case	1989	21	2	10	ST10 Cplx	10 11 4 8 8 8 2
E60725	-	UK	GBRU Archive	Case	1990	92	33	34	ST10 Cplx	10 11 4 1 8 8 2
E67643	-	UK	GBRU Archive	Case	1990	78	10	10	ST10 Cplx	10 11 4 8 8 8 2
E73339	-	UK	GBRU Archive	Case	1991	15	H-	34	ST10 Cplx	10 11 4 1 8 8 2
E89111	-	Bangladesh	Outbreak E	Case	1998	89	18	10	ST10 Cplx	10 11 4 8 8 8 2
E92356	-	UK	GBRU Archive	Case	1994	3	2	10	ST10 Cplx	10 11 4 8 8 8 2
E96390	-	UK	Outbreak B	Case	1994	O?	H-	SLV	ST10 Cplx	35 11 4 8 8 8 2
E96483	-	UK	Outbreak B	Case	1994	O?	33	34	ST10 Cplx	10 11 4 1 8 8 2
E96487	-	UK	Outbreak B	Case	1994	O?	H-	34	ST10 Cplx	10 11 4 1 8 8 2
E97470	-	UK	Outbreak C	Case	1994	86	34	10	ST10 Cplx	10 11 4 8 8 8 2
E97590	-	UK	Outbreak B	Case	1994	73	13	48	ST10 Cplx	6 11 4 8 8 8 2
E97820	-	UK	Outbreak B	Case	1994	62	30	34	ST10 Cplx	10 11 4 1 8 8 2
E97900	-	UK	Outbreak B	Case	1994	O?	H-	34	ST10 Cplx	10 11 4 1 8 8 2
H101360102	H101360102	UK	GBRU	Case	2010	O?	H?	SLV	ST10 Cplx	24 11 4 8 8 8 0
H101360104	H101360104	UK	GBRU	Case	2010	O?	H?	34	ST10 Cplx	10 11 4 1 8 8 2
H104220291	H104220291	UK	GBRU	Case	2010	O?	10	SLV	ST10 Cplx	24 11 4 8 8 8 0
H113540517	H113540517	UK	GBRU	Case	2011	O?	H-	10	ST10 Cplx	10 11 4 8 8 8 2
H122780065	H122780065	UK	GBRU	Case	2012	O?	H?	SLV	ST10 Cplx	24 11 4 8 8 8 0
H123600287	H123600287	UK	GBRU	Case	2012	168	4	10	ST10 Cplx	10 11 4 8 8 8 2
H123600288	H123600288	UK	GBRU	Case	2012	168	4	10	ST10 Cplx	10 11 4 8 8 8 2
H123620421	H123620421	UK	GBRU	Case	2012	99	52	10	ST10 Cplx	10 11 4 8 8 8 2
H131440284	H131440284	UK	GBRU	Case	2013	O?	H?	10	ST10 Cplx	10 11 4 8 8 8 2
H131720375	H131720375	UK	GBRU	Case	2013	99	H?	10	ST10 Cplx	10 11 4 8 8 8 2
H131720376	H131720376	UK	GBRU	Case	2013	100	19	10	ST10 Cplx	10 11 4 8 8 8 2
H131720388	H131720388	UK	GBRU	Case	2013	101	5	10	ST10 Cplx	10 11 4 8 8 8 2
H132320086	H132320086	UK	GBRU	Case	2013	92	H?	34	ST10 Cplx	10 11 4 1 8 8 2
H132360076	H132360076	UK	GBRU	Case	2013	175	H?	10	ST10 Cplx	10 11 4 8 8 8 2
H131920219	H131920219	UK	Spice Outbreak	Case	2013	131	27	10	ST10 Cplx	10 11 4 8 8 8 2
H131920220	H131920220	UK	Spice Outbreak	Case	2013	131	27	10	ST10 Cplx	10 11 4 8 8 8 2
H131941061	H131941061	UK	Spice Outbreak	Case	2013	55	19	10	ST10 Cplx	10 11 4 8 8 8 2
H131941064	H131941064	UK	Spice Outbreak	Case	2013	O?	21	227	ST10 Cplx	10 11 4 1 8 9 2
H131941071	H131941071	UK	Spice Outbreak	Case	2013	131	27	10	ST10 Cplx	10 11 4 8 8 8 2
H131941072	H131941072	UK	Spice Outbreak	Case	2013	131	27	10	ST10 Cplx	10 11 4 8 8 8 2
H131941073	H131941073	UK	Spice Outbreak	Case	2013	131	27	10	ST10 Cplx	10 11 4 8 8 8 2
H131941074	H131941074	UK	Spice Outbreak	Case	2013	131	27	10	ST10 Cplx	10 11 4 8 8 8 2
2336	H100160089	UK	IID2	Case	2008	114	10	10	ST10 Cplx	10 11 4 8 8 8 2
2451	H100160091	UK	IID2	Case	2008	O?	33	34	ST10 Cplx	10 11 4 1 8 8 2
651	H100700084	UK	IID2	Case	2008	151	21	48	ST10 Cplx	6 11 4 8 8 8 2

2633	H100200198	UK	IID2	Case	2008	O?	1	10	ST10 Cplx	10 11 4 8 8 8 2
2745	H100160086	UK	IID2	Case	2008	114	10	10	ST10 Cplx	10 11 4 8 8 8 2
2700	H100160087	UK	IID2	Case	2008	O?	10	SLV	ST10 Cplx	10 0 4 8 8 8 2
2466	H100200199	UK	IID2	Case	2008	92	33	34	ST10 Cplx	10 11 4 1 8 8 2
7142	H102340359	Bangladesh	GEMS Study	Control	2007-2011	89	H-	SLV	ST10 Cplx	6 11 4 8 8 8 0
7116	H10240211	Bangladesh	GEMS Study	Control	2007-2011	O?	36	SLV	ST10 Cplx	10 55 4 8 8 8 2
7201	H102400212	Bangladesh	GEMS Study	Control	2007-2011	58	51	SLV	ST10 Cplx	10 55 4 8 8 8 2
8016	H102340350	Bangladesh	GEMS Study	Case	2007-2011	O?	10	SLV	ST10 Cplx	24 11 15 8 8 8 2
7040	H102340357	Bangladesh	GEMS Study	Case	2007-2011	60	22	101	ST101 Cplx	43 41 15 18 11 7 6
E74	-	Nigeria	Okeke Study	Control	1995	-	-	476	ST12 Cplx	13 13 9 13 16 82 9
H123340465	H123340465	UK	GBRU	Case	2012	O?	H?	SLV	ST13 Cplx	0 6 5 9 9 8 2
H122780069	H122780069	UK	GBRU	Case	2012	O?	H?	152	ST152 Cplx	11 63 7 1 14 7 7
900157	H104060671	Bangladesh	GEMS Study	Case	2007-2011	34	11	223	ST155 Cplx	6 4 4 18 24 8 14
900098	H103900287	Bangladesh	GEMS Study	Case	2007-2011	126	27	155	ST155 Cplx	6 4 14 16 24 8 14
600985	H104460625	Bangladesh	GEMS Study	Case	2007-2011	77	34	SLV	ST155 Cplx	6 4 14 16 24 8 156
601191	H104560236	Bangladesh	GEMS Study	Case	2007-2011	9	21	155	ST155 Cplx	6 4 14 16 24 8 14
601307	H104580571	Bangladesh	GEMS Study	Case	2007-2011	9	21	155	ST155 Cplx	6 4 14 16 24 8 14
G48	-	Nigeria	Okeke Study	Case	1995	-	-	155	ST155 Cplx	6 4 14 16 24 8 14
E89095	-	Bangladesh	Outbreak E	Case	1998	80	27	SLV	ST155 Cplx	0 4 4 16 24 8 14
E89105	-	Bangladesh	Outbreak E	Case	1998	80	27	58	ST155 Cplx	6 4 4 16 24 8 14
8192	H102340356	Bangladesh	GEMS Study	Case	2007-2011	O?	19	58	ST155 Cplx	6 4 4 16 24 8 14
8225	H102400213	Bangladesh	GEMS Study	Control	2007-2011	15	34	155	ST155 Cplx	6 4 14 16 24 8 14
900547	H103800320	Bangladesh	GEMS Study	Case	2007-2011	O?	19	165	ST165 Cplx	10 27 5 10 12 8 2
900603	H103840256	Bangladesh	GEMS Study	Case	2007-2011	O?	19	165	ST165 Cplx	10 27 5 10 12 8 2
900500	H103840254	Bangladesh	GEMS Study	Case	2007-2011	128ab	12	SLV	ST165 Cplx	10 27 5 10 12 8 7
900820	H104080227	Bangladesh	GEMS Study	Case	2007-2011	O?	19	165	ST165 Cplx	10 27 5 10 12 8 2
7081	H102340351	Bangladesh	GEMS Study	Control	2007-2011	O?	19	165	ST165 Cplx	10 27 5 10 12 8 2
E099978	-	UK	IID1	Case	1994	80	8	SLV	ST165 Cplx	10 27 5 10 12 0 2
E100867	-	UK	IID1	Case	1994	O?	H?	1114	ST165 Cplx	10 27 5 10 12 1 2
E98527	-	UK	Outbreak A	Case	1994	19	H-	1114	ST165 Cplx	10 27 5 10 12 1 2
3029	H103320340	Bangladesh	GEMS Study	Control	2007-2011	12	4	484	ST168 Cplx	6 11 4 10 7 84 6
900696	H104060673	Bangladesh	GEMS Study	Case	2007-2011	7	4	484	ST168 Cplx	6 11 4 10 7 84 6
601235	H104580568	Bangladesh	GEMS Study	Case	2007-2011	7	4	484	ST168 Cplx	6 11 4 10 7 84 6
D09	-	Nigeria	Okeke Study	Case	1995	-	-	460	ST168 Cplx	56 11 4 10 7 8 6
E33	-	Nigeria	Okeke Study	Case	1995	-	-	484	ST168 Cplx	6 11 4 10 7 84 6
G30	-	Nigeria	Okeke Study	Case	1995	-	-	484	ST168 Cplx	6 11 4 10 7 84 6
G110	-	Nigeria	Okeke Study	Control	1995	-	-	484	ST168 Cplx	6 11 4 10 7 84 6
8089	H102340362	Bangladesh	GEMS Study	Control	2007-2011	7	4	484	ST168 Cplx	6 11 4 10 7 84 6
7172	H102340361	Bangladesh	GEMS Study	Control	2007-2011	12	4	484	ST168 Cplx	6 11 4 10 7 84 6
600983	H104200075	Bangladesh	GEMS Study	Control	2007-2011	60	H?	206	ST206 Cplx	6 7 5 1 8 18 2

G12a	-	Nigeria	Okeke Study	Case	1995	-	-	478	ST206 Cplx	56 7 5 1 8 18 2
H104480153	H104480153	UK	GBRU	Case	2010	O?	H?	206	ST206 Cplx	6 7 5 1 8 18 2
8098	H102340358	Bangladesh	GEMS Study	Control	2007-2011	11	16	206	ST206 Cplx	6 7 5 1 8 18 2
G35	-	Nigeria	Okeke Study	Case	1995	-	-	226	ST226 Cplx	10 27 5 8 8 7 2
E53	-	Nigeria	Okeke Study	Control	1995	-	-	226	ST226 Cplx	10 27 5 8 8 7 2
G159	-	Nigeria	Okeke Study	Control	1995	-	-	477	ST226 Cplx	10 127 5 8 8 7 2
H131941060	H131941060	UK	Spice Outbreak	Case	2013	111ac	4	226	ST226 Cplx	10 27 5 8 8 7 2
7207	H102340353	Bangladesh	GEMS Study	Case	2007-2011	91	9	226	ST226 Cplx	10 27 5 8 8 7 2
G137	-	Nigeria	Okeke Study	Control	1995	-	-	461	ST23 Cplx	56 4 12 1 20 13 7
G26a	-	Nigeria	Okeke Study	Case	1995	-	-	23	ST23 Cplx	6 4 12 1 20 13 7
C68	-	Nigeria	Okeke Study	Control	1995	-	-	466	ST23 Cplx	56 4 12 1 20 80 7
G55	-	NIgeria	Okeke Study	Case	1995	-	-	423	ST23 Cplx	6 4 33 1 20 12 7
D29	-	Nigeria	Okeke Study	Case	1995	-	-	278	ST278 Cplx	9 23 64 18 11 8 6
E104970	-	UK	IID1	Case	1994	33	H-	278	ST278 Cplx	9 23 64 18 11 8 6
E101396	-	UK	Outbreak D	Case	1995	98	H-	278	ST278 Cplx	9 23 64 18 11 8 6
E101406	-	UK	Outbreak D	Case	1995	98	H-	278	ST278 Cplx	9 23 64 18 11 8 6
E101621	-	UK	Outbreak D	Case	1994	98	H-	278	ST278 Cplx	9 23 64 18 11 8 6
E111613	-	UK	IID1	Control	1995	165	4	278	ST278 Cplx	9 23 64 18 11 8 6
E89107	-	Bangladesh	Outbreak E	Case	1998	O?	27	278	ST278 Cplx	9 23 64 18 11 8 6
E89115	-	Bangladesh	Outbreak E	Case	1998	162	H-	278	ST278 Cplx	9 23 64 18 11 8 6
H131920215	H131920215	UK	Spice Outbreak	Case	2013	20	H?	278	ST278 Cplx	9 23 64 18 11 8 6
H131920221	H131920221	UK	Spice Outbreak	Case	2013	20	19	278	ST278 Cplx	9 23 64 18 11 8 6
8080	H102340364	Bangladesh	GEMS Study	Case	2007-2011	125ac	9	295	ST295 Cplx	6 4 12 1 9 2 7
900545	H103800319	Bangladesh	GEMS Study	Case	2007-2011	181	16	3748	ST295 Cplx	6 133 12 1 9 2 7
900020	H103780400	Bangladesh	GEMS Study	Case	2007-2011	O?	7	841	ST295 Cplx	6 4 5 1 9 2 7
900286	H103900284	Bangladesh	GEMS Study	Case	2007-2011	O?	27	295	ST295 Cplx	6 4 12 1 9 2 7
601083	H104300180	Bangladesh	GEMS Study	Control	2007-2011	O?	28	295	ST295 Cplx	6 4 12 1 9 2 7
601110	H104500403	Bangladesh	GEMS Study	Case	2007-2011	25	7	3748	ST295 Cplx	6 133 12 1 9 2 7
601144	H104380395	Bangladesh	GEMS Study	Control	2007-2011	125ac	9	295	ST295 Cplx	6 4 12 1 9 2 7
601188	H104380398	Bangladesh	GEMS Study	Control	2007-2011	181	3	295	ST295 Cplx	6 4 12 1 9 2 7
601221	H113860299	Bangladesh	GEMS Study	Control	2007-2011	84	27	SLV	ST295 Cplx	6 4 12 1 9 2 2
601226	H104580567	Bangladesh	GEMS Study	Case	2007-2011	O?	27	SLV	ST295 Cplx	6 4 12 1 9 2 64
601251	H104580569	Bangladesh	GEMS Study	Case	2007-2011	O?	29	295	ST295 Cplx	6 4 12 1 9 2 7
7089b	H102420446	Bangladesh	GEMS Study	Control	2007-2011	O?	27	3748	ST295 Cplx	6 133 12 1 9 2 7
G02	-	Nigeria	Okeke Study	Case	1995	-	-	433	ST295 Cplx	56 4 12 1 9 2 7
G131	-	Nigeria	Okeke Study	Control	1995	-	-	433	ST295 Cplx	56 4 12 1 9 2 7
E101098	-	UK	IID1	Case	1994	O?	H?	2517	ST295 Cplx	6 4 12 1 11 2 7
E104974	-	UK	IID1	Case	1994	O?	H?	295	ST295 Cplx	6 4 12 1 9 2 7
E092836	-	UK	IID1	Case	1993	O?	H?	295	ST295 Cplx	6 4 12 1 9 2 7
E110912	-	UK	IID1	Case	1995	86	27	3570	ST295 Cplx	6 4 12 1 9 281 7

E111261	-	UK	IID1	Case	1995	O?	H?	3748	ST295 Cplx	6 133 12 1 9 2 7
E58596	-	UK	GBRU Archive	Case	1989	51	11	295	ST295 Cplx	6 4 12 1 9 2 7
E72376	-	UK	GBRU Archive	Case	1991	33	16	SLV	ST295 Cplx	0 4 12 1 9 2 7
E89112	-	Bangladesh	Outbreak E	Case	1998	69	11	SLV	ST295 Cplx	6 462 12 1 9 2 7
H131320944	H131320944	UK	GBRU	Case	2013	65	H?	841	ST295 Cplx	6 4 5 1 9 2 7
H131740867	H131740867	UK	GBRU	Case	2013	65	8	841	ST295 Cplx	6 4 5 1 9 2 7
H131920217	H131920217	UK	Spice Outbreak	Case	2013	33	16	295	ST295 Cplx	6 4 12 1 9 2 7
H131941065	H131941065	UK	Spice Outbreak	Case	2013	63	12	1664	ST295 Cplx	6 4 15 1 9 2 7
1778	H100200202	UK	IID2	Case	2008	125ac	9	295	ST295 Cplx	6 4 12 1 9 2 7
2866	H100160092	UK	IID2	Case	2008	55	4	3570	ST295 Cplx	6 4 12 1 9 281 7
7058	H102320431	Bangladesh	GEMS Study	Control	2007-2011	O?	16	295	ST295 Cplx	6 4 12 1 9 2 7
7078	H102420448	Bangladesh	GEMS Study	Control	2007-2011	O?	27	295	ST295 Cplx	6 4 12 1 9 2 7
7079	H102420450	Bangladesh	GEMS Study	Case	2007-2011	25	H-	SLV	ST295 Cplx	6 4 5 1 9 8 7
7089	H102420444	Bangladesh	GEMS Study	Control	2007-2011	181	16	3748	ST295 Cplx	6 133 12 1 9 2 7
7096	H102420447	Bangladesh	GEMS Study	Control	2007-2011	O?	12	295	ST295 Cplx	6 4 12 1 9 2 7
8002	H102400209	Bangladesh	GEMS Study	Control	2007-2011	9a	10	295	ST295 Cplx	6 4 12 1 9 2 7
8120	H102340365	Bangladesh	GEMS Study	Case	2007-2011	R	16	3748	ST295 Cplx	6 133 12 1 9 2 7
8129	H102400208	Bangladesh	GEMS Study	Control	2007-2011	O?	29	295	ST295 Cplx	6 4 12 1 9 2 7
#042	H132780234	Peru	EAEC reference strain	Case	1985	44	18	414	ST30 Cplx	18 22 20 23 5 15 4
E100877	-	UK	IID1	Case	1994	O?	H?	30	ST30 Cplx	18 22 16 23 5 15 4
E103267	-	UK	IID1	Case	1994	O?	H?	30	ST30 Cplx	18 22 16 23 5 15 4
E103621	-	UK	IID1	Case	1994	73	8	SLV	ST30 Cplx	18 22 16 23 5 15 99
E110716	-	UK	IID1	Case	1995	106	18	414	ST30 Cplx	18 22 20 23 5 15 4
E44939	-	UK	GBRU Archive	Case	1987	44	18	30	ST30 Cplx	18 22 16 23 5 15 4
E45730	-	UK	GBRU Archive	Case	1987	44	18	30	ST30 Cplx	18 22 16 23 5 15 4
E43509	-	UK	GBRU Archive	Case	1987	44	18	414	ST30 Cplx	18 22 20 23 5 15 4
900422	H103900286	Bangladesh	GEMS Study	Case	2007-2011	176	34	130	ST31 Cplx	18 22 20 6 5 5 4
900268	H103900283	Bangladesh	GEMS Study	Case	2007-2011	44	34	130	ST31 Cplx	18 22 20 6 5 5 4
900985	H104140246	Bangladesh	GEMS Study	Case	2007-2011	O?	23	130	ST31 Cplx	18 22 20 6 5 5 4
600974	H104460623	Bangladesh	GEMS Study	Case	2007-2011	15	34	SLV	ST31 Cplx	18 22 20 6 5 0 4
600988	H104460626	Bangladesh	GEMS Study	Case	2007-2011	166	16	130	ST31 Cplx	18 22 20 6 5 5 4
601010	H104480137	Bangladesh	GEMS Study	Case	2007-2011	15	23	SLV	ST31 Cplx	18 22 20 6 5 0 4
601048	H104300177	Bangladesh	GEMS Study	Control	2007-2011	O?	23	130	ST31 Cplx	18 22 20 6 5 5 4
601063	H104300179	Bangladesh	GEMS Study	Control	2007-2011	15	18	449	ST31 Cplx	18 22 94 6 5 5 4
601120	H104320309	Bangladesh	GEMS Study	Control	2007-2011	130	27	31	ST31 Cplx	18 22 17 6 5 5 4
901006	H104180060	Bangladesh	GEMS Study	Case	2007-2011	44	34	130	ST31 Cplx	18 22 20 6 5 5 4
1383	H100640645	UK	IID2	Case	2008	44	12	130	ST31 Cplx	18 22 20 6 5 5 4
E54	-	Nigeria	Okeke Study	Control	1995	-	-	31	ST31 Cplx	18 22 17 6 5 5 4
E56	-	Nigeria	Okeke Study	Control	1994	-	-	31	ST31 Cplx	18 22 17 6 5 5 4

C20	-	Nigeria	Okeke Study	Case	1995	-	-	130	ST31 Cplx	18 22 20 6 5 5 4
G22b	-	Nigeria	Okeke Study	Case	1995	-	-	130	ST31 Cplx	18 22 20 6 5 5 4
C60	-	Nigeria	Okeke Study	Control	1995	-	-	130	ST31 Cplx	18 22 20 6 5 5 4
G116	-	Nigeria	Okeke Study	Control	1995	-	-	130	ST31 Cplx	18 22 20 6 5 5 4
E68	-	Nigeria	Okeke Study	Control	1995	-	-	449	ST31 Cplx	18 22 94 6 5 5 4
G103	-	Nigeria	Okeke Study	Control	1995	-	-	449	ST31 Cplx	18 22 94 6 5 5 4
G149	-	Nigeria	Okeke Study	Control	1994	-	-	474	ST31 Cplx	18 22 1 6 5 5 4
G121a	-	Nigeria	Okeke Study	Control	1995	-	-	474	ST31 Cplx	18 22 1 6 5 5 4
C16	-	Nigeria	Okeke Study	Case	1995	-	-	512	ST31 Cplx	21 22 2 6 5 5 4
*H145-1	-	Peru	Okeke Study	Case	1995	-	-	31	ST31 Cplx	22 22 2 6 5 5 4
*44-1R	-	Thailand	Okeke Study	Case	1995	-	-	32	ST31 Cplx	23 22 2 6 5 5 4
*309-1R	-	Thailand	Okeke Study	Case	1995	-	-	33	ST31 Cplx	24 22 2 6 5 5 4
E099975	-	UK	IID1	Control	1994	15	18	31	ST31 Cplx	18 22 17 6 5 5 4
E101091	-	UK	IID1	Case	1994	134	27	31	ST31 Cplx	18 22 17 6 5 5 4
E101095	-	UK	IID1	Case	1994	134	2	31	ST31 Cplx	18 22 17 6 5 5 4
E101096	-	UK	IID1	Case	1994	134	27	31	ST31 Cplx	18 22 17 6 5 5 4
E103594	-	UK	IID1	Case	1994	11	27	31	ST31 Cplx	18 22 17 6 5 5 4
E103617	-	UK	IID1	Case	1994	130	27	31	ST31 Cplx	18 22 17 6 5 5 4
E104931	-	UK	IID1	Control	1994	O?	H?	31	ST31 Cplx	18 22 17 6 5 5 4
E104940	-	UK	IID1	Control	1994	130	27	31	ST31 Cplx	18 22 17 6 5 5 4
E104946	-	UK	IID1	Case	1994	O?	H?	31	ST31 Cplx	18 22 17 6 5 5 4
E104967	-	UK	IID1	Case	1994	O?	H?	31	ST31 Cplx	18 22 17 6 5 5 4
E104969	-	UK	IID1	Control	1994	130	25	31	ST31 Cplx	18 22 17 6 5 5 4
E104975	-	UK	IID1	Case	1994	O?	H?	SLV	ST31 Cplx	18 22 15 6 5 5 4
E105352	-	UK	IID1	Control	1994	O?	H?	130	ST31 Cplx	18 22 20 6 5 5 4
E107526	-	UK	IID1	Case	1994	O?	H?	31	ST31 Cplx	18 22 17 6 5 5 4
E107754	-	UK	IID1	Control	1995	134	25	31	ST31 Cplx	18 22 17 6 5 5 4
E097478	-	UK	IID1	Control	1994	130	27	31	ST31 Cplx	18 22 17 6 5 5 4
E097502	-	UK	IID1	Case	1994	134	27	31	ST31 Cplx	18 22 17 6 5 5 4
E108693	-	UK	IID1	Control	1995	O?	H?	449	ST31 Cplx	18 22 94 6 5 5 4
E108837	-	UK	IID1	Control	1995	130	H?	31	ST31 Cplx	18 22 17 6 5 5 4
E108839	-	UK	IID1	Case	1995	130	27	31	ST31 Cplx	18 22 17 6 5 5 4
E110852	-	UK	IID1	Case	1995	O?	H?	31	ST31 Cplx	18 22 17 6 5 5 4
E111136	-	UK	IID1	Control	1995	1	H-	31	ST31 Cplx	18 22 17 6 5 5 4
E58583	-	UK	GBRU Archive	Case	1989	77	18	SLV	ST31 Cplx	18 22 12 6 5 5 4
E71341	-	UK	GBRU Archive	Case	1991	102	27	31	ST31 Cplx	18 22 17 6 5 5 4
E89114	-	Bangladesh	Outbreak E	Case	1998	44	18	449	ST31 Cplx	18 22 94 6 5 5 4
E96485	-	UK	Outbreak B	Case	1994	134	27	31	ST31 Cplx	18 22 17 6 5 5 4
E98529	-	UK	Outbreak A	Case	1994	O?	18	31	ST31 Cplx	18 22 17 6 5 5 4
7028	H102400215	Bangladesh	GEMS Study	Control	2007-2011	44	34	DLV	ST31 Cplx	18 22 20 16 5 0 4

3017	H103320338	Bangladesh	GEMS Study	Control	2007-2011	166	15	349	ST349 Cplx	34 36 39 87 67 16 4
600950	H104180061	Bangladesh	GEMS Study	Control	2007-2011	166	15	349	ST349 Cplx	34 36 39 87 67 16 4
600990	H104460627	Bangladesh	GEMS Study	Case	2007-2011	166	15	349	ST349 Cplx	34 36 39 87 67 16 4
E60	-	Nigeria	Okeke Study	Control	1995	-	-	349	ST349 Cplx	34 36 39 87 67 16 4
7016	H102400214	Bangladesh	GEMS Study	Control	2007-2011	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
900033	H103760529	Bangladesh	GEMS Study	Case	2007-2011	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
900516	H103780408	Bangladesh	GEMS Study	Case	2007-2011	86	30	38	ST38 Cplx	4 26 2 25 5 5 19
900252	H103760528	Bangladesh	GEMS Study	Case	2007-2011	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
900002	H103780406	Bangladesh	GEMS Study	Case	2007-2011	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
900654	H103840252	Bangladesh	GEMS Study	Case	2007-2011	O?	18	SLV	ST38 Cplx	4 26 2 25 5 16 19
900978	H104140245	Bangladesh	GEMS Study	Case	2007-2011	O?	30	315	ST38 Cplx	4 26 2 25 5 8 19
900912	H104140244	Bangladesh	GEMS Study	Case	2007-2011	O?	30	315	ST38 Cplx	4 26 2 25 5 8 19
900745	H104120573	Bangladesh	GEMS Study	Case	2007-2011	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
601000	H104200076	Bangladesh	GEMS Study	Control	2007-2011	O?	30	38	ST38 Cplx	4 26 2 25 5 5 19
601029	H104560238	Bangladesh	GEMS Study	Case	2007-2011	O?	32	38	ST38 Cplx	4 26 2 25 5 5 19
601070	H104580565	Bangladesh	GEMS Study	Case	2007-2011	O?	30	38	ST38 Cplx	4 26 2 25 5 5 19
601108	H104320308	Bangladesh	GEMS Study	Control	2007-2011	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
601182	H104380397	Bangladesh	GEMS Study	Control	2007-2011	O?	30	38	ST38 Cplx	4 26 2 25 5 5 19
601225	H104400250	Bangladesh	GEMS Study	Control	2007-2011	86	30	38	ST38 Cplx	4 26 2 25 5 5 19
601264	H104580570	Bangladesh	GEMS Study	Case	2007-2011	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
G75a	-	Nigeria	Okeke Study	Case	1995	-	-	38	ST38 Cplx	4 26 2 25 5 5 19
G23	-	Nigeria	Okeke Study	Case	1995	-	-	38	ST38 Cplx	4 26 2 25 5 5 19
G29	-	Nigeria	Okeke Study	Case	1995	-	-	38	ST38 Cplx	4 26 2 25 5 5 19
G28	-	Nigeria	Okeke Study	Case	1995	-	-	426	ST38 Cplx	4 26 2 25 5 40 19
H132760800	H132760800	UK	GBRU	Case	2013	O?	H?	38	ST38 Cplx	4 26 2 25 5 5 19
669	H100640648	UK	IID2	Case	2008	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
1975	H100280375	UK	IID2	Case	2008	153	30	38	ST38 Cplx	4 26 2 25 5 5 19
7060	H102420445	Bangladesh	GEMS Study	Case	2007-2011	O?	27	38	ST38 Cplx	4 26 2 25 5 5 19
7123	H102340352	Bangladesh	GEMS Study	Control	2007-2011	21	10	315	ST38 Cplx	4 26 2 25 5 8 19
8095	H102320134	Bangladesh	GEMS Study	Control	2007-2011	86	30	SLV	ST38 Cplx	4 26 12 25 5 5 19
8130	H102340387	Bangladesh	GEMS Study	Control	2007-2011	O?	34	SLV	ST38 Cplx	4 26 2 25 5 5 7
7002	H102400216	Bangladesh	GEMS Study	Control	2007-2011	181	36	38	ST38 Cplx	4 26 2 25 5 5 19
900673	H103800323	Bangladesh	GEMS Study	Case	2007-2011	O?	18	394	ST394 Cplx	21 35 61 52 5 5 4
900088	H103760531	Bangladesh	GEMS Study	Case	2007-2011	44	40	394	ST394 Cplx	21 35 61 52 5 5 4
900416	H103900285	Bangladesh	GEMS Study	Case	2007-2011	O?	18	394	ST394 Cplx	21 35 61 52 5 5 4
600970	H104400254	Bangladesh	GEMS Study	Case	2007-2011	O?	18	394	ST394 Cplx	21 35 61 52 5 5 4
601002	H104200077	Bangladesh	GEMS Study	Control	2007-2011	O?	41	394	ST394 Cplx	21 35 61 52 5 5 4
601009	H104480136	Bangladesh	GEMS Study	Case	2007-2011	166	15	394	ST394 Cplx	21 35 61 52 5 5 4
601230	H104400251	Bangladesh	GEMS Study	Control	2007-2011	44	18	394	ST394 Cplx	21 35 61 52 5 5 4
642	H100860460	UK	IID2	Case	2008	68	17	1380	ST394 Cplx	35 35 61 52 5 5 4

G59	-	Nigeria	Okeke Study	Case	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
G10	-	Nigeria	Okeke Study	Case	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
C14	-	Nigeria	Okeke Study	Case	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
G17a	-	Nigeria	Okeke Study	Case	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
C08	-	Nigeria	Okeke Study	Case	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
E30	-	Nigeria	Okeke Study	Case	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
G67b	-	Nigeria	Okeke Study	Case	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
G108	-	Nigeria	Okeke Study	Control	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
E64	-	Nigeria	Okeke Study	Control	1995	-	-	394	ST394 Cplx	21 35 61 52 5 5 4
E62	-	Nigeria	Okeke Study	Control	1995	-	-	471	ST394 Cplx	21 125 61 52 5 5 4
E107757	-	UK	IID1	Control	1995	O?	H?	1380	ST394 Cplx	35 35 61 52 5 5 4
E105839	-	UK	IID1	Control	1994	O?	H?	1380	ST394 Cplx	35 35 61 52 5 5 4
E106507	-	UK	IID1	Case	1994	O?	H?	394	ST394 Cplx	21 35 61 52 5 5 4
E107247	-	UK	IID1	Case	1994	O?	H?	1380	ST394 Cplx	35 35 61 52 5 5 4
E108829	-	UK	IID1	Case	1995	O?	H?	1380	ST394 Cplx	35 35 61 52 5 5 4
E89102	-	Bangladesh	Outbreak E	Case	1998	44	18	394	ST394 Cplx	21 35 61 52 5 5 4
E96386	-	UK	Outbreak B	Case	1994	73	18	1380	ST394 Cplx	35 35 61 52 5 5 4
H122980178	H122980178	UK	GBRU	Case	2012	O?	18	1380	ST394 Cplx	35 35 61 52 5 5 4
1627	H000200204	UK	IID2	Case	2008	68	18	1380	ST394 Cplx	35 35 61 52 5 5 4
900657	H103800322	Bangladesh	GEMS Study	Case	2007-2011	175	28	200	ST40 Cplx	6 4 5 26 7 8 14
900618	H103820325	Bangladesh	GEMS Study	Case	2007-2011	175	7	200	ST40 Cplx	6 4 5 26 7 8 14
900114	H104060672	Bangladesh	GEMS Study	Case	2007-2011	127	11	DLV	ST40 Cplx	6 4 5 26 20 16 14
600955	H104180062	Bangladesh	GEMS Study	control	2007-2011	175	1	SLV	ST40 Cplx	6 4 5 26 7 8 156
601017	H104200078	Bangladesh	GEMS Study	Control	2007-2011	175	28	SLV	ST40 Cplx	6 4 5 26 7 8 4
601033	H104200079	Bangladesh	GEMS Study	Control	2007-2011	175	31	200	ST40 Cplx	6 4 5 26 7 8 14
601068	H104560241	Bangladesh	GEMS Study	Case	2007-2011	2	42	SLV	ST40 Cplx	6 4 5 26 7 8 156
601173	H113860296	Bangladesh	GEMS Study	Control	2007-2011	O175	31	SLV	ST40 Cplx	6 4 5 26 7 8 7
601192	H104560237	Bangladesh	GEMS Study	Case	2007-2011	175	1	200	ST40 Cplx	6 4 5 26 7 8 14
601193	H104380399	Bangladesh	GEMS Study	Control	2007-2011	175	31	200	ST40 Cplx	6 4 5 26 7 8 14
900987	H104140247	Bangladesh	GEMS Study	Case	2007-2011	175	31	200	ST40 Cplx	6 4 5 26 7 8 14
900998	H104140248	Bangladesh	GEMS Study	Case	2007-2011	175	1	200	ST40 Cplx	6 4 5 26 7 8 14
E40104	H132780233	UK	EAEC Reference strain	Case	2013	126	H?	200	ST40 Cplx	6 4 5 26 7 8 14
1150	H100720282	UK	IID2	Case	2008	111ac	21	40	ST40 Cplx	6 4 5 26 20 8 14
1171	H101200082	UK	IID2	Case	2008	126	27	200	ST40 Cplx	6 4 5 26 7 8 14
E63	-	Nigeria	Okeke Study	Control	1995	-	-	473	ST40 Cplx	56 4 5 26 20 8 14
E099967	-	UK	IID1	Case	1994	R	H?	200	ST40 Cplx	6 4 5 26 7 8 14
E099972	-	UK	IID1	Case	1994	O?	H?	200	ST40 Cplx	6 4 5 26 7 8 14
E099979	-	UK	IID1	Case	1994	75	27	200	ST40 Cplx	6 4 5 26 7 8 14
E100856	-	UK	IID1	Case	1994	126	27	200	ST40 Cplx	6 4 5 26 7 8 14

E101089	-	UK	IID1	Case	1994	126	2	200	ST40 Cplx	6 4 5 26 7 8 14
E106506	-	UK	IID1	Case	1994	O?	H?	200	ST40 Cplx	6 4 5 26 7 8 14
E107100	-	UK	IID1	Case	1994	119	27	200	ST40 Cplx	6 4 5 26 7 8 14
E107531	-	UK	IID1	Case	1994	R	H?	200	ST40 Cplx	6 4 5 26 7 8 14
E107542	-	UK	IID1	Case	1994	R	H?	200	ST40 Cplx	6 4 5 26 7 8 14
E094706	-	UK	IID1	Case	1994	111ab	H-	SLV	ST40 Cplx	6 4 5 10 20 8 14
E097298	-	UK	IID1	Case	1994	R	27	200	ST40 Cplx	6 4 5 26 7 8 14
E097501	-	UK	IID1	Case	1994	O?	H?	200	ST40 Cplx	6 4 5 26 7 8 14
E099520	-	UK	IID1	Case	1994	8	7	200	ST40 Cplx	6 4 5 26 7 8 14
E107759	-	UK	IID1	Control	1995	O?	H?	40	ST40 Cplx	6 4 5 26 20 8 14
E109907	-	UK	IID1	Control	1995	111ab	H-	40	ST40 Cplx	6 4 5 26 20 8 14
E110717	-	UK	IID1	Case	1995	O?	H?	200	ST40 Cplx	6 4 5 26 7 8 14
E111140	-	UK	IID1	Case	1995	O?	H?	200	ST40 Cplx	6 4 5 26 7 8 14
E36182	-	UK	GBRU Archive	Case	1987	111	21	40	ST40 Cplx	6 4 5 26 20 8 14
E43923	-	UK	GBRU Archive	Case	1987	126	27	200	ST40 Cplx	6 4 5 26 7 8 14
E55060	-	UK	GBRU Archive	Case	1998	126	27	200	ST40 Cplx	6 4 5 26 7 8 14
E55280	-	UK	GBRU Archive	Case	1998	126	27	SLV	ST40 Cplx	0 4 5 26 7 8 14
E57144	-	UK	GBRU Archive	Case	1989	111	21	SLV	ST40 Cplx	6 4 5 26 20 222 14
H104680397	H104680397	UK	GBRU	Case	2010	111ab	21	40	ST40 Cplx	6 4 5 26 20 8 14
H120820356	H120820356	UK	GBRU	Case	2012	111ab	21	40	ST40 Cplx	6 4 5 26 20 8 14
H122840058	H122840058	UK	GBRU	Case	2012	111ab	11	40	ST40 Cplx	6 4 5 26 20 8 14
H123160613	H123160613	UK	GBRU	Case	2012	175	28	200	ST40 Cplx	6 4 5 26 7 8 14
H123520278	H123520278	UK	GBRU	Case	2012	175	28	200	ST40 Cplx	6 4 5 26 7 8 14
H123520279	H123520279	UK	GBRU	Case	2012	175	28	200	ST40 Cplx	6 4 5 26 7 8 14
H124020363	H124020363	UK	GBRU	Case	2012	175	31	200	ST40 Cplx	6 4 5 26 7 8 14
H132360372	H132360372	UK	GBRU	Case	2013	126	27	200	ST40 Cplx	6 4 5 26 7 8 14
H120680226	H120680226	UK	Ireland Household Outbreak	Case	2012	111ab	21	40	ST40 Cplx	6 4 5 26 20 8 14
657	H100700085	UK	IID2	Case	2008	111ac	11	40	ST40 Cplx	6 4 5 26 20 8 14
1091	H100720282	UK	IID2	Case	2008	111ac	11	40	ST40 Cplx	6 4 5 26 20 8 14
1244	H100760047	UK	IID2	Case	2008	111ac	21	40	ST40 Cplx	6 4 5 26 20 8 14
2266	H100280373	UK	IID2	Case	2008	175	27	200	ST40 Cplx	6 4 5 26 7 8 14
1337	H100640644	UK	IID2	Case	2008	111ac	11	40	ST40 Cplx	6 4 5 26 20 8 14
C61	-	Nigeria	Okeke Study	Control	1995	-	-	444	ST446 Cplx	56 19 3 26 11 8 6
900644	H103820326	Bangladesh	GEMS Study	Case	2007-2011	O?	10	448	ST448 Cplx	6 6 5 16 11 8 7
G57	-	Nigeria	Okeke Study	Case	1995	-	-	46	ST46 Cplx	8 7 1 8 8 8 6
C70	-	Nigeria	Okeke Study	Control	1995	-	-	467	ST467 Cplx	64 7 4 8 8 18 6
D34	-	Nigeria	Okeke Study	Case	1995	-	-	480	ST467 Cplx	10 7 4 8 8 18 6
D32	-	Nigeria	Okeke Study	Control	1995	-	-	480	ST467 Cplx	10 7 4 8 8 18 6

G143	-	Nigeria	Okeke Study	Control	1995	-	-	467	ST467 Cplx	64 7 4 8 8 18 6
C77	-	Nigeria	Okeke Study	Control	1995	-	-	469	ST469 Cplx	6 65 5 1 9 13 6
G144a	-	Nigeria	Okeke Study	Control	1995	-	-	501	ST501 Cplx	35 132 2 27 37 5 4
C04	-	Nigeria	Okeke Study	Case	1995	86	11	507	ST501 Cplx	35 132 2 27 37 5 83
E092830	-	UK	IID1	Case	1993	86	11	501	ST501 Cplx	35 132 2 27 37 5 4
E097500	-	UK	IID1	Case	1994	73	1	501	ST501 Cplx	35 132 2 27 37 5 4
E107527	-	UK	IID1	Control	1994	6	1	73	ST73 Cplx	36 24 9 13 17 11 25
C27	-	Nigeria	Okeke Study	Case	1995	-	-	159	ST746 Cplx	59 7 4 8 12 8 2
E107758	-	UK	IID1	Case	1995	O?	H?	159	ST746 Cplx	59 7 4 8 12 8 2
E099976	-	UK	IID1	Case	1994	113	H-	746	ST746 Cplx	10 7 4 8 12 8 2
E107250	-	UK	IID1	Case	1994	O?	H?	SLV	ST746 Cplx	10 7 4 0 12 8 2
E107252	-	UK	IID1	Control	1994	O?	H?	746	ST746 Cplx	10 7 4 8 12 8 2
E89096	-	Bangladesh	Outbreak E	Case	1998	113	H-	159	ST746 Cplx	59 7 4 8 12 8 2
H104400276	H104400276	UK	GBRU	Case	2010	O?	H-	SLV	ST746 Cplx	0 7 4 8 12 8 2
H113160257	H113160257	UK	GBRU	Case	2011	O?	H-	SLV	ST746 Cplx	0 7 4 8 12 8 2
H123980248	H123980248	UK	GBRU	Case	2013	181	H?	SLV	ST746 Cplx	0 7 4 8 12 8 2
H132100889	H132100889	UK	GBRU	Case	2013	181	H?	159	ST746 Cplx	59 7 4 8 12 8 2
H131920214	H131920214	UK	Spice Outbreak	Case	2013	O?	19	746	ST746 Cplx	10 7 4 8 12 8 2
E25	-	Nigeria	Okeke Study	Case	1995	-	-	453	ST86 Cplx	99 6 33 33 24 8 7
E096617	-	UK	IID1	Case	1994	R	1	SLV	ST86 Cplx	0 6 33 33 24 8 7

Appendix 7.1 Table showing characteristics of the EAEC strains used in this study grouped by complex. **Key:** - denotes N/A (MOLIS number) or unknown (Serotyping), SLV (single locus variant), DLV (double locus variant), TLV (triple locus variant), Cplx (complex with consists of SLVs), IID (intestinal infectious disease study 1 or 2), GBRU (Gastrointestinal Bacteria Reference Unit), ST (sequence type).

All 564 EAEC strains were included in the population structure and evolutionary analysis. Three additional strains denoted by * were added for the HEp-2 cell assay and worm model. Serotyping was carried out on all strains except those from Nigeria. #EAEC prototypical 042 strain was used in all analysis

Sub-populations of strains were used in further analysis and are colour coded as follows:

Colour	No. of isolates	Analysis
Purple	27	HEp-2 assay and worm model
Blue	97	Biolog assay
Green	153	Virulence gene profiling
Pink	49	Biolog and virulence gene profiling

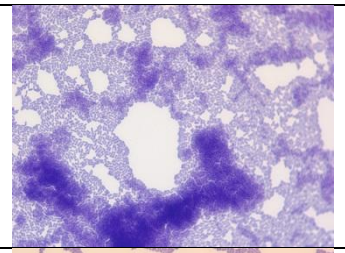
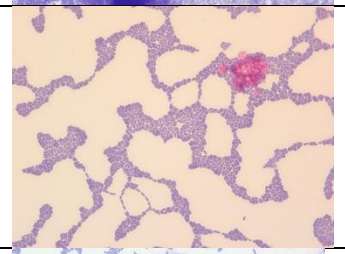
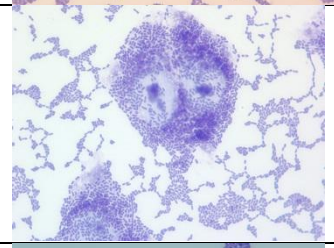
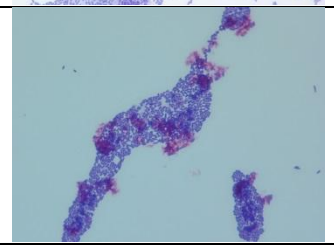
7.2 Biolog metabolite list

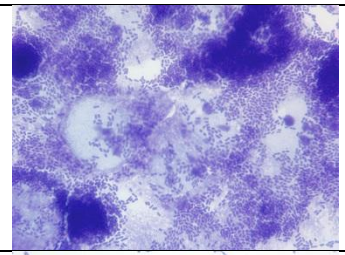
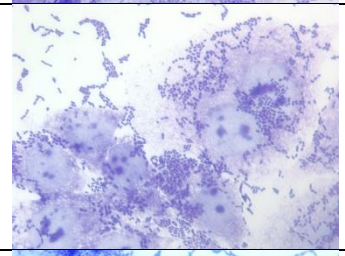
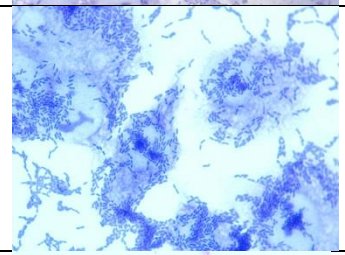
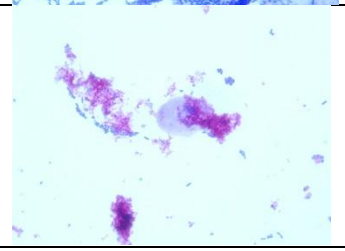
GEN III MicroPlate ™

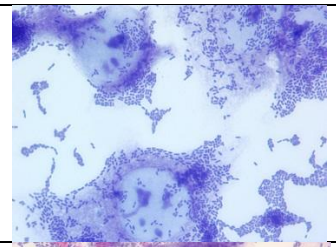
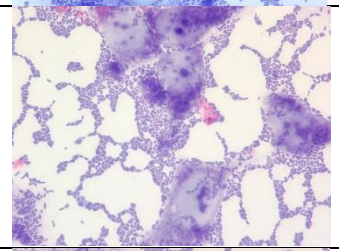
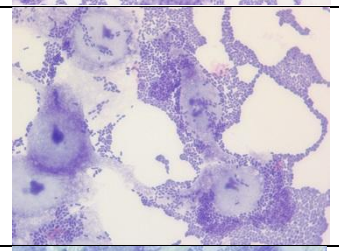
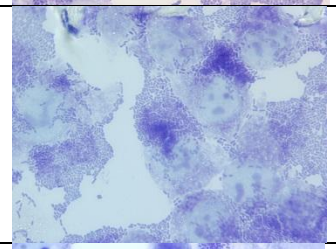
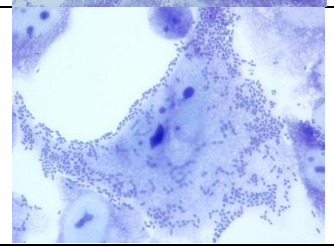
A1 Negative Control	A2 Dextrin	A3 D-Maltose	A4 D-Trehalose	A5 D-Cellobiose	A6 Gentiobiose	A7 Sucrose	A8 D-Turanose	A9 Stachyose	A10 Positive Control	A11 pH 6	A12 pH 5
B1 D-Raffinose	B2 α-D-Lactose	B3 D-Melibiose	B4 β-Methyl-D-Glucoside	B5 D-Salicin	B6 N-Acetyl-D-Glucosamine	B7 N-Acetyl-β-D-Mannosamine	B8 N-Acetyl-D-Galactosamine	B9 N-Acetyl Neuraminic Acid	B10 1% NaCl	B11 4% NaCl	B12 8% NaCl
C1 α-D-Glucose	C2 D-Mannose	C3 D-Fructose	C4 D-Galactose	C5 3-Methyl Glucose	C6 D-Fucose	C7 L-Fucose	C8 L-Rhamnose	C9 Inosine	C10 1% Sodium Lactate	C11 Fusidic Acid	C12 D-Serine
D1 D-Sorbitol	D2 D-Mannitol	D3 D-Arabitol	D4 myo-Inositol	D5 Glycerol	D6 D-Glucose-6-PO4	D7 D-Fructose-6-PO4	D8 D-Aspartic Acid	D9 D-Serine	D10 Troleandomycin	D11 Rifamycin SV	D12 Minocycline
E1 Gelatin	E2 Glycyl-L-Proline	E3 L-Alanine	E4 L-Arginine	E5 L-Aspartic Acid	E6 L-Glutamic Acid	E7 L-Histidine	E8 L-Pyroglutamic Acid	E9 L-Serine	E10 Lincomycin	E11 Guanidine HCl	E12 Niaproof 4
F1 Pectin	F2 D-Galacturonic Acid	F3 L-Galactonic Acid Lactone	F4 D-Gluconic Acid	F5 D-Glucuronic Acid	F6 Glucuronamide	F7 Mucic Acid	F8 Quinic Acid	F9 D-Saccharic Acid	F10 Vancomycin	F11 Tetrazolium Violet	F12 Tetrazolium Blue
G1 p-Hydroxy-Phenylacetic Acid	G2 Methyl Pyruvate	G3 D-Lactic Acid Methyl Ester	G4 L-Lactic Acid	G5 Citric Acid	G6 α-Keto-Glutaric Acid	G7 D-Malic Acid	G8 L-Malic Acid	G9 Bromo-Succinic Acid	G10 Nalidixic Acid	G11 Lithium Chloride	G12 Potassium Tellurite
H1 Tween 40	H2 γ-Amino-Butyric Acid	H3 α-Hydroxy-Butyric Acid	H4 β-Hydroxy-D,L-Butyric Acid	H5 α-Keto-Butyric Acid	H6 Acetoacetic Acid	H7 Propionic Acid	H8 Acetic Acid	H9 Formic Acid	H10 Aztreonam	H11 Sodium Butyrate	H12 Sodium Bromate

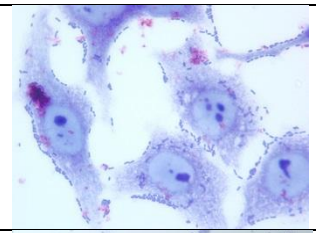
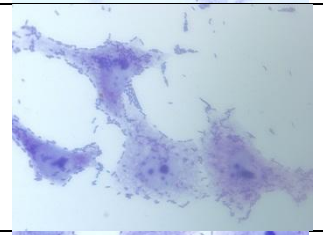
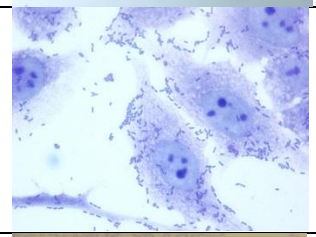
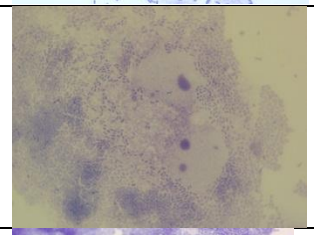
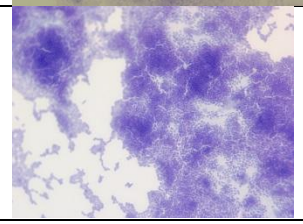
Appendix 7.2 Biolog GEN III plate listing the 95 metabolites (plus one negative control) in each well.

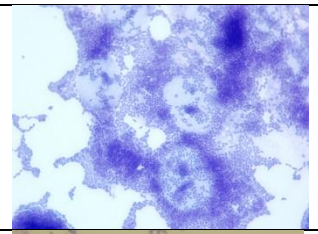
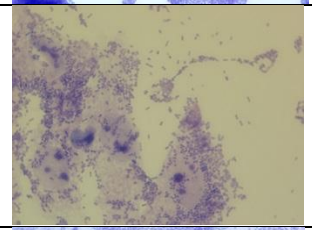
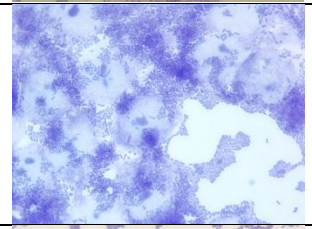
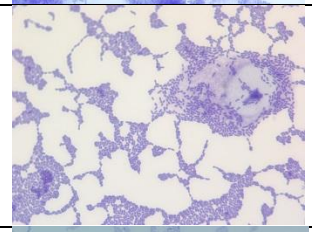
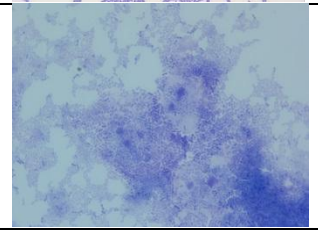
7.3 Photographs of HEp-2 Cell Assay of EAEC strains

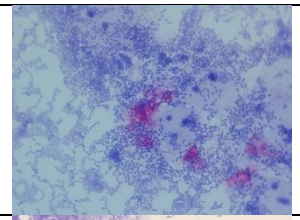
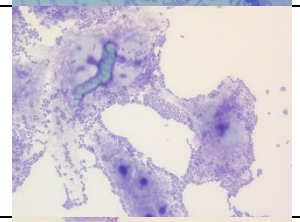
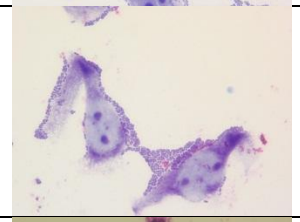
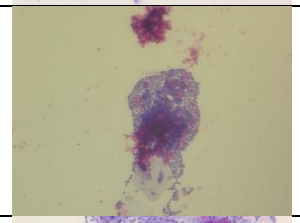
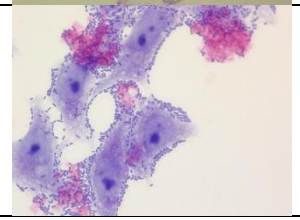
Isolate	ST	ST Cplx	HEp-2 Score	Case	HEp-2 Result
042	414	None	5 = Very Heavy	Case	
C27D Not included in analysis	159	None	0 = Cytotoxic, auto aggregative (heavy)	Case	
H145-1R	31	ST31 Cplx	4 = Heavy	Case	
44-1R	31	ST31 Cplx	0 = Cytotoxic, auto aggregative (heavy)	Case	

309-1R	31	ST31 Cplx	5 = Very Heavy	Case	
E54H	31	ST31 Cplx	3 = Medium	Control	
E56H	31	ST31 Cplx	4 = Heavy	Control	
G149H	474	ST31 Cplx	0 = Cytotoxic, autoaggregative (light)	Control	

G121aH	474	ST31 Cplx	4 = Heavy	Control	
C16D	512	ST31 Cplx	4 = Heavy	Case	
C20D	130	ST31 Cplx	4 = Heavy	Case	
C60H	130	ST31 Cplx	4 = Heavy	Control	
G116H	130	ST31 Cplx	4 = Heavy	Control	

G23D	38	ST38 Cplx	1 = Very Light	Case	
G29D	38	ST38 Cplx	3 = Medium	Case	
G28D	426	ST38 Cplx	2 = Light	Case	
G59D	394	ST394 Cplx	4 = Heavy	Case	
G10D	394	ST394 Cplx	5 = Very Heavy	Case	

C14D	394	ST394 Cplx	5 = Very Heavy	Case	
G17aD	394	ST394 Cplx	4 = Heavy	Case	
C08D	394	ST394 Cplx	5 = Very Heavy	Case	
E30D	394	ST394 Cplx	4 = Heavy	Case	
G108H	394	ST394 Cplx	5 = Very Heavy	Control	

E64H	394	ST394 Cplx	5 = Very Heavy	Control	
E33D	484	ST168 Cplx	4 = Heavy	Case	
G30D	484	ST168 Cplx	2 = Light	Case	
G110H	484	ST168 Cplx	0 = Cytotoxic, autoaggregative (light)	Control	
D09D	460	ST168 Cplx	3 = Medium	Case	

Appendix 7.3 Table showing photos of HEp-2 cell assay as evidence of how the HEp-2 score was derived.

7.4 Virulence Gene Profile of Bangladesh Strains

Strain	Somatic Type	Flagella Type	ST Type	ST Complex	Case/ Control	aggR	aat	aap	aggA	aafA	agg3A	agg4A	aaf5A	astA	pet	pic	setA	irp2	tia	aaiC	Virulence Score
7116	O?	36	DL	10	Control									1							1
7142	89	H-	SLV	10	Control									1		1	1	1	1		5
7201	58	51	SLV	10	Control		1							1				1			3
1027	O?	H-	New Allele	10	Case	1	1	1							1	1	1	1		1	8
3036	113	H-	New Allele	10	Control	1	1	1					1		1	1	1	1			8
900575	O?	10	New Allele	10	Case	1	1	1		1				1	1	1	1	1		1	10
900616	113	H-	New Allele	10	Case	1	1	1					1		1		1	1		1	8
1116	O?	33	10	10	Case			1				1		1	1		1	1	1	1	8
3042	3	2	10	10	Control	1	1	1	1					1		1	1	1			8
900008	O?	10	34	10	Case	1	1	1					1			1	1	1		1	8
601035	117	27	SLV	10	Control		1														1
601051	154	19	10	10	Control		1							1		1	1	1			5
601134	O?	10	34	10	Control	1	1	1		1						1	1	1			7
601197	3	41	10	10	Control	1	1	1	1							1	1	1			7
600978	O?	11	48	10	Case		1			1								1		1	4
601090	O?	33	10	10	Case					1						1	1	1	1	1	6
601091	O?	H-	10	10	Case					1						1	1	1	1	1	6
601096	O?	33	10	10	Case											1	1	1	1	1	5
7028	44	34	DLV	31	Control	1	1	1			1								1		5
900268	44	34	130	31	Case	1	1	1				1								1	5
900422	176	34	130	31	Case	1	1	1			1			1		1	1	1	1	1	10
900985	O?	23	130	31	Case		1	1				1						1		1	5
601048	O?	23	130	31	Control	1	1	1			1			1		1	1	1	1		9
601063	15	18	449	31	Control		1	1					1					1			4
601120	130	27	31	31	Control	1	1	1						1		1		1			6
600974	15	34	New Allele	31	Case	1	1	1			1			1		1	1	1	1	1	10
600988	166	16	130	31	Case	1	1	1			1			1		1	1	1		1	9
601010	15	23	New Allele	31	Case	1	1	1		1	1			1		1	1	1	1	1	11
8095	86	30	SLV	38	Control	1	1	1	1												4
8130	O?	34	SLV	38	Control	1	1	1			1								1		5
900654	O?	18	SLV	38	Case	1	1	1	1									1		1	6
900912	O?	30	315	38	Case	1	1	1	1									1	1	1	7
7060	O?	27	38	38	Case		1	1		1				1	1	1	1	1	1	1	10
7002	181	36	38	38	Control	1	1	1						1	1	1	1	1	1		9
7016	153	30	38	38	Control	1	1	1	1										1		5
7123	21	10	315	38	Control		1														1
900002	153	30	38	38	Case	1	1	1	1		1								1	1	7
900033	153	30	38	38	Case	1	1	1					1	1					1		7
900252	153	30	38	38	Case	1	1	1					1							1	5
900516	86	30	38	38	Case	1	1	1	1									1	1	1	7
900745	153	30	38	38	Case	1	1	1			1									1	5
900978	O?	30	315	38	Case	1	1	1	1				1						1	1	7
601000	O?	30	38	38	Control		1	1	1					1							4
601108	153	30	38	38	Control	1	1	1											1		4
601182	O?	30	38	38	Control	1	1	1										1	1		5
601225	86	30	38	38	Control	1	1	1	1									1			5
601029	O?	32	38	38	Case	1	1	1	1									1	1	1	7
601070	O?	30	38	38	Case	1	1	1	1									1	1	1	7
601264	153	30	38	38	Case	1	1	1	1	1				1	1	1	1	1	1	1	12
900618	175	7	200	40	Case	1	1	1		1				1	1	1	1	1	1	1	11
900657	175	28	200	40	Case	1	1	1		1				1	1	1	1	1		1	10
900987	175	31	200	40	Case	1	1	1		1			1	1	1	1		1		1	10
601017	175	28	SLV	40	Control	1	1	1		1				1	1	1	1	1			9

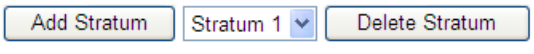
601033	175	31	200	40	Control	8
601193	175	31	200	40	Control	9
601192	175	1	200	40	Case	10
601068	2	42	SLV	40	Case	11
7040	60	22	101	101	Case	4
7071	130	35	SLV	155	Case	8
8192	O?	19	58	155	Case	4
8225	15	34	155	155	Control	8
900098	126	27	155	155	Case	5
900157	34	11	223	155	Case	4
600985	77	34	SLV	155	Case	7
601191	9	21	155	155	Case	6
601307	9	21	155	155	Case	6
900500	128ab	12	SLV	165	Case	4
900820	O?	19	165	165	Case	5
900547	O?	19	165	165	Case	5
900603	O?	19	165	165	Case	5
7172	12	4	484	168	Control	4
8089	7	4	484	168	Control	5
3029	12	4	484	168	Control	8
900696	7	4	484	168	Case	6
601235	7	4	484	168	Case	9
8098	11	16	206	206	Control	1
600983	60	H?	206	206	Control	3
7207	91	9	226	226	Case	8
7079	25	H-	SLV	295	Case	2
8080	125ac	9	295	295	Case	8
8120	R	16	SLV	295	Case	9
7058	O?	16	295	295	Control	4
7078	O?	27	295	295	Control	10
7089	181	16	SLV	295	Control	10
7089	O?	27	SLV	295	Control	10
7096	O?	12	295	295	Control	9
8002	9a	10	295	295	Control	7
8129	O?	29	295	295	Control	8
900020	O?	7	841	295	Case	9
900286	O?	27	295	295	Case	11
900545	181	16	SLV	295	Case	9
601083	O?	28	295	295	Control	8
601144	125ac	9	295	295	Control	7
601188	181	3	295	295	Control	5
601221	84	27	SLV	295	Control	10
601110	25	7	SLV	295	Case	11
601226	O?	27	SLV	295	Case	10
601251	O?	29	295	295	Case	9
3017	166	15	349	349	Control	3
600950	166	15	349	349	Control	5
600990	166	15	349	349	Case	5
900088	44	40	394	394	Case	6
900416	O?	18	394	394	Case	6
900673	O?	10	394	394	Case	5
601002	O?	41	394	394	Control	4
601230	44	18	394	394	Control	4
600970	O?	18	394	394	Case	7
601009	166	15	394	394	Case	6
900644	O?	10	448	448	Case	4
7004	O?	26	TLV	3LV	Control	10
1024	130	27	TLV	3LV	Case	7
3026	O?	27	TLV	3LV	Control	11
900753	166	15	DLV	3	Case	8
900998	175	1	200	40	Case	10
901006	44	34	130	31	Case	4

600955	175	1	SLV	40	Control	8
600982	166	15	SLV	0	Control	5
7081	O?	19	165	165	Control	4
7121	77	18	New Allele	1	Control	3
1038	25	H-	New Allele	1	Case	2
900245	161	H-	New Allele	1	Case	11
900442	121	H?	New Allele	1	Case	10
900553	51	12	New Allele	1	Case	10
601098	O?	45	New Allele	1	Control	0
7064	153	36	157	0	Control	4
7067	86	27	2166	0	Control	6
7092	125ac	8	1295	0	Control	5
1037	159	23	1490	0	Case	2
900063	O?	23	720	0	Case	7
900108	21	8	1295	0	Case	5
900512	O?	23	157	0	Case	6
900550	128	12	1326	0	Case	6
900693	44	H-	720	0	Case	6
900694	O?	24	219	0	Case	7
900732	O?	7	1891	0	Case	5
900769	69	4	1891	0	Case	8
900770	69	4	678	0	Case	9
900883	44	23	720	0	Case	6
601155	O?	30	678	0	Control	3
601180	O?	23	2186	0	Control	5
600965	126	7	SLV	0	Case	6
601176	O?	H-	720	0	Case	6
601174	78	7	1891	0	Control	4
601106	O?	23	2186	0	Control	3
7155	25	H-	SLV	2	Case	2
8016	O?	10	SLV	2	Case	2
900114	127	11	SLV	2	Case	8
900794	O?	10	SLV	2	Case	8
900851	125ac	11	SLV	2	Case	1
600961	69	38	SLV	2	Case	3
601158	2	H-	SLV	2	Case	4
601062	2	42	SLV	2	Case	4
042	44	18	414	30	Case	10

Table 7.4: Virulence gene profiling of 153 EAEC isolates from Bangladesh. Virulence score denotes the number of positive virulence genes in a given strain.

7.5 Descriptive methodology of Chi-Square for EAEC group assignment

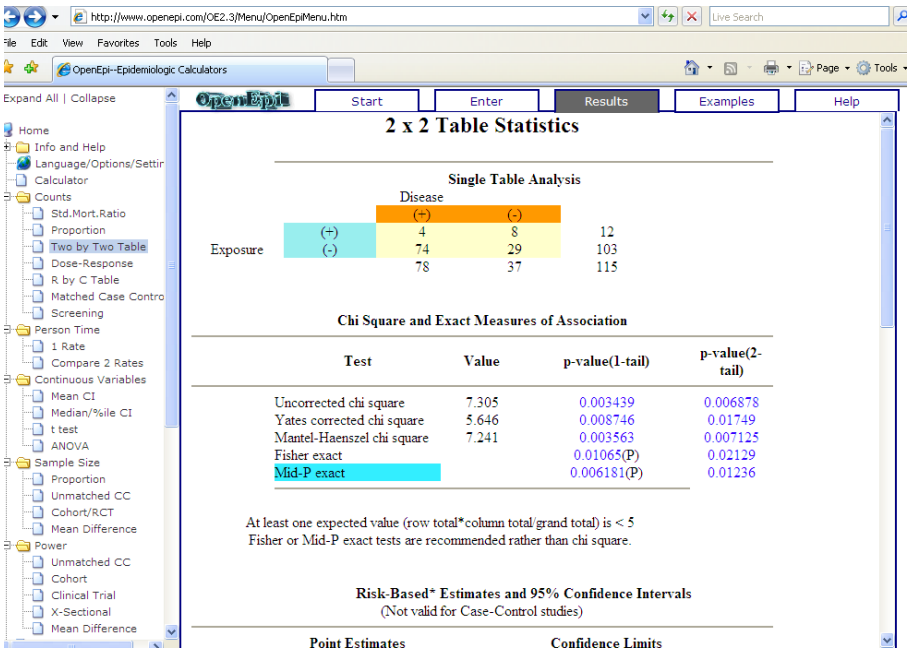
- http://wwwn.cdc.gov/epiinfo/ **then selected** Open Epi link on left hand side
- Selected two by two table under counts section on left hand side
- Select enter new data
- Put in results in table, example to test 295 group with 4 cases and 8 controls, total are 78 and 37 respectively. Put in value and minus the observed from the totals, then select calculate



Open Epi 2 x 2 Table

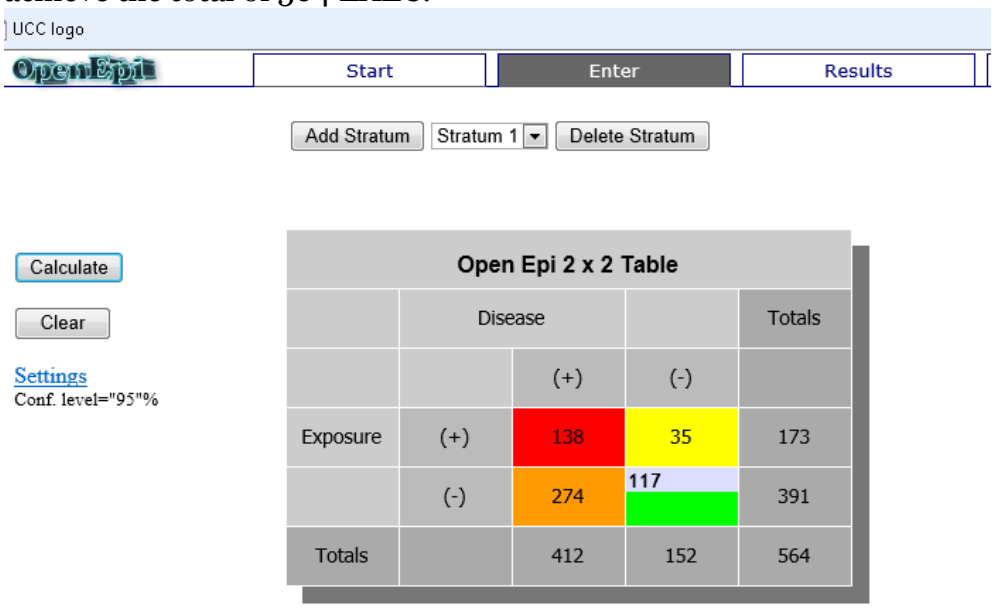
		Disease		Totals
		(+)	(-)	
Exposure	(+)	4	8	12
	(-)	74	29	103
Totals		78	37	115

- Put in value and minus the observed from the totals, then select calculate

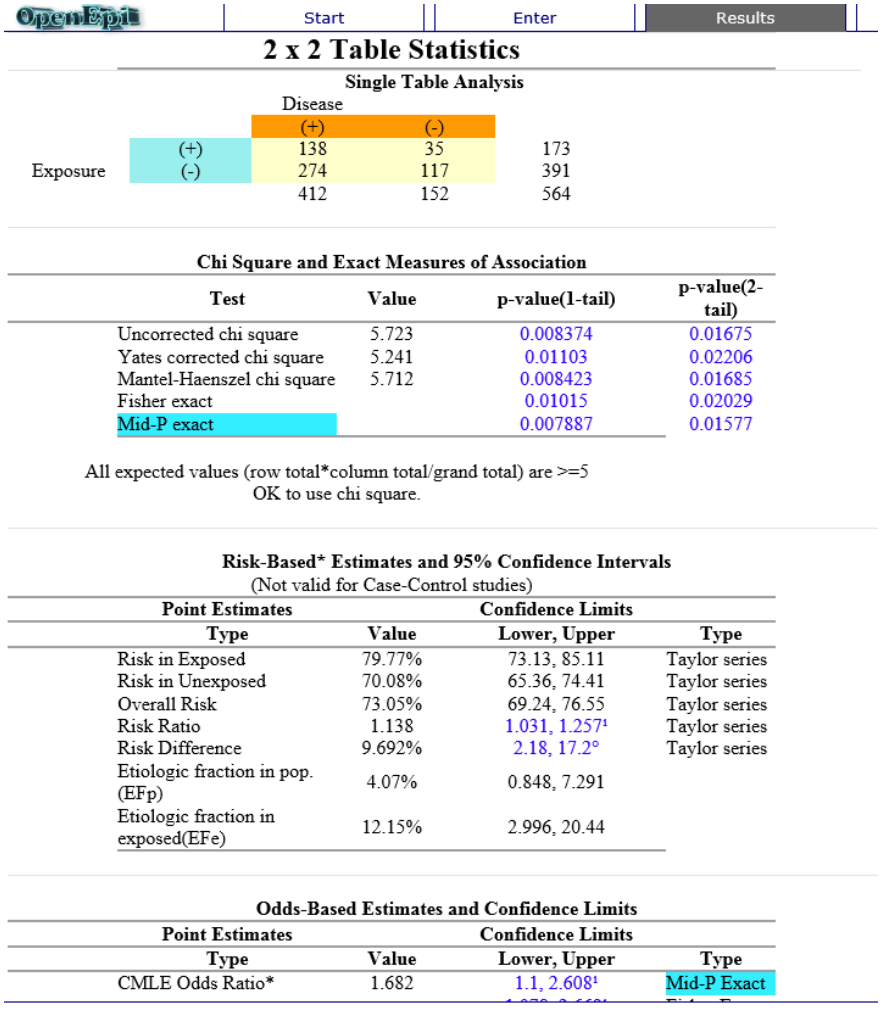


- Fishers exact test is used when the observed number is less than 5 and the totals are between 20-40. No appropriate in this case.

- Eg complex 10 with 138 cases and 35 in controls. Total of dataset is 412 cases and 152 controls, need to minus the 10 complex numbers from the total to achieve the total of 564 EAEC.



-
- Calculate

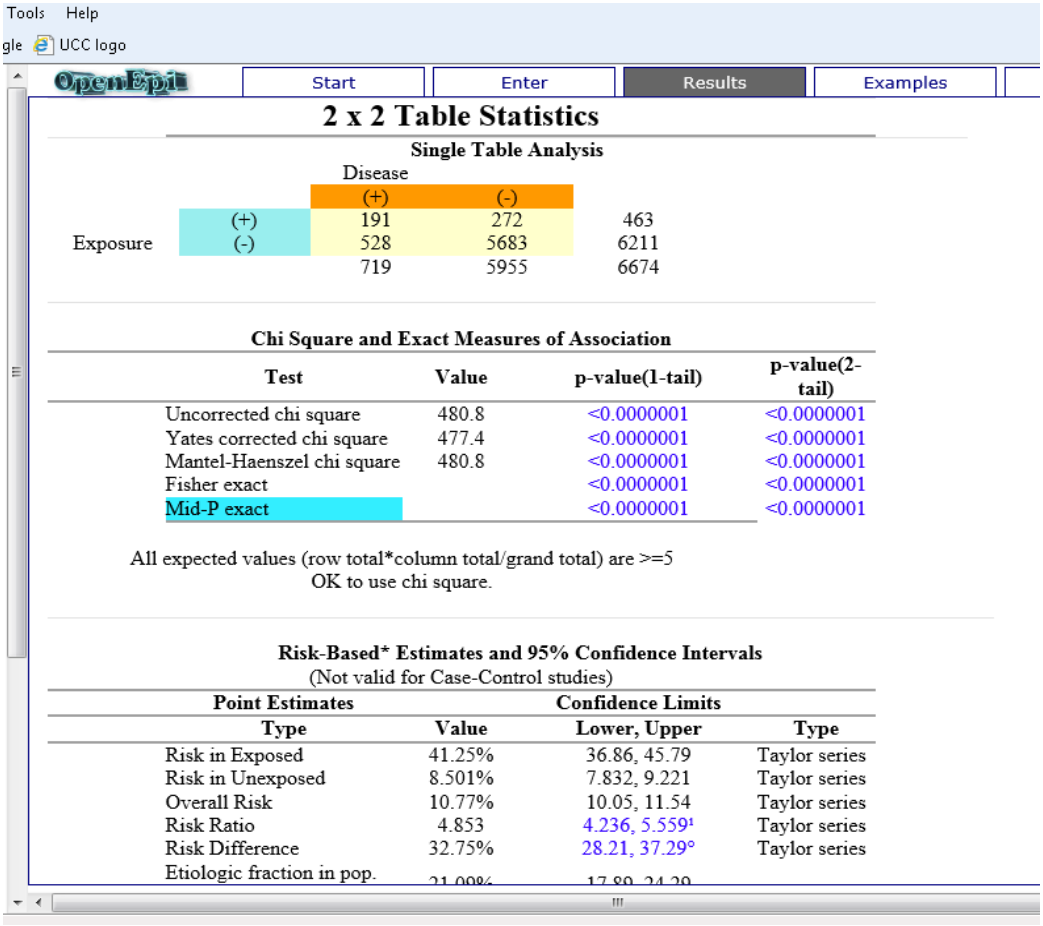


For EAEC associated groups, e.g complex 10 is 191 EAEC and 654 other E. coli

Total is 719 EAEC strains(564 from PhD + 155 from public database) and 5955 other E. coli

(6815 from PhD and public minus 719 EAEC and minus 141 shigella) from a total of 5955 E.coli strains)

e.g ST10 complex being EAEC associated, 191 EAEC in this group out of 719 and the other e.coli pathotypes have 272 in ST 10 Cplx out of 5955 E. coli.



Tools Help

gle UCC logo

OpenEpi Start Enter Results Examples

2 x 2 Table Statistics

Single Table Analysis

		Disease (+)	Disease (-)	
Exposure	(+)	191	272	463
	(-)	528	5683	6211
		719	5955	6674

Chi Square and Exact Measures of Association

Test	Value	p-value(1-tail)	p-value(2-tail)
Uncorrected chi square	480.8	<0.0000001	<0.0000001
Yates corrected chi square	477.4	<0.0000001	<0.0000001
Mantel-Haenszel chi square	480.8	<0.0000001	<0.0000001
Fisher exact		<0.0000001	<0.0000001
Mid-P exact		<0.0000001	<0.0000001

All expected values (row total*column total/grand total) are >=5
OK to use chi square.

Risk-Based* Estimates and 95% Confidence Intervals

(Not valid for Case-Control studies)

Point Estimates		Confidence Limits	
Type	Value	Lower, Upper	Type
Risk in Exposed	41.25%	36.86, 45.79	Taylor series
Risk in Unexposed	8.501%	7.832, 9.221	Taylor series
Overall Risk	10.77%	10.05, 11.54	Taylor series
Risk Ratio	4.853	4.236, 5.559[1]	Taylor series
Risk Difference	32.75%	28.21, 37.29[0]	Taylor series
Etiologic fraction in pop.	21.00%	17.89, 24.29	

7.6 Published Papers and Conference Presentations Summary

Published Papers

July 11	Enteroaggregative *Escherichia coli* O104 from an outbreak of HUS in Germany 2011, could it happen again? **Journal of Infection in Developing Countries.** *Vol. 5 Issue 6 Page 425-436. July 2011* Marie Anne Chattaway, Tim Dallman, Iruke N. Okeke and John Wain
Dec 12	Characterisation of a verocytotoxin-producing enteroaggregative *Escherichia coli* serogroup O111:H21 associated with a household outbreak in Northern Ireland **Journal of Clinical Microbiology.** *Vol. 50 Issue 12 Page 4116-4119. December 2012* Tim Dallman, Geoffrey P. Smith, Brendan O'Brien, Marie A. Chattaway, David Finlay, Kathie A. Grant, and Claire Jenkins
Sept 13	Comparative Analysis of ESBL-Positive Escherichia coli Isolates from Animals and Humans from the UK, The Netherlands and Germany **PLOS ONE.** *Vol. 8 Issue 9 Page 1-10. September 2013* Guanghui Wu, Michaela J. Day, Muriel T. Mafur, Javier Nunez-Garcia, Jackie J. Fenner, Meenaxi Sharma, Alieda van Essen-Zandbergen, Irene Rodríguez, Cindy Dierikx, Kristina Kadlec, Anne-Kathrin Schink, Marie Anne Chattaway, John Wain, Reiner Helmuth, Beatriz Guerra, Stefan Schwarz, John Threlfall, Martin J. Woodward, Neil Woodford, Nick Coldham, Dik Mevius. ***See appendix 7.6 for authorship letter***
Sept 13	Investigating the link between the presence of enteroaggregative Escherichia coli and infectious intestinal disease in the United Kingdom, 1993 to 1996 and 2008 to 2009. **Eurosurveillance.** *Vol. 18 Issue 37 Page 1-7. September 2013* Marie Anne Chattaway, Ross Harris, Claire Jenkins, Clarence Tam, John Coia, Jim Gray, Miren Iturriza-Gomara and John Wain
May 2014	An investigation of the diversity of strains of Enteroaggregative Escherichia coli isolated from cases associated with a large foodborne outbreak in the UK. **PLOS ONE**. *Vol. 9 Issue 5 Page e98103. May 2014* Tim J Dallman; Marie A Chattaway; Lauren Cowley; Michel Doumith; Rediat Tewolde; David J Wooldridge; Anthony Underwood; Derren Ready; John Wain; Kirsty Foster; Kathy A Grant and Claire Jenkins.
June 2014	Chromosomal location of blaCTX-M genes in clinical isolates of Escherichia coli from Germany, The Netherlands and The United Kingdom. **International Journal of Antimicrobial Agents.** *Vol. 50 Issue 6 Page 553-557. June 2014* Irene Rodríguez Fernández, Ph.D.; Katharina Thomas; Alieda van Essen; Anne-Kathrin Schink; Michaela Day; Marie A Chattaway; Guanghui Wu; Dik Mevius; Reiner Helmuth; Beatriz Guerra

June 2014

Establishing an enteric bacteria reference laboratory in Sierra Leone.
Journal of Infection in Developing Countries. *Vol. 8 Issue 7: Page 933-941. June 2014*
Chattaway MA, Kamara A, Rhodes F, Kaffeta K, Jambai A, Alemu W, Islam MS, Freeman MM, Welfare W, Harding D, Samba AF, Abu M, Kamanda S, Grant K, Jenkins C, Nair S, Connell S, Siorvanes L, Desai S, Allen C, Frost M, Hughes D, Jeffrey Z, Gill N, Salter M

November 2014

Evidence of an evolving extra-intestinal enteroaggregative ST38 clone.
Emerging Infectious Disease. *Vol.20 Issue 11:* DOI: 10.3201/eid2011.131845 *November 2012*
Marie Anne Chattaway, Claire Jenkins, Holly Ciesielczuk, Martin Day, Vivienne DoNascimento, Michaela Day, Irene Rodríguez, Alieda van Essen-Zandbergen, Anne-Kathrin Schink, Guanghui Wu, John Threlfall, Martin J. Woodward, Nick Coldham, Kristina Kadlec, Stefan Schwarz, Cindy Dierikx, Beatriz Guerra, Reiner Helmuth, Dik Mevius, Neil Woodford, and John Wain

November 2014

Enteroaggregative Escherichia coli Have Evolved Independently as Distinct Complexes within the *E. coli* Population with Varying Ability to Cause Disease.
PLOSONE. *9, (11) e112967 available from: PM:25415318*
Marie Anne Chattaway, Claire Jenkins, Dunstan Rajendram, Alejandro Cravatio, Kaiser Ali Talukder, Tim Dallman, Anthony Underwood ,Steve Platt, Iruka Okeke and John Wain.

January 2015

Use of whole genus genome sequence data to develop a Multi-Locus Sequencing Type tool that accurately speciates and sub-speciates within the Yersinia genus
Journal of Clinical Microbiology. *J.Clin.Microbiol.*, 53, (1) 35-42 available from: PM:25339391
Miquette Hall, Marie Chattaway, Sandra Reuter, Cyril Savin, Eckhard Strauch, Elisabeth Carniel, Thomas Connor, Inge Van Damme, Lakshani Rajakaruna, Dunstan Rajendram, Claire Jenkins, NIcholas Thomson and Alan McNally

Unpublished Papers (In draft)

February 2014 — Diversity of sequence types, phylogroups, plasmids and ESBL genes among *E.coli* from humans, animals and food in Germany, Netherlands and United Kingdom.
Submission planned for International Journal of Antimicrobial Agents.
Michaela Day; Marie A Chattaway, Neil Woodford, John Wain, Irene Rodríguez Fernández,.; Katharina Thomas; Alieda van Essen; Anne-Kathrin Schink; Guanghui Wu; Dik Mevius; Reiner Helmuth; Beatriz Guerra on behalf of the SAFEFOODERA_ESBL consortium†

May 2014 — Investigation into the misidentification of Hazard Group 3 gastrointestinal pathogens and associated health and safety risks
Submission planned for Eurosurveillance
Steve Connell, Marie Chattaway, David Powell, Elizabeth de Pinna, Claire Jenkins, Kathie Grant, Gauri Godbole, Maria Zambon

September 2014 — Virulence and Resistance of Enteroaggregative *Escherichia coli* complexes from Bangladesh
Submission planned for Journal of Infection in Developing Countries
Marie Anne Chattaway, Michaela Day, Emma White, Julia Mtwale, Kaisar Talukder, James Rogers, David Powell, Alejandro Cravioto and Claire Jenkins.

Presentations

April 2014	Invited Speaker – Oral Presentation **Public Health England, Commonwealth Secretariat , National Public Health Institute for Sierra Leone Workshop, Country Lodge, Sierra Leone** Public Health England and Laboratory Development in Sierra Leone
March 2014	Presentation/Lecture **UCL, Medical BSc (SSC10 Diseases in Developing Countries), London** Cholera and other Enteric Pathogens in Sierra Leone
Jan 2014	Presentation/Lecture **UCL, Infection & Immunity MSc course (PATHG013 Microbial Pathogenesis)** Verocytotoxic *E. coli* O157: Testing, Transmission and Outbreak Investigation
Jan 2014	Presentation/Lecture **Public Health England, Food, Water and Environment MSc Course, London** *Escherichia coli* O157: an update
Sept 2013	Poster **Public Health England, first annual conference, Warwick** Setting up an Enteric Laboratory in Sierra Leone
June 2013	Poster and 5 minute oral presentation (on the web) **MED VET NET, Conference, Copenhagen** Zoonotic disease of EAEC and EPEC via MLST transmission sourcing
Dec 2012	Oral **WHO, GOARN 13th Steering Committee Meeting, Geneva** Challenges in establishing an Enteric Laboratory in Sierra Leone
Nov 2012	Oral **WHO & Ministry of Health and Sanitation Office, Sierra Leone** Establishing a Bacteriology Laboratory at Lakka, Final report of mission
Nov 2012	Oral **Health Protection Agency, London** Establishing an Enteric Bacteriology Laboratory in Sierra Leone
Mar 2012	Poster **Society of General MicroBiology, Annual Conference, Dublin** Investigating animal reservoirs of diarrhoeagenic *E. coli*
Feb 2012	Oral **Health Protection Agency, Lunchtime lecture, London** VTEC O104 outbreak: Laboratory testing in the laboratory of gastrointestinal pathogens
Oct 2011	Poster **Health Protection Agency, Postgraduate Seminar, London** Enteroaggregative *E. coli*– is it a pathogen?

Oct 2011 Poster
American Society of Tropical Medicine and Hygiene, Annual Conference, USA
Population structure of Enteroaggregative *E. coli*

Sept 2011 Oral
Health Protection Agency, Postgraduate Seminar, London
Enteroaggregative *E. coli*– is it a pathogen?

Sept 2011 Oral
Society of General MicroBiology, Annual Conference, York
Enteroaggregative *E. coli* A previously unknown pathogen

June 2011 Oral
UCL, Divisional Colloquium, London
A novel approach to pathogen recognition in association with Enteroaggregative
E. coli

Oct 2010 Poster – Won first prize for best poster first year PhD
Health Protection Agency, Postgraduate Seminar, London
To define the population structure of Enteroaggregative *E. coli* for the development of next generation diagnostic testing

Sept 2010 Poster
Health Protection Agency, Annual conference, Warwick
To define the population structure of Enteroaggregative *E. coli* for the development of next generation diagnostic testing

---

Funding

Society for Applied MicroBiology
2010, £1500 towards MLST for PhD from Supervisory fund
Society for General MicroBiology
2011, £3000 towards secondment to USA for PhD f
Public Health England
2013, £50,000 to develop enteric laboratory in Sierra Leone from Global Fund
Public Health England
2014, £50,000 to establish enteric laboratory in Sierra Leone from Developing countries Laboratory

7.7 Enteroaggregative E. coli O104 from an outbreak of HUS in Germany 2011, could it happen again?

Chattaway MA, Dallman T, Okeke IN and Wain J . 2011. J Infect Dev Ctries 5 (2011) 425-436.

Review Article

Enteroaggregative E. coli O104 from an outbreak of HUS in Germany 2011, could it happen again?

Marie Anne Chattaway[1], Tim Dallman[1], Iruke N. Okeke[2], John Wain[1]

[1]*Laboratory of Gastrointestinal Pathogens, Health Protection Agency, London, England*
[2]*Molecular Microbiology, Haverford College, Pennsylvania, USA*

Abstract

Enterohaemorrhagic *E. coli* (EHEC) particularly O157:H7 (Sequence type 11 complex), is the best documented and most well-known of *E. coli* that cause diarrhoea. The importance of EHEC lies in the severity of disease. Outbreaks can infect thousands of people causing bloody diarrhoea and haemolytic uremic syndrome (HUS) that in turn can result in protracted illness or even death. The ability of EHEC to colonise the human gut is normally associated with the presence of genes from another group of diarrhoeagenic *E. coli*, the enteropathogenic *E. coli* (EPEC), via the locus of enterocyte effacement. However, the massive outbreak in Germany was caused by an EHEC which had acquired virulence genes from yet another group of diarrhoeagenic *E. coli*, the enteroaggregative *E. coli* (EAEC). In reality EAEC is probably the most common bacterial cause of diarrhoea but is not identified in most diagnostic laboratories. This outbreak emphasises the importance of being able to detect all diarrhoeagenic *E. coli* and not to focus on *E. coli* O157:H7 alone. Routine surveillance systems for EAEC, a once ignored global pathogen, would go a long way to reaching this goal. This review describes methods for identifying non-O157 EHEC and describes the key genetic features of EHEC and EAEC. Our aim is to provide information for laboratories and policy makers which enables them to make informed decisions about the best methods available for detecting newly emergent strains of diarrhoeagenic *E. coli*.

Key words: EAEC; EHEC; HUS; Outbreak O104:H4; ST678

J Infect Dev Ctries 2011; 5(6):425-436.

(Received 01 July 2011—Accepted 02 July 2011)

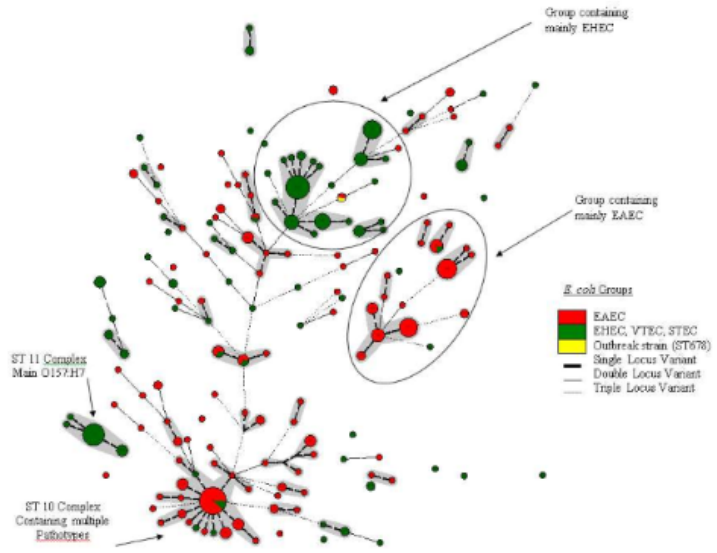
Introduction

Beginning in early May 2011, an unusually high number of haemolytic uraemic syndrome (HUS) cases were reported in Germany. The outbreak was caused by an enterohaemorrhagic *E. coli* (EHEC) which had characteristics of both a verotoxigenic *E. coli* (VTEC) (for a guide to nomenclature see Table 1) and of the less well-known diarrhoeagenic *E. coli*, enteroaggregative *E. coli* (EAEC). There is considerable expertise in diarrhoeagenic *E. coli* in Germany but even with support from the reference laboratory in Rome, the pathogen responsible for the outbreak proved challenging to characterize. Within most diagnostic laboratories the current methodology for VTEC detection is aimed at detecting sorbitol negative VTEC O157:H7 and for most European countries, the sorbitol positive outbreak strain O104:H4 could not be detected. It is therefore important that we examine the methods used by diagnostic and public health microbiology laboratories to characterise VTEC isolates and begin the process of global standardisation. A universal approach based on genomic features would be more generally applicable and transportable than current methods.

The medical care provision required to manage thousands of patients with haemolytic uremic syndrome (HUS) was a major challenge. Even the well-funded hospitals in Northern Germany were forced to loan dialysis and other medical equipment to manage the unexpected case load. Boosting diagnostic and epidemiological apparatus to improve source attribution during outbreaks is imperative to reduce the burden on already overstretched health care facilities. It is now clear that the ability to isolate and identify novel, *e.g.* non-O157:H7 VTECs, as well as known diarrhoeagenic *E. coli* (DEC) must be considered.

Strains with combinations of virulence factors from different *E. coli* pathotypes have been described before but it is the size and severity of the outbreak in Germany which has highlighted the importance and unpredictability of the consequences of genetic exchange amongst gut bacteria. This review will present what is currently known about the outbreak strain and discuss the preliminary genomic analysis

Figure 1. Spanning Tree of MLST data for EHEC and EAEC



Groupings that are highlighted represent the best known group of EHEC ST11, O157:H7, and two regions of the tree which contain almost exclusively EAEC or EHEC. Other regions of the tree contain both EHECs and EAECs which are distributed widely and evenly. The outbreak strain, *E. coli* ST678, which is positive for both VTEC and EAEC genes, is located in the region of the tree which is mainly non-ST11 EHECs. This suggests that the genes associated with VTEC or EAEC can be acquired by many different lineages of *E. coli* but that some lineages are more successful than others, probably after virulence genes are acquired, and so expand into clusters of one pathotype or the other. The combination of VTEC and EAEC genes in the ST678 *E. coli* led to a particularly virulent strain.

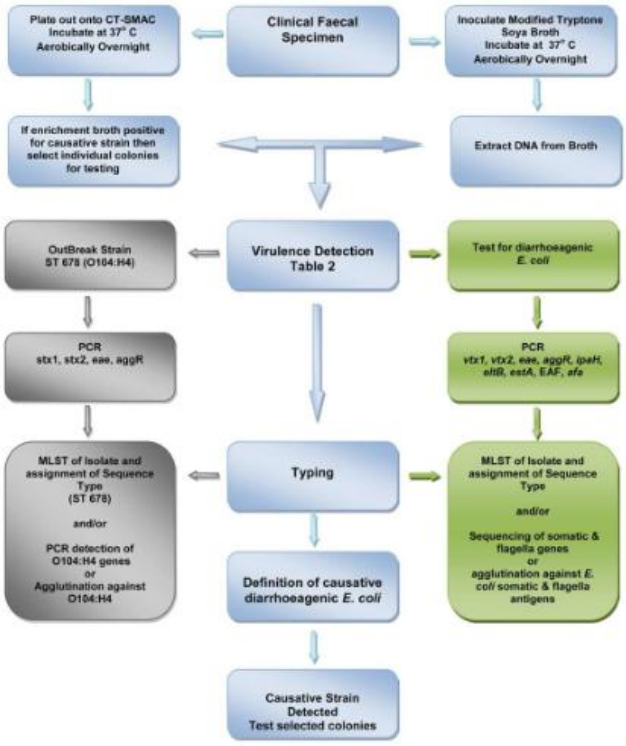
in the context of what is known about other isolates from the EAEC and VTEC pathogroups. It will also highlight an important lesson learned from this outbreak – the importance of a global epidemiological capacity, encompassing the developing world, to detect novel and emerging pathogens in addition to well-known ones.

The outbreak strain

Initial testing by German laboratories showed that the strain associated with the outbreak was of sequence type (ST) 678, serotype O104:H4, and contained genetic elements found in both EHEC (*vtx*) and EAEC (*aggR*). ST678 also contains the EAEC (55989) sequenced strain and the HUS causing O104:H4 VTEC (deposited on the public MLST database by Karch in 2001). The most closely related sequence type (the ST25 group) (Figure 1) is a VTEC O128:H2. The O128:H2 serotype (although no sequence type data is available) has been previously seen in sheep [1] and also isolated from infantile EPEC infections [2]. The group of *E. coli* most closely related to the outbreak strain are therefore a mixture of pathotypes: EAECs, EHECs and EPECs.

Although serotyping data in the MLST database is incomplete, it is clear that *E. coli* within the O104 serogroup occurs in several different unrelated STs. Therefore, this serogroup does not represent a related group of organisms and comparison of the outbreak strain with other O104 isolates might be redundant. However, including the flagella type, O104:H4 seems to describe a very closely related group of isolates, all within ST678. Members of this sequence type can belong to either the VTEC or the EAEC pathotypes (Figure 1) which can be associated with HUS in humans [3,4] but have not been commonly isolated in Europe. Possibly because detection techniques are optimised for VTEC O157:H7 in diagnostic laboratories. Mthods traditionally used to detect the commonly known VTEC O157 were not successful (O157 agglutination negative, sorbitol positive) and a combination of phenotypic and genotypic methods were necessary. The following sections describe the methods used at the reference laboratories

Figure 2. Flow chart of identification and typing procedures for diarrhoeagenic *e. coli*



Preliminary report may be generated after initial detection of the strain either from individual colonies or after enrichment. A final report of positive should only be sent after confirmation from individual colonies. Specimens should only be reported as negative after the enrichment broth has tested negative.

supplemented by published methods, and describe the potential for using genomic data.

Enrichment and isolation

Isolation directly from faecal specimens, stored at 4-8°C, was performed as soon as possible as viability of the organism decreases each day and plasmid loss may occur. An enrichment broth of Modified Tryptone Soya was inoculated with mixed faecal matter. The faeces were directly plated onto selective cefixime tellurite sorbitol MacConkey agar (CT-SMAC) and the EHEC ST678 (O104:H4) colonies grew very well producing a creamy pink morphology. The broth and agar plates were incubated aerobically at 37°C overnight. If CT-SMAC selective plates are unobtainable, other methods can be used such as exploiting the antibiotic resistant properties of this strain and using MacConkey agar supplemented with streptomycin (20 mg/ml) and/or tetracycline (10 mg/ml) [5]; commercial media is now available for this outbreak strain. Although useful for the O104:H4 outbreak strain, it is possible that other outbreaks will occur in which this supplemented media may not be appropriate.

Typing methods

Microbial typing relied on the isolation of the outbreak strain from faeces and this was straightforward because positive cases grew as an almost pure culture on CT-SMAC plates. Single colonies were picked and tested against O104:H7 antigens. Somatic antibodies (including O104) are available commercially. Molecular serotyping was performed using the O104 antigen-associated gene (*wzx*O104) and the gene encoding the H4 flagellar antigen (*fliC*H4) [6,7]. Antibodies raised against the O104 antigen are also positive with the K9 capsular antigen; therefore, O8:K9 and O9:K9 antigens can also be positive. Separate O8, O9 and O9a specific PCR has been carried out in other studies to rule out these other serotypes [8].

For serological typing of unknown isolates, screening against the whole panel of at least 185 somatic and 56 flagella *E. coli* antigens might be necessary. This is likely to be conducted by regional or international reference laboratories; the local testing laboratories can then purchase the antibodies for testing.

Alternatively, the genes that encode the specific O antigens in *E. coli* are clustered in the genome [9] and DNA sequencing can be used to predict the serotype [10].

Virulence detection

For detection of virulence factors by PCR, extraction of DNA was performed from the enrichment broth. Targets recommended for detecting the outbreak strain (Table 2 and Figure 2) include EHEC targets such as the rarely found *vtx1*, the commonly found *vtx2* and intimin (*eae*), an adhesion factor responsible for the attaching and effacing (A/E) lesions found in EHEC and EPEC [11]. The outbreak strain was *vtx2* positive and *vtx1*/*eae* negative.

Although a sub-typing scheme is available for *vtx1* and *vtx2* [12] and described for detecting the outbreak sub-type *vtx2a*, the variation at the nucleotide level is difficult to detect by PCR and needs careful optimisation. Detection using the generic *vtx2* primers and the presence of the EAEC plasmid with the absence of intimin was considered to be sufficient for the screening of the outbreak strain.

The PCR targets described for EAEC are not as stable as EHEC possibly because most are plasmid encoded; plasmids are variable and sometimes they may be lost completely during culture in the laboratory. Targets used for detecting the outbreak strain included a regulator (*agg*R) [13] of multiple EAEC virulence factors including an anti-aggregator transporter gene (*aat*) [14] and a dispersing protein (*aap*) that coats the bacterial surface [15], although this marker has also been found in other *E. coli* [16]. These gene products are linked in that they all play a role in the EAEC colonisation of the gut by aiding the translocation of dispersin across the membrane [17] and are usually found together.

Multilocus sequence typing

Multilocus sequence typing (MLST (http://www.mlst.net/) of seven gene loci (*adk*, *fum*C, *gyr*B, *mdh*, *pur*A & *rec*A) [18], define the outbreak strain as a member of ST 678. The combination of MLST and virulence marker targets (*vtx*1, *vtx*2, *eae, agg*R *etc*) is a robust and accessible test that can accurately identify strains of all *E. coli*, including unusual EHECs. Sequence type profiles should ideally be submitted to the public database so that we can start to gain an understanding of the *E. coli* pathotypes causing disease globally and enabling the assignment of new alleles and STs.

Detection of other EHEC strains

Focusing on the serotype of EHEC outbreak strains has led to a bias in laboratory testing for the detection of O157:H7. Recent research studies in both developing and developed countries have shown that non-O157 EHEC strains are prevalent and can be more dominant then O157:H7 in some geographical areas [19-22]. However, front-line laboratories still test only for VTEC O157 and so the true burden of non-O157:H7 EHEC is not known. For example, many O157 EHEC are sorbitol-negative, but the strain in this outbreak was not, nor did it react with the common antisera for EHEC, such as those recognizing O157, O26, O111 and H7 antigens.

A multiplex PCR has recently been described to specifically detect the most common toxin producing VTEC serogroups (O157, O103, O91, O113, O145, O111, and O26) [10] yet O104 had not been recognised as a potential pathogen. A microarray has also been designed to detect the most clinically relevant EHEC with additional targets for O104, O121, O118, O45 and O55 included [6]; however, microarray technology is not feasible to implement in most laboratories. A comprehensive selection of EHEC flagella antigens was included in these studies but they were selected from a historical prospective

Table 1. Nomenclature of Diarrhoeagenic *E.coli* and Enterohaemorrhagic *E. coli* toxin sub-types

Nomenclature

Abbreviation	Meaning	Description
DEC	Diarrhoeagenic *Escherichia coli*	Any defined group of *E. coli* which has been associated with the ability to cause diarrhoea
DEAC	Diffusely-adherent *Escherichia coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by a specific pattern of adherence using the HEp-2 cell assay
EAEC	Enteroaggregative *Escherichia coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by a specific pattern of aggregation using the HEp-2 cell assay
EIEC	Enteroinvasive *Escherichia coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by the presence of invasion genes also found in *Shigella*.
EPEC	Enteropathogenic *Escherichia coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea originally defined as specific serotypes and by a specific pattern of adherence using the HEp-2 cell assay but now by the presence of certain virulence factors including the locus of enterocyte effacement and associated effectors
ETEC	Enterotoxigenic *Escherichia coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by the presence of heat stable or heat labile toxins
VTEC	Verocytotoxic *Escherichia coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by the presence of a toxin gene, *vtx,* which has activity against cultured vero cells
STEC	Shiga Toxin-Producing *Escherichia coli*	A group of *E. coli* which been associated with the ability to cause diarrhoea defined by the presence of a toxin gene, *stx,* because of genetic similarity with the toxin of *Shigella dysenteriae*.
stx/vtx	Toxin genes	For *E. coli* these two gene names are synonymous – only in *Shigella dysenteriae* type 1 is *stx* used exclusively. The discussion about which name should be used revolves around the scientifically agreed use of the same gene name for genes which show homology (shared ancestry); *vtx1* and *stx* are homologues but for *vtx2*/*stx* this may not be true.
EHEC	Enterohemorrhagic *Escherichia coli*	VTEC/STEC patients that have the symptoms of bloody diarrhoea/ haemorrhagic colitis. This infection can lead to haemolytic uraemic syndrome (HUS) characterised by acute renal failure, haemolytic anaemia (anaemia due to haemolysis) and thrombocytopenia (low number of platelets).
*stx*1/*vtx*1	Toxin gene type 1	Several genetic variants including *vtx*1a, *vtx*1c, *vtx*1d
*stx*2/*vtx2*	Toxin gene type 2	Several genetic variants including *vtx*2a, *vtx*2b, *vtx*2c, *vtx*2d, *vtx*2e, *vtx*2f and *vtx*2g.

The terms *stx*/*vtx* or STEC/VTEC are entirely interchangeable and here we follow the European reference laboratories guidance and use *vtx* and VTEC.

Table 2. List of PCR targets with primers for Diarrhoeagenic *E.coli*

Target	Function	Pathotype	Primer Sequence	Reference
aat	Anti-aggregator transporter gene	EAEC	Forward: CTG GCG AAA GAC TGT ATC AT Reverse: CAA TGT ATA GAA ATC CGC TGT T	[58]
aggR	Regulator multiple EAEC virulence factors	EAEC	Forward CTA ATT GTA CAA TCG ATG TA Reverse: AGA GTC CAT CTC TTT GAT AAG	[59]
aap	Anti-aggregation protein (dispersin)	EAEC	Forward: CTT GGG TAT CAG CCT GAA TG Reverse: AAC CCA TTC GGT TAG AGC AC	[59]
elfB	Heal Liable Toxin	ETEC	Forward: TCT CTA TGT GCA TAC GGA GC Reverse: CCA TAC TGA TTG CCG CAA T	[60]
estA	Heat Stable Toxin	ETEC	Forward: AAT TTT MTT TCT GTA TTR TCT T Reverse: CAC CCG GTA CAR GCA GGA TT	[61]
Ipa-H	Invasion plasmid	EIEC	Forward: GTT CCT TGA CCG CCT TTC CGA TAC CGT C Reverse: GCC GGT CAG CCA CCC TCT GAG AGT AC	[62]
EAF	EPEC adherence factor	EPEC	Forward: CAG GGT AAA AGA AAG ATG ATA A Reverse: TAT GGG GAC CAT GTA TTA TCA	[63]
eaeA	attaching and effacing gene	EPEC/ EHEC	Forward: CTG AAC GGC GAT TAC GCG AA Reverse: CCA GAC GAT ACG ATC CAG	[64]
afa	Afrimbrial adhesion	DAEC	Forward: GCT GGG CAG CAA ACT GAT AAC TCT Reverse: CAT CAA GCT GTT TGT TCG TCC GCC G	[65]
Stx1	Shiga toxin 1	EHEC	Forward: TTTGTYACTGTSACAGCWGAAGCYTTACG Reverse: CCCCAGTTCARWGTRAGRTCMACRTC Probe : Cy5-CTGGATGATCTCAGTGGGCGTTCTTATGTAA-BHQ	[66]
eae	intimin	EHEC	Forward: CAT TGA TCA GGA TTT TTC TGG TGA TA Reverse: CTC ATG CGG AAA TAG CCG TTA Probe : Yak-ATAGTCTCGCCAGTATTCGCCACCAATACC-BHQ	[67]
Stx2	Shiga toxin 2	EHEC	Forward: TTTGTYACTGTSACAGCWGAAGCYTTACG Reverse: CCCCAGTTCARWGTRAGRTCMACRTC Probe : Yak-TCGTCAGGCACTGTCTGAAACTGCTCC-BHQ	[66]
rfbE(O157)	O157 antigen associated gene	VTEC O157	Forward: TTTCACACTTATTGGATGGTCTCAA Reverse: CGATGAGTTTATCTGCAAGGTGAT Probe : FAM-AGGACCGCAGAGGAAAGAGAGGAATTAAGG-BHQ	[66]
Additional Specific ST 678 (O104:H4) Outbreak Targets				
wzxO104	O104 antigen-associated gene	*E. coli*	Forward: TGTCGCGCAAAGAATTCAAC Reverse: AAAATCCTTTAAACTATACGCCC Probe : FAM-TTGGTTTTTTTGTATTAGCAATAAGTGGTGTC-BHQ	[6]
fliCH4	Gene encoding H4 flagella antigen	*E. coli*	Forward: GCTGGGGGTAAACAAGTCAA Reverse: CCAGTGCTTTTAACGGATCG Probe : FAM-TCTTACACTGACACCGCGTC-HEX	[7]
Stx2a	Sub-type of toxin 2a	EHEC	Forward: GCGATACTGRGBACTGTGGCC Reverse 3: CCGKCAACCTTCACTGTAAATGTG Reverse 2 : GGCCACCTTCACTGTGAATGTG	[12]

These are examples of published primers used. Ideally primers should be self designed and optimised to keep up with sequence variation.
EPEC: enteropathogenic *E. coli*; EHEC: enterohaemorrhagic *E. coli*; ETEC: enterotoxigenic *E. coli*; EIEC: enteroinvasive *E. coli*; DAEC: diffusely-adherent *E. coli*; EAEC: enteroaggregative *E. coli*.

and so emergence of unforeseen serological profiles such as H4 was unexpected. VTEC isolates can be of 60 O/H types [22] and to include all O/H types for detection on the front line is impractical. Relying on serological typing of common EHEC antigens has led to insufficient systems within the front-line laboratories to detect emerging EHEC outbreaks. The switch to molecular serotyping will overcome this problem but will always be problematic for new serotypes.

The importance of designing globally relevant tests to detect the virulence and background of circulating strains is clear and this outbreak has shown the adaptability and ability of *E. coli* to accumulate virulence genes. Therefore, several pathotypes of *E. coli* should now be considered when screening for EHEC strains to help identify emerging hybrid strains. To do this for verocytotoxin producing *E. coli* strains belonging to different serotype requires toxin assays or molecular identification of the toxin genes.

The best option for the majority of laboratories is DNA-based diagnostics for multiple DEC genes (Figure 2). Multiple genes must be sought since different strains can harbour different combinations of known virulence loci, especially in heterogeneous groups such as EAEC. Thus isolates such as the recent ST678 (O104:H4) outbreak strain can only be reliably identified using molecular methods. Molecular methods can also be used to track virulence genes in specimens or suspected sources that may no longer contain live organisms, an important feature for outbreak analyses. Moreover, the versatility of these methods means that they can be adapted when new strains appear, which is important because we cannot predict when or where a new hypervirulent *E. coli* strain will appear.

For detecting future EHEC strains, the same methodology used for the detection of the outbreak strain could be employed but with the addition of other DEC virulent targets: the invasive gene (*ipa*H) for enteroinvasive *E. coli* (EIEC); heat labile (*elt*B) and heat stable toxin (*est*A) for enterotoxigenic *E. coli* (ETEC); the enteropathogenic *E. coli* (EPEC) adherence factor (EAF) and the afrimbrial adhesion gene (*afa*) for diffusely adherent *E. coli* (DAEC) (Table 2 and Figure 2).

O104:H4 outbreak strain – genetic content

The O104:H4 outbreak was a verocytotoxin producing *E. coli* strain containing *vtx*2; however, the strain is different from many VTEC strains because it lacks both the locus for the enterocyte effacement (LEE) pathogenicity island and the EHEC virulence plasmid. The strain tested positive in initial screens for *agg*R, which encodes a transcriptional regulator of aggregative adherence genes and is located on the virulence plasmid of many EAEC strains. The strain has since been shown to possess an aggregative adherence plasmid and to demonstrate aggregative adherence; it carries an aggregative adherence plasmid as well as the verocytotoxin gene, the two elements that were of most interest to clinical microbiologists. However, there are also other multiple prophages, transposons and a number of horizontally-acquired antimicrobial resistance genes.

Enteroaggregative *E. coli* (EAEC)

Enteroaggregative *E. coli* (EAEC) is a large, diverse pathogroup of diarrhoeagenic *E. coli* (DEC) which was defined in 1987 when it was observed that some non-toxigenic strains of *E. coli* from cases of diarrhoea were not adhering to HEp-2 cells in the localised pattern typical of classical enteropathogenic *E. coli* (EPEC) but aggregated in a stacked brick formation [23,24]. Early research on EAEC linked these strains to persistent diarrhoea in children in developing countries but EAEC have since been shown to be an important cause of acute diarrhoea as well, and to be important in the etiology of intestinal infections in industrialized countries [25].

EAEC are known for their heterogeneity and although there are serotypes associated with this group, such as O44:H18, O111:H12, O125, and O126:H7 [25-29], they are not unique to EAEC. Studies have shown a wide selection of EAEC serotypes and many are untypeable [30-32]; therefore, serotyping is not a useful tool in distinguishing this problematic group.

The group contains organisms of multiple lineages [24] which harbour a virulence plasmid; because the HEp-2 assay is difficult to perform and interpret, it is detection of the virulence plasmid which forms the mainstay for identification and so diagnosis of the disease. The following are problems associated with the use of plasmid markers: plasmids have variable gene content; plasmids may be lost on sub-culture; and the plasmid may transfer and be detected in entirely unrelated bacteria which are not actually able to cause diarrhoea.

This group is a main cause of health costs in the developing world but its variable pathogenicity means that funding has not been a priority and comparative pangenome analysis of EAEC in relation

Table 3. Table of sequenced strains for Enteroaggregative *E. coli* and Enterohaemorrhagic *E. coli*

Strain Name	Accession Number	Genome size (MB)	Serotype	ST	AAF	Phylotype	VT
042	FN554766	5.35	O44:H18	414	II	D	N/A
101-1	AAMK00000000	4.98	ONT:H10	493	II	B2	N/A
55989	CU928145.2	5.15	Unknown	678	III	B1	N/A
H112180280	AFPN00000000	5.5	O104:H4	678	I	B1	*stx2*
O111	AP010960.1	5.80	O111:H-	16	N/A	B1	*stx1/stx2*
O26	AP010953.1	5.86	O26:H11	21	N/A	B1	*stx2*
O103	AP010958.1	5.48	O103:H2	17	N/A	B1	*stx1/stx2*
Sakai	BA000007.2	5.60	O157:H7	11	N/A	E	*stx1/stx2*

to other *E. coli* pathotypes and commensals has not been extensively conducted. It has therefore not been possible to define unique stable chromosomal markers for identification. One chromosomal maker (also known to be plasmid encoded) is the *pic* gene which is present in the sequenced 042 strain [33]. This gene is a multi-functional secreted protease but is not unique in the Enterobacteriaceae. Flanking sequence around this gene in EAEC and *Shigella* is different, suggesting that this gene has been acquired by horizontal transfer [34].

These problems with diagnostics have resulted in a poor understanding of this heterogeneous pathotype, which has in turn led to a lack of knowledge of its true burden and impact on human health. Despite ample evidence that EAEC is the most common DEC [35-39], it remains less well-known compared to EPEC, EIEC AND ETEC (Table 1).

Outbreaks of EAEC

There have been some reports of this organism being associated with outbreaks, the largest of which was in Japan in 1993 when 2,697 schoolchildren became ill after eating food contaminated with EAEC with their lunch [40]. Although evidence pointed to white radish sprouts in the stir-fried vegetables, the bacteria were never isolated from the most likely food source. Multiple outbreaks in association with EAEC have been reported in the United Kingdom in association with public functions such as restaurants, hotels and conference centres [41]. EAEC has also caused outbreaks in hospitals (in Serbia 19 babies were infected in a neonatal ward [42]); from well water (in India 20 cases were reported including multiple age groups [43]); and from food (seen in 24 cases in an Italian holiday resort associated with cheese made with unpasteurized sheep milk). Furthermore, they may be an animal reservoir for some EAEC strains [44]. There is no common serotype associated with EAEC outbreaks.

Genomics of EAEC

At least four EAEC genomes have been completed, are nearly completed, or are in progress (Table 3). EAEC strain 042 produced diarrhoea in three of five adult volunteers in a challenge study in which other EAEC strains tested did not produce symptoms [45]. Strain 101-1 was responsible for the largest documented EAEC outbreak prior to 2011 [40]. Interestingly, although the 101-1 did harbour some virulence genes seen in strain 042, its presumed hypervirulence as suggested by the outbreak remained enigmatic for several years. Recent data demonstrates that in addition to multiple horizontally acquired virulence genes, strain 101-1 harbours a pathoadaptive mutation [46]. Like *Shigella* and some EHEC lineages, it has lost the lysine decarboxylase or *cad* genes. Inserting these genes onto the chromosome of 101-1 attenuates the strain. [46]. The genome of strain 55989 (source) is also in progress; of the four fully or partially sequenced EAEC genomes begun in June, it is this strain that shares the most genomic sequence with the ST 678 (O104:H4) outbreak isolate whose draft genome sequence was completed in June.

Enterohemorrhagic *E. coli* (EHEC)

Enterohemorrhagic *E. coli* (EHEC) causes haemorrhage of the intestinal tract of humans. The mechanism for this is complex but, for EHEC infection, always involves a toxin [47] called verotoxin (*vtx*) or shigatoxin (*stx*) (Table 1). Originally described as a rare *E. coli* serotype in 1983 [48] causing hemorrhagic colitis, O157:H7 VTEC, a cow-adapted *E. coli*, has since expanded in the bovine population and spill-over into humans, associated with disease, is such that it is currently the

Table 4. List of Useful Links

Links to other resources

There are many useful links which describe information in relation to the outbreak strain including

- Health Protection Agency (http://www.hpa.org.uk/)
- Robert Koch Institute (http://www.rki.de/EN/Home/homepage__node.html)
- Eurosurveillance (http://www.eurosurveillance.org/)
- European food safety authority (http://www.efsa.europa.eu/)
- Centers for disease control and prevention (http://www.cdc.gov/index.htm)
- World Health Organisation (http://www.who.int/en/)
- NCBI BioProject page (http://www.ncbi.nlm.nih.gov/bioproject/68275)
- github repository for the "crowdsharing" efforts (https://github.com/ehec-outbreak-crowdsourced/)

most commonly isolated EHEC (ST11 complex). Another commonly isolated sub-type is the ST21 complex EHECs which are predominantly serotype O26:H11 but may also be O111:H- or O111:H8. The outbreak EHEC ST678 strain clusters away from these "common" EHECs but clusters, as a double locus variant, with EHECs of serotype 0128:H2. It is clear that EHEC, as with EAEC, represent a diverse group of *E. coli* which have acquired virulence genes on several different occasions (Figure 1). It is not just the virulence genes, but also the background into which the virulence genes are acquired, which results in the ability of a strain to cause disease and spread; adherence and toxin production have both been implicated for EHEC. Cases of EHEC infection normally present to health facilities as bloody diarrhoea although more severe complications can occur. The frequency of these complications is dependent on the toxin encoded; the presence of *vtx2a* has been shown to be associated with a more virulent infection [49] partly due to increased expression [50]. The ability to adhere to intestinal cells has also been shown to be associated with virulence and although EHEC, as with EPEC, normally adhere using the LEE [47], there are other mechanisms of attachment within *E. coli* and outbreaks have been caused by several different lineages of EHEC using non-LEE mediated attachment. For the outbreak strain, adherence is presumably mediated via the acquired EAEC virulence factors. It is possible that this adherence is more effective than LEE mediated adherence and so may explain why the outbreak strain caused such a virulent infection.

Outbreaks of EHEC

EHEC outbreaks are more often reported from industrialised countries than from developing countries because surveillance and reporting systems are in place. The most common type from outbreaks is O157:H7 which was responsible for one of the largest outbreaks which included 106 HUS cases from 2,764 confirmed infections in Japan in 1996 [51]. Outbreaks caused by non-O157 EHEC have for several years been highlighted as a potential risk [52] and are well documented again in some industrialised countries [53]. One of the common non-O157 VTECs in the USA is O111:H8 and one of the largest outbreaks was caused by an EHEC O111 (ST and H group not given) in the USA in 2008 causing 341 illnesses [54]. Another strain of EHEC O111:H2 (unknown ST) caused an outbreak in 1998 [55] and had features very similar to those of the German outbreak strain: *eae* negative, EAEC aggregative adherence, and associated with HUS. In 2007 there was an outbreak of EHEC in which five children were infected by two serotypes (O145 and O26) from consumption of ice-cream produced from a Belgium farm [56], perhaps showing the widespread nature of non-O157 EHECs and emphasising their potential to contaminate food handled by people. Thus highly virulent non-O157 *E. coli* has been circulating for some time but the potential impact may not be fully appreciated.

Genomics of EHEC

It is clear that EHEC, as with EAEC, is a heterogeneous group of DECs defined by a virulence factor (*vtx*). The best studied, single locus variants of ST11 share the serotype O157:H7 and show a

conserved genome containing around 1.5 Mb of horizontally acquired DNA which includes a type III secretion system and effectors, the LEE. The LEE contains around 30 coding sequences in 5 operons and encodes the ability of both EPEC and EHEC to attach to the gut and cause disease. Several (currently 24) non-LEE effectors have been described [47] for which the cellular function is being investigated. There is some variability within the EHEC ST11 (O157:H7) group in toxins (Table 1) and in the other accessory genes. This suggests that acquisition has occurred on several occasions and that the genomes of these closely related bacteria are dynamically exchanging DNA with other gut flora. It is believed that O157:H7 as a group evolved from the O55 ancestor, after the horizontal acquisition of genes encoding the O157 antigen, and then branched into two lineages O157:H7 and O157:H-. There has been little radiation in human isolates since this occurred and comparison with the cattle strains (the normal host for ST11 VTEC) suggests that it is a limited subset of cattle-adapted strains which cause infection in humans. This may be due to the source-sink nature of the population dynamics. The source is cattle which support the majority of the bacterial population whilst spill-over into the human population occurs with a restricted set of strains that have the ability to shed in high numbers from cattle and/or to amplify in the environment as well as the ability to colonise and cause disease in humans.

There are several non-O157 EHECs now described and there is sequence data available for ST/serotype: ST21/O26:H11, ST16/O111:H- and ST17/O103:H2 (Table 3). These non-O157 EHECs are from different lineages and yet contain a set of relatively conserved accessory genes [57]. Although analysis of the accessory genome suggests that selective forces within the same environment have led to the acquisition and maintenance of a similar accessory gene content (parallel evolution), there is high level clustering of several of the non-O157 EHECs suggesting a common ancestry. It seems likely that some *E. coli* lineages acquire genetic material via horizontal exchange more often than others; however, whether this is driven by a pathogenic lifestyle [18] or whether pathogens have emerged from strains with a commensal lifestyle within such lineages is not clear. What is clear is that there are many diverse *E. coli* in which *vtx* genes have been found but it is only those that can also adhere to the intestine which will remain in the *E. coli* population and come to our notice as a cause of infectious disease in humans or animals. The latest of these emergent *E. coli* caused the massive outbreak of HUS in Germany (see the useful links in Table 4).

The emergence of new pathogenic *E. coli*

The distribution of EAEC and EHEC across the tree drawn from MLST (Figure 1) suggests that both pathotypes have arisen on several occasions from several ancestral strains. However, there are clear patterns in the ancestry; the majority of VTECs are clustered around two STs: ST11 (O157:H7) and ST21 (O26:H11). The ST678 (O104:H4) strain from the outbreak clusters with other non-O157 VTECs, possibly around ST25, but most closely with the strain 55989 (also ST678 and also an EAEC). However, this is away from most EAEC isolates suggesting that the acquisition of the plasmid encoding the EAEC phenotype has occurred independently into the ST678 lineage and is not a previously widespread EAEC strain. It seems likely that the emergence of "new" VTECs will be from lineages of *E. coli* which have the ability to adhere to the gut of an animal host, which may be human, either by the mechanisms classically shown by EPEC (the LEE) or by the virulence plasmid of EAEC.

Conclusion

Although EHEC is the best documented and most severe of the DEC it is not the most common cause of *E. coli* diarrhoea. The importance of EAEC as the most common causative agent is now being realised. This may be due to improved testing rather than a recent increase in identified cases, in which case as testing improves further so will estimates of the burden of EAEC disease. The recent combination of EAEC and EHEC virulence factors in a single outbreak strain causing such severe disease emphasises the importance of being able to detect all DEC using appropriate genetic methods and not to just focus on *E. coli* O157:H7. Routine surveillance systems for EAEC, a once ignored global pathogen, would go a long way to reaching this goal.

Acknowledgements

Thank you to all members of the Gastrointestinal Infections Reference Unit at Colindale and Flemming Scheutz for valuable discussion and to the scientists who released sequence data on the outbreak strain into the public domain. Thank you to the authors of published MLST data [18,24,32] which was used for the analysis in this paper.

References

1. Beutin L, Geier D, Steinruck H, Zimmermann S, Scheutz F (1993) Prevalence and some properties of verotoxin (Shiga-like toxin)-producing Escherichia coli in seven different species of healthy domestic animals. J Clin Microbiol 31: 2483-2488.
2. Knutton S, Lloyd DR, McNeish AS (1987) Adhesion of enteropathogenic Escherichia coli to human intestinal enterocytes and cultured human intestinal mucosa. Infect Immun 55: 69-77.
3. Bae WK, Lee YK, Cho MS, Ma SK, Kim SW, Kim NH, Choi KC (2006) A case of hemolytic uremic syndrome caused by Escherichia coli O104:H4. Yonsei Med J 47: 437-439.
4. Mellmann A, Bielaszewska M, Kock R, Friedrich AW, Fruth A, Middendorf B, Harmsen D, Schmidt MA, Karch H (2008) Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic Escherichia coli. Emerg Infect Dis 14: 1287-1290.
5. Scheutz F, Nielsen EM, Frobe I, Frimodt-Moller J, Boisen N, Morabito S, Tozzoli R, Nataro J, Caprioli A (2011) Characteristics of the enteroaggregative Shiga toxin/verotoxin-producing Escherichia coli O104:H4 strain causing the outbreak of haemolytic uraemic syndrome in Germany, May to June 2011. Euro Surveill 16: 19889.
6. Bugarel M, Beutin L, Martin A, Gill A, Fach P (2010) Micro-array for the identification of Shiga toxin-producing Escherichia coli (STEC) seropathotypes associated with Hemorrhagic Colitis and Hemolytic Uremic Syndrome in humans. Int J Food Microbiol 142: 318-329.
7. EU Reference Laboratory for E.coli Department of Veterinary Public Health and Food Safety (2011) Detection and identification of Verocytotoxin-producing Escherichia coli (VTEC) O104:H4 in food by Real Time PCR. Available from: [Accessed 11 Jun 2011] http://www.iss.it/binary/vtec/cont/Lab_Proc_VTEC_O104.pdf
8. Wang L, Briggs CE, Rothemund D, Fratamico P, Luchansky JB, Reeves PR (2001) Sequence of the E. coli O104 antigen gene cluster and identification of O104 specific genes. Gene 270: 231-236.
9. Reeves PR, Hobbs M, Valvano MA, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz CR, Rick PD (1996) Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol 4: 495-503.
10. Debroy C, Roberts E, Valadez AM, Dudley EG, Cutter CN (2011) Detection of Shiga Toxin-Producing Escherichia coli O26, O45, O103, O111, O113, O121, O145, and O157 Serogroups by Multiplex Polymerase Chain Reaction of the wzx Gene of the O-Antigen Gene Cluster. Foodborne Pathog Dis 8: 651-652.
11. Law D(2000) Virulence factors of Escherichia coli O157 and other Shiga toxin-producing E. coli. J Appl Microbiol 88: 729-745.
12. WHO Collaborating Centre for Reference and Research on Escherichia and Klebsiella(2011) Identification of three vtx1 and seven vtx2 subtypes of Verocytotoxin encoding genes of Escherichia coli by conventional PCR amplification. Copenhagen: Statens Serum Institut..Available from: [Accessed 11 Jun 2011] http://www.ssi.dk/English/PublicHealth/
13. Morin N, Tirling C, Ivison SM, Kaur AP, Nataro JP, Steiner TS (2010) Autoactivation of the AggR regulator of enteroaggregative Escherichia coli in vitro and in vivo. FEMS Immunol Med Microbiol 58: 344-355.
14. Baudry B, Savarino SJ, Vial P, Kaper JB, Levine MM (1990) A sensitive and specific DNA probe to identify enteroaggregative Escherichia coli, a recently discovered diarrheal pathogen. J Infect Dis 161: 1249-1251.
15. Sheikh J, Czeczulin JR, Harrington S, Hicks S, Henderson IR, Le BC, Gounon P, Phillips A, Nataro JP (2002) A novel dispersin protein in enteroaggregative Escherichia coli. J Clin Invest 110: 1329-1337.
16. Monteiro BT, Campos LC, Sircili MP, Franzolin MR, Bevilacqua LF, Nataro JP, Elias WP (2009) The dispersin-encoding gene (aap) is not restricted to enteroaggregative Escherichia coli. Diagn Microbiol Infect Dis 65: 81-84.
17. Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V, Boutin A, Rose DJ, Blattner FR, Nataro JP (2003) The export of coat protein from enteroaggregative Escherichia coli by a specific ATP-binding cassette transporter system. J Biol Chem 278: 45680-45689.
18. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M (2006) Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol 60: 1136-1151.
19. Bonyadian M, Momtaz H, Rahimi E, Habibian R, Yazdani A, Zamani M (2010) Identification & characterization of Shiga toxin-producing Escherichia coli isolates from patients with diarrhoea in Iran. Indian J Med Res 132: 328-331.
20. Burnens AP, Boss P, Orskov F, Orskov I, Schaad UB, Muller F, Heinzle R, Nicolet J (1992) Occurrence and phenotypic properties of verotoxin producing Escherichia coli in sporadic cases of gastroenteritis. Eur J Clin Microbiol Infect Dis 11: 631-634.
21. Kehl SC(2002) Role of the laboratory in the diagnosis of enterohemorrhagic Escherichia coli infections. J Clin Microbiol 40: 2711-2715.
22. Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, Blanco J (2004) Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing Escherichia coli isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J Clin Microbiol 42: 311-319.
23. Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, Levine MM (1987) Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. Pediatr Infect Dis J 6: 829-831.
24. Okeke IN, Wallace-Gadsden F, Simons HR, Matthews N, Labar AS, Hwang J, Wain J (2010) Multi-locus sequence typing of enteroaggregative Escherichia coli isolates from Nigerian children uncovers multiple lineages. PLoS One 5: e14093.
25. Okeke IN, Nataro JP (2001) Enteroaggregative Escherichia coli. Lancet Infect Dis 1: 304-313.
26. Shazberg G, Wolk M, Schmidt H, Sechter I, Gottesman G, Miron D (2003) Enteroaggregative Escherichia coli serotype O126:H27, Israel. Emerg Infect Dis 9: 1170-1173.
27. Abe CM, Knutton S, Pedroso MZ, Freymuller E, Gomes TA (2001) An enteroaggregative Escherichia coli strain of serotype O111:H12 damages and invades cultured T84 cells and human colonic mucosa. FEMS Microbiol Lett 203: 199-205.
28. Chart H, Spencer J, Smith HR, Rowe B (1997) Magnesium ions are required for HEp-2 cell adhesion by

enteroaggregative strains of Escherichia coli O126:H27 and O44:H18. FEMS Microbiol Lett 148: 49-52.
29. Smith HR, Scotland SM, Willshaw GA, Rowe B, Cravioto A, Eslava C (1994) Isolates of Escherichia coli O44:H18 of diverse origin are enteroaggregative. J Infect Dis 170: 1610-1613.
30. Jenkins C, Tembo M, Chart H, Cheasty T, Willshaw GA, Phillips AD, Tompkins D, Smith H (2006) Detection of enteroaggregative Escherichia coli in faecal samples from patients in the community with diarrhoea. J Med Microbiol 55: 1493-1497.
31. Jenkins C, Chart H, Willshaw GA, Cheasty T, Smith HR (2006) Genotyping of enteroaggregative Escherichia coli and identification of target genes for the detection of both typical and atypical strains. Diagn Microbiol Infect Dis 55: 13-19.
32. Kaur P, Chakraborti A, Asea A (2010) Enteroaggregative Escherichia coli: An Emerging Enteric Food Borne Pathogen. Interdiscip Perspect Infect Dis 2010: 254159.
33. Chaudhuri RR, Sebaihia M, Hobman JL, Webber MA, Leyton DL, Goldberg MD, Cunningham AF, Scott-Tucker A, Ferguson PR, Thomas CM, Frankel G, Tang CM, Dudley EG, Roberts IS, Rasko DA, Pallen MJ, Parkhill J, Nataro JP, Thomson NR, Henderson IR (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative Escherichia coli strain 042. PLoS One 5: e8801.
34. Henderson IR, Czeczulin J, Eslava C, Noriega F, Nataro JP (1999) Characterization of pic, a secreted protease of Shigella flexneri and enteroaggregative Escherichia coli. Infect Immun 67: 5587-5596.
35. Weintraub A(2007) Enteroaggregative Escherichia coli: epidemiology, virulence and detection. J Med Microbiol 56: 4-8.
36. Tompkins DS, Hudson MJ, Smith HR, Eglin RP, Wheeler JG, Brett MM, Owen RJ, Brazier JS, Cumberland P, King V, Cook PE (1999) A study of infectious intestinal disease in England: microbiological findings in cases and controls. Commun Dis Public Health 2: 108-113.
37. Presterl E, Zwick RH, Reichmann S, Aichelburg A, Winkler S, Kremsner PG, Graninger W (2003) Frequency and virulence properties of diarrheagenic Escherichia coli in children with diarrhea in Gabon. Am J Trop Med Hyg 69: 406-410.
38. Knutton S, Shaw R, Phillips AD, Smith HR, Willshaw GA, Watson P, Price E (2001) Phenotypic and genetic analysis of diarrhea-associated Escherichia coli isolated from children in the United Kingdom. J Pediatr Gastroenterol Nutr 33: 32-40.
39. Araujo JM, Tabarelli GF, Aranda KR, Fabbricotti SH, Fagundes-Neto U, Mendes CM, Scaletsky IC (2007) Typical enteroaggregative and atypical enteropathogenic types of Escherichia coli are the most prevalent diarrhea-associated pathotypes among Brazilian children. J Clin Microbiol 45: 3396-3399.
40. Itoh Y, Nagano I, Kunishima M, Ezaki T (1997) Laboratory investigation of enteroaggregative Escherichia coli O:untypeable:H10 associated with a massive outbreak of gastrointestinal illness. J Clin Microbiol 35: 2546-2550.
41. Smith HR, Cheasty T, Rowe B (1997) Enteroaggregative Escherichia coli and outbreaks of gastroenteritis in UK. Lancet 350: 814-815.
42. Cobeljic M, Miljkovic-Selimovic B, Paunovic-Todosijevic D, Velickovic Z, Lepsanovic Z, Zec N, Savic D, Ilic R, Konstantinovic S, Jovanovic B, Kostic V (1996) Enteroaggregative Escherichia coli associated with an outbreak of diarrhoea in a neonatal nursery ward. Epidemiol Infect 117: 11-16.
43. Pai M, Kang G, Ramakrishna BS, Venkataraman A, Muliyil J (1997) An epidemic of diarrhoea in south India caused by enteroaggregative Escherichia coli. Indian J Med Res 106: 7-12.
44. Scavia G, Staffolani M, Fisichella S, Striano G, Colletta S, Ferri G, Escher M, Minelli F, Caprioli A (2008) Enteroaggregative Escherichia coli associated with a foodborne outbreak of gastroenteritis. J Med Microbiol 57: 1141-1146.
45. Nataro JP, Deng Y, Cookson S, Cravioto A, Savarino SJ, Guers LD, Levine MM, Tacket CO (1995) Heterogeneity of enteroaggregative Escherichia coli virulence demonstrated in volunteers. J Infect Dis 171: 465-468.
46. Hwang J, Mattei LM, VanArendonk LG, Meneely PM, Okeke IN (2010) A pathoadaptive deletion in an enteroaggregative Escherichia coli outbreak strain enhances virulence in a Caenorhabditis elegans model. Infect Immun 78: 4068-4076.
47. Wong AR, Pearson JS, Bright MD, Munera D, Robinson KS, Lee SF, Frankel G, Hartland EL (2011) Enteropathogenic and enterohaemorrhagic Escherichia coli: even more subversive elements. Mol Microbiol 80: 1420-1438.
48. Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML (1983) Hemorrhagic colitis associated with a rare Escherichia coli serotype. N Engl J Med 308: 681-685.
49. Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, Mladonicky JM, Somsel P, Rudrik JT, Dietrich SE, Zhang W, Swaminathan B, Alland D, Whittam TS (2008) Variation in virulence among clades of Escherichia coli O157:H7 associated with disease outbreaks. Proc Natl Acad Sci U S A 105: 4868-4873.
50. Abu-Ali GS, Ouellette LM, Henderson ST, Lacher DW, Riordan JT, Whittam TS, Manning SD (2010) Increased adherence and expression of virulence genes in a lineage of Escherichia coli O157:H7 commonly associated with human infections. PLoS One 5: e10167.
51. Pennington H(2010) Escherichia coli O157. Lancet 376: 1428-1435.
52. Johnson KE, Thorpe CM, Sears CL (2006) The emerging clinical importance of non-O157 Shiga toxin-producing Escherichia coli. Clin Infect Dis 43: 1587-1595.
53. Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, Strockbine NA (2005) Non-O157 Shiga toxin-producing Escherichia coli infections in the United States, 1983-2002. J Infect Dis 192: 1422-1429.
54. Piercefield EW, Bradley KK, Coffman RL, Mallonee SM (2010) Hemolytic Uremic Syndrome After an Escherichia coli O111 Outbreak. Arch Intern Med 170: 1656-1663.
55. Morabito S, Karch H, Mariani-Kurkdjian P, Schmidt H, Minelli F, Bingen E, Caprioli A (1998) Enteroaggregative, Shiga toxin-producing Escherichia coli O111:H2 associated with an outbreak of hemolytic-uremic syndrome. J Clin Microbiol 36: 840-842.
56. De SK, Buvens G, Posse B, Van den Branden D, Oosterlynck O, De ZL, Eilers K, Pierard D, Dierick K, Van Damme-Lombaerts R, Lauwers C, Jacobs R (2008) Outbreak of verocytotoxin-producing E. coli O145 and O26 infections associated with the consumption of ice cream produced at a farm, Belgium, 2007. Euro Surveill 13: 8041.

57. 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 20;106: 17939-17944.
58. Schmidt H, Knop C, Franke S, Aleksic S, Heesemann J, Karch H (1995) Development of PCR for screening of enteroaggregative Escherichia coli. J Clin Microbiol 33: 701-705.
59. Cerna JF, Nataro JP, Estrada-Garcia T (2003) Multiplex PCR for detection of three plasmid-borne genes of enteroaggregative Escherichia coli strains. J Clin Microbiol 41: 2138-2140.
60. Blomen I, Lofdahl S, Stenstrom TA, Norberg R (1993) Identification of enterotoxigenic Escherichia coli isolates, a comparison of PCR, DNA hybridisation, ELISAs and bioassays. J Microbiol Methods 17: 181-191.
61. Stacy-Phipps S, Mecca JJ, Weiss JB (1995) Multiplex PCR assay and simple preparation method for stool specimens detect enterotoxigenic Escherichia coli DNA during course of infection. J Clin Microbiol 33: 1054-1059.
62. Sethabutr O, Venkatesan M, Murphy GS, Eampokalap B, Hoge CW, Echeverria P (1993) Detection of Shigellae and enteroinvasive Escherichia coli by amplification of the invasion plasmid antigen H DNA sequence in patients with dysentery. J Infect Dis 167: 458-461.
63. Franke J, Franke S, Schmidt H, Schwarzkopf A, Wieler LH, Baljer G, Beutin L, Karch H (1994) Nucleotide sequence analysis of enteropathogenic Escherichia coli (EPEC) adherence factor probe and development of PCR for rapid detection of EPEC harboring virulence plasmids. J Clin Microbiol 32: 2460-2463.
64. Reid SD, Betting DJ, Whittam TS (1999) Molecular detection and identification of intimin alleles in pathogenic Escherichia coli by multiplex PCR. J Clin Microbiol 37: 2719-2722.
65. Le BC, Archambaud M, Labigne A (1992) Rapid and specific detection of the pap, afa, and sfa adhesin-encoding operons in uropathogenic Escherichia coli strains by polymerase chain reaction. J Clin Microbiol 30: 1189-1193.
66. Perelle S, Dilasser F, Grout J, Fach P (2004) Detection by 5'-nuclease PCR of Shiga-toxin producing Escherichia coli O26, O55, O91, O103, O111, O113, O145 and O157:H7, associated with the world's most frequent clinical cases. Mol Cell Probes 18: 185-192.
67. Nielsen EM, Andersen MT (2003) Detection and characterization of verocytotoxin-producing Escherichia coli by automated 5' nuclease PCR assay. J Clin Microbiol 41: 2884-2893.

Corresponding Author
Dr John Wain
Laboratory of Gastrointestinal Pathogens
Health Protection Agency, Colindale
61, Colindale Avenue
London NW9 5HT
United Kingdom
John.wain@hpa.org.uk

Conflict of Interests: No conflict of interests is declared.

7.8 Characterization of a verocytotoxin-producing enteroaggregative Escherichia coli serogroup O111:H21 strain associated with a household outbreak in Northern Ireland.

Dallman T, Smith GP, O'Brien B, Chattaway MA, Finlay D, Grant KA and Jenkins C. 2012. J Clin Microbiol 50 (2012) 4116-4119.

JCM
Journals.ASM.org

Characterization of a Verocytotoxin-Producing Enteroaggregative *Escherichia coli* Serogroup O111:H21 Strain Associated with a Household Outbreak in Northern Ireland

Tim Dallman,[a] Geoffrey P. Smith,[b] Brendan O'Brien,[c] Marie A. Chattaway,[a] David Finlay,[d] Kathie A. Grant,[a] and Claire Jenkins[a]

Laboratory of Gastrointestinal Pathogens, Health Protection Agency, London, United Kingdom[a]; Illumina Cambridge Ltd., Chesterford Research Park, Saffron Walden, Essex, United States[b]; Public Health Agency, Belfast, United Kingdom[c]; and Agri-Food and Biosciences Institute, Veterinary Sciences Division, Belfast, United Kingdom[d]

A strain of *Escherichia coli* O111:H21 recently isolated in the United Kingdom harbored the phage-encoded *vtx2c* gene and the aggregative adherence plasmid. Although exhibiting the same pathogenic profile as the *E. coli* O104:H4 strain linked to the outbreak in Germany, there were important differences in strain characteristics and in the epidemiological setting.

Verocytotoxin-producing *Escherichia coli* (VTEC) strains, a well-known cause of hemolytic uremic syndrome (HUS) and bloody diarrhea, are defined by the presence of one or both phage-encoded verocytotoxin genes, *vtx1* and *vtx2* (10). Enteroaggregative *E. coli* (EAggEC) strains are a heterogeneous group defined by their stacked-brick aggregating adherence to HEp-2 cells (16). Those with proven pathogenicity typically harbor a set of plasmid genes encoding aggregative adherence fimbriae (AAF), a dispersin protein (Aap), and an aggregative adherence transport (Aat) protein, all regulated by a transcription factor encoded by the *aggR* gene (17).

In May 2011, a large food-borne outbreak in Germany, where the source was identified as fenugreek seeds, was associated with a strain of *E. coli* O104:H4 (3, 5). Molecular analysis of this isolate with PCR detected a *vtx2*-positive, *eae*-negative VTEC strain which was subsequently found to possess typical molecular characteristics of EAggEC (1). Due in part to the size of the outbreak but also to the rapid availability of the genome sequence, this isolate has been described in detail in the literature (14, 18). Prior to this outbreak, two strains with the same pathogenic profile had been described: *E. coli* O111:H2 that caused an outbreak of HUS in France in 1996 (15) and *E. coli* O86:NM, associated with a case of HUS in Japan (7).

In February 2012, a stool specimen from a 3-year-old female with HUS was submitted to the Laboratory of Gastrointestinal Pathogens at the Health Protection Agency in Colindale, for analysis, having been found negative for nonsorbitol-fermenting *E. coli* O157 and other common enteric bacteria, at the frontline diagnostic laboratory. The child was extremely unwell, with severe cerebral involvement, and required admission to a pediatric intensive care unit. At the reference laboratory, the fecal specimen was tested according to the protocol described previously (8). An isolate of *E. coli* grown from the fecal specimen was found to be positive by PCR for the *vtx2* and *aggR* genes. The strain (designated 226) was identified biochemically and serotyped as *E. coli* O111:H21 (6).

Prompt notification to the local public health authority facilitated rapid implementation of control measures. Screening of household contacts revealed the child's mother and 4-year-old male sibling each carried a strain of *E. coli* O111:H21 with the same pathogenic profile. Both contacts described had diarrhea, but neither were admitted to the hospital. Enhanced surveillance was initiated, but no additional cases were identified either microbiologically or via the British Pediatric HUS Surveillance program.

TABLE 1 Comparison of key genomic properties of VTEC/EAEC O111:H21 (226) and VTEC/EAEC O104:H4 (280)

Strain characteristic	Result	
	VTEC/EAEC O111:H21 (226)	VTEC/EAEC O104:H4 (280)
Sequence type	40	678
Plasmid size	102 kb	75 kb
Shiga toxin	Stx2c	Stx2a
Plasmid-encoded EAEC genes	Aat complex genes, *aap*, *aggR*, SepA gene, and AAF/V gene	Aat complex genes, *aap*, *aggR*, SepA gene, and AAF/I gene
Plasmid incompatibility group	FII2 and FIB most closely related to FIB10	FIB25
SPATES	SepA, Sat, and Pic	SepA, SigA, and Pic
Antibiotic resistance profile	Ampicillin	Extended-spectrum beta-lactamase producer, TN21 drug resistance island

The index case lived on a farm with her parents and sibling, where the livestock consisted of dairy and beef cattle herds totaling 102 animals; there were no other ruminants. The farm yard was very compact, with the cattle sheds in close proximity to the farm house. Fecal samples were obtained from 29 cattle, including the animal with a recent history of mastitis. The agricultural samples were processed in the regional veterinary laboratory at the Agri-Food and Biosciences Institute. All farm samples were examined by a standardized immunomagnetic bead procedure using beads

Received 10 August 2012 Returned for modification 6 September 2012
Accepted 23 September 2012
Published ahead of print 3 October 2012
Address correspondence to Claire Jenkins, claire.jenkins@hpa.org.uk.

doi:10.1128/JCM.02047-12

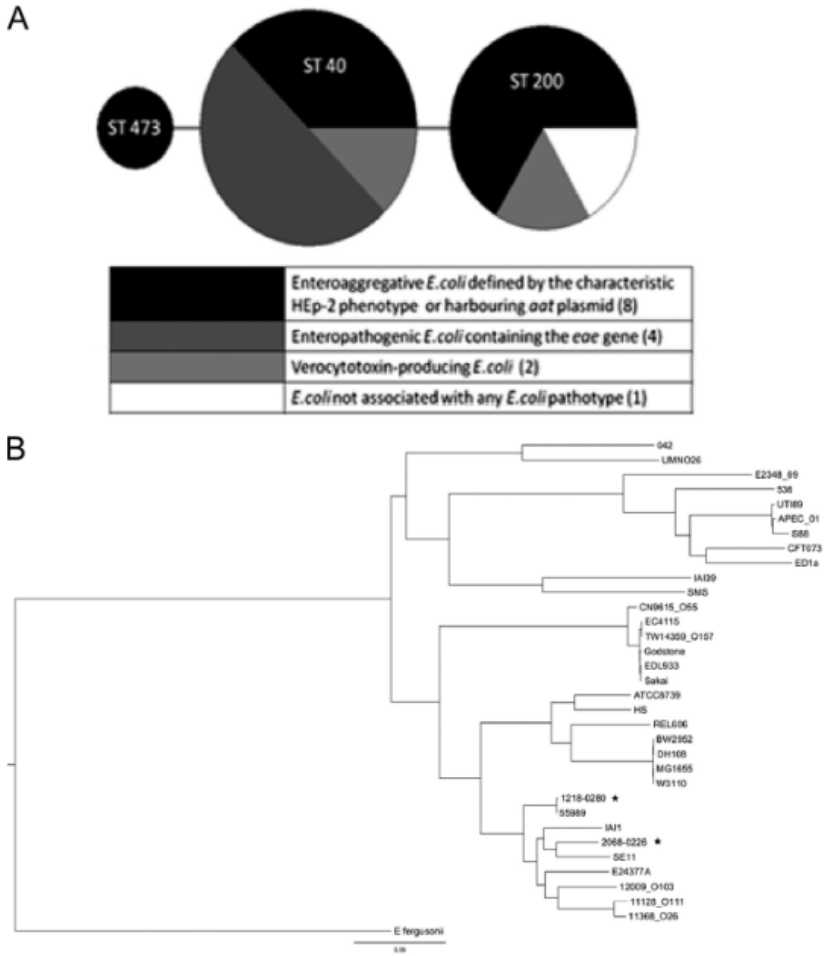


FIG 1 (A) A minimum-spanning tree of the ST40 MLST clonal complex showing the proportion of enteroaggregative *E. coli* (EAEC), enteropathogenic *E. coli* (EPEC), Shiga-toxin-producing *E. coli* (STEC), and nonpathogenic *E. coli*. (B) A maximum-likelihood tree of 43,957 polymorphic positions from 32 *Escherichia coli* genomes. Strains highlighted with a star represent Shiga-toxin-producing enteroaggregative *E. coli*.

specifically labeled with O111 antibody (Dynabeads, Life Technologies), after preenrichment in buffered peptone water for 6 h. Enriched samples were also examined by an O111 monoclonal antibody-based sandwich enzyme-linked immunosorbent assay (ELISA) developed at the Veterinary Science Division, AFBI (unpublished data). Five of the fecal samples gave a positive reaction with the sandwich ELISA, but no single colony purification of *E. coli* O111 was achieved with either assay. No food or environmental source was confirmed.

In light of the similarities in their pathogenic profile, the VTEC EAggEC O111:H21 was sequenced and compared to an isolate of VTEC EAggEC O104:H4 from a German national diagnosed in the United Kingdom but linked to the German outbreak (strain reference 280). DNA from the O111:H21 isolate was prepared for sequencing with the Nextera sample preparation method and sequenced using a standard 2- by 151-base protocol on an MiSeq instrument (Illumina) as described previously (11). FASTQ reads were deposited in the NCBI Short Read Archive, accession no. SRA055981.

Velvet version 1.1.04 (20) was used to produce *de novo* assem-

14. Mellmann A, et al. 2011. Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. PLoS One 6:e22751. doi:10.1371/journal.pone.0022751.
15. Morabito S, et al. 1998. Enteroaggregative, Shiga toxin-producing Escherichia coli O111:H2 associated with an outbreak of hemolytic-uremic syndrome. J. Clin. Microbiol. 36:840–842.
16. Nataro JP, et al. 1987. Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. Pediatr. Infect. Dis. J. 6:829–831.
17. Rasko DA, et al. 2011. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N. Engl. J. Med. 365:709–717.
18. Rohde H, et al. 2011. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. N. Engl. J. Med. 365:718–724.
19. Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.
20. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

7.9 Comparative Analysis of ESBL-Positive Escherichia coli Isolates from Animals and Humans from the UK, The Netherlands and Germany.

Wu G, Day MJ, Mafura MT, Nunez-Garcia J, Fenner JJ, Sharma M, van Essen-Zandbergen A, Rodriguez I, Dierikx C, Kadlec K, Schink AK, Wain J, Helmuth R, Guerra B, Schwarz S, Threlfall J, Woodward MJ, Woodford N, Coldham N and Mevius D. 2013.PLoS One 8 (2013) e75392.

OPEN ACCESS Freely available online

PLOS ONE

Comparative Analysis of ESBL-Positive *Escherichia coli* Isolates from Animals and Humans from the UK, The Netherlands and Germany

Guanghui Wu[1*], Michaela J. Day[2], Muriel T. Mafura[1], Javier Nunez-Garcia[1], Jackie J. Fenner[1], Meenaxi Sharma[1], Alieda van Essen-Zandbergen[3], Irene Rodríguez[4¤a], Cindy Dierikx[3], Kristina Kadlec[5], Anne-Kathrin Schink[5], John Wain[2,6], Reiner Helmuth[4], Beatriz Guerra[4], Stefan Schwarz[5], John Threlfall[2], Martin J. Woodward[1¤b], Neil Woodford[2], Nick Coldham[1], Dik Mevius[3]

1 Animal Health and Veterinary Laboratories Agency (AHVLA, Weybridge), Addlestone, United Kingdom, **2** Public Health England, London, United Kingdom, **3** Department of Bacteriology and TSEs, Central Veterinary Institute (CVI) of Wageningen, Lelystad, the Netherlands, **4** Department of Biological Safety, Federal Institute for Risk Assessment (BfR), Berlin, Germany, **5** Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut (FLI), Neustadt-Mariensee, Germany, **6** Norwich Medical School, University of East Anglia, Norwich, United Kingdom

Abstract

The putative virulence and antimicrobial resistance gene contents of extended spectrum β-lactamase (ESBL)-positive *E. coli* (n=629) isolated between 2005 and 2009 from humans, animals and animal food products in Germany, The Netherlands and the UK were compared using a microarray approach to test the suitability of this approach with regard to determining their similarities. A selection of isolates (n=313) were also analysed by multilocus sequence typing (MLST). Isolates harbouring $bla_{CTX\text{-}M\text{-}group\text{-}1}$ dominated (66%, n=418) and originated from both animals and cases of human infections in all three countries; 23% (n=144) of all isolates contained both $bla_{CTX\text{-}M\text{-}group\text{-}1}$ and $bla_{OXA\text{-}1\text{-}like}$ genes, predominantly from humans (n=127) and UK cattle (n=15). The antimicrobial resistance and virulence gene profiles of this collection of isolates were highly diverse. A substantial number of human isolates (32%, n=87) did not share more than 40% similarity (based on the Jaccard coefficient) with animal isolates. A further 43% of human isolates from the three countries (n=117) were at least 40% similar to each other and to five isolates from UK cattle and one each from Dutch chicken meat and a German dog; the members of this group usually harboured genes such as *mph*(A), *mrx*, *aac*(6')-*Ib*, *catB3*, $bla_{OXA\text{-}1\text{-}like}$ and $bla_{CTX\text{-}M\text{-}group\text{-}1}$. forty-four per cent of the MLST-typed isolates in this group belonged to ST131 (n=18) and 22% to ST405 (n=9), all from humans. Among animal isolates subjected to MLST (n=258), only 1.2% (n=3) were more than 70% similar to human isolates in gene profiles and shared the same MLST clonal complex with the corresponding human isolates. The results suggest that minimising human-to-human transmission is essential to control the spread of ESBL-positive *E. coli* in humans.

Citation: Wu G, Day MJ, Mafura MT, Nunez-Garcia J, Fenner JJ, et al. (2013) Comparative Analysis of ESBL-Positive *Escherichia coli* Isolates from Animals and Humans from the UK, The Netherlands and Germany. PLoS ONE 8(9): e75392. doi:10.1371/journal.pone.0075392

Editor: Axel Cloeckaert, Institut National de la Recherche Agronomique, France

Received June 7, 2013; **Accepted** August 12, 2013; **Published** September 26, 2013

Funding: This project is supported by United Kingdom Food Standards Agency (http://www.food.gov.uk/, project B14016) as a part of the EU-SAFEFOODERA project 08176, entitled: The role of commensal microflora of animals in the transmission of extended spectrum ß-lactamases. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: guanghui.wu@ahvla.gsi.gov.uk

¤a Current address: Department of Microbiology, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain
¤b Current address: Department of Food and Nutritional Sciences, University of Reading, Reading, United Kingdom

Introduction

Escherichia coli is a commensal organism in people and animals but is also a causative agent of diarrhoea and extra-intestinal infections. It is responsible for an estimated 120 million cases of community-acquired urinary tract infections (UTI) diagnosed worldwide annually. It can also cause neonatal meningitis, pneumonia and surgical site infections. The sepsis-associated mortalities due to *E. coli* are estimated at 868,000 per year globally [1]. In England, Wales and Northern Ireland, *E. coli* has been the most common cause of bacteraemia for most years since 1990, with year-on-year increases to 27,055 reports in 2010 [2].

Since circa 2003, there has been a rapid and global increase in the occurrence of *E. coli* with resistance to oxyimino-cephalosporins due to the production of extended-spectrum β-

lactamases (ESBLs). These isolates have emerged in both community and healthcare settings, are often also resistant to other antimicrobial agents, including fluoroquinolones, aminoglycosides and sulphonamides and resistant isolates have been associated with treatment failures [3]. Bacteraemia caused by ESBL producers can be associated with increased mortality, primarily because multi-resistance undermines the efficacy of empiric therapies, which are prescribed before the antimicrobial susceptibility of the infecting organism is known [4]. The CTX-M types of β-lactamases are the dominant family of ESBLs in *E. coli*, with particular subtypes associated with different geographic regions. However, the CTX-M-15 ESBL is pandemic and is often disseminated with the O25:H4-ST131 *E. coli* clone [5-7].

The occurrence of ESBL-positive *E. coli* in animals is also showing a general tendency to increase in some countries, among those bacteria isolated from the poultry gut as well as among those that contaminate food products [8-13]. The rise of community-acquired urinary tract infections caused by *E. coli* resistant to 3rd - or 4th - generation cephalosporins has been linked to international travel of people to countries of high prevalence [13-17] and to reservoirs of resistant bacteria in food-producing animals, especially poultry [13,18]. To explore the genetic relatedness of ESBL- and/or plasmid-mediated (p) AmpC-β-lactamase-producing *E. coli*, we used virulence and resistance gene microarrays as a convenient and rapid tool to investigate isolates from Germany, The Netherlands and the UK obtained from humans, food producing animals and animal food products. A subset of these isolates was also characterised by multi-locus sequence typing (MLST) to assist in elucidating the clonal relationship of isolates.

Materials and Methods

Sources of isolates

This study sought to compare ESBL/pAmpC-producing *E. coli* from the animal gut flora, animal-derived food products and from cases of human infections, especially from urinary tract infections (UTI) where the ESBL-producing isolates were often identified. For this, *E. coli* isolates that showed resistance to both ampicillin and cefotaxime, based on EUCAST criteria (www.eucast.org), and that had been isolated between 2005 and 2009 were included in this study. The isolates were from existing strain collections of the UK (AHVLA and Public Health England, formerly the Health Protection Agency), Germany (FLI and BfR) and The Netherlands (CVI) and had been obtained as part of national antimicrobial resistance surveillance programmes or from participants' routine diagnostic or reference laboratory activities. The UK poultry isolates were from a structured survey [19] and those from cattle were derived from scanning surveillance of clinical diagnostic submissions to the 14 AHVLA regional laboratories across England and Wales [20]. The German isolates were from the collection at the National Reference Laboratory for *E. coli* (NRL-E. *coli*) and the NRL for Antimicrobial Resistance (NRL-AR) of the BfR (for food and animal isolates) and the National Reference Centre for Salmonella and other enterics (NRZ, human isolates, Robert Koch Institute). In addition, all putative ESBL-producing *E. coli* isolates from the BfT-Germ\Vet collection were included [21]. The Dutch human isolates were selected from a national ESBL-prevalence study conducted in Public Health Laboratory Services in 2009 [13] and the animal isolates were selected from the collection of the NRL-AR (CVI). No field samples were specifically collected for this study. *E. coli* isolates (n= 629) from animals (n=295), animal-derived foods (n=59), humans (n=274) and an unknown source (n=1), originating from Germany (n=84), The Netherlands (n =254) and the UK (n =291) were analysed using microarrays (Table 1). MLST analysis was performed on 313 isolates isolated from different countries and host species.

Microarray analysis

The principle and methodology of the microarray have been described previously [22-24] and the probe content was described by Geue et al. [25] and can be found at: http://alere-technologies.com/fileadmin/Media/Paper/Ecoli/Supplement_Geue__layout_E_coli.xlsx. In addition, a description of the probes that generated positive signals among isolates can be found in Table S1. The bla_{CTX-M} family probes identified genes only to group level (i.e. groups 1, 2, 8/25, 9) [26], and the bla_{SHV} and bla_{TEM} probes did not identify their precise alleles. The virulence probes allowed the detection of major pathotypes of *E. coli* such as enteropathogenic *E. coli* (EPEC), verotoxigenic *E. coli* (VTEC), enterotoxigenic (ETEC), extra-intestinal pathogenic *E. coli* (ExPEC) from humans and animals including avian pathogenic *E. coli* (APEC), and uropathogenic *E. coli* (UPEC) as described previously [23]. All probes were present on the array in duplicate, except probes for *sul3*, *tet*(A-G), bla_{TEM} which were singletons.

Hybridisation signals were normalised to the 50th percentile of the signal intensities of control genes: *ihfA*, *gad*, *gapA*, *hemL* and *dnaE*. Normalised signal intensities with values >0.1 were in 95% in agreement with positive PCR results for $bla_{CTX-M-group-1}$ (n=248) and $bla_{CTX-M-group-9}$ (n=95) and 100% agreement with four $bla_{CTX-M-group-2}$, bla_{TEM} (n=11) and bla_{SHV} (n=5) [27]; a gene was therefore considered present if the complementary array probe(s) produced normalised signal intensity that was >0.1. Cluster analysis of isolates was performed with Jaccard coefficients and Unweighted Pair Group Method with Arithmetic Mean (UPGMA, Bionumerics 5.1, Gent, Belgium). Isolates were grouped where stated to form larger clusters at different similarity levels; when the similarity level is not indicated, a 40% similarity threshold was used.

Analysis of the diversity of isolates

To assess the diversity of isolates from each country and host species, different similarity cut-off points were used to group isolates into larger clusters to facilitate their comparison. The number of isolates within each cluster from each host species and country was counted and the Simpson's Index of Diversity (1 -D) was calculated [28]. The distribution of isolates from each country and species among those clusters was analysed. Where possible, genes specifically associated with, and hence potentially indicative of major clusters, were identified.

Table 1. Distribution of selected β-lactamase genes in *E. coli* from different hosts[1] and countries.

Host	Country	Percentage (no. of positive isolates)						
		Isolates (n)	$bla_{CTX\text{-}M\text{-}group\text{-}1}$	$bla_{CTX\text{-}M\text{-}group\text{-}9}$	bla_{SHV}	$bla_{CMY\text{-}611}$[2]	bla_{CMY_11}[3]	$bla_{MOX\text{-}cmy\_9}$[4]
	All	274	74% (204)	8% (23)	11% (31)	0.7% (2)	3% (8)	7% (18)
	UK	152	74% (112)	6% (9)	16% (25)	0	5% (7)	10% (15)
Human	Germany	14	93% (13)	7% (1)	0	7% (1)	7% (1)	7% (1)
	Netherlands	108	73% (79)	12% (13)	6% (6)	1% (1)	0	2% (2)
	All	157	55% (86)	1% (1)	5% (8)	22% (35)	22% (34)	3% (5)
	UK	10	100%(10)	0	0	0	0	0
Chicken	Germany	11	18% (2)	0	0	72% (8)	73% (8)	0
	Netherlands	136	54% (74)	1% (1)	6% (8)	20% (27)	19% (26)	4% (5)
	All	133	63% (84)	31% (41)	1% (1)	3% (4)	2% (3)	1% (1)
	UK	95	57% (54)	43% (41)	0	2% (2)	2% (2)	0
Cattle	Germany	32	81% (26)	0	0	0	0	3% (1)
	Netherlands	6	67% (4)	0	17% (1)	33% (2)	17% (1)	0
	All	35	63% (22)	46% (16)				3% (1)
Turkey	UK	32	63% (20)	50% (16)	0	0	0	3% (1)
	Germany	2	50% (1)	0	0	0	0	0
	Netherlands	1	100% (1)	0	0	0	0	0
	All	17	94% (16)					
	UK	1	100% (1)	0	0	0	0	0
Pig	Germany	13	92% (12)	0	0	0	0	0
	Netherlands	3	100% (3)	0	0	0	0	0
All[5]		629	66% (418)	13% (81)	6% (40)	7% (44)	7% (48)	4% (25)

[1] The precise sources of strains can be found in Table S1. Those from meat products were counted as if they were from the corresponding animals.

[2] blaCMY-611, probe name and the sequence was found in blaCMY-2 (AB212086.1[93:118])and blaCMY-13 (AY339625.2[4669:4694:r])

[3] blaCMY_11, probe name and the sequence was found in blaCMY-2 (AB212086.1[569:592])and blaCMY-13 (AY339625.2[4195:4218:r]

[4] bla MOX-cmy_9, probe name and the sequence was in bla MOX-cmy_9 (AF357599.1[379:402])

[5] There were also 6 isolates labelled as from poultry and 4 from cats/dogs from Germany, one UK isolate from sheep and 2 from unknown German animals.

doi: 10.1371/journal.pone.0075392.t001

E. coli MLST typing

In order to estimate the relative importance of clonal spread versus horizontal gene transfer in the dissemination of ESBL genes, and to assess the association between virulence and resistance genes with genetic backbones of the host bacteria, isolates (n=313) representing different countries of origin or different host species were analysed by MLST according to the scheme described by Wirth et al. [29] following the guidelines given at http://mlst.ucc.ie/mlst/dbs/Ecoli

Results

Pathotypes of *E. coli* isolates and the presence of major ESBL/ pAmpC β-lactamase genes.

Virulence genes and antimicrobial resistance genes identified by microarrays are listed in Table S1. Only 0.5% of isolates (n=3) were ETEC and harboured the *fim41* (encoding F41 fimbria) and *sta1* (encoding a heat-stable toxin) genes [30]. Fewer than 7.5% of the isolates (n=46) were EPEC (that contained *eae* encoding intimin) or VTEC, harbouring both *eae* and *stx* genes (the latter gene encoding Shiga toxins).

The majority of isolates were either commensal or ExPEC and typically harbouring *prfB* (encoding P-related fimbria, n=241, 38% of the isolates) and/or *iroN* (encoding the enterobactin siderophore receptor protein) or microcin genes; 13% of isolates (n=80) contained *tsh* (encoding a temperature-sensitive haemagglutinin), a gene often associated with APEC [31]. Hybrid pathotypes of ExPEC and EPEC were also identified: three isolates (0.5%) harboured both *prfB* and *eae*. In addition, about 20% of Dutch chicken isolates (n=28) harboured EPEC-associated genes [30] and nearly half (n=13) also carried *tsh* (Table S1).

As expected, $bla_{CTX\text{-}M\text{-}group\text{-}1}$ predominated amongst the CTX-M genes and was detected in 418 of 629 (66%) isolates (Table 1). Additionally, $bla_{OXA\text{-}1\text{-}like}$ genes were found in combination with $bla_{CTX\text{-}M\text{-}group\text{-}1}$ genes in about 23% of the isolates (n=144), mainly from humans (88%, n=127) and UK cattle (10%, n=15), but also in single isolates from Dutch chicken meat (ESBL428) and from a German pig (ESBL 154). While $bla_{CTX\text{-}M\text{-}group\text{-}1}$ genes were widespread in all host species and in all three countries, $bla_{CTX\text{-}M\text{-}group\text{-}9}$ genes were more common among cattle and turkey isolates from the UK (Table 1). Forty isolates were positive for bla_{SHV}, among which 6 and 25 were from Dutch and UK humans respectively, one and 8 from Dutch cattle and chickens respectively. More than half of the isolates (n=399) were bla_{TEM} positive, which were from all species and countries studied here. However, the array does not distinguish the ESBL and non-ESBL forms of bla_{SHV} and bla_{TEM} genes.

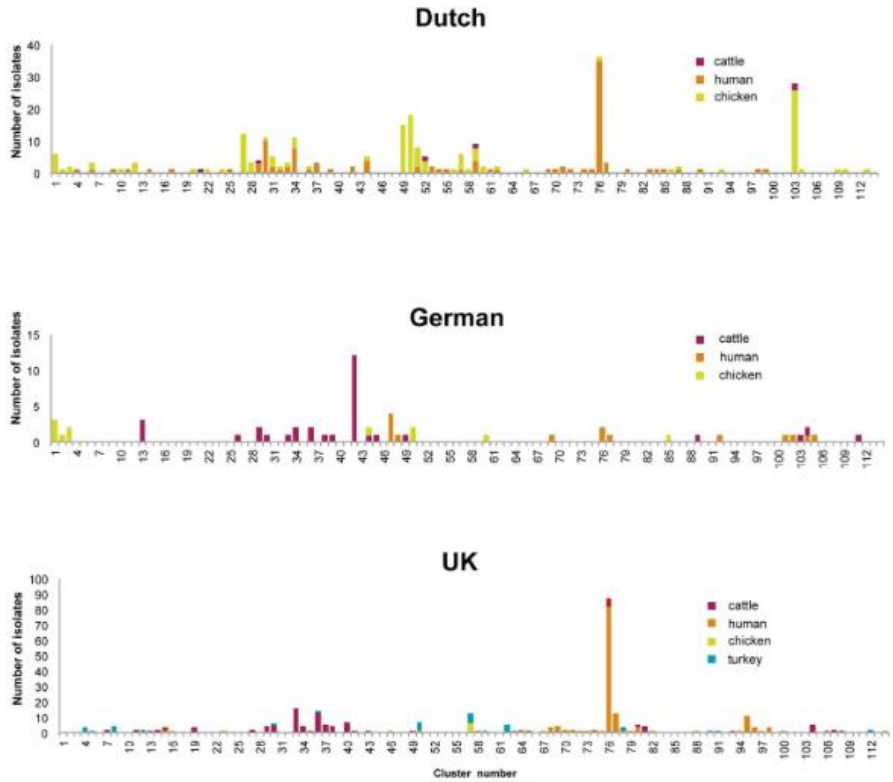


Figure 1. Isolates were grouped into 114 clusters at the 40% similarity level.

doi: 10.1371/journal.pone.0075392.g001

Analysis of the diversity among the isolates

Isolates were clustered initially using the Jaccard coefficient and UPGMA (Bionumerics 5.1) based on the presence or absence of virulence and antimicrobial resistance genes. To identify isolates that may be related, similar isolates were grouped together into larger clusters. Isolates were divided into 12 clusters at 15% similarity and four of those clusters contained a single isolate, but approximately 76% of the 629 isolates were grouped in a single cluster (not shown). At 25% similarity, isolates were divided into 33 clusters of which 9 contained a single isolate, but 24%, 17% and 26% of the isolates were in three clusters respectively (data not shown). At 40% similarity, isolates were divided into 114 clusters (Figure 1). When raising the similarity threshold, the number of defined clusters increased further so that at 90% similarity, isolates were separated into 556 clusters, with 484 (87%) of them being singletons and only 8 clusters including isolates from different species.

To calculate the diversity of isolates, a series of cut-off points from 25% to 90% were used and the number of isolates from each category (country and species) in each cluster was counted and used to calculate the index of diversity and to plot the results (Figure 2). This approach showed that isolates from The Netherlands and Germany were very diverse and the differences were most pronounced between the 25% and 50% similarity levels. Only 14 human isolates from Germany were included; their high diversity was probably due to the inclusion of EPECs, whereas the human isolates from The Netherlands and the UK originated exclusively from urinary tract infections. With the exception of turkey isolates, the UK isolates showed the lowest overall diversity (Figure 2). As only three turkey isolates from Germany and The Netherlands were included in

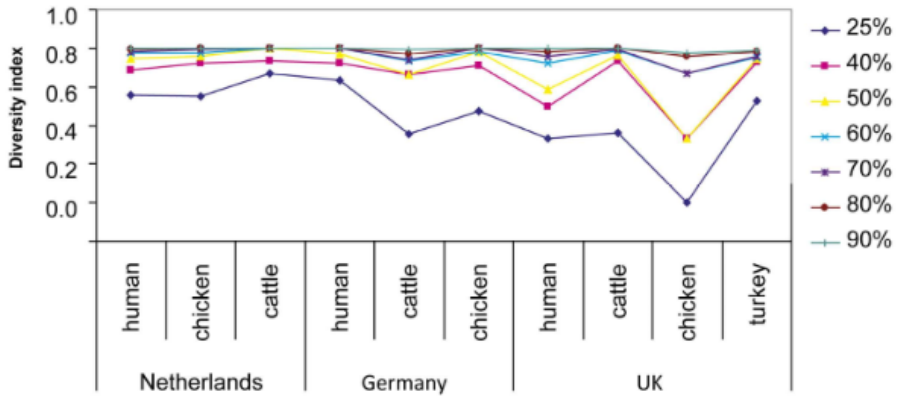


Figure 2. Simpson's Index of Diversity (1-D) of isolates from different categories (country and host species) at the similarity levels from 25% to 90%.

doi: 10.1371/journal.pone.0075392.g002

this study, a valid comparison with those from the UK could not be made.

Comparison of isolates from different host species and countries

While microarrays detect virulence and antimicrobial resistance genes that are often acquired through horizontal gene transfer, MLST provides information on the genetic backbone of host strains. In this study, MLST analysis was performed on 313 isolates in order to seek relationships between array clustering and multilocus sequence types (ST, Figure 3). The results showed that the major array clusters (e.g. 27, 30, 34, 42, 49, 50, 76, 103) all contained isolates of numerous STs, and that isolates of major STs (ST131 n=45, ST10 n=25, ST88 n=24, ST405 n=11, ST69 n=10, ST58 n=10) were divided among several very different array clusters (Figure 3). The largest array cluster, 76, consisted of 124 isolates of which 41 were typed by MLST [humans (n=36), UK cattle (n=3), Dutch chicken meat (n=1) and German dog (n=1)]. Among them, 44% (n=18, all from humans) belonged to ST131, 22% (n=9, all from humans) to ST405, 12% (n=5) to the ST10 complex and 5% (n=2) to the ST23 complex. Conversely, of 44 ST131 isolates only 43% were grouped in cluster 76. In the following sections, isolates from each host species were analysed together.

At the 40% similarity level, human isolates (n=274) were scattered among 69 clusters, but 43% (n=117) were in cluster 76 (Figure 1), including those from the UK (82/152), The Netherlands (33/108) and Germany (2/14). The isolates in this cluster were defined as the "common human type" and usually contained genes *mph*(A), encoding a macrolide 2'-phosphotransferase; *mrx*, encoding a putative protein; *aac*(6')-*Ib*, encoding an aminoglycoside 6'-N-acetyltransferase; *catB3*, encoding a chloramphenicol acetyltransferase as well as the $bla_{\text{OXA-1-like}}$ and $bla_{\text{CTX-M-group-1}}$ β-lactamase genes. A small number of the human isolates from the UK (13%, n=20) and The Netherlands (1.9%, n=2) were grouped in clusters 94 to 100 which harboured less commonly detected resistance genes such as bla_{SHV} and/or $bla_{\text{CTX-M-group-2}}$. In addition, 33.6% of isolates from cases of human infection in the Netherlands were in clusters 29-34, 36, 37, 39, 42 and 44 (Figure 1). All in all, a total of 32% (n=87) of human isolates were in 40/69 clusters, which consisted of only human isolates, and a further 43% in cluster 76 in which 95% were human isolates. On the other hand, 19 human isolates (7% of the total human isolates) were at least 70%, 13 human isolates (4.7%) were at least 80%, and five human isolates (1.8%) were at least 90% similar to animal isolates in array profiles based on the Jaccard coefficient.

Chicken isolates (n=157) were found among 40 clusters at the 40% similarity level; 20% of the total were in 13 of these clusters, which consisted only of chicken isolates. Dutch chicken isolates harbouring $bla_{\text{CTX-M-group-1}}$ often also carried *intI1* (integrase gene of class 1 integron), while those from the UK often carried *intI2* (integrase gene of class 2 integron). The Dutch chicken meat isolate ESBL428 (ST88) found in cluster 76 shared the ST23 clonal complex (which is the second largest clonal complex within the *E. coli* MLST database) with a Dutch human isolate ESBL7 (ST 90), however, the similarity in array profiles between these two isolates was just over 40%. In addition, 21% (n=29) of Dutch chicken isolates shared 13 array clusters with human isolates. Also, 19% (n=26) of Dutch chicken isolates were in cluster 103 together with one German and two Dutch cattle isolates (Figure 1). The isolates in this cluster harboured genes often found among EPEC, such as *eae* and genes encoding type III secretion system, and half of those isolates also carried *tsh*.

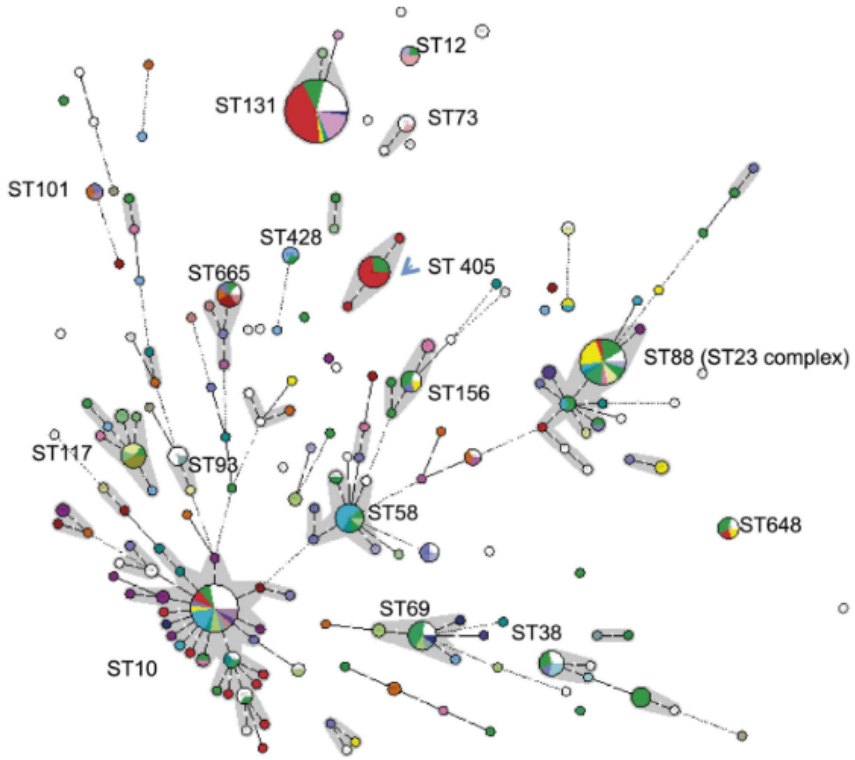


Figure 3. Minimal spanning tree constructed based on the MLST profiles of the 313 isolates (Bionumerics V6) and coloured according to the cluster numbers: red, cluster 76; purple, cluster 103; light blue, cluster 30; dark green, cluster 42; green, no array data and white, a mixture of rare cluster number.

doi: 10.1371/journal.pone.0075392.g003

Cattle isolates (n=133) were found among 41 clusters at the 40% similarity level and 23% (n=31) of the isolates were in 16/41 clusters that consisted of only cattle isolates. About 61% (n=81) of cattle isolates (from all three countries) shared clusters with less common human types (mostly Dutch) and 5% (n=5) of the cattle isolates from the UK were found in cluster 76. Two of the isolates in this cluster (ESBL569, ST167 and ESBL528, ST1284, both were within the ST10 clonal complex, the largest clonal complex within the *E. coli* MLST database) were at least 70% or 80% similar in array profiles to human isolates (ESBL73, ST617 from The Netherlands; ESBL760, ST167 and ESBL691, a double locus variant of ST1284, both were from the UK) that also belonged to this ST clonal complex. The UK cattle isolate ESBL528 shared more than 90% similarity in array profiles with the Dutch human isolate ESBL106. In array cluster 30, ESBL517 from UK cattle and ESBL101 from a Dutch human patient were both ST58 and were at least 80% similar in their array profiles. In terms of the differences in genes they harboured, $bla_{CTX\text{-}M\text{-}group\text{-}9}$ was found in cattle isolates from the UK, but not in those from The Netherlands or Germany (Table 1). Among the six cattle isolates from The Netherlands, diverse genes including ESBL/AmpC genes bla_{CMY}, bla_{SHV}, and virulence genes *f17A*, *f17G* and *cif*, were detected. Cattle isolates from Germany (in cluster 42) harboured different genes as compared to cattle isolates from the UK (Figure 1).

Turkey isolates (n=35) were found in 16 clusters and 25.7% (n=9) of them were in five clusters which consisted of only turkey isolates, while 54% (from UK, n=17; Germany, n=1 and The Netherlands, n=1) of those shared clusters with human

Table 2. Genes that show significant differential presence among isolates from different species.

Genes	The proportion of isolates where the gene is present (The binomial 95% exact confidence interval)				
More common among human isolates	**Human**	**Chicken**	**Pig**	**Turkey**	**Cattle**
aac6	0.482 (0.42,0.54)	0.006 (0,0.03)	0 (0,0.19)	0.057 (0,0.19)	0.075 (0.03,0.13)
aac(6')-Ib	0.507 (0.44,0.56)	0.006 (0,0.03)	0 (0,0.19)	0.086 (0.01,0.23)	0.083 (0.04,0.14)
$bla_{OXA-1-like}$	0.485 (0.42,0.54)	0.006 (0,0.03)	0.059 (0,0.28)	0 (0,0.1)	0.165 (0.1,0.24)
catB3	0.456 (0.39,0.51)	0.006 (0,0.03)	0 (0,0.19)	0 (0,0.1)	0.06 (0.02,0.11)
iha	0.504 (0.44,0.56)	0.019 (0,0.05)	0 (0,0.19)	0.086 (0.02,0.23)	0.211 (0.14,0.29)
sat	0.482 (0.42,0.54)	0 (0,0.02)	0 (0,0.19)	0.057 (0,0.19)	0.008 (0,0.04)
prfB	0.613 (0.55,0.67)	0.076 (0.04,0.13)	0.176 (0.03,0.43)	0.029 (0,0.14)	0.421 (0.33,0.51)
Less common among human isolates					
lpfA	0.084 (0.05,0.12)	0.268 (0.2,0.34)	0.353 (0.14,0.61)	0.286 (0.14,0.46)	0.256 (0.18,0.33)
More common among bovine isolates					
floR	0.026 (0.01,0.05)	0.057 (0.02,0.1)	0 (0,0.19)	0 (0,0.1)	0.331 (0.25,0.41)
espP	0.011 (0,0.03)	0.013 (0,0.04)	0.059 (0,0.28)	0.029 (0,0.14)	0.489 (0.4,0.57)
mcmA	0.113 (0.07,0.15)	0.019 (0,0.05)	0.059 (0,0.28)	0.057 (0,0.19)	0.444 (0.35,0.53)

doi: 10.1371/journal.pone.0075392.t002

isolates from The Netherlands and the UK. All these turkey isolates were different from the "common human type". Six turkey isolates were more than 70% similar and four were more than 80% similar to human isolates. One UK turkey isolate (ESBL 626) was 90% similar to the Dutch human isolate ESBL 80.

The 17 pig isolates, 13 of which were from Germany, were found in seven clusters. No cluster was observed that contained only pig isolates. All but one pig isolate (94%) clustered with human isolates. The majority (71%, n=12) harboured genes *mph*(A) and *mrx*; both genes were typically found among the "common human type". Two pig isolates shared more than 70% similarity with human isolates, but their MLST profiles (single locus variant of ST348) were dissimilar to the respective human isolates ESBL116 (ST453) and ESBL80 (a double locus variant of ST167).

Genes that were unevenly distributed among different host species and countries

Genes found significantly more or less common among isolates from humans and cattle in comparison with other species are summarized in Table 2. Antimicrobial resistance genes $bla_{OXA-1-like}$, *aac(6')-Ib*, *catB3* and virulence genes *iha* (encoding an adhesion), *sat* (encoding serine protease precursor), *prfB* were frequently found among human isolates, especially among the UK human isolates from urinary tract infections (Table S2). The long polar fimbria gene (*lpfA*) was less commonly represented among human isolates, especially among those from the UK (3.3%) as compared with those from The Netherlands (14.8%). The gene *floR*, which confers resistance to florfenicol, was seen among 33% of the cattle isolates, but was rare among isolates of other species (Table 2). Florfenicol is a fluorinated derivative of chloramphenicol approved for use in cattle in Europe since 1995 [32]. The *espP* gene that encodes serine protease autotransporter and *mcmA* that encodes for microcin M protein were significantly more common among cattle isolates than other species (Table 2). Genes that were more or less common in *E. coli* isolates from each country are listed in Tables S2 and S3.

Discussion

In this study, we assessed the use a microarray approach to study gene relatedness among ESBL carrying *E. coli*. Many of the genes represented on the array are readily transmissible being encoded on mobile genetic elements. Therefore, we might anticipate greater diversity using this approach than other standard methods such as PFGE and MLST that focus on the genetic backbone of a strain. One advantage of the array approach that in this study detected genes encoding important phenotypes, in this case resistance and virulence, is the ability to identify known and novel (potentially emerging) gene associations.

A considerable number of human *E. coli* isolates (32%) did not share array clusters with animal isolates and a further 43% were concentrated in one cluster (cluster 76), which consisted mainly of human isolates. These results are consistent with a recent observation that the bla_{CTX-M}-carrying *E. coli* isolates from UK chicken and turkey were different from the human-associated ST131 CTX-M-15 type in the UK [19]. In another study, 11% of human isolates were found to contain poultry-associated ESBL genes, plasmids and MLST types in The Netherlands [13], however, this study did not investigate virulence and additional antimicrobial resistance genes. In this study, only 1.2% (3/258) of characterized *E. coli* from animals were similar (as judged by gene content and clonal complex) to isolates from humans.

Nearly half of the human isolates investigated in this study were located in clusters surrounding cluster 76 and many of those belonged to ST131. Therefore, those isolates are likely to be similar to *E. coli* ST131-O25:H4-B2 that harboured $bla_{CTX-M-15}$ plasmids (as $bla_{CTX-M-15}$ belongs to $bla_{CTX-M-group-1}$) [33,34]. Moreover, ST131 was found among 11 different array clusters, which is consistent with the reports that O25:H4-B2- ST131

isolates carry various bla_{CTX-M} types on a number of different plasmid types (Inc FII, FIA-FIB, FIA, FII-FIA, I1, N and Y for example) [35,36]. The O25:H4-ST131 clone has been found around the globe [5-7] and its spread between humans has been facilitated by international travel [37]. ST405 was identified as another important resistant sequence type among human isolates that shared similar virulence and antimicrobial resistance genes with many ST131 isolates.

As many human isolates from the three countries were highly similar to each other, the widespread human-to-human transmission of ESBL-producing *E. coli* is a strong possibility. Nevertheless the potential threat posed by animals or animal food products as sources for human ESBL-positive isolates cannot be ignored [13]. When clonally related CTX-M-producing *E. coli* were identified in community settings, common infection sources such as food or water were often suspected [38,39]. The dissemination of resistance genes in *E. coli* may occur through multiple routes. It is the opinion of EFSA that "transmission of ESBL genes, plasmids and clones from poultry to humans is most likely to occur through the food chain" [40], however the data presented in a recent review showed that considerable differences in ESBL types between poultry and humans in Europe exist, leaving the question open as to what extent livestock has contributed to the spread of ESBLs in humans [41]. The ESBL genes are often located on plasmids and are therefore likely to disseminate *via* horizontal gene transfer, as described recently for the transmission of a $bla_{CTX-M-14}$-carrying plasmid [42]. This study did not seek to confirm routes of transmission between man and animals or animals and man, but to investigate whether organisms similar at the genetic level were present in different animal species and man, indicating possible epidemiological links. Such links might be consistent with food-borne transmission from animal to man, but equally, might indicate exposure of animals to human faecal bacteria in sewage or flooding incidents or by other routes.

This work demonstrated that microarray analysis is a useful tool for detecting the genetic diversity of a large number of ESBL-producing *E. coli* isolates although we recognize one weakness regarding the inability to differentiate group and sub-groups of ESBL in sufficient depth. The amount of data generated by this small array platform is more manageable compared with the whole genome glass slide microarrays [43,44] or whole genome sequencing. The whole genome arrays provide more information on genes located on the chromosome, but less information on virulence genes and mobile genetic elements, such as plasmids, where the ESBL and pAmpC-β-lactamase genes are often located. The associations of some genes with *E. coli* from certain animal species also suggest that a targeted approach can provide useful epidemiological information.

The isolates analysed in this study were highly diverse especially those from The Netherlands and Germany. The results do not necessarily reflect the diversity of isolates in that country and that species in general, because the isolates were collected from different surveillance programmes and between 2005 and 2009. Some of the observed diversity might therefore reflect differences in the sampling strategies used and in the time periods when samples were taken. There is also the need to be particularly cautious when only very small numbers of isolates were analysed (e.g. those from Germany) and indeed a recommendation for future work is the application of agreed standardised sampling protocols.

The clusters defined by the arrays showed poor congruence with MLST data, i.e. isolates of different, unrelated STs were found within the same array cluster, and isolates sharing the same ST were found in different array clusters; this observation is consistent with a previous study on ExPEC isolates from animals, where no correlation between virulence and antimicrobial resistance genes and genetic backbone of the host strain was found [45]. This was anticipated given the mobility and potential transience of many of the genes represented on the array.

In conclusion, the microarray analysis of ESBL- and AmpC-producing *E. coli* isolates showed a high diversity in virulence and antimicrobial resistance gene contents. The array profiles of the majority of isolates from humans were generally different from those isolated from animals, while many human isolates from the three countries were highly similar in both array profiles and MLST types. Thus, approaches to minimize human-to-human transmission are essential for controlling the spread of ESBL-positive *E. coli*. From public health perspective, ESBL-positive *E. coli* from animals may represent a reservoir of virulence and resistance genes rather than being the direct cause of infections in humans.

Supporting Information

Table S1. Microarray results of isolates used in this study.
(XLSX)

Table S2. Genes that show significant differential presence among human isolates from The Netherlands and the UK.
(XLS)

Table S3. Genes that show significant differential presence among isolates from different countries.
(XLS)

Acknowledgements

We also thank Dr. E. Tietze (Robert Koch Institute, Wernigerode, Germany) for the German human isolates and Dr. Maurine Leverstein – van Hall (University Medical Center of Utrecht, The Netherlands, on behalf of the national ESBL surveillance group) for the Dutch human isolates. We thank Janine Beutlich from BfR, Kees Veldman from CVI and Chris Teale from AHVLA for comments and help. We also thank Sarah Brown from AHVLA for helping in preparing the figures.

Author Contributions

Conceived and designed the experiments: DM NC JT NW MJW CD IR BG RH A-KS KK JW. Performed the experiments: MTM MJD GW JF MS DM CD AvE IR SS A-KS KK NW JT JW.

Analyzed the data: GW MJD MTM JN JF DM CD AvE IR BG SS A-KS KK. Wrote the manuscript: GW NC NW DM. Revising the article critically: SS MJW MJD JN JT MTM JF MS CD RH BG IR A-KS KK.

References

1. Russo TA, Johnson JR (2003) Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect 5: 449-456. doi:10.1016/S1286-4579(03)00049-2. PubMed: 12738001.
2. HPA website. Available: http://www.hpa.org.uk/web/HPAwebFile/HPAweb_C/1296686942137 Accessed 2013 Aug. 23.
3. Coque TM, Baquero F, Canton R (2008) Increasing prevalence of ESBL-producing Enterobacteriaceae in Europe. Euro Surveill 13(47): 19044,19021958.
4. Schwaber MJ, Carmeli Y (2007) Mortality and delay in effective therapy associated with extended-spectrum beta-lactamase production in Enterobacteriaceae bacteraemia: A systematic review and meta-analysis. J Antimicrob Chemother 60: 913-920. doi:10.1093/jac/dkm318. PubMed: 17848376.
5. Lau SH, Kaufmann ME, Livermore DM, Woodford N, Willshaw GA et al. (2008) UK epidemic *Escherichia coli* strains A-E, with CTX-M-15 β-lactamase, all belong to the international O25:H4-ST131 clone. J Antimicrob Chemother 62: 1241-1244. doi:10.1093/jac/dkn380. PubMed: 18779256.
6. Livermore DM, Hawkey PM (2005) CTX-M: changing the face of ESBLs in the UK. J Antimicrob Chemother 56: 451-454. doi:10.1093/jac/dki239. PubMed: 16006451.
7. Hawkey PM, Jones AM (2009) The changing epidemiology of resistance. J Antimicrob Chemother 64 Suppl 1: i3-10. doi:10.1093/jac/dkp256. PubMed: 19675017.
8. Hasman H, Mevius D, Veldman K, Olesen I, Aarestrup FM (2005) β-Lactamases among extended-spectrum β-lactamase (ESBL)-resistant *Salmonella* from poultry, poultry products and human patients in The Netherlands. J Antimicrob Chemother 56: 115-121. doi:10.1093/jac/dki190. PubMed: 15941775.
9. Hunter PA, Dawson S, French GL, Goossens H, Hawkey PM et al. (2010) Antimicrobial-resistant pathogens in animals and man: prescribing, practices and policies. J Antimicrob Chemother 65 Suppl 1: i3-17. doi:10.1093/jac/dkp433. PubMed: 20045808.
10. Liebana E, Batchelor M, Hopkins KL, Clifton-Hadley FA, Teale CJ et al. (2006) Longitudinal farm study of extended-spectrum β-lactamase-mediated resistance. J Clin Microbiol 44: 1630-1634. doi:10.1128/JCM.44.5.1630-1634.2006. PubMed: 16672386.
11. Liebana E, Gibbs M, Clouting C, Barker L, Clifton-Hadley FA et al. (2004) Characterization of β-lactamases responsible for resistance to extended-spectrum cephalosporins in *Escherichia coli* and *Salmonella enterica* strains from food-producing animals in the United Kingdom. Microb Drug Resist 10: 1-9. doi:10.1089/107662904323047745. PubMed: 15140388.
12. Roest HI, Liebana E, Wannet W, van Duynhoven Y, Veldman KT et al. (2007) [Antibiotic resistance in *Escherichia coli* O157 isolated between 1998 and 2003 in The Netherlands]. Tijdschr Diergeneeskd 132: 954-958. PubMed: 18225714.
13. Leverstein-van Hall MA, Dierikx CM, Cohen Stuart J, Voets GM, van den Munckhof MP, et al. (2011) Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. Clin Microbiol Infect 17: 873-880. doi:10.1111/j.1469-0691.2011.03497.x. PubMed: 21463397.
14. Vincent C, Boerlin P, Daignault D, Dozois CM, Dutil L et al. (2010) Food reservoir for *Escherichia coli* causing urinary tract infections. Emerg Infect Dis 16: 88-95. doi:10.3201/eid1601.091118. PubMed: 20031048.
15. Johnson JR, Sannes MR, Croy C, Johnston B, Clabots C et al. (2007) Antimicrobial drug-resistant *Escherichia coli* from humans and poultry products, Minnesota and Wisconsin, 2002-2004. Emerg Infect Dis 13: 838-846. doi:10.3201/eid1306.061576. PubMed: 17553221.
16. Laupland KB, Church DL, Vidakovich J, Mucenski M, Pitout JD (2008) Community-onset extended-spectrum β-lactamase (ESBL) producing *Escherichia coli*: importance of international travel. J Infect 57: 441-448. doi:10,1016/j.jinf,2008,09.034. PubMed: 18990451.
17. Tängdén T, Cars O, Melhus A, Löwdin E (2010) Foreign travel is a major risk factor for colonization with *Escherichia coli* producing CTX-M-type extended-spectrum β-lactamases: a prospective study with Swedish volunteers. Antimicrob Agents Chemother 54: 3564-3568. doi: 10.1128/AAC.00220-10. PubMed: 20547788.
18. Overdevest I, Willemsen I, Rijnsburger M, Eustace A, Xu L et al. (2011) Extended-spectrum β-lactamase genes of *Escherichia coli* in chicken meat and humans, The Netherlands. Emerg Infect Dis 17: 1216-1222. doi:10.3201/eid1707.110209. PubMed: 21762575.
19. Randall LP, Clouting C, Horton RA, Coldham NG, Wu G et al. (2011) Prevalence of *Escherichia coli* carrying extended-spectrum β-lactamases (CTX-M and TEM-52) from broiler chickens and turkeys in Great Britain between 2006 and 2009. J Antimicrob Chemother 66: 86-95. doi:10.1093/jac/dkq396. PubMed: 21098542.
20. Snow LC, Warner RG, Cheney T, Wearing H, Stokes M et al. (2012) Risk factors associated with extended spectrum β-lactamase *Escherichia coli* (CTX-M) on dairy farms in North West England and North Wales. Prev Vet Med 106: 225-234. doi:10.1016/j.prevetmed.2012.03.009. PubMed: 22552330.
21. Schink AK, Kadlec K, Schwarz S (2011) Analysis of $bla_{(CTX-M)}$-carrying plasmids from *Escherichia coli* isolates collected in the BfT-GermVet study. Appl Environ Microbiol 77: 7142-7146. doi:10.1128/AEM.00559-11. PubMed: 21685166.
22. Wu G, Mafura M, Carter B, Lynch K, Anjum MF et al. (2010) Genes associated with *Escherichia coli* isolates from calves with diarrhoea and/or septicaemia. Vet Rec 166: 691-692. doi:10.1136/vr.b4791. PubMed: 20511653.
23. Anjum MF, Mafura M, Slickers P, Ballmer K, Kuhnert P et al. (2007) Pathotyping *Escherichia coli* by using miniaturized DNA microarrays. Appl Environ Microbiol 73: 5692-5697. doi:10.1128/AEM.00419-07. PubMed: 17630299.
24. Batchelor M, Hopkins KL, Liebana E, Slickers P, Ehricht R et al. (2008) Development of a miniaturised microarray-based assay for the rapid identification of antimicrobial resistance genes in Gram-negative bacteria. Int J Antimicrob Agents 31: 440-451. doi:10.1016/j.ijantimicag.2007.11.017. PubMed: 18243668.
25. Geue L, Schares S, Mintel B, Conraths FJ, Müller E et al. (2010) Rapid microarray-based genotyping of enterohemorrhagic Escherichia coli serotype O156:H25/H-/Hnt isolates from cattle and clonal relationship Analysis. Appl Environ Microbiol 76: 5510-5519. doi:10.1128/AEM.00743-10. PubMed: 20581183.
26. Bonnet R (2004) Growing group of extended-spectrum β-lactamases: the CTX-M enzymes. Antimicrob Agents Chemother 48: 1-14. doi: 10.1128/AAC.48.1.1-14.2004. PubMed: 14693512.
27. Dierikx CM, van Duijkeren E, Schoormans AH, van Essen-Zandbergen A, Veldman K et al. (2012) Occurrence and characteristics of extended-spectrum-β-lactamase- and AmpC-producing clinical isolates derived from companion animals and horses. J Antimicrob Chemother 67: 1368-1374. doi:10.1093/jac/dks049. PubMed: 22382469.
28. Simpson's Diversity Index. Available: http://www.countrysideinfo.co.uk/simpsons.htm. Accessed 2013 Aug. 23
29. Wirth T, Falush D, Lan R, Colles F, Mensa P, et al (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 60: 1136-1151. doi:10.1111/j.1365-2958.2006.05172.x. PubMed: 16689791.
30. Nataro JP, Kaper JB (1998) Diarrheagenic *Escherichia coli*. Clin Microbiol Rev 11: 142-201. PubMed: 9457432.
31. Dho-Moulin M, Fairbrother JM (1999) Avian pathogenic *Escherichia coli* (APEC). Vet Res 30: 299-316. PubMed: 10367360.
32. Cloeckaert A, Baucheron S, Flaujac G, Schwarz S, Kehrenberg C et al. (2000) Plasmid-mediated florfenicol resistance encoded by the *floR* gene in *Escherichia coli* isolated from cattle. Antimicrob Agents Chemother 44: 2858-2860. doi:10.1128/AAC.44.10.2858-2860.2000. PubMed: 10991873.
33. Coque TM, Novais A, Carattoli A, Poirel L, Pitout J et al. (2008) Dissemination of clonally related *Escherichia coli* strains expressing extended-spectrum β-lactamase CTX-M-15. Emerg Infect Dis 14: 195-200. doi:10.3201/eid1402.070350. PubMed: 18258110.
34. Mesko Meglic K, Koren S, Palepou MF, Karisik E, Livermore DM et al. (2009) Nationwide survey of CTX-M-type extended-spectrum β-lactamases among *Klebsiella pneumoniae* isolates in Slovenian hospitals. Antimicrob Agents Chemother 53: 287-291. doi:10.1128/AAC.00746-08. PubMed: 19001116.
35. Dhanji H, Doumith M, Rooney PJ, O'Leary MC, Loughrey AC et al. (2011) Molecular epidemiology of fluoroquinolone-resistant ST131 *Escherichia coli* producing CTX-M extended-spectrum β-lactamases in nursing homes in Belfast, UK. J Antimicrob Chemother 66: 297-303. doi:10.1093/jac/dkq463. PubMed: 21131323.

36. Dhanji H, Murphy NM, Akhigbe C, Doumith M, Hope R et al. (2011) Isolation of fluoroquinolone-resistant O25b:H4-ST131 *Escherichia coli* with CTX-M-14 extended-spectrum β-lactamase from UK river water. J Antimicrob Chemother 66: 512-516. PubMed: 21172781.
37. Pitout JD, Campbell L, Church DL, Gregson DB, Laupland KB (2009) Molecular characteristics of travel-related extended-spectrum-β-lactamase-producing *Escherichia coli* isolates from the Calgary Health Region. Antimicrob Agents Chemother 53: 2539-2543. doi:10.1128/AAC.00061-09.
38. Pitout JD, Gregson DB, Church DL, Elsayed S, Laupland KB (2005) Community-wide outbreaks of clonally related CTX-M-14 β-lactamase-producing *Escherichia coli* strains in the Calgary health region. J Clin Microbiol 43: 2844-2849. doi:10.1128/JCM.43.6.2844-2849.2005. PubMed: 15956407.
39. Pitout JD, Laupland KB (2008) Extended-spectrum β-lactamase-producing Enterobacteriaceae: an emerging public-health concern. Lancet Infect Dis 8: 159-166. doi:10.1016/S1473-3099(08)70041-0. PubMed: 18291338.
40. (BIOHAZ) EPoBH (2011) Scientific Opinion on the public health risks of bacterial strains producing extended-spectrum β-lactamases and/or AmpC β-lactamases in food and food-producing animals EFSA Journal 9: 2322.
41. Ewers C, Bethe A, Semmler T, Guenther S, Wieler LH (2012) Extended-spectrum β-lactamase-producing and AmpC-producing *Escherichia coli* from livestock and companion animals, and their putative impact on public health: a global perspective. Clin Microbiol Infect 18: 646-655. doi:10.1111/j.1469-0691.2012.03850.x. PubMed: 22519858.
42. Stokes MO, Cottell JL, Piddock LJV, Wu G, Wootton M et al. (2012) Detection and Characterisation of $bla_{CTX-M-14}$ IncK pCT like plasmid vectors in *Escherichia coli* isolates from humans, turkeys and cattle in England and Wales. J Antimicrob Chemother 67(7):1639-44
43. Wu G, AbuOun M, Hackl E, La Ragione RM, Fookes M et al. (2010) Epidemic multidrug-resistant (MDR-AmpC) *Salmonella enterica* serovar Newport strains contain three phage regions and a MDR resistance plasmid. Environ Microbiol Rep 2: 228-235. doi:10.1111/j.1758-2229.2009.00095.x. PubMed: 23766073.
44. Wu G, Carter B, Mafura M, Liebana E, Woodward MJ et al. (2008) Genetic diversity among *Escherichia coli* O157:H7 isolates and identification of genes linked to human infections. Infect Immun 76: 845-856. doi:10.1128/IAI.00956-07. PubMed: 18070900.
45. Wu G, Ehricht R, Mafura M, Stokes M, Smith N et al. (2012) *Escherichia coli* isolates from extraintestinal organs of livestock animals harbour diverse virulence genes and belong to multiple genetic lineages. Vet Microbiol 160: 197-206. doi:10.1016/j.vetmic.2012.05.029. PubMed: 22766078.

OPEN ACCESS Freely available online

PLOS ONE

Correction

Correction: Comparative Analysis of ESBL-Positive *Escherichia coli* Isolates from Animals and Humans from the UK, The Netherlands and Germany

The *PLOS ONE* Staff

Dr. Marie Chattaway should be included as an author for this article. She should be listed as the 12th author, and her affiliation is: 2. Public Health England, London, United Kingdom. The contributions of this author are as follows: Performed the Experiments, analyzed the data.

Reference

1. Wu G, Day MJ, Mafura MT, Nunez-Garcia J, Fenner JJ, et al. (2013) Comparative Analysis of ESBL-Positive Escherichia coli Isolates from Animals and Humans from the UK, The Netherlands and Germany. PLoS ONE 8(9): e75392. doi:10.1371/journal.pone.0075392.

University of East Anglia

Professor John Wain B.Sc., M.Sc., Ph.D., F.R.C.Path.

14 January 2014

Re: G Wu, MJ Day, et al. Comparative Analysis of ESBL-Positive Escherichia coli Isolates from Animals and Humans from the UK, The Netherlands and Germany. (2013) PloS one 8 (9), e75392

This is to confirm that Marie Chattaway established the Multi Locus Sequence Typing methodology used for her PhD work on *E. coli* using the Safefoodera project isolates. She spent considerable time both extracting DNA and processing sequence data. In recognition of this she has now been added to the Safefoodera consortium and is to be included, where appropriate, in the writing of any further papers by the consortium. Her role in the generation and analysis of data for the above paper should have been sufficient to warrant authorship but the researchers who submitted the manuscript were not made aware of her contribution. In my opinion this manuscript can be included as publications arising from this thesis.

Yours sincerely

Professor John Wain,

Medical Microbiology Research
University of East Anglia
NRP Innovation Centre, Norwich Research Park,
Colney Lane, Norwich, Norfolk, NR4 7GJ, UK
Telephone: +44 1603 567597

7.10 Investigating the link between the presence of enteroaggregative Escherichia coli and infectious intestinal disease in the United Kingdom, 1993 to 1996 and 2008 to 2009.

Chattaway M, Harris R, Jenkins C, Tam C, Coia J, Gray J, Iturriza-Gomara M and Wain J. Euro.Surveill 18 (2013)37

RESEARCH ARTICLES

Investigating the link between the presence of enteroaggregative *Escherichia coli* and infectious intestinal disease in the United Kingdom, 1993 to 1996 and 2008 to 2009

M A Chattaway[1], R Harris[2], C Jenkins[1], C Tam[3], J E Coia[4], J Gray[5], M Iturriza-Gomara[6], J Wain (J.Wain@uea.ac.uk)[1,3]
1. Gastrointestinal Bacteria Reference Unit, Public Health England, Colindale, London, United Kingdom
2. Statistics, Modelling and Economics Department, Public Health England, Colindale, United Kingdom
3. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, United Kingdom
4. Scottish Salmonella, Shigella and Clostridium difficile Reference Laboratory, Glasgow, United Kingdom
5. Microbiology, Norwich and Norfolk University hospital and the Norwich Medical School, University of East Anglia, Norfolk, United Kingdom
6. Clinical Infection, Microbiology and Immunology, University of Liverpool, United Kingdom

Citation style for this article:
Chattaway MA, Harris R, Jenkins C, Tam C, Coia JE, Gray J, Iturriza-Gomara M, Wain J. Investigating the link between the presence of enteroaggregative Escherichia coli and infectious intestinal disease in the United Kingdom, 1993 to 1996 and 2008 to 2009. Euro Surveill. 2013;18(37):pii=20582. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20582

Article submitted on 24 January 2013 / published on 12 September 2013

There are an estimated 17 million human diarrhoea cases annually in the United Kingdom. In 2008 and 2009, enteroaggregative *E. coli* (EAEC) were identified in 1.9% of stools. However, it remains unclear whether there is a causal link between presence of EAEC and disease. This study used bacterial load, the presence of co-infections and demographic data to assess if EAEC was independently associated with intestinal infectious disease. Quantitative real-time PCR data (Ct values) generated directly from stool specimens for several pathogen targets were analysed to identify multiple pathogens, including EAEC, in the stools of cases and healthy controls. Sensitivity and specificity using Ct value (60% and 60%) was not useful for identifying cases or controls, but an independent association between disease and EAEC presence was demonstrated: multivariate logistic regression for EAEC presence (odds ratio: 2.41; 95% confidence interval: 1.78–3.26; p<0.001). The population-attributable fraction was 3.3%. The group of bacteria known as EAEC are associated with gastrointestinal disease in at least half of the cases with EAEC positive stools. We conclude that the current definition of EAEC, by plasmid gene detection, includes true pathogens as well as non-pathogenic variants.

Introduction

Measuring the burden of infectious disease is essential for the rational design of public health intervention strategies and for the allocation of resources. For intestinal infectious diseases (IID) there is a massive global burden; the World Health Organization (WHO) estimates around 2 billion cases every year [1]. Detailed surveillance studies have shown that there are up to 17 million sporadic community cases of IID and one million general practitioner (GP) consultations annually in the United Kingdom (UK) [2]. Routine investigations of IID in the UK include salmonellosis, shigellosis, campylobacteriosis, cholera, infection with verotoxin-producing *Escherichia coli* O157 (VTEC), rotavirus, norovirus and parasitic infections and yet no cause is identified for over half of the laboratory-investigated diarrhoeal episodes [3]. One, often undiagnosed, potential pathogen is enteroaggregative *E. coli* (EAEC). In England, this pathotype of *E. coli*, defined by the ability to aggregate to HEp-2 cells [4], has been associated with cases of gastrointestinal infection [2,5,6] at a level comparable to *Salmonella* [6,7]. EAEC gained notoriety during a recent outbreak in Germany and France caused by an *E. coli* strain that was both a verotoxin-producing and enteroaggregative [8]. This outbreak was unusual due to the scale of morbidity and mortality, high even for VTEC infection, and the acquisition of the EAEC plasmid which may have played an important role in adherence to the human gut; the *E. coli* strain that caused the outbreak lacked the attachment and effacement (*eae*) gene for intimate adherence to human gut epithelium normally associated with severe disease caused by VTEC [9]. The emergence of this hybrid pathogen has been described before in 1996, when an O111:H2 strain had caused an outbreak of haemolytic uraemic syndrome (HUS) in France [10], in 1999, when an O86:H strain associated with HUS was isolated in Japan [11], and most recently in 2011, when an O111:H21 strain was associated with a family outbreak in Ireland [12]. All of these cases were associated with severe disease. It is likely that there are more cases of IID caused by EAEC and VTEC hybrids, but the EAEC pathotype is not routinely looked for.

Although EAEC itself has been associated with disease globally [13-19] including outbreaks (most notably a large outbreak in Japan involving 2,697 children [20]), a considerable proportion of healthy controls in case–control studies (16–31%) also harbour this pathotype [21-23]. Furthermore, research data describing the association of genetic factors with virulence are contradictory [21,24,25]. The reliability of virulence factors to identify EAEC for diagnostic purposes is therefore unclear [16]. The situation is further complicated by the presence of co-infections in IID [7]. When multiple pathogens are present in a diarrhoeic stool, defining which are causing the symptoms can be problematic, and as diagnostic tools improve, mixed infections in the gut are being recognised more frequently [26]. This is especially true in studies looking at EAEC infection; in Peru, for instance, multiple pathogens are found in 40% of infants with diarrhoea and with EAEC in their stool [27].

The successful completion of two IID burden studies in the UK [2,6] using quantitative PCR, presented the opportunity to investigate the causal link between gastrointestinal disease and the presence of EAEC in the stool. We estimated bacterial load for EAEC and the presence of co-infection in a well-defined population in the UK and tested the independent association between EAEC presence and disease.

Methods

Datasets

Data from two IID studies were used in this analysis: the IID1 case–control study (August 1993–January 1996) [6,7,28] and the IID2 case-only study (April 2008–March 2009) [29]. The data had been generated by testing stool samples by real-time PCR for the presence of a range of pathogens and recording the number of PCR cycles (Ct) needed before detection of product, to give a semi-quantitative estimate of bacterial load. The EAEC probe was the anti-aggregation protein transporter gene CVD432/*aatD* [30].

Cases of IID were defined in the same way in both studies as having had more than one loose stool, or clinically significant vomiting, over a two week period with no underlying non-infectious cause, followed by a symptom-free period of three weeks [2]. Healthy controls (IID-free) were only recruited in IID1 and were selected from the study cohort, matched for age and sex, and asked to submit a stool specimen.

The dataset for the IID1 case–control study contained 4,664 stool specimens (2,443 cases, 2,221 controls); EAEC was detected, by PCR, in 113 cases and in 38 controls but real-time Ct values (for the EAEC probe) were only available for 102 cases and 31 controls; in this study, all 151 positive cases were used for descriptive comparisons, and the 133 with Ct values for quantitative analysis.

The dataset for the IID2 case-only study [29] contained PCR Ct values from 3,966 stools (all of which were from individuals with diarrhoea); EAEC was detected in 83 of them. These data were used for burden estimations and comparisons of demographic data for cases; there had been no controls recruited in IID2 case-only study, and so IID1 data only were used for comparison of cases with controls.

Statistical methods

One aim of this study was to assess the methods for estimating burden of EAEC in England from the current IID2 study results. However, no controls were recruited to the IID2 study and so a receiver-operating characteristic (ROC) analysis was constructed from the case–control data (IID1), and used to look for a cut-off between case and control in the Ct values. We compared the distribution of Ct values from EAEC-positive cases and controls using Student's t-test.

It is clear that the relationship between presence of EAEC and disease is not absolute and so several methods were used to investigate the association of EAEC with disease:

Carriage rates of EAEC in healthy controls, compared to other pathogens

For each infection, the chi-squared test was used to test if the distribution of the pathogen between cases and controls was as expected by chance.

Association of disease with individual pathogens in persons with multiple pathogens in their stool

For all EAEC-positive individuals with multiple pathogens (both cases and controls), we tested whether individual pathogens were equally distributed between cases and controls using chi-squared tests for independence. Because norovirus was the most common pathogen, we also compared by chi-squared test co-infection in all individuals positive for EAEC and all individuals positive for norovirus to see if the presence of other individual pathogens was dependent on infection with EAEC or norovirus.

Independent association of EAEC presence with disease

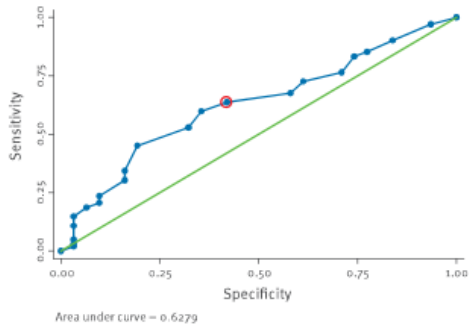
A logistic regression of univariate and multivariate analysis was carried out using case or control as outcome, and infecting agent and age as independent variables. In this way we assessed the independent association between EAEC and disease, while controlling for other pathogens. Model results were then used to calculate the population attributable fraction (PAF):

$$PAF = P_e (RR_e - 1) / RR_e,$$

where P_e is the proportion of cases with the exposure (EAEC) and RR_e the relative risk of disease. This form allows for confounding of the exposure if an adjusted RR is used, as recommended in Rockhill et al. [31]. In that case, adjusted odds ratios (OR) are substituted

FIGURE 1

Receiver-operating characteristic analysis of Ct values for enteroaggregative *Escherichia coli* from gastrointestinal disease cases (n=102) and controls (n=31) , United Kingdom, August 1993–January 1996



The red circle at Ct value 31 indicates the cut-off value which was chosen at the point where sensitivity and specificity were equivalent.

into this equation to give an approximate, adjusted PAF.

Results

Defining diagnostic cut-off values for Ct values in EAEC infection

In order to investigate the link between Ct value and disease, the sensitivity and specificity of the Ct value was assessed in EAEC-positive specimens from the case–control study (dataset IID1); Ct values were obtained and included 102 cases and 31 controls. Figure 1 shows the resulting ROC curve, and Figure 2 the distribution of Ct values in cases and controls. The cut-off was chosen to balance sensitivity and specificity and was set at a Ct value of 31 (Figure 1). The ratio of false positives versus false negatives with this cut-off point was 1.09 (95% confidence interval (CI): 0.79–1.53) (Figure 2), so the total number of test-positives, although not a good diagnostic for the individual, was a reasonable estimate of the total number of cases. Importantly however, in the population studied, there was a significant association between bacterial load and disease state (p=0.039), and further investigations were carried out using the point of <40 to indicate presence of EAEC.

Descriptive statistics

To test if the analysis of data from the IID1 case–control study remained relevant in 2009, we compared the demographic data from the two periods. There was no significant difference between the rate of EAEC in the IID1 case–control study (1993–96) and IID2 case-only study (2008–09), with 1.4% and 1.9%, respectively; individuals with EAEC present in their stool were distributed evenly across all age groups in both IID1 and IID2 (chi-squared p value for non-independence: 0.253). For EAEC-positive individuals, there was no significant difference in age between cases and controls (p=0.237). We therefore believe that the epidemiology did not change significantly for EAEC infection between the two periods. Cases tended to be slightly older than controls in IID1 (mean age of cases: 30.1 years, standard deviation (SD): 24.7 years; mean age of controls 28.7 years, SD: 23.9 years; p value for difference: 0.051).

Investigation of the association of EAEC presence with disease

Carriage rates of EAEC, compared to other pathogens, in healthy controls

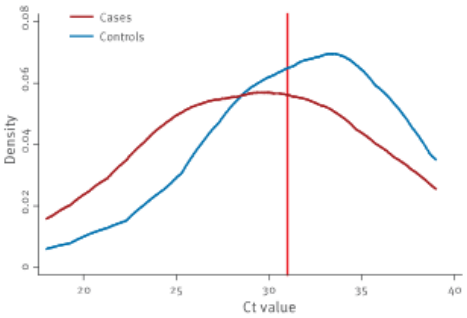
Submitting a stool specimen that was positive for EAEC was positively associated with having disease (Figure 3). However, one quarter of all EAEC positive individuals were asymptomatic (38/151).

Association of disease with individual pathogens in persons with multiple pathogens in their stool

The presence of co-infection was almost three times higher in EAEC-positive cases (74/113, 66%) than in EAEC-positive asymptomatic controls (9/38, 24%)_ (Figure 4). Cases had more multiple co-infections

FIGURE 2

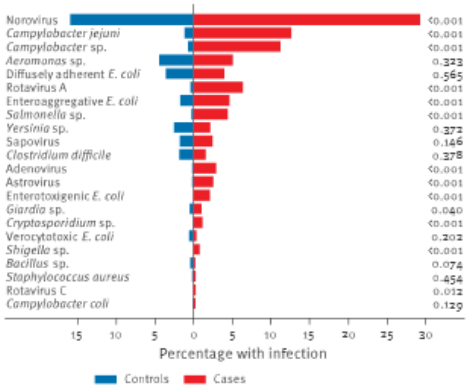
Distribution of Ct values for curve analysis of enteroaggregative *Escherichia coli* in gastrointestinal disease cases (n=102) and controls (n=31), United Kingdom, August 1993–January 1996



Fitted Curve distribution of Ct values. The red line indicates the cut-off point where the ratio of false positives versus false negatives with this cut-off point was closest to equivalent 1.09; 95% confidence interval: 0.79–1.53.

FIGURE 3

Organisms present in stool samples from gastrointestinal disease cases (n=2,221) and controls (n=2,243) in the IID1 study, United Kingdom, August 1993–January 1996



Submitting a stool specimen that was positive for enteroaggregative *Escherichia coli* (EAEC) was positively associated with having disease. EAEC was found in <2% of controls, indicating that EAEC is not a ubiquitous commensal organism.

The p values are indicated on the right.

(38/113, 34%) than controls (1/38, 3%) (chi-square test, p<0.001).

Investigation of the independent association of EAEC presence with disease

The logistic regression of EAEC status (but not Ct value) in univariate analysis gave an OR of 2.55 (95% CI: 1.91–3.39, p≤0.001); in multivaritate analysis, the OR was 2.41 (95% CI: 1.78–3.26, p≤0.001). This means that among IID cases, the odds of EAEC infection were 2.5 times higher compared with asymptomatic controls. The resulting adjusted PAF was 0.033% (95% CI: 0.024–0.039), suggesting that around 3.3% of cases of IDD in the UK were attributable to EAEC. This confirmed that EAEC was an independent cause of IID.

A comparison of co-infections with the most common cause of IID, norovirus, is presented in Figure 5.

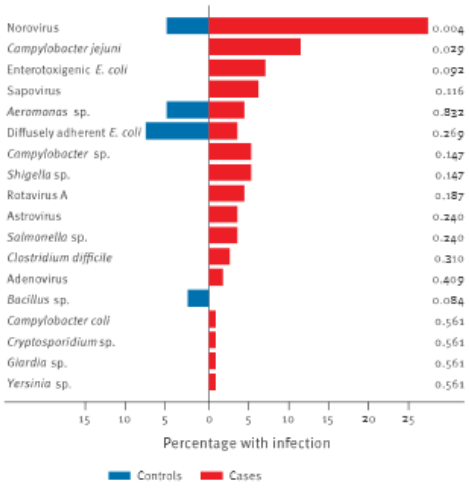
Discussion

Although described as a pathogenic group of *E. coli*, it is well documented that EAEC may be associated with asymptomatic infection [21-23]. In this study we asked the question how much disease EAEC is responsible for. In an attempt to remove healthy carriers from the case definition (a lower bacterial load might be expected in carriers than in cases), we analysed data from a PCR-based case–control study (IID1). Using the Ct value as an indicator of bacterial load, we were only able to define a cut-off with 60% sensitivity and specificity. These values suggest that estimation of bacterial load by the Ct value of a quantitative PCR for virulence factors is not a useful diagnostic test for EAEC infection.

However, there was a strong association between higher load (low Ct) and being a case, so we tried to define more accurately in which positive individuals EAEC was the causal agent of diarrhoea. The bacterial load data revealed the presence of two overlapping normally distributed data sets for EAEC: one representing the load in health (controls) and one in disease (cases) (see Figure 2). We further addressed any possible confounding effects of age (i.e. acquired immunity) and co-infection using logistic regression confirmed by univariate analysis; the results showed that an individual was 2.5 times more likely to be a case than a control if they had EAEC. Therefore we concluded that EAEC was

FIGURE 4

Co-infection with enteroaggregative *Escherichia coli* in gastrointestinal disease cases (n=113) and controls (n=38) in the IID1 study, United Kingdom, August 1993–January 1996



EAEC: enteroaggregative *E.coli*.

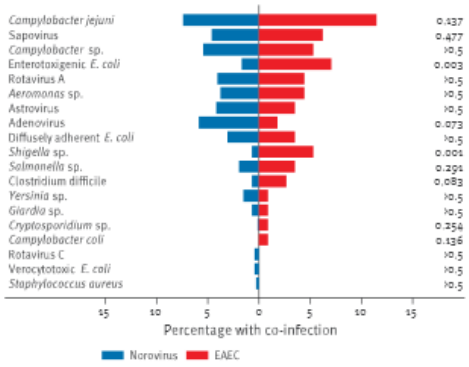
There were a higher variety of co-infection types, a higher percentage of co-infections and more multiple co-infection in EAEC-positive cases than in EAEC-positive controls.

Note: organisms designated sp. include all species of that genus (except *Campylobacter* sp. which list *C. jejuni* and *C. coli* separately), *Staphylococcus aureus* refers to all *S. aureus* >10⁶/g.

The p values are indicated on the right.

FIGURE 5

Comparison of co-infections with enteroaggregative *Escherichia coli* (n=113) or norovirus (n=715), United Kingdom, August 1993–January 1996



EAEC: enteroaggregative *E.coli*.

Co-infection with EAEC was more common than with norovirus (66% versus 43%).

The p values for individual agents are indicated on the right.

independently associated with disease and we investigated the factors influencing this association.

Our results suggest that EAEC is common in the absence of disease. This situation is similar for gastrointestinal viral infection where post-infection levels of virus particles, although reduced, persist up to 56 days after symptoms have cleared [32,33]. Another possibility is pre-existing immunity to the infection at the time of exposure, which could result in reduced viral replication and a failure to develop symptoms. If pre-existing immunity was the cause of symptomless EAEC carriage we would expect to find an age distribution where adults are less frequently infected (older individuals have a higher chance of exposure and therefore a higher chance of immunity). The age distribution was even across the age groups and, as seen in the ROC analysis, the association between bacterial load and symptoms was not strong. Therefore we investigated an alternative explanation, the presence of a co-infecting pathogen.

The presence of increased co-infection in cases raises the possibility that the co-infecting pathogen rather than the EAEC, or a combination of both, is causing disease. To test this hypothesis we took norovirus, an infectious agent known to be present in both symptomatic and asymptomatic infection, as a comparator. As norovirus was a very common infection, we removed cases infected simultaneously with both norovirus and EAEC from the calculation: there were slightly more co-infections in EAEC-positive cases than in norovirus-positive cases (66% versus 43%). For EAEC co-infection, 12.6% were explained by enterotoxigenic *E. coli* (ETEC) and *Shigella* co-infections (Figure 5). This suggests that a proportion of EAEC cases can be explained by other pathogens (ETEC and *Shigella* are associated almost exclusively with symptomatic infection), but by no means all cases.

The logistic regression of co-infection univariate and multivaritate was statistically significant and again confirmed that EAEC was independently associated with disease; the odds of disease were 2.4 times higher if EAEC was present than if not and were still highly significant after controlling for co-infections. The PAF adjustments indicated that EAEC would be responsible for disease in 3.3% of cases, a significant proportion in gastrointestinal disease, higher than for *Salmonella* [2]. Although age was an independent predictor for disease overall, controlling for age did not change the association of disease with EAEC, and there was no interaction between EAEC and age.

This study did not directly address causality over association, but we believe that bacterial variation best explains the observed association of EAEC with disease for the following reasons. There are two common arguments for EAEC being found in high levels in healthy individuals: (i) Low levels of EAEC are present in a symptomless commensal relationship in the human gut and only increase to detectable levels after infection with a true pathogen because adherence of EAEC to the gut epithelium is stronger than for other commensals; an independent association of EAEC with disease argues against this for at least half of the infections in this study. (ii) Post-infection immunity leads to carriage in apparently healthy individuals; lack of any detectable trends in age distribution and no clear association between pathogen load and disease, as seen in norovirus infection [34], suggest that acquired immunity against EAEC does not protect against infection and is therefore unlikely to lead to symptomless carriage. Transient passage, as with plant viruses, is also unlikely, as there is no known reservoir for exposure to EAEC from outside the human gut.

It seems therefore clear that some, but not all, EAEC cause disease. The explanation for this may be that EAEC are defined by in vitro phenotype rather than by the ability to cause disease: non-pathogenic EAEC, able to agglutinate cells in the laboratory but unable to cause disease in the human host, are found in controls and in co-infections with true pathogens, but pathogenic variants are found as the sole pathogen detected in diarrhoeic stools. Attempts to define genetic markers for EAEC using alternative probes still do not define those EAEC capable of causing disease: the presence of the *aat* (anti-aggregative transporter) [35] or *aggR* (a transcriptional activator) [13,18,35] does not correlate

precisely with disease, but rather with the ability to agglutinate cells in the laboratory.

It may be that the genetic factors used for EAEC diagnostics are not true virulence factors and that they rather encode the ability to adhere to human intestinal cells and allow colonisation (especially during infection with a true pathogen). It is likely that a combination of the EAEC-associated adherence factors and a true virulence factor allows EAEC to cause primary infection. This was seen in the German ST678 (O104:H4) outbreak [36], where the EAEC adherence genes were present in the same bacterial host as the Shiga-like toxin gene (*stx*). We suggest that an appropriate diagnostic test for pathogenic EAEC should look for the EAEC plasmid genes and other virulence factors. More work is still needed to define those other virulence factors in diarrhoeagenic EAEC.

The main limitation of this study is the lack of controls in the IID2 study. Although there were 20 years between the IID1 and IID2 studies, the demographic data for cases suggest that the epidemiology has not changed during that period. Although there may have been some change in co-infection rates, we believe the data to be relevant in 2013. Another limitation, but also a strength, of the study is the range of infectious agents identified. Small numbers in some groups of cases with co-infections (six cases or less for EAEC co-infections with *C. difficile, Yersinia, Giardia, Cryptosporidium*, rotavirus C, VTEC and *Staphylococcus*) mean that the ability to detect statistical differences between cases and controls was limited. However, the study allowed us, for the first time, to explore the association between EAEC and all potential co-infecting agents as well as the more common pathogens norovirus (n=29) and *Campylobacter* (n=12).

Conclusion

This study highlights the importance of EAEC as a pathogenic group of bacteria which caused disease in more than 1% of all IID cases in the UK in 2008–09. The EAEC group is most likely to be a mixture of pathotypes which needs to be split into rational subgroups before tests for detection and typing can be implemented. Detailed studies of the genetic content of EAECs from case–control studies are warranted.

Acknowledgements

We would like to thank the study teams for allowing us access to the data generated from the two IID studies and in particular Dr Corrine Amar from the Laboratory of Gastrointestinal Pathogens (LGP) Unit, Health Protection Agency (HPA), London and Sarah O'Brien, University of Liverpool, Liverpool. We would also like to thank Prof Iruka Okeke from the Molecular Microbiology Department, Haverford College, Pennsylvania, USA and Kathie Grant from LGP, HPA, London for valuable advice with this manuscript.

Conflict of interest

None declared.

Author's contributions

Conception, design of study, interpretation of data, drafting and revising manuscript: MA Chattaway and J Wain; Acquisition and analysis of statistical data: R Harris; Drafting manuscript and interpretation of data: T Clarence, M Iturriza-Gomara, Claire Jenkins and John Coia.

References

1. World Health Organization (WHO). Diarrhoeal disease. Fact Sheet no 330. Geneva: WHO; Apr 2013. Available from: http://www.who.int/mediacentre/factsheets/fs330/en/
2. Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, Hunter PR, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut. 2012;61(1):69-77. http://dx.doi.org/10.1136/gut.2011.238386. PMid:21708822. PMCid:PMC3230829.
3. Tam CC, O'Brien SJ, Tompkins DS, Bolton FJ, Berry L, Dodds J, et al. Changes in causes of acute gastroenteritis in the United Kingdom over 15 years: microbiologic findings from 2 prospective, population-based studies of infectious intestinal disease. Clin Infect Dis. 2012;54(9):1275-86. http://dx.doi.org/10.1093/cid/cis028. PMid:22412058.
4. Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, Levine MM. Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. Pediatr Infect Dis J. 1987;6(9):829-31. http://dx.doi.org/10.1097/00006454-198709000-00008. PMid:3313248.
5. Nataro JP, Mai V, Johnson J, Blackwelder WC, Heimer R, Tirrell S, et al. Diarrheagenic Escherichia coli infection in Baltimore, Maryland, and New Haven, Connecticut. Clin Infect Dis. 2006;43(4):402-7. http://dx.doi.org/10.1086/505867. PMid:16838226.
6. Tompkins DS, Hudson MJ, Smith HR, Eglin RP, Wheeler JG, Brett MM, et al. A study of infectious intestinal disease in England: microbiological findings in cases and controls. Commun Dis Public Health. 1999;2(2):108-13. PMid:10402744.
7. Amar CF, East CL, Gray J, Iturriza-Gomara M, Maclure EA, McLauchlin J. Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996). Eur J Clin Microbiol Infect Dis. 2007;26(5):311-23. http://dx.doi.org/10.1007/s10096-007-0290-8. PMid:17447091.
8. Rubino S, Cappuccinelli P, Kelvin DJ. Escherichia coli (STEC) serotype O104 outbreak causing haemolytic syndrome (HUS) in Germany and France. J Infect Dev Ctries. 2011;5(6):437-40. http://dx.doi.org/10.3855/jidc.2172. PMid:21727641.
9. Jenkins C, Willshaw GA, Evans J, Cheasty T, Chart H, Shaw DJ, et al. Subtyping of virulence genes in verocytotoxin-producing Escherichia coli (VTEC) other than serogroup O157 associated with disease in the United Kingdom. J Med Microbiol. 2003;52(Pt 11):941-7. http://dx.doi.org/10.1099/jmm.0.05160-0. PMid:14532337.
10. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Prior K, et al. Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. PLoS One. 2011;6(7):e22751. http://dx.doi.org/10.1371/journal.pone.0022751. PMid:21799941. PMCid:PMC3140518.
11. Iyoda S, Tamura K, Itoh K, Izumiya H, Ueno N, Nagata K, et al. Inducible stx2 phages are lysogenized in the enteroaggregative and other phenotypic Escherichia coli O86:HNM isolated from patients. FEMS Microbiol Lett. 2000;191(1):7-10. http://dx.doi.org/10.1111/j.1574-6968.2000.tb09311.x. PMid:11004392.
12. Dallman T, Smith GP, O'Brien B, Chattaway M, Finlay D, Grant KA, et al. Characterization of a verocytotoxin-producing enteroaggregative Escherichia coli serogroup O111:H21 strain associated with a household outbreak in Northern Ireland. J Clin Microbiol. 2012;50(12):4116-9. http://dx.doi.org/10.1128/JCM.02047-12. PMid:23035193. PMCid:PMC3502983.
13. Antikainen J, Tarkka E, Haukka K, Siitonen A, Vaara M, Kirveskari J. New 16-plex PCR method for rapid detection of diarrheagenic Escherichia coli directly from stool samples. Eur J Clin Microbiol Infect Dis. 2009;28(8):899-908. http://dx.doi.org/10.1007/s10096-009-0720-x. PMid:19238467.

14. Cennimo D, Abbas A, Huang DB, Chiang T. The prevalence and virulence characteristics of enteroaggregative Escherichia coli at an urgent-care clinic in the USA: a case-control study. J Med Microbiol. 2009;58(Pt 4):403-7. http://dx.doi.org/10.1099/jmm.0.005793-0. PMid:19273633.
15. Jenkins C, Tembo M, Chart H, Cheasty T, Willshaw GA, Phillips AD, et al. Detection of enteroaggregative Escherichia coli in faecal samples from patients in the community with diarrhoea. J Med Microbiol. 2006;55(11):1493-7. http://dx.doi.org/10.1099/jmm.0.46683-0. PMid:17030907.
16. Okeke IN, Nataro JP. Enteroaggregative Escherichia coli. Lancet Infect Dis. 2001;1(5):304-13. http://dx.doi.org/10.1016/S1473-3099(01)00144-X
17. Okhuysen PC, Dupont HL. Enteroaggregative Escherichia coli (EAEC): a cause of acute and persistent diarrhea of worldwide importance. J Infect Dis. 2010;202(4):503-5. http://dx.doi.org/10.1086/654895. PMid:20594108.
18. Toma C, Lu Y, Higa N, Naksone N, Chinen I, Baschkier A, et al. Multiplex PCR assay for identification of human diarrheagenic Escherichia coli. J Clin Microbiol. 2003;41(6):2669-71. http://dx.doi.org/10.1128/JCM.41.6.2669-2671.2003. PMid:12791900. PMCid:PMC156568.
19. Usein CR, Tatu-Chitoiu D, Ciontea S, Condei M, Damian M. Escherichia coli pathotypes associated with diarrhea in Romanian children younger than 5 years of age. Jpn J Infect Dis. 2009;62(4):289-93. PMid:19628907.
20. Itoh Y, Nagano I, Kunishima M, Ezaki T. Laboratory investigation of enteroaggregative Escherichia coli O untypeable:H10 associated with a massive outbreak of gastrointestinal illness. J Clin Microbiol. 1997;35(10):2546-50. PMid:9316905. PMCid:PMC230008.
21. Huang DB, Mohamed JA, Nataro JP, DuPont HL, Jiang ZD, Okhuysen PC. Virulence characteristics and the molecular epidemiology of enteroaggregative Escherichia coli isolates from travellers to developing countries. J Med Microbiol. 2007;56(10):1386-92. http://dx.doi.org/10.1099/jmm.0.47161-0. PMid:17893178.
22. Okeke IN, Lamikanra A, Czeczulin J, Dubovsky F, Kaper JB, Nataro JP. Heterogeneous virulence of enteroaggregative Escherichia coli strains isolated from children in Southwest Nigeria. J Infect Dis. 2000;181(1):252-60. http://dx.doi.org/10.1086/315204. PMid:10608774.
23. Regua-Mangia AH, Gomes TA, Vieira MA, Irino K, Teixeira LM. Molecular typing and virulence of enteroaggregative Escherichia coli strains isolated from children with and without diarrhoea in Rio de Janeiro city, Brazil. J Med Microbiol. 2009;58(4):414-22. http://dx.doi.org/10.1099/jmm.0.006502-0. PMid:19273635.
24. Boisen N, Scheutz F, Rasko DA, Redman JC, Persson S, Simon J, et al. Genomic characterization of enteroaggregative Escherichia coli from children in Mali. J Infect Dis. 2012;205(3):431-44. http://dx.doi.org/10.1093/infdis/jir757. PMid:22184729. PMCid:PMC3256949.
25. Jenkins C, van Ijperen C, Dudley EG, Chart H, Willshaw GA, Cheasty T, et al. Use of a microarray to assess the distribution of plasmid and chromosomal virulence genes in strains of enteroaggregative Escherichia coli. FEMS Microbiol Lett. 2005;253(1):119-24. http://dx.doi.org/10.1016/j.femsle.2005.09.040. PMid:16243450.
26. Lindsay B, Ramamurthy T, Sen Gupta S, Takeda Y, Rajendran K, Nair GB, et al. Diarrheagenic pathogens in polymicrobial infections. Emerg Infect Dis. 2011;17(4):606-11. PMid:21470448. PMCid:PMC3377398.
27. Ochoa TJ, Ecker L, Barletta F, Mispireta ML, Gil AI, Contreras C, et al. Age-related susceptibility to infection with diarrheagenic Escherichia coli among infants from Periurban areas in Lima, Peru. Clin Infect Dis. 2009;49(11):1694-702. http://dx.doi.org/10.1086/648069. PMid:19857163. PMCid:PMC2779581.
28. Sethi D, Wheeler JG, Cowden JM, Rodrigues LC, Sockett PN, Roberts JA, et al. A study of infectious intestinal disease in England: plan and methods of data collection. Commun Dis Public Health. 1999;2(2):101-7. PMid:10402743.
29. O'Brien SJ, Rait G, Hunter PR, Gray JJ, Bolton FJ, Tompkins DS, et al. Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). BMC Med Res Methodol. 2010;10:39. http://dx.doi.org/10.1186/1471-2288-10-39. PMid:20444246. PMCid:PMC2886083.
30. Baudry B, Savarino SJ, Vial P, Kaper JB, Levine MM. A sensitive and specific DNA probe to identify enteroaggregative Escherichia coli, a recently discovered diarrheal pathogen. J Infect Dis. 1990;161(6):1249-51. http://dx.doi.org/10.1093/infdis/161.6.1249. PMid:2189007.
31. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health. 1998;88(1):15-9. http://dx.doi.org/10.2105/AJPH.88.1.15. PMid:9584027. PMCid:PMC1508384.
32. Atmar RL, Opekun AR, Gilger MA, Estes MK, Crawford SE, Neill FH, et al. Norwalk virus shedding after experimental human infection. Emerg Infect Dis. 2008;14(10):1553-7. http://dx.doi.org/10.3201/eid1410.080117. PMid:18826818. PMCid:PMC2609865.
33. Partridge DG, Evans CM, Raza M, Kudesia G, Parsons HK. Lessons from a large norovirus outbreak: impact of viral load, patient age and ward design on duration of symptoms and shedding and likelihood of transmission. J Hosp Infect. 2012;81(1):25-30. http://dx.doi.org/10.1016/j.jhin.2012.02.002. PMid:22445090.
34. Phillips G, Lopman B, Tam CC, Iturriza-Gomara M, Brown D, Gray J. Diagnosing norovirus-associated infectious intestinal disease using viral load. BMC Infect Dis. 2009;9:63. http://dx.doi.org/10.1186/1471-2334-9-63. PMid:19442278. PMCid:PMC2698835.
35. Denno DM, Shaikh N, Stapp JR, Qin X, Hutter CM, Hoffman V, et al. Diarrhea Etiology in a Pediatric Emergency Department: A Case Control Study. Clin Infect Dis. 2012;55(7):897-904. http://dx.doi.org/10.1093/cid/cis553. PMid:22700832.
36. Chattaway MA, Dallman T, Okeke IN, Wain J. Enteroaggregative E. coli O104 from an outbreak of HUS in Germany 2011, could it happen again? J Infect Dev Ctries. 2011;5(6):425-36. http://dx.doi.org/10.3855/jidc.2166. PMid:21727640.

7.11 An investigation of the Diversity of Strains of Enteroaggregative Escherichia coli isolated from cases associated with a large multi-pathogen foodborne outbreak in the UK

Dallman, T.J., Chattaway, M.A., Cowley, L.A., Doumith, M., Tewolde, R., Wooldridge, D.J., Underwood, A., Ready, D., Wain, J., Foster, K., Grant, K.A., & Jenkins, C. PLOSONE 9, (5) e98103 available from: PM:24844597

OPEN ACCESS Freely available online

PLOS | ONE

An Investigation of the Diversity of Strains of Enteroaggregative *Escherichia coli* Isolated from Cases Associated with a Large Multi-Pathogen Foodborne Outbreak in the UK

Timothy J. Dallman[1], Marie A. Chattaway[1], Lauren A. Cowley[1], Michel Doumith[2], Rediat Tewolde[3], David J. Wooldridge[4], Anthony Underwood[3], Derren Ready[5], John Wain[6], Kirsty Foster[7], Kathie A. Grant[1], Claire Jenkins[1]*

1 Gastrointestinial Bacteria Reference Unit, Public Health England, London, United Kingdom, **2** Antimicrobial Resistance and Healthcare Associated Infections Reference Unit, Public Health England, London, United Kingdom, **3** Applied Laboratory and Bio- Informatics Unit, Public Health England, London, United Kingdom, **4** Department for Bio-analysis and Horizon Technologies, Next Generation Sequencing, Public Health England, London, United Kingdom, **5** Public Health Laboratory London, Barts & The London Hospital Trust, London, United Kingdom, **6** Innovation Centre, University of East Anglia, Norwich, United Kingdom, **7** North East Public Health England Centre, Newcastle upon Tyne, United Kingdom

Abstract

Following a large outbreak of foodborne gastrointestinal (GI) disease, a multiplex PCR approach was used retrospectively to investigate faecal specimens from 88 of the 413 reported cases. Gene targets from a range of bacterial GI pathogens were detected, including *Salmonella* species, *Shigella* species and Shiga toxin-producing *Escherichia coli*, with the majority (75%) of faecal specimens being PCR positive for *aggR* associated with the Enteroaggregative *E. coli* (EAEC) group. The 20 isolates of EAEC recovered from the outbreak specimens exhibited a range of serotypes, the most frequent being O104:H4 and O131:H27. None of the EAEC isolates had the Shiga toxin (stx) genes. Multilocus sequence typing and single nucleotide polymorphism analysis of the core genome confirmed the diverse phylogeny of the strains. The analysis also revealed a close phylogenetic relationship between the EAEC O104:H4 strains in this outbreak and the strain of *E. coli* O104:H4 associated with a large outbreak of haemolytic ureamic syndrome in Germany in 2011. Further analysis of the EAEC plasmids, encoding the key enteroaggregative virulence genes, showed diversity with respect to FIB/FII type, gene content and genomic architecture. Known EAEC virulence genes, such as *aggR*, *aat* and *aap*, were present in all but one of the strains. A variety of fimbrial genes were observed, including genes encoding all five known fimbrial types, AAF/1 to AAF/V. The AAI operon was present in its entirety in 15 of the EAEC strains, absent in three and present, but incomplete, in two isolates. EAEC is known to be a diverse pathotype and this study demonstrates that a high level of diversity in strains recovered from cases associated with a single outbreak. Although the EAEC in this study did not carry the *stx* genes, this outbreak provides further evidence of the pathogenic potential of the EAEC O104:H4 serotype.

Citation: Dallman TJ, Chattaway MA, Cowley LA, Doumith M, Tewolde R, et al. (2014) An Investigation of the Diversity of Strains of Enteroaggregative *Escherichia coli* Isolated from Cases Associated with a Large Multi-Pathogen Foodborne Outbreak in the UK. PLoS ONE 9(5): e98103. doi:10.1371/journal.pone.0098103

Editor: Axel Cloeckaert, Institut National de la Recherche Agronomique, France

Received December 30, 2013; **Accepted** April 28, 2014; **Published** May 20, 2014

Funding: These authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: claire.jenkins@phe.gov.uk

Introduction

The Enteroaggregative *Escherichia coli* (EAEC) group is a large, diverse group of diarrhoeagenic *E. coli* originally defined by their adherence to HEp-2 cells in a stacked brick formation [1]. Generally, EAEC are detected and identified using PCR targeting EAEC associated virulence genes that are predominately plasmid encoded, including a regulator of multiple plasmid virulence factors (*aggR*), the anti-aggregation transporter gene (*aat*) and the gene encoding dispersin (*aap*) [2–4]. AggR also activates the expression of the chromosomal *aai* genes encoding a Type VI Secretion System (T6SS) [5].

Virulence gene content associated with EAEC is highly variable between different strains, as illustrated in studies aimed at genotyping EAEC from a variety of clinical sources, healthy control groups and outbreaks [6–9]. In these studies, strains show inconsistent presence and concordance of EAEC virulence genes by PCR in specimens from symptomatic and asymptomatic cases. These data suggest that the full genetic component of this phenotype is not yet fully understood and, although most of these genes are found on the aggregative virulence plasmid, their inheritability is complex.

Early research on EAEC linked these strains to persistent diarrhoea in children in developing countries but EAEC have since been shown to be a significant cause of acute diarrhoea and important in the aetiology of intestinal infections in industrialized countries [10]. Two independent, large, prospective studies of diarrhoea aetiology conducted in the UK (1993–1996) and USA (2002–2004) reported a similar EAEC prevalence in patients with diarrhoea: 4.6% (160/3506) and 4.5% (37/823) in the UK and

Table 1. Phenotypic and genotypic characteristic of the 20 strains of EAEC isolated from faecal specimens linked to the outbreak.

Isolate Number	Isolate ID	Serotype	MLST	Plasmid type	Fimbrial type	Antibiotic resistance profile (ESBL *)
1	216/13	O104:H4	678	FIB_25 FII_48	I	AMP/SUL/STR/TET/TMP/NAL
2	218/13	O104:H4	678	FIB_25 FII_48	I	AMP/SUL/STR/TET/TMP/NAL
3	1062/13	O104:H4	678	FIB_25 FII_48	I	AMP/SUL/STR/TET/TMP/NAL
4	1063/13	O104:H4	678	FIB_25 FII_48	I	AMP/SUL/STR/TET/TMP/NAL
5	1070/13	O104:H4	678	FIB_25 FII_48	I	AMP/SUL/STR/TET/TMP/NAL
6	0219/13	O131:H27	10	FIB_5 FII_17	I	AMP/SUL/STR/TET/TMP/NAL
7	0220/13	O131:H27	10	FIB_5 FII_17	I	AMP/SUL/STR/TET/TMP/NAL
8	1071/13	O131:H27	10	FIB_5 FII_17	I	AMP/SUL/STR/TET/TMP/NAL
9	1072/13	O131:H27	10	FIB_5 FII_17	I	AMP/SUL/STR/TET/TMP/NAL
10	1073/13	O131:H27	10	FIB_5 FII_17	I	AMP/SUL/STR/TET/TMP/NAL
11	1074/13	O131:H27	10	FIB_5 FII_17	I	AMP/SUL/STR/TET/TMP/NAL
12	0215/13	O20:H19	278	FIB_5 FII_17	IV	AMP/SUL/TMP/CAZ/CTX/CPR/CEF *
13	0221/13	O20:H19	278	FIB_5 FII_17	IV	AMP/SUL/TMP/CAZ/CTX/CPR/CEF *
14	0222/13	O19a:H30	38	FIB_33 FII_1	III	AMP/SUL/STR/TMP/NAL/CAZ/CTX/CEF
15	1061/13	O55:H19	10	FIB_33 FII_1	III	AMP/SUL/STR/TEM/NAL
16	1065/13	O63:H12	1664	FIB_33 FII_1	III	AMP/SUL/STR/TET/TMP/NAL/CIP/CAZ/CTX/CEF
17	0214/13	O?:H19	746	FII	III	NAL
18	0217/13	O33:H16	295	FII_9	II	AMP/CHL/STR/TET/TMP/NAL
19	1064/13	O?:H21	227	FII_9	II	AMP/SUL/STR/TET/TMP/NAL/CAZ/CTX/CPR/CEF *
20	1060/13	O111:H4	226	rep8	-	AMP/CHL/COL/TET/NAL

Key: AMP Ampicillin; CAZ ceftazidime; CTX Cefataxime; CEF Ceftiofur; CPR cefpirome; NAL nalidixic acid; STR Streptomycin; SUL Sulphonamide; TET tetracycline; TMP Trimethoprim.

doi:10.1371/journal.pone.0098103.t001

US studies, respectively, and 1.7% in control subjects from both studies [11,12]. Clinical symptoms include watery diarrhoea, often with mucus, low grade fever, abdominal pain, nausea and vomiting [10].

Several EAEC foodborne outbreaks of gastroenteritis have been documented, notably in Japan, the UK and Italy [13–15]. Recently, a strain of enteroaggregative Shiga toxin-producing *E. coli* O104:H4 was identified as the cause of a foodborne outbreak of bloody diarrhoea and haemolytic ureamic syndrome (HUS) in Germany and France [16–20]. Case-control, cohort and trace back studies implicated fenugreek sprouts from Egypt as the source of the infection [21]. Detailed and timely microbiological outbreak investigations were followed by whole genome sequencing of strains of *E. coli* O104:H4 by various international groups [17,19,22–23].

In March 2013, a large outbreak of GI disease occurred in the North East of England and cases were linked to a food festival. Four hundred and thirteen cases reported illness including symptoms of persistent diarrhoea and abdominal pain immediately following the event, and a total of 592 cases were identified following an on-line questionnaire. One hundred and ten specimens were submitted to the regional Public Health England and local hospital laboratories. Using traditional culture methods, *Salmonella enterica* serotype Agona was isolated from 25 cases and 4 further cases had other *Salmonella* species. Cohort and trace back studies implicated a contaminated, fresh curry leaves from Pakistan as the source of the infection.

The low number of cases testing positive for *Salmonella* species raised the suspicion that this was a multi-pathogen outbreak and further testing using a pan pathogen PCR was requested by the Outbreak Control Team. Subsequently, strains of EAEC harbouring *aggR* were isolated from PCR positive faecal specimens. The aim of this study was to use whole genome sequencing to explore the genomic diversity of the 20 strains of EAEC harbouring *aggR* by determining their phylogenetic relationship, plasmid type and virulence gene content and to assess the likely contribution of each strain type to the reported symptoms of GI disease.

Material and Methods

Microbiology

Retrospectively, 88 faecal specimens from cases associated with the outbreak were tested for the presence of other bacterial GI pathogens using a multiplex GI pathogens PCR [24]. Although the faecal specimens had been stored for over 10 weeks at 4°C, an attempt was made to isolate the pathogens detected by the multiplex PCR by testing individual colonies for the *stx*, *ipaH* and *aggR* target genes, associated with Shiga toxin-producing *Escherichia coli* (STEC), *Shigella* species and EAEC respectively. For faecal specimens positive for *aggR*, 20 colonies were picked from bacterial growth on MacConkey or Sorbitol MacConkey agar plates and retested using the same PCR. Those colonies harbouring the *aggR* genes were identified biochemically as *E. coli* and serotyped using antisera raised in rabbits to the *E. coli* somatic O antigens.

Library preparation and whole genome sequencing

DNA was extracted for sequencing using the Wizard kit (Promega UK). Paired-end libraries were generated using the Illumina Nextera XT sample preparation kit. Automated

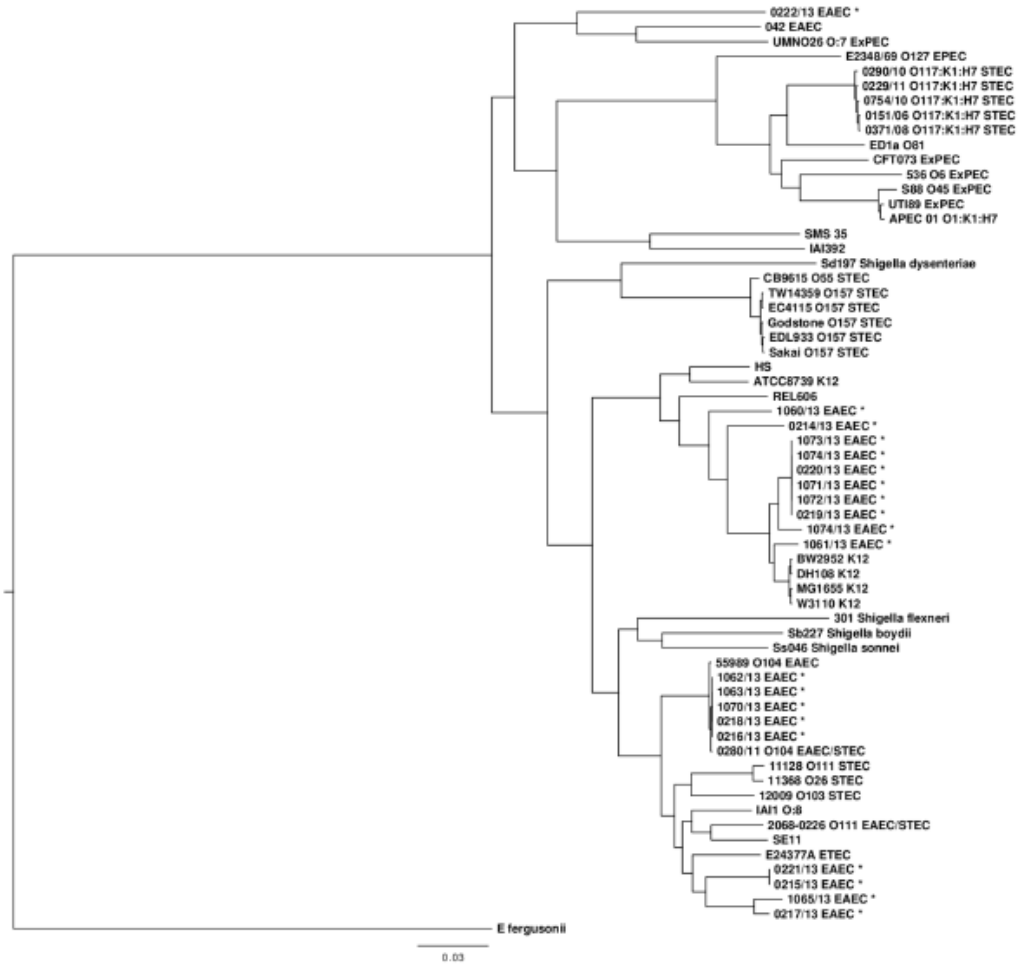


Figure 1. Whole genome chromosomal phylogeny of strains of *E. coli* and *Shigella* spp using previously published sequences and showing the 20 strains of EAEC isolates during this outbreak (highlighted in red). EAEC Enteroaggregative *E. coli*; EHEC Enterohaemorrhagic *E. coli*; ExPEC Extraintestinal Pathogenic *E. Coli*; EPEC Enteropathogenic *E. coli*; ETEC Enterotoxigenic *E. coli*.
doi:10.1371/journal.pone.0098103.g001

platforms were used for sample preparation, library generation and quality checks. Assessment of fragment sizes was performed on the Perkin Elmer Labchip GX after fragmentation and clean-up. After normalisation, samples were pooled by hand and library quantification was performed using the KAPA library quantification kit for Illumina sequencing, on an ABI Viia7. Libraries were diluted to 15 pM and denatured at 96°C on a heat block for 2 minutes before being placed on ice for 5 minutes. Denatured libraries were spiked with 5% phiX and loaded on a Rapid flowcell by the Illumina cBot instrument. Paired-end sequencing was performed on the Illumina HiSeq 2500 instrument running HCS 2.0.10.0 using the TruSeq Rapid SBS kit (200 cycle) and TruSeq Paired-end rapid cluster kit. The following cycle parameters were used for sequencing: Read 1: 101, Index read 1: 8, Index read 2: 8 and Read 2: 101. RTA version 1.17.21.3 was used for generation of base call files.

Spades version 2.5.1 [25] was used to produce *de novo* assemblies of the sequenced paired-end fastq file. The number of contigs produced ranged from 221 to 552 per sample with N50s from 48338 to 192731 nucleotides.

Phylogenetic analysis

Illumina reads were mapped to the reference EAEC strain 55989 using BWA-SW [26]. The Sequence Alignment Map output from BWA was sorted and indexed to produce a Binary Alignment Map (BAM) using Samtools [27]. GATK2 [28] was

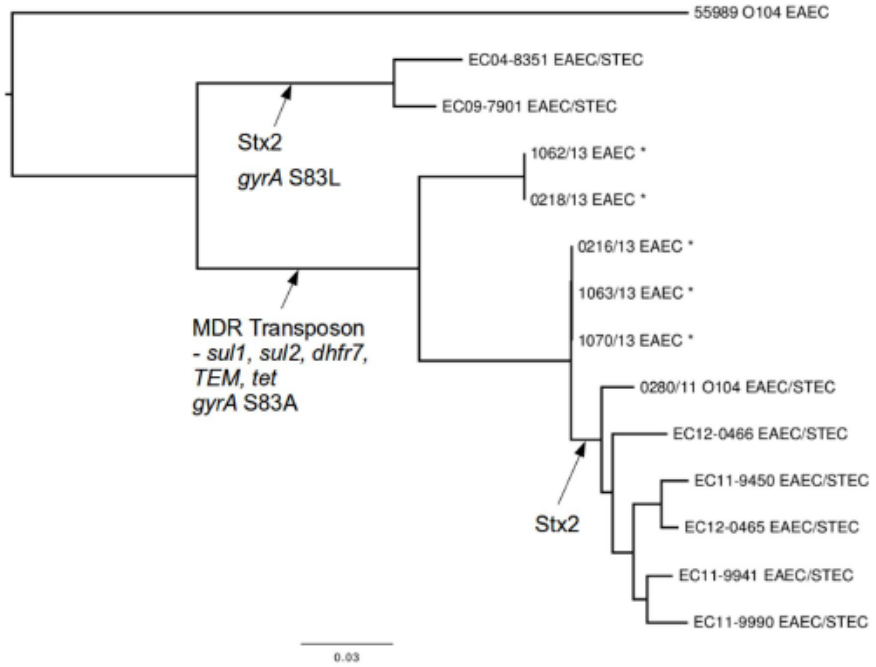


Figure 2. Chromosomal phylogeny of sequenced EAEC ST687 (strains including five strains isolated during this study marked *) represented as a maximum-likelihood tree. Previously published genome sequences included 280/11 isolated from a case linked to the outbreak in Germany in 2011 and 55989 isolated in the late 1990s in the Central African Republic. Other strains were previously described in Grad *et al.* 2013.
doi:10.1371/journal.pone.0098103.g002

used to create a Variant Call Format (VCF) file from each of the BAMs, which were further parsed to extract only single nucleotide polymorphism (SNP) positions which were of high quality in all genomes (MQ>30, DP>10, 128 GQ>30, Variant Ratio >0.9). Pseudosequences of polymorphic positions were used to create approximate maximum likelihood trees using FastTree [29] under the General time reversible (GTR) model of nucleotide evolution.

Multilocus sequence typing (MLST)

MLST types were identified by mapping the reads against all *E. coli* allele variants held in the Achtman MLST database (www.mlst.ucc.ie/mlst/dbs/Ecoli) using a modification of the SRST software [30].

Plasmid FIB/FII typing

Plasmid incompatibility groups were determined using the specific sequences for plasmid replicon types defined by Carattoli *et al.* 2005 [31]. These sequences were searched for using blastn against the assembled genomes. Retrieved IncF and IncI replicon sequences were extracted *in silico* and further characterised to sequence type level according to the new scheme described in the plasmid MLST database (pMLST: www.pubmlst.or/plasmid/)

BRIG analysis

Assembled genomes were loaded into BRIG as concentric rings [32] and compared against the pAA reference genome using blastn. pAA annotations from genbank file were added in the final ring.

Mapping of known EAEC virulence genes

Illumina reads were mapped to a panel of putative EAEC virulence factors (*aggR, aatA, aatB, aatC, aatP, aap, SepA, IDI* and *aaiC*) using BWA-SW [26]. The number of reads that mapped to each position was calculated using Samtools mpileup [27]. The aggregative fimbrial adhesion type was determined based on mapping to each of the five variable fimbrial subunits AAF/I to AAFV (*aggA, aafA, agg3a, hdaA, aaf5a*) [33–37].

Determination of the presence or absence of AAI operon T6SS components using BLAST

AAI operon T6SS coding genes were extracted from the reference strain 55989 genbank file (http://www.ncbi.nlm.nih.gov/nuccore/NC_011748.1) and made into a BLAST database. Each of the assembled genomes was queried against the database using blastn to recover whether it had significant hits for each component of the AAI.

Data Submission

The short read sequence data has been deposited in the NCBI Short Read Archive under the BioProject PRJNA245029

Table 2. Number of mapped reads to the EAEC virulence genes in each outbreak isolate.

Isolate Number	Isolate ID	Serotype	AggR	aatA	aap	sepA	IDI	aaiC
1	216/13	O104:H4	1001	759	752	387	631	1780
2	218/13	O104:H4	1752	1783	874	443	1278	2213
3	1062/13	O104:H4	1932	1593	1525	805	1360	2195
4	1063/13	O104:H4	1483	1145	1133	733	1056	1165
5	1070/13	O104:H4	1554	1345	866	434	897	3219
6	0219/13	O131:H27	1799	799	540	128	591	2335
7	0220/13	O131:H27	1792	739	563	172	584	2683
8	1071/13	O131:H27	3501	1705	1054	347	1123	2228
9	1072/13	O131:H27	5185	2300	1607	366	1472	5146
10	1073/13	O131:H27	3337	1599	973	273	999	1879
11	1074/13	O131:H27	5160	2455	1357	378	1566	1943
12	0215/13	O20:H19	1826	1545	831	271	781	1841
13	0221/13	O20:H19	757	615	366	136	264	1390
14	0222/13	O19:a:H30	685	568	493	118	393	**0**
15	1061/13	O55:H19	1813	1859	1206	355	1127	1598
16	1065/13	O63:H12	2339	2313	1448	383	1558	**2**
17	0214/13	O2:H19	2380	1963	1901	**32**	1533	**270**
18	0217/13	O33:H16	794	620	716	**10**	827	2842
19	1064/13	O2:H21	3008	3045	2159	**1**	2291	**0**
20	1060/13	O111:H4	**0**	**0**	**0**	**0**	**0**	**193**

Where the average coverage is less than 10% across the length of the gene the cell in the table is highlighted in bold.
doi:10.1371/journal.pone.0098103.t002

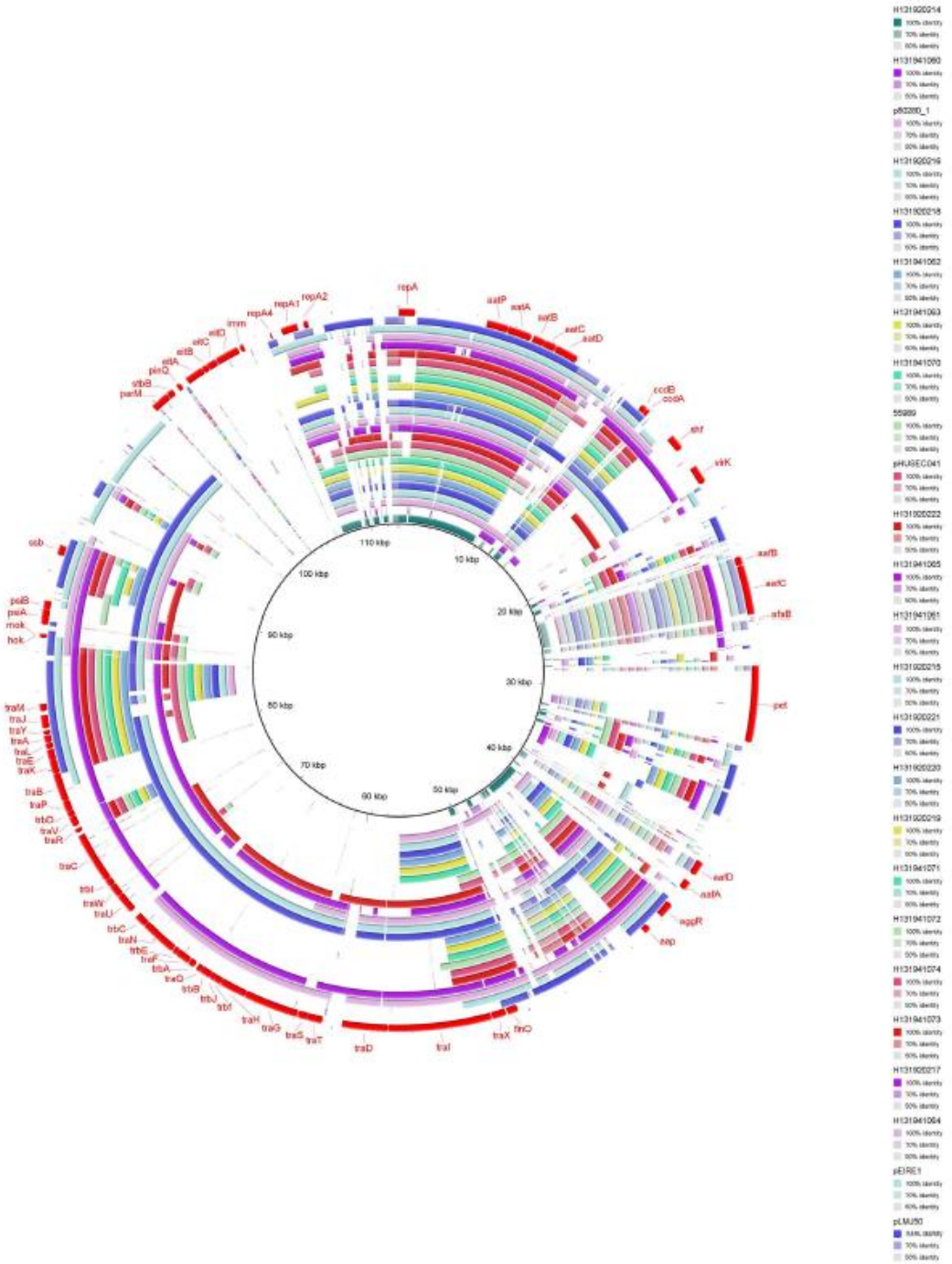

repA
repA2
repA1
repA4
impm
eilD
eilC
eilB
eilA
pinQ
stbB
parM
astP
astA
astB
astC
astD
ccdB
ccdA
sbf
virK
sab
psiB
psiA
mok
hok
traM
traJ
traY
traA
traL
traE
traK
traB
traP
trbD
traV
traR
traC
trbI
traW
traU
trbC
traN
trbE
traF
trbA
traO
trbB
trbJ
trbF
traH
traG
traS
traT
traD
traI
traX
finO
astB
astC
sfaB
pet
astD
astA
eppR
eap
110 kbp
10 kbp
100 kbp
20 kbp
90 kbp
30 kbp
80 kbp
40 kbp
70 kbp
50 kbp
60 kbp
H131900214
H131941060
p82260_1
H131900219
H131900218
H131941062
H131941063
H131941070
55989
pHUSEC041
H131900222
H131941065
H131941061
H131900215
H131900221
H131900220
H131900219
H131941071
H131941072
H131941074
H131941073
H131900217
H131941064
pERE1
pLMU50

Figure 3. Assembled genomes displayed as concentric rings using BRIG and BLASTed against pAA genbank file as a reference. Coloured bars represent regions of homology. The darker shades represent a high percentage similarity, lighter shades represent lower levels of similarity and the absence of colour signifies absence of the gene.
doi:10.1371/journal.pone.0098103.g003

Results

Detection of multiple GI pathogens by PCR

Retrospectively, 88 specimens from cases associated with the outbreak were tested for the presence of other bacterial GI pathogens using a multiplex GI pathogens PCR [24]. A variety of bacterial GI pathogens were detected by PCR from 88 of the stored faecal specimens from cases associated with the outbreak including *Salmonella* (3 cases), STEC (5 cases) and *Shigella* (29 cases). The *aggR* gene was identified in 65 (75%) specimens. Twenty strains of EAEC harbouring *aggR* were isolated from the 65 PCR positive faecal specimens. No STEC or *Shigella* species were isolated.

Phylogeny of EAEC isolated from the cases associated with the outbreak

Ten different serotypes and nine MLSTs were identified among the EAEC isolated from the outbreak cases (Table 1). The most commonly observed serotypes were O131:H27 (6), O104:H4 (5) and O20:H19 (2), and the most frequently identified STs, corresponding with these serotypes, were ST10, ST678 and ST278 respectively.

SNP analysis confirmed that the strains were phylogenetically diverse between serotypes (Figure 1). Strains belonging to the same serotype clustered on the same branch of the tree, however, even within the same serotype, isolates were phylogenetically distinct. Figure 2 shows a phylogeny based on 3115 core SNPs of 14 strains of *E. coli* O104:H4 and illustrates the relationship between the EAEC O104:H4 strains in this study with sporadic strains of *E. coli* O104:H4 and the strain associated with the *E. coli* O104:H4 outbreak in Germany in 2011 [22,23]. Although none of the EAEC O104:H4 isolates in this study had the *stx* gene they share a common ancestor with the German outbreak strain 280/11 and the sporadic *stx* harboring enteroaggregative strains characterised by Grad *et al* (23). All five strains of *E. coli* O104:H4 isolated during this study share the MDR genomic island conferring resistance to ampicillin, the sulphonamides, streptomycin and tetracycline, and the S83A gyrA mutation in common with German outbreak strain and the closely related EAEC/STEC sporadic isolates from France. The EAEC O104:H4 strains isolated in this study are phylogenetically integrated with strains of EAEC/STEC suggesting either multiple gain or gain then loss of the *stx* phage within the O104:H4 serotype.

None of the EAEC O104:H4 isolates in this study had the *stx* gene or carried the extended spectrum beta lactamase (ESBL) plasmid characteristic of the 280/11 strain, although three other strains isolated during this study were identified phenotypically and genotypically as being ESBL-producers (Table 1).

Replicon types of the EAEC plasmids encoding the key enteroaggregative virulence genes (pEAEC)

Multiple replicon types were observed with multiple combinations of FII and FIB proteins, with all but three plasmids having both the FIB and FII replicon types (Table1). Plasmids of type FIB5_FII17 were carried by strains belonging to two serotypes, O131:H27 and O20:H19. The plasmid type FIB25_FII48 harboured by the strains of EAEC O104:H4 was the same FIB/FII type described in the strains of *E. coli* O104:H4 linked to the 2011 German outbreak (Table 1).

pEAEC encoded virulence genes and genomic architecture

Several plasmid encoded genes associated with EAEC have been described in previous studies. These include the transcriptional activator *aggR*, the anti-aggregation transporter locus *aat*, the anti-aggregative dispersin protein *aap* (2–4), the aggregative adherence fimbriae (AAF) (30–34), the serine protease autotransporter toxin *SepA* [38] and the recently described putative isopentenyl isomerase (IDI) enzymes [39]. Table 2 shows the number of reads that mapped to these targets in each outbreak isolate. All of the strains, apart from *E. coli* O111:H4 designated 1060/13, had sequence reads that mapped to *aggR*, *aat*, *aap* and the putative IDI enzymes. This isolate originally tested positive with the *aggR* PCR subsequently tested negative following storage on Dorset Egg medium at room temperature. It is likely that this isolate lost the EAEC plasmid during storage. The serine protease *sepA* was present in 16 of the 20 EAEC strains isolated from the cases associated with the outbreak. Whilst the pEAEC virulence gene complement was conserved, the genomic context in terms of flanking IS elements was highly variable across the different plasmids (Figure 3).

Five types of pEAEC associated AAF have been described [33–37] and all five fimbriae types were identified in the strains analysed during this study. Strains of EAEC belonging to serotypes O104:H4 and O131:H27 had AAF/I fimbriae, as seen in the aggregative plasmid of *E. coli* O104:H4 linked to the 2011 German outbreak. Those strains belonging to serotype O20:H19 had the Type IV fimbriae (HdaA) [36]. AAF/II, AAF/III and AAF/V fimbriae were detected in five strains belonging to five different serotypes but three harbouring the same plasmid type, FIB33_FII1 (Table 2).

AAI operon encoding the putative T6SS

A 117 kb pathogenicity island, first described in the chromosome of EAEC 042, has been implicated as an EAEC pathogenicity factor. Twenty-five contiguous genes (*aaiA–Y*) in this island were previously shown to be transcriptionally activated by the plasmid encoded AggR protein and encoded for a T6SS [5]. In the EAEC strains isolated from the outbreak cases described in this study, the AAI operon was present in its entirety in the strains belonging to serotypes O104:H4, O131:H27, O20:H19 and O55:H19, whilst the island was absent in the strains belonging to the serotypes O19a:H30, O?:H21 and O63:H12 (Figure 3). Table 3 shows the distribution of the putative T6SS genes, *aaiA* to *aaiN*, in the outbreak strains. In the EAEC strains designated 1060/13 and 0214/13 (serotypes O111:H4 and O?:H19 respectively), a contig with 84% identity to the AAI operon and no homology to the NCBI non-redundant database was identified. In the EAEC O?:H19 isolate this homologue to *aai* was co-located on a contig with a plasmid addiction system suggestive of a non-chromosomal location in these strains.

Discussion

Historically, outbreaks have been associated with strains of a single pathogen exhibiting similar, if not identical, phenotypic and genotypic characteristics. However, the multiplex PCR approach to detection of GI pathogens directly from faecal specimens has provided good evidence that many individual cases of diarrhoea

Table 3. The distribution of the putative T6SS genes *aaiA* to *aaiN* in the outbreak strains.

	Serotypes								
Gene	O104:H4	O33:H27	O20:H19	O55:H19	O111:H4	O7:H19	O19a:H30	O63:H12	O7:H21
aaiA	X	X	X	X	X	X			
aaiB	X	X	X	X	X	X			
aaiC	X	X	X	X					
aaiD	X	X	X	X	X	X			
aaiE	X	X	X	X	X	X			
aaiF	X	X	X	X	X	X			
aaiG	X	X	X	X	X	X			
aaiH	X	X	X	X					
aaiI	X	X	X	X					
aaiJ	X	X	X	X	X	X			
aaiK	X	X	X	X	X	X			
aaiL	X	X	X	X	X	X			
aaiM	X	X	X	X	X	X			
aaiN	X	X	X	X	X	X			

X indicates the genes was present on the genome.
doi:10.1371/journal.pone.0098103.t003

and outbreaks of GI disease are associated with multiple pathogens [40–41]. Although there was clear microbiology evidence that established GI pathogens, such as *Salmonella* and *Shigella* species, played a significant part, the symptoms described by the cases and the presence of *aggR* in 75% of the specimens retrospectively tested by PCR, suggested that certain serotypes of the EAEC isolated, contributed to the GI disease associated with this outbreak. However, the variety of EAEC serotypes identified in the 20 strains isolated presented a complex picture.

Initially, it was suggested that the variation in serotype in the outbreak strains was masking a closer phylogenetic relationship. However, the phylogenetic tree created by comparing SNPs in the core genome showed that, although strains of the same serotype were relatively closely related, those of different serotypes were diverse. EAEC belong to several lineages with different evolutionary histories demonstrating independent acquisition of the plasmids encoding EAEC virulence genes [42]. Conversely, strains with a recent common ancestor, e.g. those that share an MLST sequence type, may have different pathotypes [43]. For example, *E. coli* O104:H4 ST678 has been shown to be STEC and EAEC [19]. The pathotype distribution is explained with multiple loss/gain events of pathogenicity elements.

Although strains of EAEC have been shown to harbour a wide diversity of plasmids that encode the enteroaggregative phenotype even in conserved chromosomal backgrounds [19], it was considered possible that similar plasmids would be found in the different strains of EAEC linked to this outbreak, given their spatial and temporal association. However, analysis of the plasmid genomes showed that they demonstrated a high level of variation in replicon type, gene content and genomic architecture. Some plasmid similarity was seen within strains of the same MLST and serotype but wide diversity was observed between different MLST and serotypes. The interspersing of different plasmids in the phylogeny suggests that the aggregative phenotype (specifically the presence of *aggR*, *aat* and *aap*) has been acquired by several different replicons of F-plasmids on multiple occasions. This level of strain and plasmid diversity has not previously been identified in isolates of EAEC from the same outbreak, although EAEC outbreaks involving more than one serotype and variation in pEAEC have been described previously [14,44]

Generally, *aggR*, *aat* and *aap* were conserved between strains of EAEC linked to this outbreak but a variety of fimbrial genes were identified. The presence of AAF is required for mediating the aggregative adherence seen in EAEC. To date five non-homologous AAF fimbiral structural proteins have been described and a representative of each was identified in strains belonging to this outbreak.

aai4-P comprise a T6SS apparatus for *aaiC* and was the first example of a conserved chromosomal aggregative genotype whose expression is under the control of a conserved plasmid encoded pathogenicity factor *AggR* [5]. In this study, five isolates all harbouring *aggR* and *aat*, had a missing or an incomplete AAI operon. This raises a question regarding the pathogenic potential of the *aggR*-positive but AAI operon deficient strains, in relation to those *aggR*-positive strains with complete AAI cassettes. Previous prevalence studies detecting the *aaiA* show its presence in between 26 and 44% of phenotypic aggregative *E. coli* [45,46]. Animal models for investigating EAEC virulence have been described previously [47] and further virulence studies are required. Interestingly, two strains (EAEC O111:H4 and O?:H19) had incomplete AAI operons encoding a T6SS. These regions were homologous to those found in serotypes O104:H4, O131:H27, O20:H19 and O55:H19 but with a different *aaiC* component. In addition, there was some evidence that it may be plasmid-encoded.

It was suggested that certain strains of EAEC may have been carried asymptomatically by the cases before the outbreak occurred. Hwoever, it was not possible to compare the serotypes isolates following the outbreak with the serotypes of strains of EAEC currently circulating in England as there are very little data on domestically acquired strains of EAEC. Surveillance data indicates that the majority of strains of EAEC isolate in England are from cases of travellers' diarrhoea [12,50].

One hypothesis is that not all the strains of EAEC associated with this outbreak had the same level of pathogenicity and that only certain EAEC serotypes isolated contributed to the symptoms described. For example, a complete AAI operon may increase the pathogenic potential of strains of *E. coli* harbouring the pEAEC. Other studies have suggested that strains harbouring different fimbrial types maybe more pathogenic that others. Nüesch-Inderbinen *et al.* (2013) [48] showed a statistically significant association of the agg3C gene with the asymptomatic state. The presence of AAF/I and AAF/II have been associated with symptomatic cases [6,49].

Importantly, although colonies of EAEC O104:H4 were isolated from only five outbreak cases, all faecal specimens were retrospectively tested by PCR for the presence of the O104 O-antigen gene (*wzx*O104) [51]. The PCR detected *wzx*O104 in faecal specimens from 36 cases. The EAEC serotype O104:H4, with and without *stx2*, has been previously identified as a cause of GI disease. Furthermore, although the EAEC O104:H4 in this study did not carry the *stx* genes, this outbreak provides further evidence of the pathogenic potential of this EAEC serotype. EAEC is known to be a diverse pathotype and this study demonstrates this diversity can be seen within a single outbreak.

Acknowledgments

We would like to acknowledge Jennifer Collins and Michelle Payne at Newcastle PHE Regional Laboratory for their microbiological expertise. We would also like to thank Martin Day for producing the antibiotic resistance data. We are grateful to Dr Rohini Manuel and colleagues at Public Health Laboratory London for supporting the use of the GI multiplex PCR for outbreak investigations.

Author Contributions

Conceived and designed the experiments: TD LC MD RT AU KF KG CJ. Performed the experiments: TD MC LC DW DR KF CJ. Analyzed the data: TD LC CJ. Contributed reagents/materials/analysis tools: TD LC MD JW DR KF. Wrote the paper: TD KF KG CJ.

References

1. Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, et al. (1987) Patterns of adherence of diarrheagenic *Escherichia coli* to HEp-2 cells. Pediatr Infect Dis J 6: 829–31.
2. Nataro JP, Yikang D, Yingkang D, Walker K (1994) AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli*. J Bacteriol 176: 4691–9
3. Sheikh J, Czeczulin JR, Harrington S, Hicks S, Henderson IR, et al. (2002) A novel dispersin protein in enteroaggregative *Escherichia coli*. J Clin Invest 110: 1329–37.
4. Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V, et al. (2003) The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. J Biol Chem 278: 45680–9.
5. Dudley EG, Thomson NR, Parkhill J, Morin NP, Nataro JP (2006) Proteomic and microarray characterization of the AggR regulon identifies a pheU

pathogenicity island in enteroaggregative *Escherichia coli*. Mol Microbiol 61: 1267–82.
6. Jenkins C, Chart H, Willshaw GA, Cheasty T, Tompkins DS (2007) Association of putative pathogenicity genes with adherence characteristics and fimbrial genotypes in typical enteroaggregative Escherichia coli from patients with and without diarrhoea in the United Kingdom. Eur J Clin Microbiol Infect Dis 26: 901–6
7. Jenkins C, Tembo M, Chart H, Cheasty T, Willshaw GA, et al. (2006a) Detection of enteroaggregative *Escherichia asli* in faecal samples from patients in the community with diarrhoea. J Med Microbiol 55: 1493–7
8. Boisen N, Scheutz F, Rasko DA, Redman JC, Persson S, et al. (2012) Genomic characterization of enteroaggregative *Escherichia coli* from children in Mali. J Infect Dis 205: 431–44.
9. Okeke IN, Wallace-Gadsden F, Simons HR, Matthews N, Labar AS, et al. (2010) Multi-locus sequence typing of enteroaggregative *Escherichia asli* isolates from Nigerian children uncovers multiple lineages. PLoS One 5:e14093.
10. Okeke IN, Nataro JP (2001) Enteroaggregative *Escherichia coli*. Lancet Infect Dis 1(5):304–13. Review
11. Nataro JP, Mai V, Johnson J, Blackwelder WC, Heimer R, et al. (2006) Diarrheagenic *Escherichia asli* infection in Baltimore, Maryland, and New Haven, Connecticut. Clin Infect Dis 43: 402–7.
12. Wilson A, Evans J, Chart H, Cheasty T, Wheeler JG, et al. (2001) Characterisation of strains of enteroaggregative *Escherichia coli* isolated during the infectious intestinal disease study in England. Eur J Epidemiol 17: 1125–30.
13. Itoh Y, Nagano I, Kunishima M, Ezaki T (1997) Laboratory investigation of enteroaggregative *Escherichia coli* O untypeable:H10 associated with a massive outbreak of gastrointestinal illness. J Clin Microbiol 35: 2546–50
14. Smith HR, Cheasty T, Rowe B (1997) Enteroaggregative *Escherichia coli* and outbreaks of gastroenteritis in UK. Lancet 350: 814–5.
15. Scavia G, Staffolani M, Fisichella S, Striano G, Colletta S, et al. (2008) Enteroaggregative *Escherichia coli* associated with a foodborne outbreak of gastroenteritis. J Med Microbiol 7: 1141–6.
16. Frank C, Werber D, Cramer JP, Askar M, Faber M, et al. (2011) Epidemic profile of Shiga-toxin-producing *Escherichia asli* O104:H4 outbreak in Germany. N Engl J Med 365: 1771–80.
17. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, et al. (2011) Prospective genomic characterization of the German enterohemorrhagic *Escherichia asli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One. 6:e22751.
18. Bielaszewska M, Mellmann A, Zhang W, Köck R, Fruth A, et al. (2011) Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. Lancet Infect Di 11: 671–6.
19. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, et al. (2011) Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med. 365: 709–17.
20. King LA, Nogareda F, Weill FX, Mariani-Kurkdjian P, Loukiadis E, et al. (2012) Outbreak of Shiga toxin-producing Escherichia coli O104:H4 associated with organic fenugreek sprouts, France, June 2011. Clin Infect Dis. 54: 1588–94.
21. Buchholz U, Bernard H, Werber D, Böhmer MM, Remschmidt C, et al. (2011) German outbreak of *Escherichia asli* O104:H4 associated with sprouts. N Engl J Med. 365: 1763–70
22. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, et al. (2012) Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci U S A. 109: 3065–70.
23. Grad YH, Godfrey P, Cerquiera GC, Mariani-Kurkdjian P, Gouali M, et al. (2013) Comparative genomics of recent Shiga toxin-producing Escherichia coli O104:H4: short-term evolution of an emerging pathogen. MBio. 4:e00452–12.
24. Moran-Gilad J, Chand M, Brown C, Shetty N, Morris G, et al. (2012) Microbiological aspects of public health planning and preparedness for the 2012 Olympic Games. Epidemiol Infect 140: 2142–51.
25. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19: 455–77
26. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589–95.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–9.
28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297–303.
29. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One 5: e9490.
30. Inouye M, Conwa TC, Zobel J, Holt KE (2012) Short read sequence typing (SRST): multi-locus sequence types from short reads. BMC Genomics 13: 338
31. Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, et al. (2005) Identification of plasmids by PCR-based replicon typing. J Microbiol Methods 63: 219–28.
32. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA (2011). BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC genomics* 12, 402.
33. Nataro JP, Deng Y, Maneval DR, German AL, Martin WC, et al. (1992) Infect Aggregative adherence fimbriae I of enteroaggregative *Escherichia coli* mediate adherence to HEp-2 cells and hemagglutination of human erythrocytes. Immun 60: 2297–304.
34. Czeczulin JR, Balepur S, Hicks S, Phillips A, Hall R, et al. (1997) Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia asli*. Infect Immun 65: 4135–45
35. Bernier C, Gounon P, Le Bouguénec C (2002) Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative *Escherichia asli* as a sensitive probe for detecting the AAF-encoding operon family. Infect Immun 70: 4302–11.
36. Boisen N, Struve C, Scheutz F, Krogfelt KA, Nataro JP (2008) New adhesin of enteroaggregative *Escherichia asli* related to the Afa/Dr/AAF family. Infect Immun 76: 3281–92.
37. Dallman T, Smith GP, O'Brien B, Chattaway MA, Finlay D, et al. (2012) Characterization of a verocytotoxin-producing enteroaggregative *Escherichia coli* serogroup O111:H21 strain associated with a household outbreak in Northern Ireland. J Clin Microbiol 50: 4116–9.
38. Boisen N, Ruiz-Perez F, Scheutz F, Krogfelt KA, Nataro JP (2009) Short report: high prevalence of serine protease autotransporter cytotoxins among strains of enteroaggregative *Escherichia asli*. Am J Trop Med Hyg 80: 294–301
39. Morin N, Santiago AE, Ernst RK, Guillot SJ, Nataro JP (2013) Characterization of the AggR regulon in enteroaggregative *Escherichia coli*. Infect Immun 81: 122–32.
40. Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, et al. (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Gut. 61: 69–77.
41. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, et al. (2013) Lancet 382: 209–22.
42. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, et al. (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc Natl Acad Sci U S A. 106: 17939–44.
43. Wirth T, Falush D, Lan R, Colles F, Mensa P, et al. (2006) Sex and virulence in *Escherichia asli*: an evolutionary perspective. Mol Microbiol 60: 1136–51.
44. Spencer J, Smith HR, Chart H (1999) Characterization of enteroaggregative *Escherichia coli* isolated from outbreaks of diarrhoeal disease in England. Epidemiol Infect 123: 413–21.
45. Jenkins C, Chart H, Willshaw GA, Cheasty T, Smith HR (2006b) Genotyping of enteroaggregative *Escherichia coli* and identification of target genes for the detection of both typical and atypical strains. Diagn Microbiol Infect Dis 55: 13–9.
46. Jenkins C, van Ijperen C, Dudley EG, Chart H, Willshaw GA, et al. (2005) Use of a microarray to assess the distribution of plasmid and chromosomal virulence genes in strains of enteroaggregative *Escherichia asli*. FEMS Microbiol Lett 253: 119–24.
47. Hwang J, Mattei LM, VanArendonk LG, Meneely PM, Okeke IN (2010) A pathoadaptive deletion in an enteroaggregative *Escherichia asli* outbreak strain enhances virulence in a *Caenorhabditis elegans* model. Infect Immun 78: 4068–76.
48. Nüesch-Inderbinen MT, Hofer E, Hächler H, Beutin L, Stephan R (2013) Characteristics of enteroaggregative *Escherichia asli* isolated from healthy carriers and from patients with diarrhoea. Med Microbiol doi: 10.1099/jmm.0.065177-0. [Epub ahead of print]
49. Okeke IN, Lamikanra A, Czeczulin J, Dubovsky F, Kaper JB (2000) Heterogeneous virulence of enteroaggregative *Escherichia coli* strains isolated from children in Southwest Nigeria. J Infect Dis 181: 252–60.
50. Perry N, Jenkins C, Cheasty T, Wain J (2010) Diarrhoeagenic Escherichia coli from routine diagnostic faecal samples in England and Wales. J Med Microbiol. 59: 870–2.
51. Scheutz F, Nielsen EM, Frimodt-Møller J, Boisen N, Morabito S, et al. (2011) Characteristics of the enteroaggregative Shiga toxin/verotoxin-producing *Escherichia coli* O104:H4 strain causing the outbreak of haemolytic uraemic syndrome in Germany, May to June 2011. Euro Surveill 16(24).

7.13 Establishing an Enteric Microbiology Reference Laboratory in Sierra Leone. *In Press*

Chattaway MA, Kamara A, Rhodes F, Kaffeta K, Jambaik A, Wondimagegnehu A,Mohanned SI, Freeman M, Welfare W, Harding D, Ahmed SF, Abu M, Kamanda S, Grant K, Jenkins C, Nair S, Connell S, Siorvanes L, Desai S, Allen C, Frost M, Hughes D, Jeffrey Z, Gill N & Salter M.2014. J.Infect.Dev.Ctries. 8(7) 933-941

Perspective

Establishing an enteric bacteria reference laboratory in Sierra Leone

Marie Anne Chattaway[1], Abdul Kamara[2], Fay Rhodes[2], Konneh Kaffeta[2], Amara Jambai[3], Wondimagegnehu Alemu[4], Mohammed Sirajul Islam[5], Molly M Freeman[6], William Welfare[1], Doris Harding[2], Ahmed F Samba[2], Musu Abu[2], Sylvester Kamanda[2], Kathie Grant[1], Claire Jenkins[1], Satheesh Nair[1], Steve Connell[1], Lisa Siorvanes[1], Sarika Desai[1], Collette Allen[1], Margaret Frost[1], Daniel Hughes[1], Zonya Jeffrey[1] Noel Gill[1], Mark Salter[1]

[1] *Public Health England, United Kingdom*
[2] *Central Public Health Reference Laboratory, Sierra Leone*
[3] *Directorate of Prevention and Control, Sierra Leone*
[4] *World Health Organisation, Country Office, Uganda*
[5] *Global Outbreak Response Alert Network consisting of International Centre of Diarrhoeal Disease Research, Bangladesh*
[6] *Centers for Disease Control and Prevention, United States*

Abstract

In 2012, Sierra Leone experienced its worst cholera outbreak in over 15 years affecting 12 of the country's 13 districts. With limited diagnostic capability, particularly in bacterial culture, the cholera outbreak was initially confirmed by microbiological testing of clinical specimens outside of Sierra Leone. During 2012 – 2013, in direct response to the lack of diagnostic microbiology facilities, and to assist in investigating and monitoring the cholera outbreak, diagnostic and reference services were established in Sierra Leone at the Central Public Health Reference Laboratory focusing specifically on isolating and identifying *Vibrio cholerae* and other enteric bacterial pathogens. Sierra Leone is now capable of confirming cholera cases by reference laboratory testing.

Key words: Cholera; Sierra Leone; enteric disease; *Salmonella*; *Shigella*.

J Infect Dev Ctries 2014; 8(7):933-941. doi:10.3855/jidc.5074

(Received 04 April 2014 – Accepted 09 June 2014)

Background to cholera and diarrhoeal disease in developing countries

Diarrhoeal disease is a major global public health problem and affects populations in the developing world especially, causing illness and death among young children [1,2]. Despite improving trends in mortality rates, 1 in 10 deaths during the first five years of life are from diarrhoeal diseases and a total of approximately 800,000 deaths occur each year worldwide, mostly in sub Saharan Africa and South Asia [1,2].

Cholera is a non-invasive diarrhoeal disease caused by the rod shaped Gram negative bacterium *Vibrio cholerae*. Epidemics are caused by serogroups O1 and O139, of which the O1 serogroup can be further subdivided into serotypes Ogawa and Inaba, and biotypes classical and El-Tor [3]. Transmission of cholera is via the faecal-oral route. Due to its short incubation period (2 hours to 5 days) many epidemics happen in an explosive manner. Most cholera episodes are mild to moderate and clinically indistinguishable from other causes of acute diarrhoea [4].

There are an estimated 3-5 million cases of cholera that occur globally every year resulting in 100,000 – 120,000 deaths mainly in Africa and South Asia. The mortality rate is 6.3 per 100,000 people at risk in endemic countries. Reported cases probably only represent 5-10% of the true number of annual worldwide cases [4,5].

The 7th cholera pandemic arrived in Africa in the early 1970s and since then has become endemic and a public health issue in many African countries. Approximately 1.3 million cases were reported to WHO between 2005-2012 with over 30,000 deaths [6-14].

Many developing countries lack the capability to perform basic microbiological testing for gastrointestinal pathogens which has major implications for disease diagnosis, outbreak detection, disease monitoring and the ability to assess the effectiveness of interventions and accurately measure burden of disease. All of these are important in tackling the causes of diarrhoeal disease.

Background to cholera epidemiology in Sierra Leone

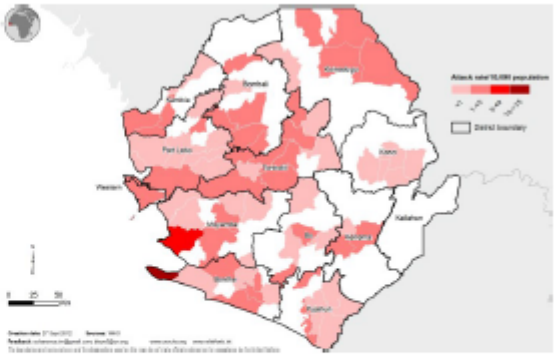
Sierra Leone, a West African Country of just over 6 million people (http://hdrstats.undp.org), experienced its worst cholera outbreak in 15 years in 2012. The Human Development Index for Sierra Leone in 2012 was 0.359, 177th out of 187 countries and territories. In Sierra Leone, life expectancy at birth is only 48 years, heavily influenced by one of the highest under 5 mortality rates (182 per 1000 live births WHO Global Health Observatory 2012 [15].

Sierra Leone has a surveillance system in place for cholera but data were previously based on clinicians reporting diarrhoeal disease and suspected cholera cases based on typical clinical presentation. Sierra Leone adopted the Integrated Disease Surveillance and Response (IDSR) strategy in 2004. The surveillance case definitions for cholera used in Sierra Leone are outlined in Table 1. An increase in the weekly number of cases of acute diarrhoea and vomiting (996 cases in 2011 as compared to 266 cases in 2010) was noted by the Ministry of Health and Sanitation (MoHS) in November 2011 in the Western Area (the most populous district that includes Freetown, the capital city). Initial investigations suggested *E. coli* was the causative agent.

In February 2012, the situation worsened (2134 cases for the three districts: Port Loko, Kambia and Pujehun), resulting in a joint investigation by MoHS and WHO. *Vibrio cholerae* O1 Ogawa was confirmed by the WHO regional laboratory in Burkina Faso and an outbreak of cholera was declared in the coastal district of Kambia.

With the onset of the rainy season, both the number of diarrhoeal cases and the districts affected began to increase (8200 cases between weeks 32, 33, and 34). Samples collected in the Western Area were again confirmed by the WHO regional laboratory in Burkina Faso as *V. cholerae* O1 Ogawa. On the 17th August 2012, the President of Sierra Leone declared the cholera outbreak as a public health emergency.

Figure 1. Spread of cholera in Sierra Leone in 2012 at the peak of the outbreak.



By the end of 2012, there were 22, 969 cases and 299 deaths (Case Fatality ratio (CFR) = 1.30), affecting 12 out of the 13 districts in Sierra Leone. The Western Area, the most populated district, reported more than 50% of the cases. Figure 1 shows the distribution of cases at the peak of the outbreak in week 38.

Cholera outbreak in Sierra Leone – The need for Laboratory testing and systems

Due to a lack of trained, enteric microbiology laboratory staff in Sierra Leone, initially, it was necessary to send clinical specimens from suspected cholera cases to outside of the country. However, this arrangement could not be sustained long term and highlighted the urgent need for Sierra Leone to develop an independent cholera testing facility. Once the outbreak was confirmed as cholera, the initial focus was on identifying and characterising isolates of *V. cholerae,* eventually other common bacterial enteric pathogens prevalent in developing countries, such as *Salmonella* and *Shigella* species, were also included.

Table 1. Surveillance case definitions for Cholera in Sierra Leone

Category	Case Definition
Suspected case (when there is not an outbreak)	Any person aged 5 years of age or more who develops severe dehydration or dies from acute watery diarrhoea
Suspected case (when there is an outbreak)	Any person aged 5 years of age or more with acute watery diarrhoea, with or without vomiting
Confirmed case	A suspected case in which *Vibrio cholerae* O1 or O139 has been isolated in the stool

Methodology

In order to develop a reliable and accurate diagnostic and reference laboratory service for enteric pathogens, a two phase process of training and testing was developed. This incorporated technical testing in line with standard laboratory safety guidelines and a quality based system according to ISO 15189 – Strengthening Laboratory Management Towards Accreditation (SLMTA). The latter being a tool kit developed to promote immediate and measurable improvement in laboratories in developing countries [16] and recommended by the African Society for Laboratory Medicine (http://www.aslm.org).

The initial training (Phase 1) was based on emergency procedures where staff were trained on basic techniques in a short period of time. Training was concentrated on developing skills required for the detection and identification of *V. cholerae* in order to identify epidemic strains of cholera. Laboratory tests carried out in Central Public Health Reference Laboratory (CPHRL), Lakka, between November 2012 and February 2013 were based on Phase 1 training methods.

Phase 2 training involved a six month microbiological training programme of diagnostic and reference testing of a wider range of enteric bacterial pathogens including *V. cholerae*, *E. coli* (including O157) *Salmonella spp* (including *S.* Typhi) and speciation of *Shigella boydii*, *S. dysentariae*, *S. flexneri* and *S. sonnei*. A set of known enteric bacterial strains was used as positive controls to quality control all media and reagents and also used during training sessions to compare tests results against known reactions.

Testing of samples at CPHRL from March 2013 onwards was based on Phase 2 training methods.

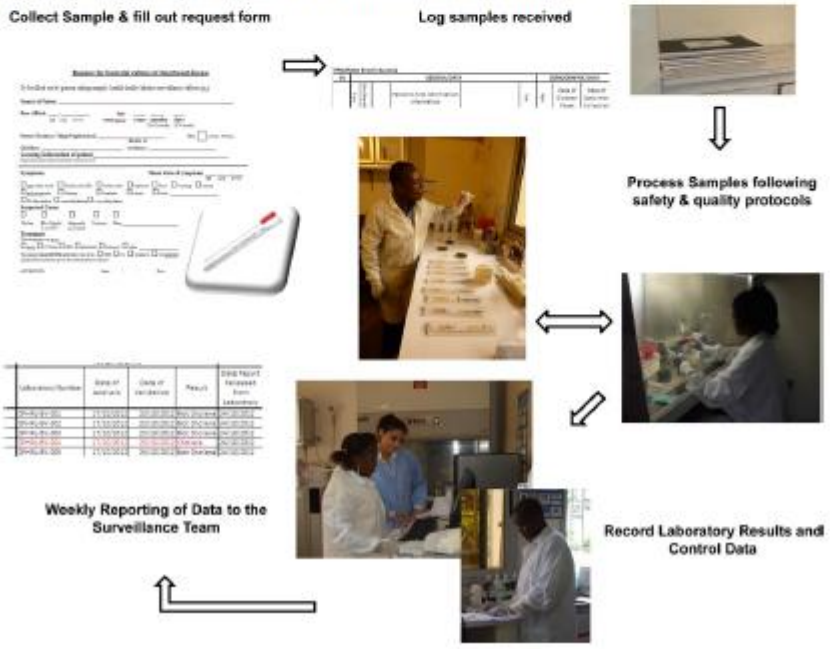
Microbiological Set Up

Processing and Reporting Systems

Clinicians were asked to notify suspected cholera cases to District Surveillance Officers and to take a rectal swab using Cary Blair swabs (VWR Jencons, Lutterworth, UK). A specimen request form was devised to accompany these specimens.

Cary Blair was chosen as the most appropriate transport media as it maintains the viability of *V. cholerae* and other enteric bacteria without refrigeration for a number of days. Rectal swabs were transported to the CPHRL, assigned a unique laboratory reference number prior to microbiological analysis. Results were reported weekly to the Directorate of Prevention and Control (DPC)

Figure 2. Processing system flow chart (rectal swab collection to reporting)



surveillance team. All processes were recorded on controlled documents and batches of media and reagents were tested with internal quality controls (IQCs) to ensure accuracy of results. All records were recorded in duplicate on both paper based and electronic systems. A flow chart of the laboratory processing from rectal swab collection to reporting is summarised in Figure 2

Microbiological Testing

Swabs were plated out for single colonies on relevant selective agar media (details below) and also used to inoculate selective enrichment broths designed to inhibit the growth of other bacteria whilst enriching the pathogen of interest. Inoculated broths were incubated aerobically overnight at 37°C and then plated onto the relevant selective media. Following incubation these plates were examined, if direct plating failed to yield a suspect pathogen or yielded insufficient colonies for testing, colonies from the enrichment broths were used. Colonies of interest from the selective media were sub-cultured onto Tryptone Soya Agar (TSA) (Oxoid, Basingstoke, UK) for further testing. Suspected *V. cholerae* samples were isolated and identified using standard techniques including alkaline peptone water (APW), Thiosulfate Citrate Bile Salts Sucrose (TCBS) agar, oxidase test and O1/ O139/Ogawa/Inaba serotyping [17] (Supplementary Figure 1). Suspected *Salmonella* species were isolated and identified using standard techniques including selenite enrichment broth, Xylose Lysine Deoxycholate (XLD) agar, oxidase test and *Salmonella* serotyping (Supplementary Figure 2). Suspected *Shigella* species were isolated and identified using standard techniques, including MacConkey (MAC) agar, oxidase test and *Shigella* serotyping (Supplementary Figure 2). *E. coli* were also isolated from MAC and slide agglutinations performed with O157 antisera [18,19].

Confirmed isolates were archived either on *Vibrio* stabs or dorset egg agar slopes (PHE Media, London, UK) at room temperature and also on microbeads and at -20°C (Prolab, Wirral, UK). For Phase 2 testing, colonies positive by slide agglutinations were inoculated into API 20E strips, incubated overnight at 37°C and the profile interpreted according to the manufacturer's instructions (bioMérieux, Craponne, France).

During Phase 2 testing, positive and negative controls were used to test every batch of media agar plates and all reagents used for biochemical and serological testing of isolates. This is particularly important in developing countries where electricity can be sporadic, impacting the appropriate storage of laboratory reagents and media ingredients and ultimately compromising their quality.

Quality Assurance

All systems adopted by CPHRL were in accordance with Stepwise Laboratory Improvement Process Towards Accreditation (SLIPTA) using SLIPTA guidance [20]. Training records were devised for all members of staff to record their training activities. All documents; risk assessments and standard operating procedures were assigned CPHRL quality controlled version reference numbers.

All media received and used had batch numbers and expiry dates recorded, freshly prepared media were assigned batch numbers and quality tested with positive controls. All reagents used were also tested with controls and results recorded for every batch of tests performed. All equipment was monitored and temperatures recorded. Any errors were recorded and investigated, tests were repeated if necessary. Data were recorded on quality controlled laboratory books and electronic databases.

Safety

Risk assessments and a safety manual were written for the procedures and implemented before any laboratory work was carried out. Hand washing before leaving the enteric laboratory was strictly implemented and personal protective equipment (PPE) including laboratory coats and gloves were used for all laboratory procedures. A microbiological safety cabinet was used for handling primary clinical specimens; potential aerosol producing manipulation and any suspected containment level 3 organisms (*E. coli* O157). Good laboratory practice when handling clinical specimens and microbiological cultures was established as the standard in the laboratory and was continuously assessed for compliance. Safety material including PPE, liquid soap, paper towels, safety signs, first aid kits, safety labels, biohazard incineration bags and disposal bins was provided before staff started work in the enteric laboratory.

Training

First phase (4 weeks): An intense multi-method approach was used for training staff in laboratory procedures including, one to one practical sessions, group discussions, lectures, homework and culminated with three day practical and theory exams. A basic enteric microbiological technique certificate was

awarded to laboratory staff who successfully completed the examinations.

Second phase (6 months): An advanced six month development programme / curriculum was developed building on the multi-method approach and final three day advanced theory and practical exam. Four biomedical scientists from Public Health England (PHE) reference and hospital laboratories were seconded to CPHRL for four weeks each, over a six month period to carry out laboratory training. In between these secondments CPHRL staff were assigned to carry out weekly blind internal quality assessments of isolates to test and build on skills. An advanced enteric microbiological technique certificate was awarded to staff who completed the curriculum and passed both the theory and practical exams.

Communications

Update reports were communicated between PHE and CPHRL via email and when necessary by telephone. Communications between CPRHL and DPC were established via weekly electronic reporting of results from CPHRL and presented at the weekly cholera task force meetings.

Results

Microbiological Testing

During Phase 1 testing (using selective media, oxidase and serotyping for *V. cholerae*), 120 samples were processed between November 2012 and February 2013 (17 from the Western Area, 38 from the Southern Area, 44 from the Eastern Area and 21 from the Northern Area). Four presumptive *V. cholerae* O1 colonies were isolated from different samples. Clinical strain EB23 contained three yellow TCBS colonies, two presumptive *V. cholerae* colonies were correctly identified, though only one was O1 positive, the third colony was *Aeromonas hydrophilia*. Two of three further presumptive isolates were correctly identified; the third was identified as *Vibrio fluvialis* (Table 2).

During Phase 2 testing (with the addition of API 20e kits and testing for multiple enteric pathogens), 258 samples were processed between March 2013 and August 2013 (137 from the Western Area, 32 from the Southern Area, 18 from the Eastern Area and 71 from the Northern Area). Eight presumptive isolates of *V. cholerae*, *E. coli*, *Shigella* and *Salmonella* were identified by CPHRL and confirmed as being correct at the Gastrointestinal Bacteria Reference Unit, PHE (Table 2).

The addition of biochemical identification methods enabled more accurate identification of pathogens and the ability to eliminate commensal flora. All of the original samples sent were retested and only the true positives were selected by CPHRL for confirmation testing. Accuracy of identification by CPHRL improved with the correct identification of all isolates sent to PHE.

Safety and Quality Assurance

Safe systems of practice were successfully set up and staff continually worked in a safe manner with the emphasis on the importance of hand washing when working with enteric samples. Quality systems at CPHRL have been put in place and will continue to improve according to ISO 15189 guidelines with a view to gaining accreditation.

Training

Phase one: Four staff from CPHRL were trained in the basic identification of *V. cholerae* and passed both the theory and practical exams. Eight staff were trained in quality and safety in the laboratory and all passed the theory exam.

Phase two: Between the first and second phase testing there was a two month gap during which the CPHRL lab was without continuous external expertise and support and dedicated CPHRL staff were not assigned to the enteric section. This led to a number of issues for example, lack of supply source for consumables for the enteric laboratory and major interruptions to the electricity supply to the laboratory which affected the continuous running of fridges and incubators. It also became evident that laboratory skills had been lost and routine testing of clinical samples ceased.

To rectify this problem, PHE in collaboration with the MoHS and WHO provided volunteer staff from the UK, consumables and financial support for a six month period. These resources facilitated the assignment of two dedicated staff to run the enteric section, the ability to access locally available consumables and funding became available to provide fuel to run the electricity generator at CPHRL. Two staff were trained in isolation and identification of multiple enteric organisms and passed the advanced theory exam as well as correctly identifying all isolated organisms. Training for the advanced practical exam was severely affected by the lack of a constant electrical supply and staff were unable to complete this aspect of the programme. A further eight staff were trained in quality and safety in the laboratory and all passed the advanced theory exam.

Table 2. Clinical and epidemiological information about enteric bacteria isolated between November 2012 and September 2013 by the Central Public Health Reference Laboratory in Sierra Leone with confirmation testing by Public Health England.

Date received at PHE	CPHRL Ref	District	CPHRL Phase +	Clinician presumptive ID	Clinical Symptoms	CPHRL ID	Reference ID by PHE
19.03.2013	EB0023-1*	Eastern/ Kenema City	1	Cholera	Not stated	*Vibrio cholerae* (Weak reaction with O1 Sera)	*Vibrio cholerae* Serotype: non O1, O139
19.03.2013	EB0023-2*	Eastern/ Kenema City	1	Cholera	Not stated	*Vibrio cholerae* (Weak reaction with O1 Sera)	*Aeromonas hydrophilia*
19.03.2013	EB0023-3*	Eastern/ Kenema City	1	Cholera	Not stated	*Vibrio cholerae* Serotype: O1 Ogawa	*Vibrio cholerae* Serotype: O1 Ogawa Biotype: El Tor
19.03.2013	EB0054	Southern/ Bo	1	Cholera & *Shigella*/*E. coli* O157	RWS,BS,V,AP, SD	*Vibrio cholerae* (Weak reaction with O1 Sera)	*Vibrio fluvialis*
19.03.2013	EB0091	Northern/ Kambia	1	Cholera	RWS,D,F, V,AP,H	*Vibrio cholerae* Serotype: O1 Inaba	*Vibrio cholerae* Serotype O1 Inaba Biotype: El Tor
19.03.2013	EB095	Northern/ Tonkolili	1	Unknown	RWS,D,F, V,AP	*Vibrio cholerae* Serotype O1 Ogawa	*Vibrio cholerae* Serotype: O1 Ogawa Biotype: El Tor
30.08.2013	EB271	Western	2	Cholera	WS,V	*E. coli*	*E. coli* Serotype: O unidentifiable
30.08.2013	EB91	Northern/ Kambia	2	Cholera	RWS,D,F, V,AP,H	*Vibrio cholerae* Serotype: O1 Inaba	*Vibrio cholerae* Serotype: O1 Inaba Biotype: El Tor
30.08.2013	EB095	Northern/ Tonkolili	2	Unknown	RWS,D,F, V,AP	*Vibrio cholerae* Serotype O1 Ogawa	*Vibrio cholerae* Serotype: O1 Ogawa Biotype: El Tor
30.08.2013	EB23	Eastern/ Kenema City	2	Cholera	Not stated	*Vibrio cholerae* Serotype: O1 Ogawa	*Vibrio cholerae* Serotype: O1 Ogawa Biotype: El Tor
30.08.2013	EB279	Western/ Western Urban	2	Salmonella	BS, WS, D, F, AP, SD	*Shigella flexneri*	*Shigella flexneri* Serotype: 2b
30.08.2013	EB255	Western/ Western Urban	2	Unknown	D,F,V, AP, R	*Salmonella sp*	*Salmonella Havana* Serotype: I13,23f,g
02.09.2013	EB274	Western/ Western Urban	2	Unknown	WS,F,V, N ND	*Salmonella sp*	*Salmonella Java* Serotype: I4,5,12:b:1 2 Phage type: RDNC
02.09.2013	EB254	Western/ Western Urban	2	Unknown	WS,D,F V,R,SD	*Salmonella sp*	*Salmonella Enteritidis* Serotype: I9,12:g m Phage type: RDNC

*Multiple colonies were tested some of which gave which cross-reactive agglutinations with the *Vibrio cholerae* O1 sera.
+Phase 1 (November 2012- February 2013) involved testing using the oxidase test and *Vibrio cholerae* serology for presumptive identification.
Phase 2 (March – August 2013) involved the addition of the API 20e tests to confirm for reference standard identification
Clinical Symptoms: RWS-rice watery stools, BS-bloody stools, WS-watery stools, D-diarrhoea, F-fever, V-vomiting, N-nausea, AP-abdominal pain, H-headache, R-rashes, ND-no dehydration, SD-some dehydration, SVD-Severe dehydration
Phage Type: RDNC – Recognised but does not conform (i.e. reacts with phages but does not match a recognised reference pattern for typing)

Communications

The DPC and CPHRL in Sierra Leone set up a communication system where samples from each district are collected by district surveillance officers and sent to CPHRL to confirm suspected cholera cases. CPHRL reports the results of confirmed cases of cholera weekly to the surveillance team.

Discussion

The aim of this project was to set up a functional diagnostic and reference laboratory services for enteric bacterial pathogens in Sierra Leone. Efforts were initially focussed on establishing emergency services for *V. cholerae* testing in response to the 2012 cholera epidemic, but it became clear there was an ideal opportunity to develop laboratory testing for other common gastrointestinal bacteria pathogens. A number of key learning issues were identified during the initial set up and training programme which may be valuable to others in similar situations

This project highlighted the importance of continuous training and support following the emergency response to the cholera outbreak. Assessment of the impact of a two month gap between the first and second phases of the training program emphasised how quickly initial training skills can be lost without continuous practice and the importance of engaging and monitoring the development of laboratory staff once the emergency period is over. The commitment to further support the initial training provided the staff opportunities to develop problem solving skills including the recognition of ambiguous laboratory testing results. For example following Phase 1 training, staff were able to identify the cholera epidemic strain but were reporting false positive serogroup O1 agglutination results due to cross reactions with the anti-sera. Following Phase 2, training staff were able to differentiate between weak and strong agglutination reactions.

Microbiological testing of *V. cholerae* is still in its infancy in Sierra Leone and therefore action to improve sanitation in hot spot areas will still continue to be based on current surveillance systems until a reliable network of laboratory confirmation is in place.

Sierra Leone has diarrhoeal surveillance, but with the exception of *V. cholerae*, is not pathogen specific. Request forms sent to CPHRL did sometimes indicate other pathogens such as *Shigella* or *Salmonella spp.* as being the causative agent but this is likely to be based on the clinician's knowledge of pathogens and/or assumptions of typical association (i.e. bloody diarrhoea was assumed to be associated with *Shigella spp.*in some cases).

The main issue encountered in setting up new laboratory systems was frequent disruptions to the electricity power supply. This resulted in reagents becoming inactivated. The introduction of positive control organisms ensured that reagents were able to be validated on a regular basis, before being used for every batch of tests. The inconsistent power supply also prevented clinical samples and laboratory tests from being incubated at the optimal temperature for bacterial growth, which had adverse consequences on the isolation and identification of enteric bacteria.

A further issue was the quality of the clinical samples collected. Whilst rectal swabs were requested to be taken and training and instructions were provided, it is possible that anal swabs were sometimes collected instead. Ideally, faecal samples should be tested within 24 hours, however, this was rarely possible due to transport and other logistical issues. Cary Blair swabs in transport media were used for specimen collection; these can be kept at room temperature (usually around 23°C) for several days and maintain enteric pathogen viability. In Sierra Leone, room temperature can be equivalent to the optimal temperature for bacteria growth (ie > 30°C) enabling commensal or contaminating bacteria to proliferate. The distance between specimen collection and the CPHRL together with the lack of paved roads meant that sometimes specimens were not tested within a few days but after longer periods. Often patients were treated with antibiotics and it is possible that clinical specimens were taken after treatment. The highest rate of *V. cholerae* recovery was from samples where no antibiotic treatment was given (data not shown). These difficulties might explain cases where the patient had classical rice water stools but *V. cholerae* was not isolated (see EB54 in Table 2). An instruction sheet regarding sample collection and negative results was provided to clinicians to encourage sample collection prior to antibiotic administration. It is important to note that microbiological results were not intended for patient management (as the patient would have been treated before the sample reaches CPHRL) but for national surveillance purposes. The collection of clinical specimens from diarrhoeal patients was a new concept in Sierra Leone and had not been performed routinely prior to the cholera outbreak. It is therefore important to feedback results to the clinicians to enhance clinical knowledge and encourage further engagement.

Conclusions

CPHRL has a functioning enteric diagnostic laboratory and is now capable of identifying *V. cholerae*, *E. coli* (including serogroup O157), *Salmonella sp* and speciation of *Shigella boydii, S. dysentariae, S. flexneri and S. sonnei*. Using microbiological identification is vital in supporting clinical definitions of cholera to obtain accurate number of cholera cases and important to develop for future diagnostics, surveillance and target hotspots to implement control and preventative methods.

CPHRL is working towards ISO15189 accreditation which is a necessary requirement if the laboratory is to continue to work at a high standard and produce accurate, reliable results. The next steps are for CPHRL to gain full accreditation status and to start rolling out training to hospital laboratories in other districts. Ultimately, the goal is to have a network of laboratories capable of performing front line presumptive identification of enteric bacteria and subsequently sending those isolates to CPHRL for reference confirmation. Ideally, reference centres could also be established in the Eastern, Southern and Northern areas of Sierra Leone so that clinical specimens can be processed in a timely manner thus improving the isolation of bacterial enteric pathogens.

Acknowledgements

Acknowledgements to the staff from the following organisations for their support and advice:

Ministry of Health and Sanitation (MoHS), Sierra Leone Foday Dafae, Sahr Gevao, Donald Taqi, Roland Conteh

World Health Organisation (WHO), Sierra Leone - Teniin Gakuruh, Pamela Mitula, Michael Ndolie, Ishata Conteh, Alex Chimbaru, Charles Mugero, Naba Kamara

Public Health England (PHE), UK Layi Onakunle, Elizabeth De Pinna, Anthony Kessel, Maria Zambon, Kenny Yap, Monica Sethi, Dawn Hedges, Vivienne DoNascimento

Centers for Diseases Control and Prevention (CDC), USA – Ishata Wurie (Based in Sierra Leone)

Global Outbreak and Response Alert Network (GOARN), Geneva - Pat Dury, Khristeen Umali Dalangin.

UK Department for International Development (DFID), Sierra Leone - Susan Mshana, Uzoamaka Gilpin and Martin Walshe

University of Haverford, USA - Iruka Okeke

Funding

This project was funded by the Health Protection Agency and its successor, Public Health England. Thank you to WHO for providing the transportation and support as well as the MoHS and other organisations who have contributed in developing laboratory testing. Thank you to WHO, DFID and other organisations who funded building of CPHRL and provided equipment, thereby laying a foundation to which this project could build upon.

Thank you to the MoHS in Sierra Leone for their enthusiasm and input, to the healthcare workers who are taking the samples and to the WHO Regional Office for Africa (AFRO).

Thank you to all of the other organisations contributing to fighting cholera in Sierra Leone and improving healthcare systems in developing countries. Thank you to the UN Office for the Coordination of Humanitarian Affairs and WHO for permission to print their data (Figure 1).

References

1. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Rudan I, Campbell H, Cibulskis R, Li M, Mathers C, Black RE (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 379: 2151-2161.
2. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acacio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 382: 209-222.
3. Kaper JB, Morris JG, Jr., Levine MM (1995) Cholera. Clin Microbiol Rev 8: 48-86.
4. WHO (2014) WHO CHOLERA. Fact Sheet. 107: http://www.who.int/mediacentre/factsheets/fs107/en/. Accessed 4th June 2014.
5. Ali M, Lopez AL, You YA, Kim YE, Sah B, Maskery B, Clemens J (2012) The global burden of cholera. Bull World Health Organ 90: 209-218A.
6. WHO(2006) Weekly epidemiological Record. 81: 297-308. http://www.who.int/cholera/statistics/en/. Accessed 4th June 2014.
7. WHO(2007) Weekly epidemiological Record. 82: 273-284. http://www.who.int/cholera/statistics/en/. Accessed 4th June 2014.
8. WHO(2008) Weekly epidemiological Record. 83: 269-284. http://www.who.int/cholera/statistics/en/. Accessed 4th June 2014.
9. WHO(2009) Weekly epidemiological Record. 84: 309-324. http://www.who.int/cholera/statistics/en/. Accessed 4th June 2014
10. WHO(2010) Weekly epidemiological Record. 85: 293-308. http://www.who.int/cholera/statistics/en/. Accessed 4th June 2014.
11. WHO (2011) Weekly epidemiological Record. 86: 325-340. http://www.who.int/cholera/statistics/en/. Accessed 4th June 2014.
12. WHO(2012) Weekly epidemiological Record. 87: 289-304. http://www.who.int/cholera/statistics/en/. Accessed 4th June 2014.

13. WHO(2013) Weekly epidemiological Record. 88: 321-336. http://www.who.int/cholera/statistics/en/. Accessed 4th June 2014.
14. WHO(2013) Outbreak Bulletin. 10: 1-8. Available: http://www.afro.who.int/en/clusters-a-programmes/dpc/epidemic-a-pandemic-alert-and-response/outbreak-news/3775-outbreak-bulletin-vol-2-issue-10-07-february-2013.html. Accessed 4th June 2014.
15. WHO(2014) Global Health Observatory Data Repository. MD4:Child health: Under-five mortality. Data by Country Available: http://apps.who.int/gho/data/node.main.525. Accessed 4th June 2014
16. Yao K, McKinney B, Murphy A, Rotz P, Wafula W, Sendagire H, Okui S, Nkengasong JN (2010) Improving quality management systems of laboratories in developing countries: an innovative training approach to accelerate laboratory accreditation. Am J Clin Pathol 134: 401-409.
17. Pal SC (1992) Laboratory Diagnosis 1st edition.New York: Plenum Publishing Corporation 229-251p.
18. Kauffman F(1966) The bacteriology of Enterobacteriaceae.1st edition. Copenhagen: Munksgaard 49-305p.
19. Edwards PR, Ewing WH (1962) Identification of Enterobacteriaceae. 2nd edition. Minneapolis: Burgess Publishing Company 25-167p.
20. WHO Regional Office for Africa (2011) WHO Guide for the Stepwise Laboratory Quality Improvement Process Towards Accreditation (SLIPTA). Available: http://www.afro.who.int/en/clusters-a-programmes/hss/blood-safety-laboratories-a-health-technology/blt-highlights/3859-who-guide-for-the-stepwise-laboratory-improvement-process-towards-accreditation-in-the-african-region-with-checklist.html. Accessed 4th June 2014.

Corresponding author
Marie Anne Chattaway
Gastrointestinal Bacteria Reference Unit, Public Health England,
61 Colindale Avenue, London, NW9 5EQ, UK
Phone: +44 (0) 208 327 6172
Email: marie.chattaway@phe.gov.uk

Conflict of interests: No conflict of interests is declared.

7.14 Evidence of an emerging extra-intestinal enteroaggregative ST38 clone. *In Press*

Chattaway MA, Jenkins C, Ciesielczuk H, DoNaccimento VN, Day M, Rodriguez I, Essen-Zanderbergen A, Chink AK, Wu G, Threlfall J, Woodford N and Wain J . 2014. Emerg.Infect.Dis. 20, (11) 1935-1937 available from: PM:25340736

Janaína Oliveira Luna, Sérgio Cabral Pinto, Fábio Fraga Schwingel, Concepta McManus, and Luiza Seixas

Author affiliations: University of Brasília, Brasília, Brazil (C.B. de Melo, A.R. Souza, C. McManus, L. Seixas); Ministry of Agriculture, Livestock and Food Supply (MAPA), Brasília, Brazil (M.E.P. de Sa); MAPA, Galeão Airport, Rio de Janeiro, Brazil (P.R Campani); MAPA, Guarulhos Airport, São Paulo, Brazil (J.O. Luna); MAPA, Confins International Airport, Belo Horizonte/Confins, Brazil (S. Cabral Pinto); MAPA, Brasília International Airport (BSB), Brasília, Brazil (F.F. Schwingel); and MAPA, Pedro Leopoldo, Brazil (A.M. de Oliveira, P.M.P.C. Mota)

DOI: http://dx.doi.org/10.3201/eid2011.131422

References

1. Schneider H. Good governance of national Veterinary Services. Rev. Sci. Tech. 2011;30:325–38 [cited 2011 Apr].
2. Hartnett E, Adkin A, Seaman M, Cooper J, Watson E, Coburn H, et al. A quantitative assessment of the risks from illegally imported meat contaminated with foot and mouth disease virus to Great Britain. Risk Anal. 2007;27:187–202. http://dx.doi.org/10.1111/j.1539-6924.2006.00869.x
3. Brückner GK. Ensuring safe international trade: how are the roles and responsibilities evolving and what will the situation be in ten years' time? Rev Sci Tech. 2011;30:317–24.
4. de Melo CB, de Sa MEP, Alves FF, McManus C, Aragão LF, Belo BB, et al. Profile of international air passengers intercepted with illegal animal products in baggage at Guarulhos and Galeão airports in Brazil. SpringerPlus. 2014;3:69. http://dx.doi.org/10.1186/2193-1801-3-69
5. International Air Transport Association. 3.6.2 Division 6.2—infectious substances. 2011 Jan 1 [cited 2011 Aug 10]. http://www.iata.org/whatwedo/cargo/dgr/Documents/DGR52_InfectiousSubstances(DGR362).pdf
6. Millar BC, Jiru X, Moore JE, Earle JAP. A simple and sensitive method to extract bacterial, yeast and fungal DNA from blood culture material. J Microbiol Methods. 2000;42:139–47. http://dx.doi.org/10.1016/S0167-7012(00)00174-3.
7. Dias NL. *Staphylococcus aureus* identification, evaluation of the enterotoxigenic potential and methicillin resistance by the PCR technique in dulce de leche samples in the Sete-Lagoas microregion, in the State of Minas Gerais; Brazil (dissertation) [in Portuguese]. Belo Horizonte, Minas Gerais, Brazil: Federal University of Minas Gerais, 2010.
8. Irenge LM, Walravens K, Govaerts M, Godfroid J, Rosseels V, Huygen K, et al. Development and validation of a triplex real-time PCR for rapid detection and specific identification of *M. avium* sub sp. *paratuberculosis* in faecal samples. Vet Microbiol. 2009;136:166–72. http://dx.doi.org/10.1016/j.vetmic.2008.09.087.

Address for correspondence: Cristiano Barros de Melo, Universidade de Brasília (UnB-FAV), Campus Darcy Ribeiro, ICC Sul, Asa Norte, Brasília, DF 70910-900 Brazil; email: cristianomelo@unb.br

Evidence of Evolving Extraintestinal Enteroaggregative *Escherichia coli* ST38 Clone

To the Editor: Several clones of extended-spectrum β-lactamase (ESBL)–producing extraintestinal pathogenic *Escherichia coli* (ExPEC) have globally expanded their distribution, including multilocus sequence types (MLSTs) ST38, ST131, ST405, and ST648 (*1*). ExPEC infections often originate from the patient's own intestinal flora, although the degree of overlap between diarrheagenic *E. coli* and ExPEC pathotypes is unclear. Relatively little is known about antimicrobial drug resistance in the most common diarrheagenic *E. coli* groups, including enteroaggregative *E. coli* (EAEC), and bacterial gastroenteritis is generally managed without use of antimicrobial drugs.

The ability of diarrheagenic *E. coli* to cause extraintestinal infections has been shown in previous studies: a study among children in Nigeria linked EAEC to uropathogenic clonal group A (*2*), and a study in Brazil showed that EAEC markers were present in 7.1% of the *E. coli* isolates from urinary tract infections (*3*). Neither of these studies identified clonal lineages of EAEC specifically associated with extraintestinal infections.

We conducted this study to establish the presence and characteristics of ESBL-producing EAEC in a well-defined collection of ESBL-producing isolates (*4*). The isolates were from human and animal sources in Germany, the Netherlands, and the United Kingdom. The study was conducted at Public Health England during January–April 2013.

DNA from 359 ESBL isolates (*4*) was screened for the presence of the EAEC transport regulator gene (*aggR*), located on the EAEC plasmid, by using a real-time PCR assay and the following primers and probe: AggR_F 5′-CCATTTATCGCAATCAGATTAA-3′ AggR_R 5′-CAAGCATCTACTTTTGATATTCC-3′, AggR_P Cy5-CAGCGATACATTAAGACGCCTAAAGGA-BHQ. The amplification parameters were 50°C for 2 min, 95°C for 2 min, and 40 cycles at 95°C for 10 s and at 60°C for 20 s. Isolates positive for *aggR* were confirmed to be *E. coli* by using the Omnilog GenIII MicroPlate (Biolog, Hayward, CA, USA). Serotyping was done by using standard methods (*5*).

The phylogroup was determined for each isolate, and isolates were then assigned to 1 of the 4 major *E. coli* groups: A, B1, B2, and D (*6*). A microarray was used to detect ESBL genes, such as bla_{CTX-M}, at the group level, as previously described (*4*). The antimicrobial drug susceptibilities of EAEC isolates were determined by using the agar incorporation method, as described in the British Society for Antimicrobial Chemotherapy guidelines (*7*).

Virulence factors associated with intestinal and extraintestinal infection

Table. Characteristics of human-derived ESBL-producing enteroaggregative *Escherichia coli* isolates from sources in Germany, the Netherlands, and the United Kingdom*

Isolate	Serotype†	ST	Cplx‡	Country	Source	Phylotype	*aggR*§	Plasmidic ESBL
ESBL-723	OR:H30	38	38	UK	Urine	D	+	CTX-M-15
ESBL-746	O125ac:H30	38	38	UK	Urine	D	+	CTX-M-15
ESBL-884	O19a:H30	38	38	UK	Urine	D	+	CTX-M-14
ESBL-831	O19a:H30	38	38	UK	Urine	D	+	CTX-M-14
ESBL-815	O19a:H30	38	38	UK	Blood	D	+	CTX-M-15
ESBL-26	O153:H30	38	38	Netherlands	Urine	D	+	CTX-M-51
ESBL-221	O92:H33	34	10	Germany	Feces	A	+	CTX-M-3
ESBL-45	O?:H26	58	155	Netherlands	Urine	B1	+/–	CTX-M-14
ESBL-46	O?:H-	694	None	Netherlands	Urine	A	+/–	CTX-M-15
ESBL-48	O15:H1	545	None	Netherlands	Urine	D	+/–	CTX-M-1
ESBL-64	O?:H23	224	None	Netherlands	Urine	B1	+/–	CTX-M-1

*All isolates were collected in 2009 (*4*). ESBL, extended-spectrum β-lactamase. ST, sequence type.
†H- not motile; O?, O unidentifiable; R, rough reaction.
‡Cplx–ST complex comprising single-locus variants.
§*aggR*, enteroaggregative *E. coli* regulatory gene; +, positive in screen and isolates; –, negative in screen and isolates; +/–, positive in screen but negative in isolates, indicating unstable plasmid.

(*8*) and with EAEC were investigated as previously described (*9*). We assigned a virulence score (total number of virulence factor genes detected; maximum possible score 22) and a resistance score (total number of drug classes; maximum score 11) to each isolate.

We isolated 11 EAEC from humans. Eight of the EAEC were isolated from urine specimens, and 1 was isolated from a blood culture; 63% belonged to phylogroup D (Table). EAEC ST38, the most common (55%) ST, was significantly associated with extraintestinal sites in the subset of 140 human isolates (Fisher exact test, $p<0.0001$).

In this study, we identified multidrug-resistant EAEC isolates belonging to ST38; the isolates had various somatic antigens and bla_{CTX-M} genes (Table). The multiple somatic antigens, variety of antimicrobial drug–resistance scores, and variety of gene complements in this successful ST indicate multiple acquisitions of virulence markers, rather than clonal expansion from a single source (Table; online Technical Appendix Figure, http://wwwnc.cdc.gov/EID/article/20/11/13-1845-Techapp1.pdf).

In the MLST public database, which contained 5,143 *E. coli* entries in June 2013, ST38 is predominantly associated with urinary tract infections, but in-house MLST studies at the Gastrointestinal Bacteria Reference Unit, Public Health England, have shown that ST38 is a successful EAEC group. The presence of EAEC virulence factors, such as aggregative adherence fimbria AAFI and *aggR*, can mediate adherence of *E. coli* to bladder epithelial cells, but the virulence factors do not impart uropathogenic properties to all EAEC isolates (*10*). The ST38 strain described here probably originated from the gut and independently acquired the 2 phenotypes (uropathogenic *E. coli* [UPEC] and EAEC), which would suggest the emergence of a UPEC/EAEC hybrid strain. It seems likely that an ST38 *E. coli* strain adapted to EAEC plasmid carriage (a change that would help survival in the gut through increased adherence) has acquired UPEC virulence factors, facilitating the exploitation of an extraintestinal niche, the urinary tract.

Despite the characterization of numerous virulence factors, no single genetic feature currently defines EAEC or UPEC isolates. Because the EAEC ST38 strain had 4–7 ExPEC-associated virulence factors, we suggest that, on the basis of epidemiologic, microbiological, and molecular characteristics, the EAEC ST38 described in this study should be considered an ExPEC associated with uropathogenic infections. It is possible that the multidrug-resistant EAEC ExPEC group has expanded globally but is currently underreported. We therefore urge testing for the EAEC genotype in all clinical studies of *E. coli* pathotypes.

Our findings show the potential for EAEC, previously considered a gut pathogen, to cause extraintestinal infection. We suggest that the UPEC/EAEC pathotype may be an evolving clonal group. In particular, a single sequence type, ST38, was associated with multidrug resistance and with urinary tract infection in humans.

Acknowledgments

We thank Dawn Hedges for serotyping the organisms, Andy Lawson for designing the *aggR* primers, Danielle Hall for help with PCR screening, and Daniele Meunier for her valuable insight.

This study was funded by the Public Health England Gastrointestinal Bacteria Reference Unit.

Marie Anne Chattaway, Claire Jenkins, Holly Ciesielczuk, Martin Day, Vivienne DoNascimento, Michaela Day, Irene Rodríguez, Alieda van Essen-Zandbergen, Anne-Kathrin Schink, Guanghui Wu, John Threlfall, Martin J. Woodward, Nick Coldham, Kristina Kadlec, Stefan Schwarz, Cindy Dierikx, Beatriz Guerra, Reiner Helmuth, Dik Mevius, Neil Woodford, and John Wain

Author affiliations: Public Health England, Colindale, London, UK (M.A. Chattaway, C. Jenkins, H. Ciesielczuk, Martin Day, V. DoNascimento, Michaela Day, J. Threlfall, N. Woodford, J. Wain); Hospital Universitario Ramón y Cajal, Madrid, Spain (I. Rodríguez); University of East Anglia, Norfolk, UK (J. Wain); Federal Institute for Risk Assessment, Berlin, Germany (I. Rodríguez, B. Guerra, R. Helmuth); Central Veterinary Institute of Wageningen, Lelystad, the Netherlands (A. van Essen-Zandbergen, C. Dierikx, D. Mevius); Friedrich-Loeffler-Institut, Neustadt-Mariensee, Germany (A.-K. Schink, K. Kadlec, S. Schwarz); and Animal Health and Veterinary Laboratories Agency, Weybridge, UK (G. Wu, M.J. Woodward, N. Coldham)

DOI: http://dx.doi.org/10.3201/eid2011.131845

References

1. Pitout JD. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. Front Microbiol. 2012;3:9. http://dx.doi.org/10.3389/fmicb.2012.00009
2. Wallace-Gadsden F, Johnson JR, Wain J, Okeke IN. Enteroaggregative *Escherichia coli* related to uropathogenic clonal group A. Emerg Infect Dis. 2007;13:757–60. http://dx.doi.org/10.3201/eid1305.061057
3. Abe CM, Salvador FA, Falsetti IN, Vieira MA, Blanco J, Blanco JE, et al. Uropathogenic *Escherichia coli* (UPEC) strains may carry virulence properties of diarrhoeagenic *E. coli*. FEMS Immunol Med Microbiol. 2008;52:397–406. http://dx.doi.org/10.1111/j.1574-695X.2008.00388.x
4. Wu G, Day MJ, Mafura MT, Nunez-Garcia J, Fenner JJ, Sharma M, et al. Comparative analysis of ESBL-positive *Escherichia coli* isolates from animals and humans from the UK, the Netherlands and Germany. PLoS ONE. 2013;8:e75392. http://dx.doi.org/10.1371/journal.pone.0075392
5. Gross RJ, Rowe B. Serotyping of *Escherichia coli*. In: Sussman M, editor. The virulence of *Escherichia coli*. Cambridge (UK): Cambridge University Press; 1985. p. 345–60.
6. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol. 2000;66:4555–8. http://dx.doi.org/10.1128/AEM.66.10.4555-4558.2000
7. Howe RA, Andrews JM. BSAC standardized disc susceptibility testing method (version 11). J Antimicrob Chemother. 2012;67:2783–4. http://dx.doi.org/10.1093/jac/dks391
8. Johnson JR, Stell AL. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. J Infect Dis. 2000;181:261–72. http://dx.doi.org/10.1086/315217
9. Jenkins C, Chart H, Willshaw GA, Cheasty T, Smith HR. Genotyping of enteroaggregative Escherichia coli and identification of target genes for the detection of both typical and atypical strains. Diagn Microbiol Infect Dis. 2006;55:13–9. http://dx.doi.org/10.1016/j.diagmicrobio.2005.10.019
10. Boll EJ, Struve C, Boisen N, Olesen B, Stahlhut SG, Krogfelt KA. Role of enteroaggregative *Escherichia coli* virulence factors in uropathogenesis. Infect Immun. 2013;81:1164–71. http://dx.doi.org/10.1128/IAI.01376-12

Address for correspondence: Marie Anne Chattaway, Public Health England, Gastrointestinal Bacteria Reference Unit, Colindale, NW9 5EQ, UK; email: marie.chattaway@phe.gov.uk

Resolution Threshold of Current Molecular Epidemiology of Diphtheria

"The fox who longed for grapes, beholds with pain

The tempting clusters were too high to gain;

Grieved in his heart he forced a careless smile,

And cried, 'They're sharp and hardly worth my while.'"

(Aphra Behn, 1687, after Aesop's *The Fox and the Grapes*)

To the Editor: Diphtheria is an extremely rare disease in Europe but remains a major health issue in developing countries (*1–3*). In recent years, steady progress has been made toward understanding the factors of pathogenicity of its causative agent (*Corynebacterium diphtheriae*). In contrast, remarkable advances in its basic genomics have not been sufficiently translated into the molecular epidemiology of diphtheria. A recent report by Zasada (*4*) offers an apt opportunity to take a new look at this issue.

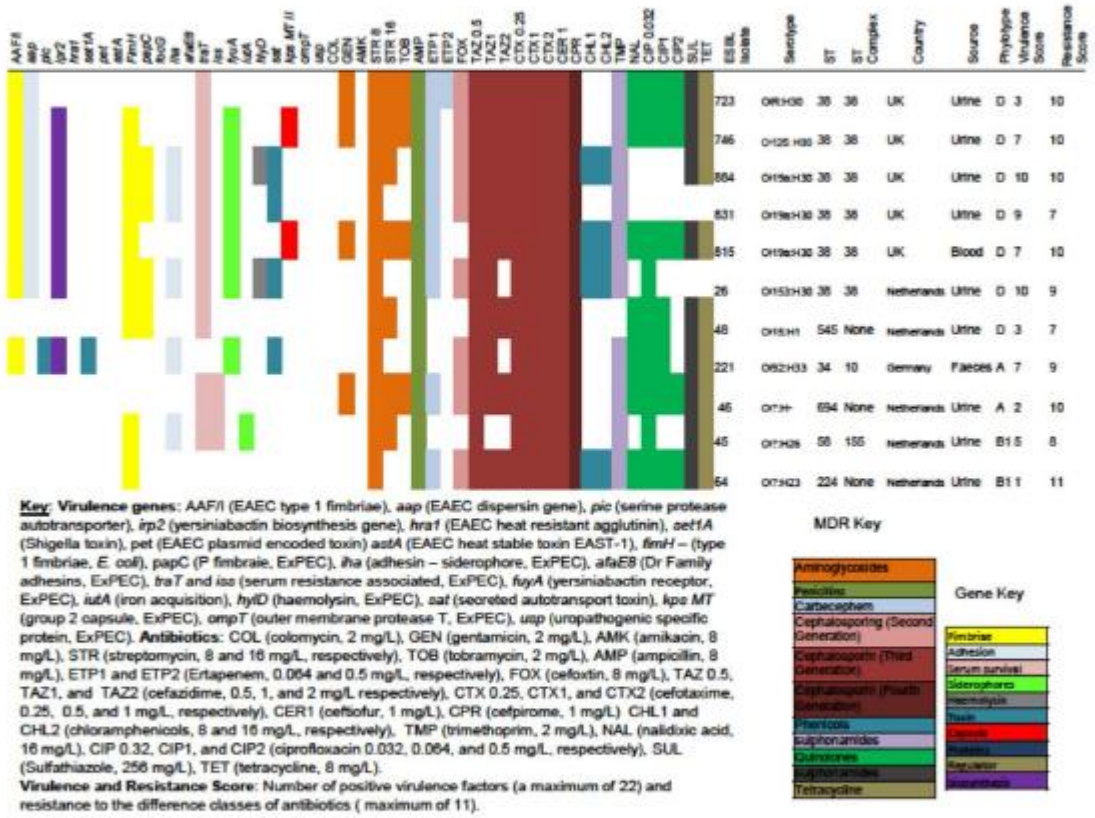
The current genotyping repertoire of *C. diphtheriae* includes several methods but those most frequently used are classical ribotyping and pulsed-field gel electrophoresis (PFGE). More recently, a multilocus sequence typing (MLST) scheme for *C. diphtheriae* was developed (*5*). Compared with ribotyping, PFGE, and other methods based on analysis of banding profiles, MLST results are digital, unambiguous, and portable. MLST discrimination of 150 isolates from 18 countries and spanning 50 years was "in accordance with previous ribotyping data, and clonal complexes associated with disease outbreaks were clearly identified by MLST" (*5*).

In the report by Zasada (*4*), all 3 recommended methods (PFGE, MLST, and ribotyping) were used to genotype 25 nontoxigenic *C. diphtheriae* isolates from Poland. The author concluded that these isolates "represent a single clone despite isolation … in different part of the country over a 9-year period" and raised the question of whether a single clone of *C. diphtheriae* is circulating in Poland (*4*). These isolates are related genetically, but do they represent a truly single clone or might they be further discriminated? Their circulation in Poland may be caused by their high pathogenicity, but also (or instead) it might reflect their endemic, historical prevalence in this country. I believe that these questions are unlikely to be answered by the internationally agreed-upon methods for *C. diphtheriae* typing because of their insufficient resolution: the discriminatory power of MLST does not exceed that of ribotyping (*5*).

Article DOI: http://dx.doi.org/10.3201/eid2011.131845

Evidence of Evolving Extraintestinal Enteroaggregative *Escherichia coli* ST38 Clone

Technical Appendix



Key: **Virulence genes**: AAF/I (EAEC type 1 fimbriae), *aap* (EAEC dispersin gene), *pic* (serine protease autotransporter), *irp2* (yersiniabactin biosynthesis gene), *hra1* (EAEC heat resistant agglutinin), *aet1A* (Shigella toxin), pet (EAEC plasmid encoded toxin) *astA* (EAEC heat stable toxin EAST-1), *fimH* – (type 1 fimbriae, *E. coli*), papC (P fimbraie, ExPEC), *iha* (adhesin – siderophore, ExPEC), *afaE8* (Dr Family adhesins, ExPEC), *traT* and *iss* (serum resistance associated, ExPEC), *fuyA* (yersiniabactin receptor, ExPEC), *iutA* (iron acquisition), *hylD* (haemolysin, ExPEC), *sat* (secreted autotransport toxin), *kps MT* (group 2 capsule, ExPEC), *ompT* (outer membrane protease T, ExPEC), *uap* (uropathogenic specific protein, ExPEC). **Antibiotics**: COL (colomycin, 2 mg/L), GEN (gentamicin, 2 mg/L), AMK (amikacin, 8 mg/L), STR (streptomycin, 8 and 16 mg/L, respectively), TOB (tobramycin, 2 mg/L), AMP (ampicillin, 8 mg/L), ETP1 and ETP2 (Ertapenem, 0.064 and 0.5 mg/L, respectively), FOX (cefoxtin, 8 mg/L), TAZ 0.5, TAZ1, and TAZ2 (cefazidime, 0.5, 1, and 2 mg/L respectively), CTX 0.25, CTX1, and CTX2 (cefotaxime, 0.25, 0.5, and 1 mg/L, respectively), CER1 (ceftiofur, 1 mg/L), CPR (cefpirome, 1 mg/L) CHL1 and CHL2 (chloramphenicols, 8 and 16 mg/L, respectively), TMP (trimethoprim, 2 mg/L), NAL (nalidixic acid, 16 mg/L), CIP 0.32, CIP1, and CIP2 (ciprofloxacin 0.032, 0.064, and 0.5 mg/L, respectively), SUL (Sulfathiazole, 256 mg/L), TET (tetracycline, 8 mg/L).
Virulence and Resistance Score: Number of positive virulence factors (a maximum of 22) and resistance to the difference classes of antibiotics (maximum of 11).

Technical Appendix Figure. Virulence factors and antimicrobial drug resistance gene content of Enteroaggregative *Escherichia coli* (EAEC) isolates, grouped by phylogroup.

7.15 Enteroaggregative *Escherichia coli* have evolved independently as distinct groups within the *E. coli* population with varying ability to cause disease.

Chattaway MA, Jenkins C, Rajendram D, Cravatio A, Talukder KA, Dallman T, Underwood A, Platt S, Okeke IN and Wain J . 2014. PLOSONE, 9, (11) e112967 available from: PM:25415318

PLOS ONE

RESEARCH ARTICLE

Enteroaggregative *Escherichia coli* Have Evolved Independently as Distinct Complexes within the *E. coli* Population with Varying Ability to Cause Disease

Marie Anne Chattaway[1]*, Claire Jenkins[1], Dunstan Rajendram[2], Alejandro Cravioto[3,4], Kaisar Ali Talukder[3], Tim Dallman[1], Anthony Underwood[5], Steve Platt[5], Iruka N. Okeke[6], John Wain[7]

1. Gastrointestinal Bacteria Reference Unit, Public Health England, London, United Kingdom, **2.** Genomic Service Unit, Public Health England, London, United Kingdom, **3.** International Vaccine Institute, Gwanak-gu, Seoul, Republic of Korea, **4.** Centre for Food and Water Borne Diseases, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh, **5.** Bioinformatics, PHE, London, United Kingdom, **6.** Haverford College, Haverford, Pennsylvania, United States of America, **7.** Norwich Medical School, University of East Anglia, Norwich, United Kingdom

*marie.chattaway@phe.gov.uk

Abstract

Enteroaggregative *E. coli* (EAEC) is an established diarrhoeagenic pathotype. The association with virulence gene content and ability to cause disease has been studied but little is known about the population structure of EAEC and how this pathotype evolved. Analysis by Multi Locus Sequence Typing of 564 EAEC isolates from cases and controls in Bangladesh, Nigeria and the UK spanning the past 29 years, revealed multiple successful lineages of EAEC. The population structure of EAEC indicates some clusters are statistically associated with disease or carriage, further highlighting the heterogeneous nature of this group of organisms. Different clusters have evolved independently as a result of both mutational and recombination events; the EAEC phenotype is distributed throughout the population of *E. coli*.

OPEN ACCESS

Citation: Chattaway MA, Jenkins C, Rajendram D, Cravioto A, Talukder KA, et al. (2014) Enteroaggregative *Escherichia coli* Have Evolved Independently as Distinct Complexes within the *E. coli* Population with Varying Ability to Cause Disease. PLoS ONE 9(11): e112967. doi:10.1371/journal.pone.0112967

Editor: Dongsheng Zhou, Beijing Institute of Microbiology and Epidemiology, China

Received: July 7, 2014

Accepted: October 16, 2014

Published: November 21, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files

Funding: This study was supported by Gastrointestinal Bacteria Reference Unit, Public Health England, International Centre for Diarrhoeal Disease Research, Bangladesh, Society for Applied Microbiology (Supervisor fund), the US National Science Foundation (Awards #1329248 and #0948460) and Society for General Microbiology (Presidents fund). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The definition of EAEC varies in studies which either use its aggregative adherence (AA) phenotype on HEp-2 cells [1], the CVD432 probe [2] or PCR to detect the anti-aggregative transporter (*aat*) gene [3] or the EAEC regulatory gene (*aggR*) [4] or a combination of phenotype and genotype. Enteroaggregative *E. coli* (EAEC)

have been associated with diarrhoea in epidemiological studies and outbreaks. Investigations of EAEC are based on identification of a group of bacteria (EAEC) assumed to be pathogenic as they were isolated from symptomatic cases. However, not all *E. coli* which contain EAEC virulence factors are pathogenic [5,6] and so associations between EAEC and virulence are not clear. A comprehensive study looking at the relationship between phylogeny from case or healthy carriage in multiple countries has not been performed and there has been limited analysis of EAEC at the population level. The most detailed study on EAEC population analysis was in Nigeria and was carried out to find an association with EAEC complexes and disease in children under 5 with links to virulence genes, resistance and plasmid groups [7]. Results indicated that the range of sequence types (STs) associated with EAEC is very large and disease, only within a specific age-group, was linked to ST10, an ST associated with multiple *E. coli* pathotypes. There were no reported associations between disease and, virulence genes, resistance profiles, nor plasmid compatibility groups.

Serogrouping (typing of the somatic antigen only) and serotyping (typing of the somatic and flagella antigen) is used extensively for characterising and classifying *E. coli* and *Salmonella enterica*. For both species serogroup is not discriminatory enough to be a useful strain typing tool but serotype can be more robust. For *Salmonella*, serotype is strongly associated with sequence type [8]. Serotyping therefore can give a robust typing scheme although conversion between serotypes can occur by horizontal genetic exchange [8] and so distort the relationship within serotypes. The relationship between serotype and the EAEC phenotype is not defined; here we describe a comprehensive examination of the relationship between phylogeny/serotype/sequence type and whether the strain was isolated from a patient with diarrhoea (case) or a healthy control.

We addressed the questions, are certain EAEC lineages more likely to be associated with disease and have all EAEC evolved from a common ancestor? The study used globally sourced EAEC isolates from three major case control studies and analysed chromosomal core sequence data to look for an association between bacterial background and disease.

Materials and Methods

Bacterial Strains

Three case control studies, sporadic and outbreak cases of 564 EAEC spanning over 29 years (1985–2013) were used in this study (Table 1). All of these strains were included to encompass a representation of EAEC in the global community (including UK travellers) over the past three decades. EAEC were defined as having the *aat* gene/CVD432 probe reaction [2,3], and/or the *aggR* regulatory gene [6] and/or the aggregative adherence (AA) phenotype [1] where the phenotypic test was available (Table 1). Isolates included strains from multiple studies including the UK (273), Bangladesh (169), Nigeria (121) and the

Table 1. Summary of 564 EAEC strains analysed in this study.

Country	Source	Year Range	Case	Control	Reference
Peru	∝042 prototypical strain	1985	1	0	[27]
UK	#GBRU Archive Clinical strains	1985–1995	17	0	This Study∝
UK	∝IID1 Case/Control Study	1993–1996	121	36	[28]
UK	∝GBRU Outbreak A	1994	2	0	[29]
UK	∝GBRU Outbreak B	1994	8	0	[29]
UK	∝GBRU Outbreak C	1994	1	0	[29]
UK	∝GBRU Outbreak D	1995	3	0	[29]
Bangladesh	∝GBRU Outbreak E	1998	12	0	This Study∝
Nigeria	∝Nigeria Case/Control Study	1999	66	55	[7]
UK	#IID2 case study	2008–2009	25	0	[5]
Bangladesh	∝GEMS Case/Control Study	2007–2011	97	61	[30,31]
Germany	#O104:H4 VTEC Outbreak	2011	1	0	[26]
UK	#O111:H2 Household Outbreak	2012	1	0	[32]
UK	#GBRU Clinical Strains	2009–2013	38	0	This Study∝
UK	#GBRU Spice Outbreak	2013	19	0	[33]

Selection of EAEC strains used in this study including the year the strain was isolated and its geographical location. ∝Strains from this study not previously described include archived clinical strains received by GBRU for typing between 1985–1995, Outbreak E of enteroaggregative *E. coli* that occurred in Bangladesh in 1998, recent clinical strains received by GBRU for typing between 2009–2013. #EAEC were defined as having the *aat* and/or *aggR* gene. ∝Other EAEC strains were defined as having the *aat* gene/CVD432 probe reaction and/or the aggregative adherence (AA) phenotype.

doi:10.1371/journal.pone.0112967.t001

prototypical 042 EAEC reference strain from Peru (1) (Table 1). Due to the varying definition of EAEC, all strains were included irrespective of phenotypic and genotypic definition to prevent any bias that may affect the analysis. Where an EAEC outbreak was related to one ST and serotype, only one representative strain has been included.

Nigeria isolates were previously analysed [7] All other EAEC strains were plated onto blood agar plates (PHE Media) to test for purity and archived onto Dorset Eggs (PHE Media) and stored at room temperature and also archived on beads [Prolab] and stored at −80 ℃.

Identification and Serotyping

Identification of UK and Bangladesh enteroaggregative *Escherichia coli* (EAEC) strains (443 strains) was confirmed phenotypically using biochemical profiling of media tubes [9] by the Gastrointestinal Bacteria Reference Unit of PHE at Colindlae. Typical metabolic profiles of *E. coli* included positive reactions for glucose, gas, lactose, mannitol, lysine, ornithine, mucate, sodium acetate and indole. Serotyping of the somatic and flagella antigen [10] was carried out on the heat stable lipopolysaccharide (Somatic or O) antigens and the flagellar (H) antigens. Strains which reacted with all antigens were termed rough and those that did not react with any were termed 'O?' or 'H?'. Nigerian strains had previously been identified and published [7], strains were not accessible for serotyping.

Multi-locus sequence typing and analysis of EAEC

Nigerian sequence data was provided by Okeke *et al* as previously published [7]. Genomic DNA Extraction of all other *E. coli* isolates was carried out using the Wizard Genomic DNA purification kit (Promega). PCR amplification of seven Multilocus sequence typing (MLST) gene targets; *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA* [11] was carried out followed by PCR purification of the amplicons using the ExoSAP-IT PCR cleanup method (Amersham Biosciences UK Ltd). Purified PCR fragments from the seven MLST gene targets were sequenced with both forward and reverse sequencing primers using the ABI prism Bigdye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and detected and analysed on the 3730XL ABI Genetic Analyser (Applied Biosystems). Sequence data was analysed and checked for quality and alleles trimmed for analysis, any ambiguous results were repeated (BioNumerics v6.1). Allele numbers and sequence types (ST) were calculated and deposited in the publically accessible *E. coli* MLST database (http://mlst.warwick.ac.uk.). Phylogenetic inference of the EAEC complexes ancestral allelic profiles and strain interrelatedness were made using minimum spanning trees (BioNumerics v6.1). A complex (Cplx) included any single locus variants (SLV) of an allele in relation to a ST.

Selection of EAEC Disease and Carriage complexes and statistical analysis

As of 18th December 2013, the data available in the public database indicates there were 155 EAEC (121 *Okeke et al* Nigerian study used in this study excluded) out of 6110 *E. coli* entries, accounting for 2.4% of the database. There were 1164 entries of defined diarrhoeagenic pathotypes (see below for description) of *E. coli* which EAEC accounts for 13 % (155/1164). From the 564 strains used in this study, a complex was considered a successful representation if it contained 4 or more strains which would account for a minimum of 2.5% (4/155) of the known EAEC deposited in the public database. The majority of the MLST data associated with these isolates has been previously published [7,11].

From the EAEC dataset used in this study, complexes containing four or more EAEC were deemed successful (i.e. strains which have continued to proliferate over time in the population) of which there were 17 complexes. The 17 assigned complexes were then tested using a fishers exact test [12] for the significance of the complexes being associated with disease or carriage in relation to the entire dataset (564 strains). Statistical tests of significance were conducted using the Fisher's exact test on Epi-Info version 2.3.1 (http://www.openepi.com).

The public database was compared against each of the 17 complexes to rule out complexes with a high association with other pathotypes [11]. Pathotypes included diarrheagenic types including enterotoxigenic, verocytotoxic, entero-pathogenic, enteroinvasive and diffusely adherent *E. coli* (ETEC, VTEC, EPEC, EIEC and DAEC respectively). Extra-intestinal pathogenic *E. coli* (ExPEC) including wounds, meningitis, external sources (ExPEC_Vag) and urinary pathogenic *E. coli* (UPEC). Antibiotic resistance *E. coli* (ESBL, CTX-M-15,

NMEC, AmpC CYM-2, c CMY-2, NDM-1, ESBL CTX-M-32 & OXA-48). Other pathotypes included avian pathogenic *E. coli* (APEC), non-pathogenic commensal strains and *E. coli* with no defined pathotype. EAEC complexes were assessed based on the public database and data from this study and tested using a fisher exact test [12] (open epi version 2.3.1) for significance of the complexes being associated with EAEC.

ClonalFrame Analysis

Clonal Frame analysis was carried out (http://www.xavierdidelot.xtreemhost.com/clonalframe.htm) on all EAEC isolates to investigate the relationships of the different sequence type complexes. ClonalFrame is a Bayesian method of constructing evolutionary histories that takes both mutation and recombination into account [13]. The Graphic User Interface in the ClonalFrame programme was used to construct 75% majority-rule consensus trees, mutational (theta) and recombination rates. Other analysis including the measure of the frequency at which recombination occurs relative to mutation (ρ/θ). The relative effect of recombination on the genetic diversification of populations, ratio r/m in which the ratio of rates at which nucleotides become substituted as a result of recombination and mutation [14] was also used. Finally, the external to internal branch length ratio was computed which gave the inferred expected values against the coalescent and actual ratios. Analysis was split into assessing the Bangladesh and Nigeria case control studies and UK clinical data set for comparison against the entire dataset.

Placing EAEC in the *E. coli* phylogeny

Multi-locus sequence analysis (MLSA) was performed by concatenating MLST sequence alleles of the EAEC from this dataset and all sequence types representative of the *E. coli* phylogeny. These were aligned and clustered (MEGA V 5.1) and the genetic relationship of isolates designated as was assessed in the context of all *E. coli* using a neighbour joining tree phylogeny (MEGA V 5.1 and FigTree V 1.4). Phylogrouping PCR was carried out on the 17 main groups of EAEC [15] and labelled on the phylogeny.

Results

Serotype and complex distribution within the EAEC population structure

From the 564 EAEC strains studied, there were 126 different sequence types, including additional not previously described sequence types of which 57 were single locus variants (SLV), 20 double locus variants (DLV) and two were triple locus variants (TLV).

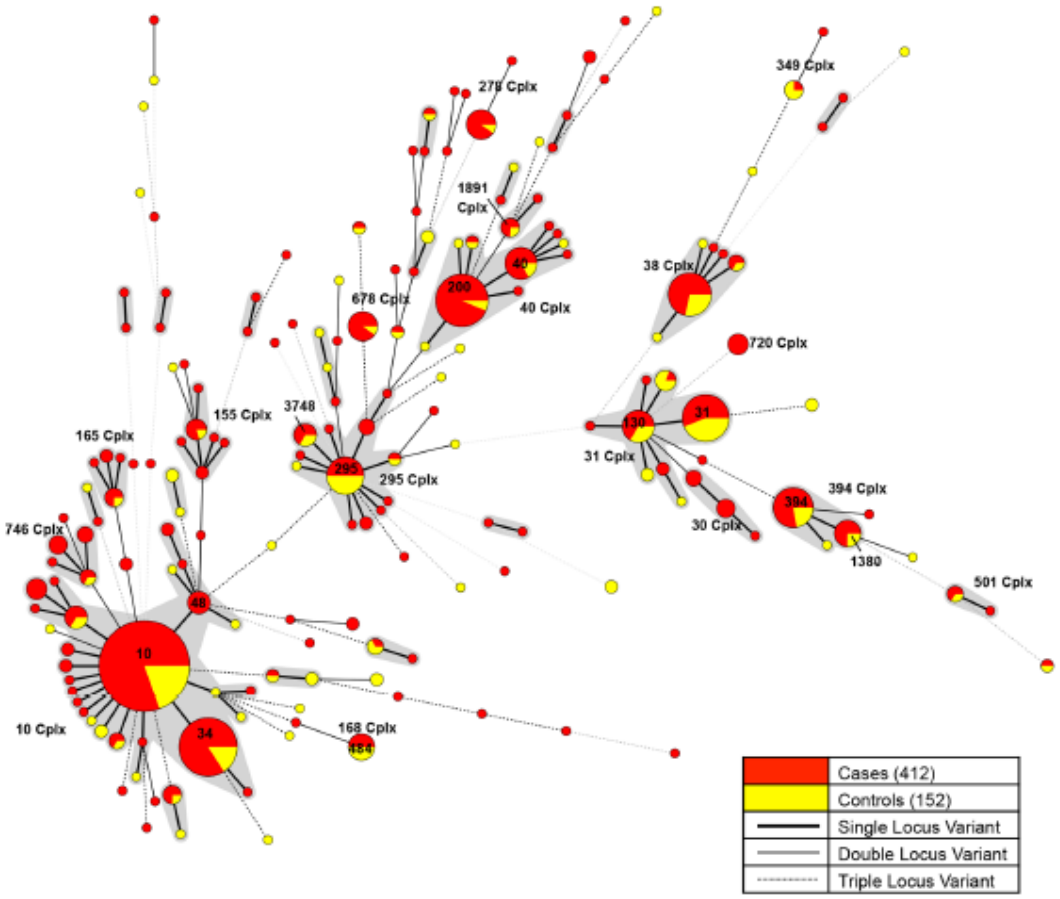


Figure 1. Minimal spanning tree of 564 enteroaggregative *E. coli*. Minimum spanning tree of the 564 EAEC used in this study colour coded by isolates from cases (red) and controls (yellow). Complexes shaded in grey consist of single locus variants (SLV). Sequence types and complex (Cplx) are labelled as numbers.

doi:10.1371/journal.pone.0112967.g001

There were 17 main complexes (Figure 1) containing 4 or more strains of EAEC totalling 358 strains with the top five complexes (Cplx) including ST10 Cplx (39%, 141/358), ST31 Cplx and ST40 Cplx (12%, 42/358), ST394 Cplx (7%, 26/358) and ST295 Cplx and ST38 Cplx (6%,21/358). There were 35 isolates (6.2%, 35/564) that contained one or more new alleles (40 new alleles in total) not previously described. All new alleles were deposited to the public database (http://mlst.ucc.ie/mlst/dbs/Ecoli) for a new allele and/or ST assignment.

Most EAEC serotypes were heterogeneous with respect to ST and dispersed throughout the population structure (Figure S1): Some serotypes were

We addressed the questions, are certain EAEC lineages more likely to be associated with disease and have all EAEC evolved from a common ancestor? The study used globally sourced EAEC isolates from three major case control studies and analysed chromosomal core sequence data to look for an association between bacterial background and disease.

predominantly associated with STs (O7:H4-ST484, O104:H4-ST678, O111:H21-ST40, O125ac:H9-ST295, O153:H30-ST38,) while others were found in multiple STs (O44:H18-ST449, ST414, ST30, O126:H27-ST200 & SLV, ST155, O166:H15-ST349 & SLV/DLV, ST130, ST394,). There were no mutually exclusive ST and serotypes found in the EAEC population structure (Table S1).

EAEC complexes associated with disease and carriage

The population structure of EAEC was heterogeneous containing 17 complexes (either single ST or complexes) of successful lineages containing 4 or more EAEC (Figure 1, Table 2).

There was a 2.71 ratio of case isolates to controls in this study. Complexes with a higher ratio in cases were deemed associated with cases and complexes with a higher ratio in controls were deemed associated with controls, complexes that were below this ratio were deemed to be not associated with cases or controls. This resulted in eleven complexes being associated with disease (ST10, 30, 40, 155, 165, 278, 501, 678, 720, 746 and 1891, Cplx), two complexes associated with carriage (ST31 and 349 Cplx) and four complexes neither associated with disease or carriage (ST,38, 168, 295 and 394 Cplx).

The disease complexes and carriage complexes were combined and statistical analysis showed both of the disease and carriage complexes were statistically significant ($P=<0.001$ and $P=0.001$ respectively) (Table 2).

Individual complexes were then tested for statistical association with disease or carriage which showed ST10 Cplx and ST40 Cplx were independently statistically significantly ($P=0.01$ & 0.03 respectively) associated with disease. ST31 was independently statistically significantly (Fishers chi-square, $p=0.005$) associated with carriage (due to the fact that there was a higher ratio of controls).

Situating the 17 successful EAEC complexes identified in this study within the global *E. coli* phylogeny as represented in the public database (Table 3) showed that with the exception of ST155 Cplx, all complexes were significantly associated with being EAEC pathotype ($P\leq0.01$).

Evolutionary Events leading to successful EAEC disease complexes

ClonalFrame analysis showed that EAEC mutation and recombination rates varied across the complexes and Countries (Table 4 & 5). Complex ST10 Cplx had the highest mutation rate (4.05) and recombination rate (1.2) whereas ST295 Cplx the lowest mutation rate (0.02) and lowest recombination rate (0.002). However, both of these complexes had a similar mutation to recombination ratio. Recombination had the greatest impact (on the diversification of the lineages) on ST40 Cplx (12) and ST394 Cplx (10). Recombination occurred 1.7 times more often than mutation rate among isolates from Bangladesh and Nigeria whereas among strains isolated in the UK, recombination and mutation rate was almost equal. The entire dataset recombination events occurred 1.3 times more often than mutational events.

Table 2. Assessment of EAEC complexes associated with cases or controls.

Group	ST complex	UK	Nigeria	Bangladesh	Case	Control	Total	Total % of EAEC	CASE: CONTROL %	P Value
Group 1	10	128	24	21	138	35	173	30.7	8020	0.01
Group 2	40	39	1	12	44	8	52	9.2	8515	0.03
Group 3	31	27	11	12	28	22	50	8.9	5644	0.005
Group 4	295	13	2	21	24	12	36	6.4	6733	0.24
Group 5	38	3	4	21	19	9	28	5.0	6832	0.33
Group 6	394	9	10	8	20	7	27	4.8	7426	0.56
Group 7	746	9	1	1	10	1	11	2.0	9010	0.16
Group 8	155	0	1	9	9	1	10	1.8	9010	0.2
Group 9	678	8	0	2	9	1	10	1.8	9010	0.2
Group 10	278	7	1	2	9	1	10	1.8	9010	0.2
Group 11	168 (ST484)	0	4	5	5	4	9	1.6	5644	0.2
Group 12	30	7	0	0	8	0	8	1.4	1000	0.08
Group 13	165	3	0	5	7	1	8	1.4	8317	0.32
Group 14	1891	0	0	5	4	1	5	0.9	8020	0.59
Group 15	720	0	0	5	5	0	5	0.9	1000	0.21
Group 16	501	2	2	0	3	1	4	0.7	7525	0.71
Group 17	349	0	1	3	1	3	4	0.7	2575	0.06
Totals	-	248	62	132	343	107	442	-	-	-
Whole Data Set	-	273	121	169	412	152	564	-	-	-

Assessment of the successful EAEC complexes (>4 strains) as to the association with cases or controls and showing the data of EAEC numbers according to complex size, Country and association with case or control. Groups are in order of complex size from the largest to smallest. Probability (Fishers exact test) of the group being significantly associated with case or control is tabulated at the end.

doi:10.1371/journal.pone.0112967.t002

The geographical location of the place of isolation of an EAEC strain bears no significance in its phylogeny grouping (with the exception of small geographical specific STs possibly due to sampling bias) and successful EAEC ST were distributed globally (Figure S2) The impact of recombination in the diversification of the sample set relative to mutation showed the greatest impact in the Bangladesh strain set, and the least impact in the strains from the UK. This data suggest that recombination may play an important role in the evolution of EAEC (Table 4 & 5).

External to Internal Branch Length Ratio gave coalescent expectations indicating that all EAEC irrespective of location and including the entire dataset were significantly different ($p=<0.001$) from the inferred value (Table 4).

Evolution of EAEC in the context of the *E. coli* population

Of the five main branches of *E. coli* phylogeny, EAEC are most prominent on branches 1, 2 and 3 (Figure 2) consisting of phylogroups D, A and B1 respectively. ST30, 31, 38, and 394 Cplxs which are grouped together by MLST population structure (Figure 1) are all located on branch 1 of the *E. coli* phylogeny. The other large successful complexes are dispersed throughout branch 2 and 3. ST10 Cplx shows that some SLVs on the MLST structure are separate in the context of the

Table 3. Assessment of EAEC associated with other pathotypes.

Group	ST complex	EAEC (This study)	EAEC Public	EPEC	ETEC	STEC	EIEC	DAEC	Commensal	No Pathotype	Other pathotypes	Total DEC	Total E.coli/ Inc. EAEC	Other E.coli/ total	%EAEC: DEC	%EAEC: E.coli	Total EAEC	P value
Group 1	10	149	42	17	22	4	0	0	5	141	83	234	463	272	81.6	41.3	191	<0.001
Group 2	40	51	8	4	0	3	0	0	0	2	0	66	68	9	89.4	86.8	59	<0.001
Group 3	31	39	19	0	0	0	0	0	0	6	8	58	72	14	100.0	80.6	58	<0.001
Group 4	295	34	1	3	0	0	0	0	1	0	2	38	41	6	92.1	85.4	35	<0.001
Group 5	38	24	4	0	0	1	0	0	0	10	27	29	66	38	96.6	42.4	28	<0.001
Group 6	394	17	11	0	0	0	0	0	0	3	2	28	33	5	100.0	84.8	28	<0.001
Group 7	746	10	0	0	4	0	0	0	0	0	1	14	15	5	71.4	66.7	10	<0.001
Group 8	155	9	2	1	3	1	1	0	2	27	22	17	68	57	64.7	16.2	11	0.11
Group 9	678	10	0	0	0	0	0	0	0	1	0	10	11	1	100.0	90.9	10	<0.001
Group 10	278	9	0	0	0	0	1	0	0	0	0	10	10	1	90.0	90.0	9	<0.001
Group 11	168 (ST484)	5	4	0	0	0	1	2	0	10	8	12	30	21	75.0	30.0	9	0.003
Group 12	30	8	2	1	0	0	0	0	0	0	0	11	11	1	90.9	90.9	10	<0.001
Group 13	165	8	0	1	7	3	0	0	0	6	1	19	26	18	42.1	30.8	8	0.005
Group 14	1891	5	0	0	0	1	0	0	0	0	0	6	6	1	83.3	83.3	5	<0.001
Group 15	720	5	0	0	0	1	0	0	0	0	1	6	7	2	83.3	71.4	5	<0.001
Group 16	501	2	1	0	0	0	0	0	0	0	0	3	3	0	100.0	100.0	3	<0.001
Group 17	349	3	1	2	0	0	0	0	1	0	2	6	9	5	66.7	44.4	4	0.01

Assessment of the successful EAEC complexes (>4 strains), as to the association with the complexes being associated with EAEC or other *E. coli* pathotypes in the public database (all data from 18.12.2013) including commensal, diarrhoeagenic and extra-intestinal *E. coli*. Nigerian dataset is included under the public database, UK and Bangladesh dataset is included under EAEC (This study). See methods for description of pathotypes included. Total EAEC included is 598 strains (443 from this study plus 155 EAEC from public database strains), other *E. coli* total is 6076 strains (6674 minus 598 EAEC and minus 141 *Shigella* isolates included in the public database). Probability (Fishers exact test) of the group being significantly associated with EAEC or other pathotypes is tabulated at the end.

doi:10.1371/journal.pone.0112967.t003

Table 4. Mutation and Recombination rates of dataset by geographical source and all Sequence types found in dataset.

Parameters	Bangladesh N=169 (108 Cases, 61 Controls)	Nigeria N=121 (66 cases, 55 controls)	UK N=254 (228 cases, 36 controls)	All ST N=199 (138 cases, 61 controls)
Mutation Rate (theta 0) Mutational rate & assumed to be constant on the branches of topology	mean: 15.03, credibility_region: 6.95–26.14	mean: 120.79, credibility_region: 69.29–33.00	mean: 70.13, credibility_region: 49.35–94.01	mean: 16.01, credibility_region: 8.64–23.71
Recombination rate (R) recombination rate & assumed constant on branches of topology	mean: 22.58, credibility_region: 14.05–33.46	mean: 31.38, credibility_region: 19.68–43.37	mean: 15.66, credibility_region: 9.84–22.31	mean: 89.53, credibility_region 64.21–121.96
view rho over theta (p/0) How often recombination occurs relative to mutations	mean: 1.65, credibility_region: 0.77–3.14	mean: 1.68, credibility_region: 0.78–3.80	mean: 1.048907, credibility_region: 0.50–1.987	mean: 1.317856, credibility_region: 0.76–2.07
view r over m (r/m) The impact of how important the effect of recombination was in the diversification of the sample relative to mutation	mean: 4.38, credibility_region: 2.38–8.05	mean: 4.10, credibility_region: 2.13–8.09	mean: 2.60, credibility_region: 1.44–4.39	mean: 2.87, credibility_region 1.94–4.24
External to Internal Branch Length Ratio Gives the inferred expected values against the coalescent and actual rations. It they are significantly apart then it shows there was a genetic event such as recombination that led to these values.	mean: 0.73, interval: 0.54–0.94 **Significance: 0.00**	mean: 0.56, interval: 0.40–0.76 **Significance: 0.01**	mean: 0.67, interval: 0.50–0.88 **Significance: 0.00**	mean: 0.90, interval: 0.72–1.06 **Significance: 0.00**

ClonalFrame mutation and recombination rates shown as well the impact of recombination over mutation in the diversification of the data and also the significance of the expected value over the inferred value as to whether the data evolved over a period of time (not significant) or due to a large genetic event (significant). This analysis was applied to the different geographical locations, and all 564 EAEC ST found in this study.

doi:10.1371/journal.pone.0112967.t004

E. coli phylogeny though still closely related. ST295 Cplx which is linked to ST10 Cplx by ST48 is on the opposite end of branch 2 and therefore evolutionary distant. The smaller successful complexes with only 4 EAEC were found at the end of branch 4 which contained a mixture of phylogroups A and D. None of the main EAEC

Table 5. Mutation and Recombination rates of dataset by ST complex.

Parameters	ST10 Cplx & DLV	ST38 Cplx & DLV	ST40 Cplx & DLV	ST295Cplx & DLV	ST394Cplx & DLV	ST31 & ST 130Cplx & DLV
Mutation Rate (theta 0)	mean: 4.04, credibility_region: 2.097–6.31	mean: 0.28, credibility_region: 0.02–1.00	mean: 0.94, credibility_region: 0.02–2.62	mean: 0.02, credibility_region: 0.00–1.87	mean: 0.23, credibility_region: 0.00–1.00	mean: 0.65, credibility_region:0.13–1.48
Recombination rate (R)	mean: 1.24, credibility_region: 0.41–2.84	mean: 0.08, credibility_region: 0.00–0.38	mean: 0.61, credibility_region: 0.00–1.90	mean: 0.00, credibility_region: 0.00–0.01	mean: 0.10, credibility_region: 0.00–0.46	mean: 0.37, credibility_region: 0.03–0.97
view rho over theta (p/0)	mean: 0.33, credibility_region: 0.09–0.82	mean: 0.68, credibility_region: 0.00–3.60	mean: 5.55, credibility_region: 0.00–46.86	mean: 0.57, credibility_region: 0.00–4.49	mean: 4.07, credibility_region: 0.00–33.12	mean: 1.07, credibility_region: 0.04–5.63
view r over m (r/m)	mean: 1.20, credibility_region: 0.39–2.66	mean: 3.55, credibility_region: 0.01–19.63	mean: 12.00, credibility_region: 0.00–102.35	mean: 0.91, credibility_region: 0.00–7.04	mean: 10.39, credibility_region: 0.00–74.56	mean: 4.27, credibility_region: 0.24–20.06
External to Internal Branch Length Ratio	mean: 0.48, interval:0.28–0.72 **Significance:0.02**	mean: 0.77, interval:0.30–1.51 **Significance:0.15**	mean: 0.64, interval:0.30–1.20 **Significance:0.09**	mean: 0.64, interval:0.29–1.25 **Significance:0.15**	mean: 0.64, interval:0.23–1.32 **Significance:0.24**	mean: 0.56, interval:0.27–1.143 **Significance:0.12**

ClonalFrame mutation and recombination rates shown as well the impact of recombination over mutation in the diversification of the data and also the significance of the expected value over the inferred value as to whether the data evolved over a period of time (not significant) or due to a large genetic event (significant). This analysis was applied to the large main complexes including single locus variants (SLV) and double locus variants (DLV).

doi:10.1371/journal.pone.0112967.t005

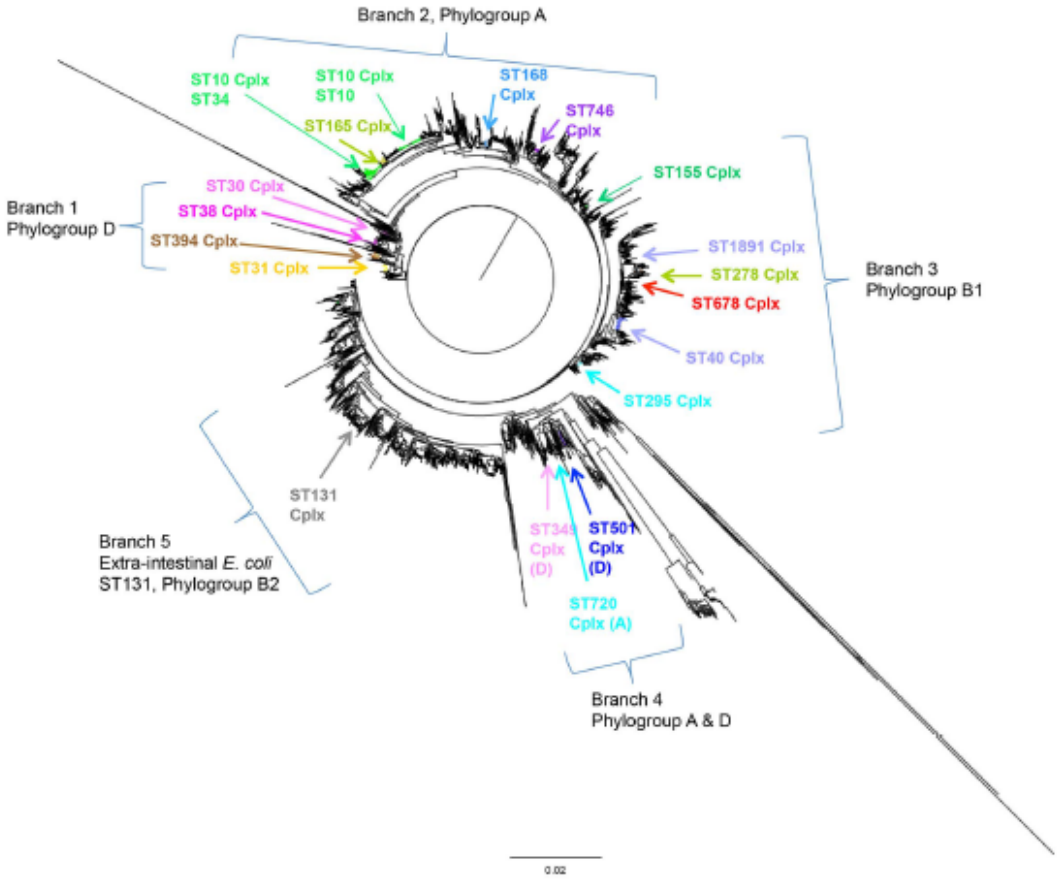


Figure 2. Neighbour joining tree of all *E. coli* and enteroaggregative *E. coli* in this study. Neighbour joining tree of concatenated MLVA of the 564 EAEC used in this and all ST across the *E. coli* population structure. Phylogeny is separated into four main branches. EAEC is distributed throughout the *E. coli* phylogeny as shown in branches 1–4 containing phylogroups, A, B1 and D. The main EAEC complexes was not found in branch 5, phylogroup B2 associated with extra-intestinal infections.

doi:10.1371/journal.pone.0112967.g002

complexes (Table 2) were found in branch 5 of the *E coli* phylogeny which is generally associated with extra-intestinal infections such as ST131 belonging to phylogroup B2.

Discussion

Serotyping does not always correlate with genetic relatedness and cannot be used to infer genetic background

Although there were serotypes exclusively from cases (O3:H2, O44:H18, O104:H4, O111:H21, O126:H27and O134:H27), In this study we found no link between a

sequence type and a single serotype Although some serotypes were associated with single clonal complexes, they were not mutually exclusive and high recombination rates in some lineages meant that a given serotype could also be distributed in different complexes (Figure S1, Table S1).

Since the development of sequence based typing, such as MLST, the use of traditional typing methods, such as serotyping as a means of population structure [16,17] have come under close scrutiny. Other studies have also shown that the same serogroups are found in genetically unrelated strains of *E. coli* indicating possible horizontal gene transfer [18] of the cassette encoding the serogroup genes. In this study we were looking for lineages of EAEC and so we used MLST as the primary typing method. and we conclude, as others have, that serotyping is not a suitable method for determining ancestral relatedness of EAEC.

There are successful multiple lineages of EAEC complexes that are globally distributed

We have shown a statistically significant association of certain sequence type complexes of enteroaggregative *E. coli* with disease or carriage. These complexes represent independent lineages which were spread throughout the entire *E. coli* population (Figure 2) and included the EAEC published complexes in the public database: ST10 Cplx, ST40 Cplx, ST38 Cplx, ST394 Cplx and ST349 Cplx [7]. Prototypical EAEC strains 042 (from Peru) and 17–2 (from Chile) belong to ST31 Cplx and ST10 Cplx respectively, which were prominent in this study. This study also identified MLST complexes that were not currently represented in the public database as associated with the aggregative phenotype including ST130 Cplx, ST295 Cplx, ST484 Cplx, ST678 and ST720 Cplx. This data represents a snapshot of EAEC, from three different countries, and the addition of strains across the globe will expand the number STs associated with EAEC. It should be noted that the public database is biased towards *E. coli* of clinical interest such as pathogenic and antibiotic resistant strains with little representation of commensal strains and it is likely that not all isolates were tested for the aggregative phenotype. A larger, better defined, population of *E. coli* as a whole is needed to comprehensively define the distribution of EAEC in MLST complexes.

Although there are some MLST complexes/STs restricted to one country, these contain small numbers and all of the complexes with larger numbers of isolates are distributed throughout the phylogeny indicating a global distribution of the major clusters (Figure S2) most likely due to human travel. The independent appearance of the EAEC phenotype in discrete complexes across phylogeny (homoplasy), supports the observation of others [19] and suggests convergent evolution - the EAEC phenotype therefore confers a biological advantage in certain bacterial genetic backgrounds.

Multiple genetic events have led to the independent evolution of EAEC

In order to understand the genetic events which led to the formation of different EAEC associated MLST complexes Clonal Frame analysis of the branching events for each node was carried out. Variation in the frequency of recombination or mutation which occurred in all of the seven loci at different time points was seen indicating multiple genetic events over time. The relative frequency of recombination as compared to mutation (ρ/θ) for the entire data set was 1.31 and is comparable to the rates proposed by Wirth et al [11] and Touchon et al [20] but higher than computed rates for the *E. coli* species via MLST including those that estimated recombination at approaching zero [21].

The parameters of rates and impact are based on the Markov model [22] which assumes that horizontal gene transfer events are equally probable between any pair of lineages, irrespective of phylogenetic and ecological proximity [23]. Our analysis clearly showed that this isn't the case and that (in this dataset) recombination rates vary within the EAEC pathotype between different lineages, the most ancestral being ST10 Cplx with the least impact of recombination in comparison to the other lineages (Table 5).

Multiple successful complexes (Figure 1) vary in mutation and recombination rate (Table 4) and are distributed throughout the *E. coli* population (Figure 2). These complexes have clearly evolved independently through multiple genetic events that have led to the phenotypic congruency of this pathotype. The selection of strains with a biological advantage has resulted in different, apparent, mutation/recombination rates suggests that certain bacterial backgrounds allow the advantage to be expressed - possibly influenced by the ability to retain the EAEC plasmid. Fast radiation of the complexes after population bottlenecks and frequent recombination seems a likely explanation for this pattern [11]. This may explain why the main gastrointestinal EAEC complexes were not found in the extra-intestinal *E. coli* phylogeny branch.

Evolutionary events of EAEC

Although EAEC strains share the common phenotype of aggregative adherence, this and earlier research (Okeke et al 2010) demonstrates that the phenotype is convergent - has arisen in different lineages and been selected by survival in the human host. The selective advantage of aggregative adherence would allow *EAEC* strains to colonize the human gut during episodes of diarrhoea from other causes Lineages of EAEC found to be non-pathogenic are possibly strains that have developed exceptional colonization ability but not the ability to actually cause disease. Other lineages however, are associated with the ability to cause disease. Outbreak investigations and the strong association of some lineages with disease in this study point to multiple EAEC, but distinct, lineages that cause disease. Distinct sub-populations within a species may emerge because of differential local adaptation or genetic drift [14]. This concept may be applied to successful EAEC complexes which represent clusters of closely related genotypes and can be termed

ecotypes [24] and will differ in their homologous recombination events because of adaptive evolution or environmental constraints [14]. This is supported by the variable recombination rate in different complexes which may have evolved from different environments. The variable recombination rate from each country will depend on the complexes found from the sample size tested. For EAEC isolates from UK residents the low impact of recombination may be because EAEC infection is related to travel and would therefore include EAEC found in multiple countries.

Virulent pathotypes have been shown to recombine more than non-pathogens pointing towards the theory that that virulence is the driving force for more frequent recombination [11]. This is shown with ST40 Cplx which is statistically associated with disease ($p=0.03$) and had the highest impact of recombination on diversification. However ST10 Cplx, also statistically associated with disease ($p=0.01$), had the highest rate of mutation among the complexes and the impact of recombination was almost equal to mutation (11.2). This indicates that both types of genetic events are important in the evolution of pathogenic EAEC but that local variation occurs.

Our data analysis of the concatenated MLST sequences showed that the external to internal branch length ratio of the phylogeny was significantly higher than expected (Table 4). This means that the inferred genealogy is consistent with an expansion of the population size by acquisition of a fitness advantage early in the history of the sample [22]. For example, one suggestion is that the ancestral ST10 Cplx already had the background mutations to be able to acquire and retain the EAEC plasmid and so the external to internal branch length ratio is as expected. This fits in with previous studies where a specific genetic background is required to acquire and express virulence factors in *E. coli* [25]. Other complexes with unexpected external to internal branch length ratio, such as ST40 Cplx, needed recombination and/or mutation events to allow the stable retention of the advantageous EAEC plasmid. A recently reported example of how acquisition of this EAEC plasmid can increase fitness is the ST678 (O104) VTEC German outbreak [26]. This is a VTEC strain that didn't have the characteristic *eae* gene (attachment and effacement loci for intimate adherence) but did have the plasmid encoded *aat* gene cluster associated with adherence. This strain was particularly virulent, with high HUS rates, but had the same toxin type as many other VTEC strains, the difference, presumably, being its strong ability to adhere and hence introduce more toxin. This basic mechanism of attachment could be the fitness advantage that this relatively new pathotype, EAEC, has harboured and then successfully expanded.

Conclusions

This study has clearly shown the complexity of the evolution of EAEC, while it is evident that the same lineages prevail in multiple global locations, indicative of clonal expansion, whilst other lineages are ecologically adapting through a process

of convergent evolution. This would account for the inconsistent impact rates of recombination between different geographical locations and different complexes. The collection of organisms given the "pathotype" EAEC has evolved as multiple independent lineages with some complexes associated with disease, but not all. This is important as a non-disease causing EAEC still has the ability to acquire other virulence factors and the combination of aggregative adherence and virulence can cause severe outbreaks. The presence of the *aggR* genes as an indicator of aggregative adherence ability should therefore be considered when diagnosing gastrointestinal disease.

Supporting Information

Figure S1. Minimal Spanning Tree of 443 enteroaggregative *E. coli* serotyped. Minimum Spanning Tree of 443 EAEC serotyped in this study. Tree is colour coded by serotypes containing 3 or more isolates. Serotypes shown in one or two strains were coloured white. Complexes shaded in grey consist of single locus variants (SLV). Sequence types are labelled as numbers.
doi:10.1371/journal.pone.0112967.s001 (TIF)

Figure S2. MSTree Geographical location. Minimal spanning tree of the 564 EAEC used in this study colour coded by isolates from Bangladesh (red), Nigeria (purple) and UK (green) and the prototypical O42 strain from Peru (yellow). Complexes shaded in grey consist of single locus variants (SLV). Trees shows that complexes are mainly distrusted in at least two countries with only a few small complexes and singletons geographically specific. Sequence types and complex (Cplx) are labelled as numbers.
doi:10.1371/journal.pone.0112967.s002 (TIF)

Table S1. Strain list used in this study. Table of strains used in this study listing the year the strain was isolated, the Country the strain was isolated from, somatic and flagella typing results (serotyping), sequence type and complex the strain belongs to. NT: Not tested, Novel sequence types consisted of either single locus variants (SLV), double locus variants (DLV) or triple locus variants (TLV) of known sequence types.
doi:10.1371/journal.pone.0112967.s003 (PDF)

Acknowledgments

Thank you to Mark Achtman, University of Warwick, UK and his group for development, management and access and to the public database as well as the scientists who have submitted their data on multiple studies. Acknowledgements to the scientists involved in the case control studies, diagnostic and reference laboratories.
Thank you to Dawn Hedges and Martha Valencia for Serotyping and Yoshini Taylor and Vivienne DoNascimento for identification at the Gastrointestinal Bacteria Reference Unit (GBRU), PHE. Thank you to Michaela Day for help with

the phylotyping and PCR at the Antimicrobial Resistance and Healthcare Associated Infections Reference Unit, PHE.
This research was undertaken at Public Health England, UK and University of Haverford, USA.

Author Contributions

Conceived and designed the experiments: JW CJ INO. Performed the experiments: MAC DR. Analyzed the data: MAC. Contributed reagents/materials/analysis tools: AC KAT TD AU SP. Wrote the paper: MAC CJ DR AC KAT TD AU SP INO JW.

References

1. **Nataro JP, Kaper JB, Robins-Browne R, Prado V, Vial P, et al.** Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. Pediatr Infect Dis J. 1987;6:829–831.
2. **Baudry B, Savarino SJ, Vial P, Kaper JB, Levine MM.** A sensitive and specific DNA probe to identify enteroaggregative Escherichia coli, a recently discovered diarrheal pathogen. J Infect Dis. 1990;161:1249–1251.
3. **Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V, et al.** The export of coat protein from enteroaggregative Escherichia coli by a specific ATP-binding cassette transporter system. J Biol Chem. 2003;278:45680-45689. 10.1074/jbc.M306413200 [doi];M306413200 [pii].
4. **Nataro JP, Yikang D, Yingkang D, Walker K.** AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative Escherichia coli. J Bacteriol. 1994;176:4691–4699.
5. **Chattaway M, Harris R, Jenkins C, Tam C, Coia J, et al.** Investigating the link between the presence of enteroaggregative Escherichia coli and infectious intestinal disease in the United Kingdom, 1993 to 1996 and 2008 to 2009. Euro Surveill. 2013.18.
6. **Jenkins C, Chart H, Willshaw GA, Cheasty T, Smith HR.** Genotyping of enteroaggregative Escherichia coli and identification of target genes for the detection of both typical and atypical strains. Diagn Microbiol Infect Dis. 2006;55:13–19.
7. **Okeke IN, Wallace-Gadsden F, Simons HR, Matthews N, Labar AS, et al.** Multi-locus sequence typing of enteroaggregative Escherichia coli isolates from Nigerian children uncovers multiple lineages. PLoS One. 2010;5:e14093. 10.1371/journal.pone.0014093 [doi].
8. **Achtman M, Wain J, Weill FX, Nair S, Zhou Z, et al.** Multilocus sequence typing as a replacement for serotyping in Salmonella enterica. PLoS Pathog. 2012;8:e1002776. 10.1371/journal.ppat.1002776 [doi];PPATHOGENS-D-12-00064 [pii].
9. **Castellani A, Chalmers AJ.** Family I. Enterobacteriaceae. Genus I. *Escherichia*. In: Scheutz F, Strockbine NA, , editors. Bergy's Manual of Systematic Bacteriology.USA: Springer. pp. 2005.607–624.
10. **Gross RJ, Rowe B.** Serotyping of Escherichia coli. In: Sussman M, editors. The virulence of Escherichia coli.Cambridge: CambridgeUniversity Press. pp. 1985.345–360.
11. **Wirth T, Falush D, Lan R, Colles F, Mensa P, et al.** Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol. 2006;60:1136–1151.
12. **Fisher RA.** On the interpretation of X^2 from contingency tables, and the calculation of P. J Roy Stat Soc 1922;85:87–94.
13. **Didelot X, Falush D.** Inference of bacterial microevolution using multilocus sequence data. Genetics. 2007;175:1251–1266. genetics.106.063305 [pii];10.1534/genetics.106.063305 [doi].
14. **Vos M, Didelot X.** A comparison of homologous recombination rates in bacteria and archaea. ISME J. 2009;3:199–208. ismej200893 [pii];10.1038/ismej.2008.93 [doi].

15. **Doumith M, Day MJ, Hope R, Wain J, Woodford N.** Improved multiplex PCR strategy for rapid assignment of the four major Escherichia coli phylogenetic groups. J Clin Microbiol. 2012;50:3108–3110. JCM.01468–12 [pii];10.1128/JCM.01468-12 [doi].

16. **Hartl DL, Dykhuizen DE.** The population genetics of Escherichia coli. Annu Rev Genet. 1984;18:31–68. 10.1146/annurev.ge.18.120184.000335 [doi].

17. **Tenaillon O, Skurnik D, Picard B, Denamur E.** The population genetics of commensal Escherichia coli. Nat Rev Microbiol. 2010;8:207–217. nrmicro2298 [pii];10.1038/nrmicro2298 [doi].

18. **Beutin L, Tao J, Feng L, Krause G, Zimmermann S, et al.** Sequence analysis of the Escherichia coli O15 antigen gene cluster and development of a PCR assay for rapid detection of intestinal and extraintestinal pathogenic E. coli O15 strains. J Clin Microbiol. 2005;43:703–710. 43/2/703 [pii];10.1128/JCM.43.2.703-710.2005 [doi].

19. **Czeczulin JR, Whittam TS, Henderson IR, Navarro-Garcia F, Nataro JP.** Phylogenetic analysis of enteroaggregative and diffusely adherent Escherichia coli. Infect Immun. 1999;67:2692–2699.

20. **Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, et al.** Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet. 2009;5:e1000344. 10.1371/journal.pgen.1000344 [doi].

21. **Perez-Losada M, Browne EB, Madsen A, Wirth T, Viscidi RP, et al.** Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect Genet Evol. 2006;6:97–112. S1567-1348(05)00027-4 [pii];10.1016/j.meegid.2005.02.003 [doi].

22. **Didelot X, Falush D.** ClonalFrame User Guide Version 2006;1.1. 1–33.

23. **Galtier N.** A model of horizontal gene transfer and the bacterial phylogeny problem. Syst Biol. 2007;56:633–642. 780913211 [pii];10.1080/10635150701546231 [doi].

24. **Cohan FM.** What are bacterial species? Annu Rev Microbiol. 2002;56:457–487. 10.1146/annurev.micro.56.012302.160634 [doi];012302.160634 [pii].

25. **Escobar-Paramo P, Clermont O, Blanc-Potard AB, Bui H, Le BC, et al.** A specific genetic background is required for acquisition and expression of virulence factors in Escherichia coli. Mol Biol Evol. 2004;21:1085–1094. 10.1093/molbev/msh118 [doi];msh118 [pii].

26. **Chattaway MA, Dallman T, Okeke IN, Wain J.** Enteroaggregative E. coli O104 from an outbreak of HUS in Germany 2011, could it happen again? J Infect Dev Ctries. 2011;5:425–436.

27. **Nataro JP, Baldini MM, Kaper JB, Black RE, Bravo N, et al.** Detection of an adherence factor of enteropathogenic Escherichia coli with a DNA probe. J Infect Dis. 1985;152:560–565.

28. **Wilson A, Evans J, Chart H, Cheasty T, Wheeler JG, et al.** Characterisation of strains of enteroaggregative Escherichia coli isolated during the infectious intestinal disease study in England. Eur J Epidemiol. 2001;17:1125–1130.

29. **Spencer J, Smith HR, Chart H.** Characterization of enteroaggregative Escherichia coli isolated from outbreaks of diarrhoeal disease in England. Epidemiol Infect. 1999;123:413–421.

30. **Panchalingam S, Antonio M, Hossain A, Mandomando I, Ochieng B, et al.** Diagnostic microbiologic methods in the GEMS-1 case/control study. Clin Infect Dis. 2012;55 Suppl 4::S294–S302. cis754 [pii];10.1093/cid/cis754 [doi].

31. **Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, et al.** The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis. 2012;55 Suppl 4::S232–S245. cis753 [pii];10.1093/cid/cis753 [doi].

32. **Dallman T, Smith GP, O'Brien B, Chattaway MA, Finlay D, et al.** Characterization of a verocytotoxin-producing enteroaggregative Escherichia coli serogroup O111:H21 strain associated with a household outbreak in Northern Ireland. J Clin Microbiol. 2012;50:4116–4119. JCM.02047-12 [pii];10.1128/JCM.02047-12 [doi].

33. **Dallman TJ, Chattaway MA, Cowley LA, Doumith M, Tewolde R, et al.** An Investigation of the Diversity of Strains of Enteroaggregative Escherichia coli Isolated from Cases Associated with a Large Multi-Pathogen Foodborne Outbreak in the UK. PLoS One. 2014;9:e98103. 10.1371/journal.pone.0098103 [doi];PONE-D-13-54960 [pii].

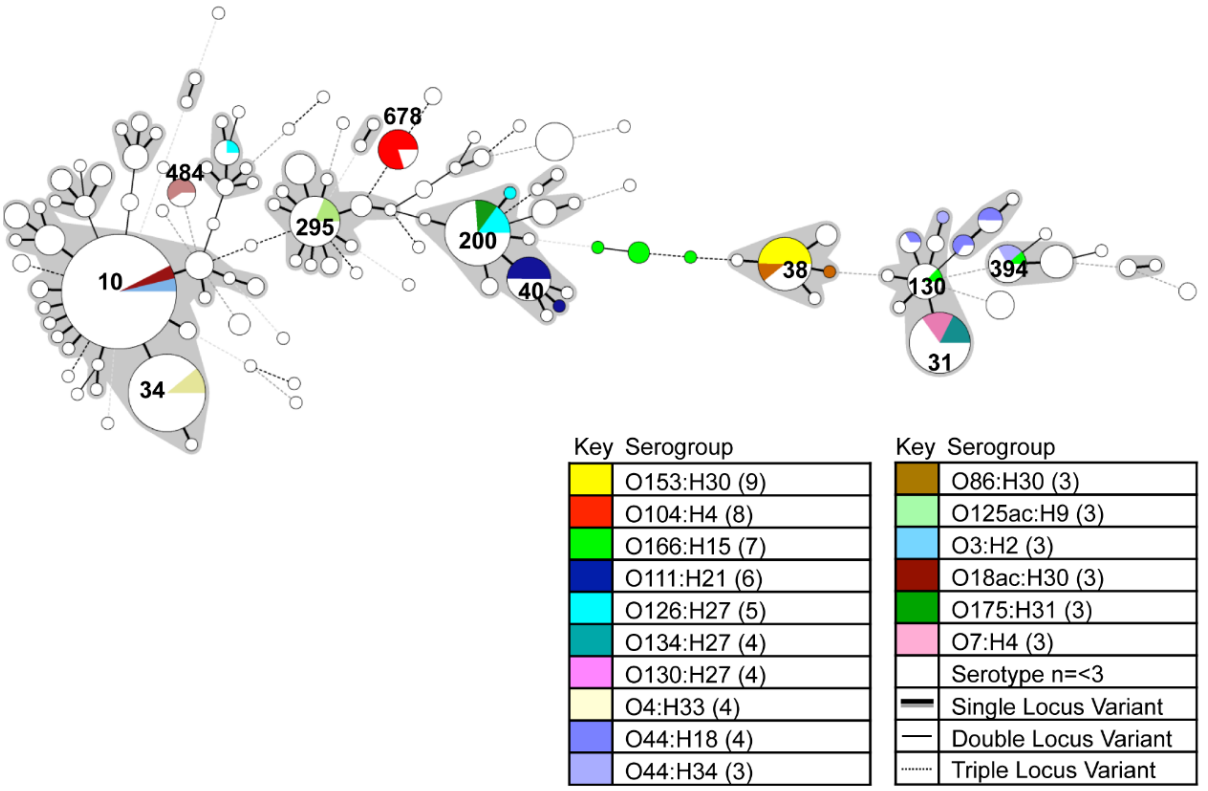
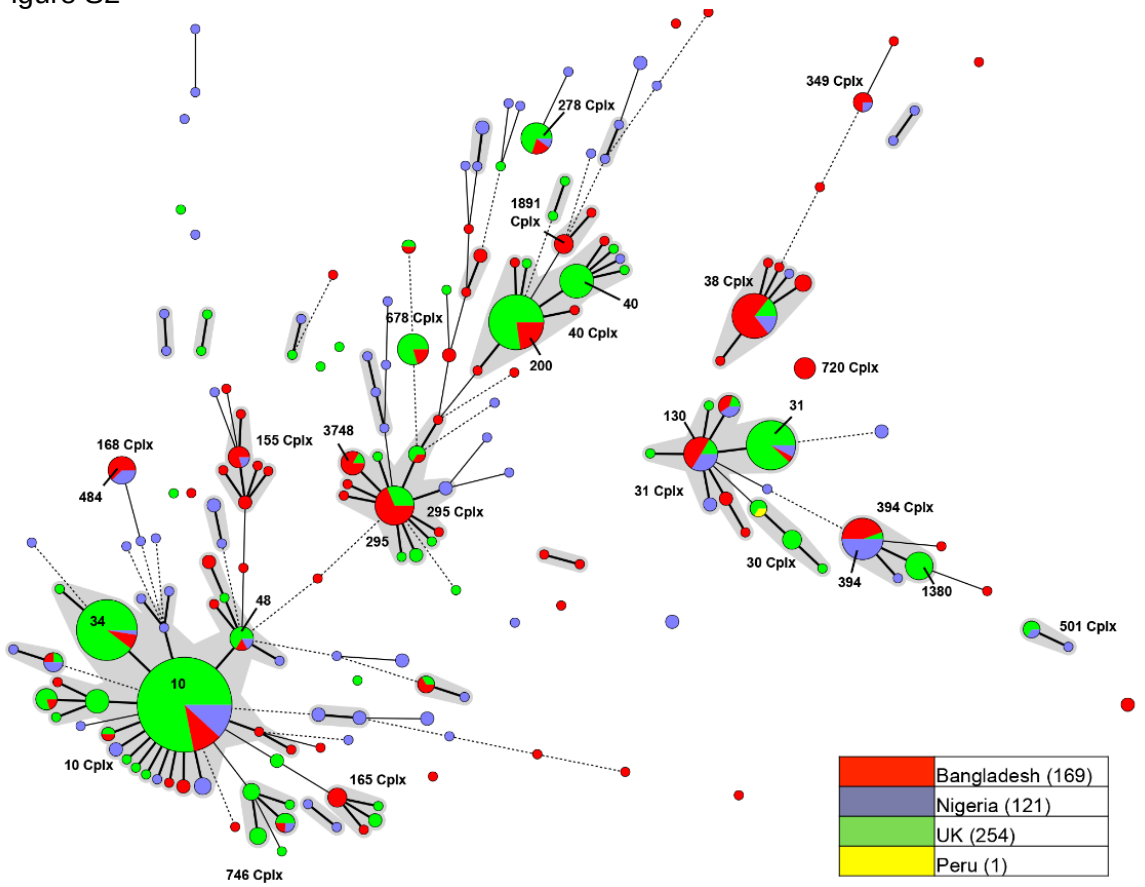

678
484
295
200
10
40
38
130
394
34
31
Key Serogroup
O153:H30 (9)
O104:H4 (8)
O166:H15 (7)
O111:H21 (6)
O126:H27 (5)
O134:H27 (4)
O130:H27 (4)
O4:H33 (4)
O44:H18 (4)
O44:H34 (3)
Key Serogroup
O86:H30 (3)
O125ac:H9 (3)
O3:H2 (3)
O18ac:H30 (3)
O175:H31 (3)
O7:H4 (3)
Serotype n=<3
Single Locus Variant
Double Locus Variant
Triple Locus Variant

Figure S1

Figure S2


278 Cplx
349 Cplx
1891 Cplx
40
40 Cplx
200
678 Cplx
38 Cplx
720 Cplx
130
31
31 Cplx
394 Cplx
30 Cplx
394
1380
501 Cplx
168 Cplx
484
155 Cplx
3748
295 Cplx
295
48
34
10
10 Cplx
165 Cplx
746 Cplx
Bangladesh (169)
Nigeria (121)
UK (254)
Peru (1)

7.16 Use of whole-genus genome sequence data to develop a multilocus sequence typing tool that accurately identifies Yersinia isolates to the species and subspecies levels

Hall, M., Chattaway, M.A., Reuter, S., Savin, C., Strauch, E., Carniel, E., Connor, T., Van, D., I, Rajakaruna, L., Rajendram, D., Jenkins, C., Thomson, N.R., & McNally, A. J.Clin.Microbiol. 53, (1) 35-42 available from: PM:25339391

Use of Whole-Genus Genome Sequence Data To Develop a Multilocus Sequence Typing Tool That Accurately Identifies *Yersinia* Isolates to the Species and Subspecies Levels

Miquette Hall,[a] Marie A. Chattaway,[b] Sandra Reuter,[c] Cyril Savin,[d] Eckhard Strauch,[e] Elisabeth Carniel,[d] Thomas Connor,[f] Inge Van Damme,[g] Lakshani Rajakaruna,[b] Dunstan Rajendram,[h] Claire Jenkins,[b] Nicholas R. Thomson,[c] Alan McNally[a]

Pathogen Research Group, Nottingham Trent University, Nottingham, United Kingdom[a]; Gastrointestinal Bacteria Reference Unit, Public Health England, Colindale, London, United Kingdom[b]; Pathogen Genomics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom[c]; French *Yersinia* Reference Laboratory and WHO Collaborating Centre, Institut Pasteur, Paris, France[d]; Bundesinstitut für Risikobewertung, Berlin, Germany[e]; Cardiff School of Biosciences, Cardiff University, Cardiff, United Kingdom[f]; Department of Veterinary Public Health and Food Safety, University Ghent, Ghent, Belgium[g]; Genomic Service Unit, Public Health England, London, United Kingdom[h]

The genus *Yersinia* is a large and diverse bacterial genus consisting of human-pathogenic species, a fish-pathogenic species, and a large number of environmental species. Recently, the phylogenetic and population structure of the entire genus was elucidated through the genome sequence data of 241 strains encompassing every known species in the genus. Here we report the mining of this enormous data set to create a multilocus sequence typing-based scheme that can identify *Yersinia* strains to the species level to a level of resolution equal to that for whole-genome sequencing. Our assay is designed to be able to accurately subtype the important human-pathogenic species *Yersinia enterocolitica* to whole-genome resolution levels. We also report the validation of the scheme on 386 strains from reference laboratory collections across Europe. We propose that the scheme is an important molecular typing system to allow accurate and reproducible identification of *Yersinia* isolates to the species level, a process often inconsistent in nonspecialist laboratories. Additionally, our assay is the most phylogenetically informative typing scheme available for *Y. enterocolitica*.

The Gram-negative *Yersinia* is one of the most important and well-studied bacterial genera, consisting of three human pathogens. *Y. pestis* is the causative agent of bubonic and pneumonic plague and is a recently diverged clone of *Yersinia pseudotuberculosis* (1), which alongside *Y. enterocolitica* is a zoonotic gastrointestinal pathogen (2). The remaining species are not associated with human disease and are considered to be environmental organisms, with the exception of the common fish pathogen *Y. ruckeri* (2) and the insecticidal species *Y. entomophaga*. Of the human-pathogenic species, *Y. enterocolitica* is the most common etiological agent of human disease, and in Germany and Scandinavia, the numbers of cases of human intestinal yersiniosis caused by *Y. enterocolitica* rival those caused by *Salmonella* (3). *Y. enterocolitica* is in itself a very diverse species that is classically subdivided into nonpathogenic, low-pathogenic, and high-pathogenic biotypes based on virulence in a mouse infection model (4). Biotype 1A isolates are considered nonpathogenic, which is concordant with a lack of the major *Y. enterocolitica* virulence factors such as pYV, invasin, YadA, and Ail (5), although there are numerous reports of biotype 1A human carriage (6, 7). Biotype 1B isolates are high pathogenic, which is concordant with carriage of the high-pathogenicity island, but isolation from human disease cases is very rare with the exception of notable outbreaks such as the recent emergence in Poland (8). Biotype 2 to 4 isolates are low pathogenic and are globally the most common causes of human gastrointestinal yersiniosis (4). Biotype 5 isolates are also considered low pathogenic but have only been isolated from wild hare populations and are very rare in nature (5).

From a clinical perspective, the isolation and subsequent identification of *Yersinia* and in particular *Y. enterocolitica* to the species and subspecies levels can be challenging, with recent publications striving to improve the efficacy of selective culturing of *Yersinia* from clinical and environmental samples (9). Once isolated, strains are most commonly identified to the species level by comparing the differential utilization of a panel of 17 biochemical substrates (4, 10). Further subdivision of *Y. enterocolitica* into biotypes is also performed based on utilization of a further 12 substrates. In both cases, the interpretation of such biochemical typing may often be subjective and affected by environmental factors such as temperature of incubation (4, 10). There is also further subdivision based on classical serotyping. As such, the identification of *Yersinia* to the species and subspecies levels can be very problematic for nonspecialist laboratories with misidentification at the species level and subtyping level not an uncommon occurrence, as exemplified by recent assignment of new species by molecular methods following inconclusive species determination by biochemical methods (11, 12).

Recent work by our group definitively characterized the phy-

Received 20 August 2014 **Returned for modification** 30 September 2014
Accepted 13 October 2014

Accepted manuscript posted online 22 October 2014

Citation Hall M, Chattaway MA, Reuter S, Savin C, Strauch E, Carniel E, Connor T, Van Damme I, Rajakaruna L, Rajendram D, Jenkins C, Thomson NR, McNally A. 2015. Use of whole-genus genome sequence data to develop a multilocus sequence typing tool that accurately identifies *Yersinia* isolates to the species and subspecies levels. J Clin Microbiol 53:35–42. doi:10.1128/JCM.02395-14.

Editor: N. A. Ledeboer

Address correspondence to Alan McNally, alan.mcnally@ntu.ac.uk.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.02395-14.

doi:10.1128/JCM.02395-14

logeny of the entire *Yersinia* genus using genome sequences of 241 strains encompassing the full diversity within the genus (13). Particular attention was given to *Y. enterocolitica*, of which 94 strains encompassing all biotypes and serotypes were sequenced. The whole-genus phylogeny was constructed using 84 housekeeping genes that are located on 23 syntenic blocks, regions of DNA-containing genes conserved across the genus, and showed the presence of 14 species clusters as determined by Bayesian analysis of population structure (BAPS) software (13). The resulting phylogeny also accurately distinguished *Y. pestis* as a distinct clone of *Y. pseudotuberculosis* and phylogenetically split *Y. enterocolitica* on the basis of high-pathogenic, low-pathogenic, and nonpathogenic biotypes (13). A core genome single nucleotide polymorphism (SNP)-based phylogeny provided greater resolution for *Y. enterocolitica* and showed that the phylogenetic separation within the low-pathogenic strains is concordant with serotype and not biotype, which is almost certainly due to difficulties in interpreting variable reactions discriminating biotypes 2 and 3 (13).

Given our significant findings on the true phylogenetic structure of the entire *Yersinia* genus and the fact that this phylogeny can be determined from housekeeping genes present on conserved syntenic blocks, we sought to determine if a standard, seven-gene multilocus sequence typing (MLST) scheme could be developed from a subset of those genes. Such a scheme would then be able to rapidly and with complete accuracy identify any member of the *Yersinia* genus to the species and subspecies levels upon the initial isolation. There is a well-established MLST scheme available for *Y. pseudotuberculosis* (14) that has been used to delineate the population structure of the species complex (15); however, this scheme has not been designed to be robust across the genus. Similarly, there have been attempts to create MLST schemes for *Y. enterocolitica* (16–18); however, these have not been informed by genomic data and their suitability for identification to the species and subspecies levels is questionable compared to that of our previous whole-genome phylogeny study (13). Here, we present the design and validation of a new pan-*Yersinia* MLST scheme that provides identification to the species level that is completely concordant with our previous whole-genome phylogeny (13). Furthermore, it accurately differentiates *Y. pestis* and *Y. similis* from *Y. pseudotuberculosis* and, more significantly, the scheme subtypes low-pathogenic *Y. enterocolitica* on the basis of serotype in complete concordance with whole-genome phylogeny of the species. We propose that the pan-*Yersinia* MLST scheme is an invaluable tool in the identification of *Yersinia* to the species and subspecies levels from clinical samples and that the classification of low-pathogenic *Y. enterocolitica* on the basis of phylogenetically distinct serotypes be adopted.

MATERIALS AND METHODS

Strains. The initial design and development of the MLST scheme utilized *de novo* assembled genome sequences of 171 *Yersinia* strains that were part of our previously published work (13). This strain collection was made up of the *Yersinia* species as follows: 1 *Y. aldovae*, 2 *Y. aleksiciae*, 3 *Y. bercovieri*, 58 *Y. enterocolitica*, 22 *Y. frederiksenii*, 16 *Y. intermedia*, 9 *Y. kristensenii*, 1 *Y. massiliensis*, 10 *Y. mollaretii*, 1 *Y. pekkanenii*, 3 *Y. pestis*, 33 *Y. pseudotuberculosis*, 5 *Y. rohdei*, 3 *Y. ruckeri*, and 4 *Y. similis*. The 171 strains are a subset of the 241 sequenced in our previous study and were chosen because their assembled genomes contained no ambiguous base calls or contig breaks in the syntenic blocks our study design focused on.

Selection of phylogenetically informative genes within conserved syntenic blocks. To establish the level of genetic diversity of each of the common housekeeping genes, GenBank files of each of the 23 syntenic blocks from *Y. enterocolitica* 8081, *Y. pseudotuberculosis* IP32953, and *Y. pestis* D106004 were created using Artemis (19). The sequences of each of the conserved housekeeping genes were then extracted and aligned in MEGA 5.0 (20), as these represent the three human-pathogenic species that are located at diametrically opposite ends of the genus phylogeny. The genes that had a level of SNPs between 10 and 25% were retained for further analysis. The sequences of the remaining genes were used to create individual gene maximum likelihood trees using MEGA 5.0 and compared to the *Yersinia* phylogeny (13). Seven genes that were able to closely match the branching order and clearly discriminate between the species clusters, with <2% strain displacement, and that were disseminated across the syntenic blocks were chosen. Pan-*Yersinia* gene primers for the seven selected genes were designed based on the multiple alignments.

PCR and sequence analysis. The culture was grown overnight in 1.5 ml LB broth at 25°C with shaking, and genomic DNA was extracted using the GenElute bacterial genomic DNA kit (Sigma-Aldrich), following the manufacturer's instructions. A temperature gradient PCR was used to establish the optimum annealing temperatures for the primers. The result was optimized by carrying out the PCR on representative strains of all the species for each primer pair as follows: initial denaturation at 94°C for 5 min; 30 cycles of denaturation at 94°C for 30 s, annealing temperature dependent upon the primer set for 30 s; elongation at 72°C for 30 s; and final elongation at 72°C for 5 min. PCRs were carried out using the GoTaq Flexi DNA polymerase kit (Promega) and deoxynucleoside triphosphates (dNTPs) (Promega) as follows: 5 μl 1.5 mM $MgCl_2$, 5 μl 10× PCR buffer, 2 μl 10 μM dNTPs, 0.3 μl 5 U/μM *Taq* DNA polymerase, 40 μl sterilized distilled water, 0.5 μl 10 pmol forward and reverse primers, 1 μl ~10 ng/μl DNA. The amplification product was then cleaned using ExoSAP-IT (Affymetrix) and Sanger sequenced in duplicate to obtain independent forward and reverse reactions.

The sequence data obtained for each gene were aligned and trimmed to a uniform length, using MEGA 5.0. Each unique sequence was identified using the Web tool Non-redundant databases (http://pubmlst.org/analysis/) and allocated a specific allele number. All of the sequence and isolate data were uploaded to the publically available MLST database (http://pubmlst.org/yersinia) using the BIGSdb genomics platform (21).

Phylogenetic and population analysis of MLST data. The freely available software START (22) was used to calculate the ratio of nonsynonymous (*dN*) to synonymous (*dS*) nucleotide substitutions to determine the level of selective pressure acting upon each MLST gene. START was also used to determine that the GC content in the MLST genes was comparable to that of the whole-genome GC content. To detect recombination within the *Y. enterocolitica* MLST data, SplitsTree 4.2 (23) was used to compute the pairwise homoplasy index (PHI). An MLST database and Web interface were created for the scheme (http://pubmlst.org/yersinia/), and the sequence data for all seven loci from all 171 individual strains were input to assign allele numbers. From these sequences, types were ascribed to each unique allele combination occurring in the data set. The designated allele numbers were visualized by creating minimum spanning trees using the goeBURST Full MLST algorithm in PHYLOViZ (24). Maximum likelihood phylogenies were created by concatenation of the sequence of the seven loci and alignment with ClustalW in MEGA 5.0, before the phylogeny was determined with the GTR gamma model in RAxML 7.2.8–2 (25).

RESULTS

Selection of genes and validation of a pan-*Yersinia* MLST scheme on *in silico* genome sequence data. The sequences of 73 genes conserved across the genus (Table 1) from 171 *de novo* assembled genomes were used to create individual gene phylogenies. Additionally, the alignments were used to identify regions of high similarity in each gene that would permit the design of universal primers capable of amplifying the gene across the genus. From this

TABLE 1 The 73 housekeeping genes selected for investigation for use in the genus MLST scheme

Syntenic block	Relative location on each syntenic block in relation to the *Y. enterocolitica* 8081 reference genome:		Size (bp)	Housekeeping gene(s) in each block
	Beginning	End		
1	0	107030	107,030	*asnA, dfp, tpiA, glnA*
2	108300	191330	83,030	*sthA*
3	191500	202555	11,055	*rhlB, rho*
4	282830	313630	30,800	*udp, aarF, hemB*
5	879520	980830	101,310	*pcm, recA*
6	1039400	1251800	212,400	*gloB, nadB, guaA, nrdF, nrdE*
7	1802900	1991400	188,500	*purB, ptsG, phoQ, phoP, purT, pip, tmk, icdA*
8	2027865	2087750	59,885	*kduD1*
9	2108500	2142600	34,100	ND[a]
10	2154500	2325300	170,800	*tyrR*
11	2447240	2499700	52,460	*topB, ansA, dadA, nhaB, fnaR, xthA*
12	2554700	2591553	36,853	*minD, zwf, aspS, znuC, znuA, znuB, minC, rnd, msbB*
13	2602500	2630230	27,730	*kdsA, prfA, hemA*
14	2640950	2668185	27,235	*chaA*
15	2668285	2709585	41,300	ND
16	2709700	2800900	91,200	ND
17	2854263	3294700	440,437	*folE, nadA, udk, sfcA, glnS*
18	3313400	3544778	231,378	*proB, rosA, hemH, adk*
19	3610900	3712864	101,964	*thyA, tas, lgt, galR, lysS, prfB*
20	3726200	3761260	35,060	*tktA, speA, gshB, endA*
21	3960000	4238800	278,800	*rfaE, pyrB, parC, gcp, uxaC*
22	4245400	4464400	219,000	ND
23	4504400	4561500	57,100	*fdoI, fdhE, glnQ*
Total size of syntenic blocks			2,639,427	
Total size of *Y. enterocolitica* 8081 genome			4,615,899	

[a] ND, no housekeeping genes present in the syntenic block.

analysis, seven optimal gene loci were selected based on their ability to mirror the genome-informed phylogeny and the ability to design primers that would work across the genus (Table 2), as well as their separation across the syntenic blocks (Fig. 1).

There was a high level of diversity shown across the seven selected MLST regions, averaging around 60 alleles and 40% polymorphic sites for each (Table 3). The *dN*/*dS* ratios were far below 1 for each MLST region, suggesting that the nucleotide substitutions are not a result of selective pressure. The average GC content found in the MLST gene regions corresponds to that of the *Yersinia* chromosomes, which ranges from 46.9% in *Y. frederiksenii* to 49.0% in *Y. mollaretii* (data accessible at the xBASE website http://www.xbase.ac.uk/taxon/Yersinia). The PHI test also failed to detect any recombination within the MLST amplicons from the *Y. enterocolitica* data set.

Pan-*Yersinia* MLST scheme is phylogenetically informative to genome sequence level. A maximum likelihood phylogeny of the concatenated MLST data obtained from the 171 genome-sequenced strains was constructed. The resulting tree showed accurate phylogenetic separation of all of the species identified by the 84-gene tree approach taken in our previous work (Fig. 2) with 100% concordance between the two phylogenies and identical

TABLE 2 Primer sequences, the sizes of the amplified regions, and the annealing temperature for the final seven selected MLST genes

MLST gene	Primer		PCR product length (bp)	MLST region length (bp)	Annealing temperature (°C)
	Forward	Reverse			
aarF	5′-TTCCATGCAGATATGCACC-3′	3′-CCACTCACTAATAGTGTAGC-5′	650	500	52
dfp	5′-GATCCGGTACGCTTTATCAG-3′	3′-CATAACGGCTGACAATCTCG-5′	547	455	59
galR	5′-ATTGGTAACGGTTACCATG-3′	3′-GTTGGGCTGAACATATTGGT-5′	648	500	59
glnS	5′-GAATCATGTATCCGTGATG-3′	3′-GCACAGAAATAACCTTCAC-5′	557	442	56.5
hemA	5′-ATGACTCTGCTCGCATTAGG-3′	3′-CGGTTGGCAATAATCATATG-5′	602	490	54
speA	5′-ATGTCTGATGATAACTTGATT-3′	3′-CAGATAAACTTTATGGCCC-5′	550	452	55.5
rfaE	5′-ATGAAAGTCACTCTGCCTGA-3′	3′-ATCACTGCCTTTAGGATC-5′	509	429	55.5

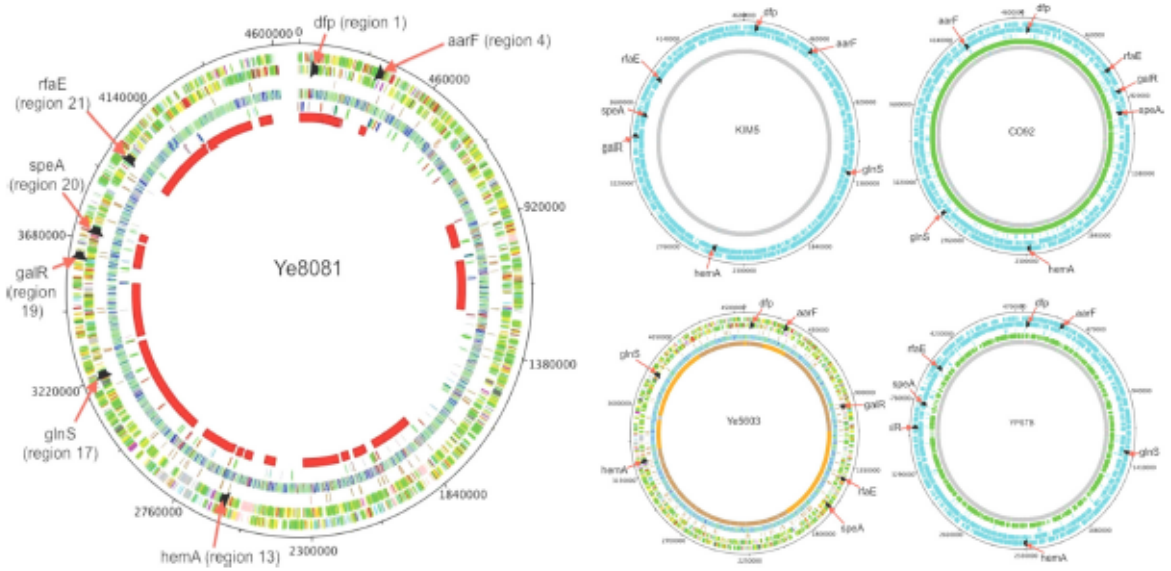


FIG 1 Diagram showing the positioning of the seven selected loci and their native syntenic block on reference genomes across the genus. Ye8081, *Y. enterocolitica* bioserotype 1B/O:8; KIM5, *Y. pestis* Medievalis; CO92, *Y. pestis* Orientalis; Ye5603, *Y. enterocolitica* bioserotype 4/O:3; YPSTB, *Y. pseudotuberculosis* YPIII serotype III.

separation into 14 distinct species clusters as determined by BAPS. The *Y. pseudotuberculosis* complex was accurately split with distinct clades containing *Y. pestis* and *Y. similis* within the larger *Y. pseudotuberculosis* complex, showing that the scheme is capable of differentiating accurately within this lineage. Closer investigation of the *Y. enterocolitica* complex showed that the MLST scheme also differentiates on the basis of high-pathogenic, low-pathogenic, and nonpathogenic groups, and within the low-pathogenic group differentiates on the basis of serotype into defined phylogroups as observed when the whole-genome phylogeny is used. As such, the pan-*Yersinia* MLST scheme provides a completely robust mechanism by which to accurately assign any *Yersinia* isolate to a defined species cluster and further subtype without any additional growth requirements beyond initial isolation.

Validation of the pan-*Yersinia* MLST scheme on reference laboratory isolate collections. To validate the *in silico* results for our genus-wide typing scheme, we performed MLST on a further 214 *Yersinia* strains archived in the national *Yersinia* reference laboratories of Belgium, Germany, United Kingdom, and France (see Table S1 in the supplemental material). The concatenated MLST sequence data for all 385 strains were then used to construct a maximum likelihood phylogeny and compare the results of the classical biochemical typing and subtyping with those for our phylogenetic approach (Fig. 3). The phylogeny once again shows unambiguous separation of strains into the previously designated species clusters, with 97.83% of strains tested being assigned to the corresponding species cluster based on their biochemical typing. Included here are strains of *Y. wautersii*, a newly proposed species which is a sublineage of *Y. pseudotuberculosis*. Two strains biochemically defined as *Y. pseudotuberculosis* by the reference laboratories with the *Y. similis* subgroup and a further 6 isolates were assigned to species clusters in disagreement with their classical biochemical typing designation by the reference laboratories.

To validate the *in silico* results showing that our MLST scheme was able to successfully subtype *Y. enterocolitica*, we separately analyzed the MLST data for the 188 *Y. enterocolitica* isolates contained within the entire data set generated here (Fig. 4). Our phylogeny perfectly assigns every strain to a defined phylogroup on the basis of serotype as previously reported with whole-genome SNP-based phylogeny. There are no ambiguous phylogroup assignations on the basis of serotype, although, as with the whole-genome study, biotype is not phylogenetically robust. To allow an easy comparator for use of the scheme, we assigned which species cluster and/or *Y. enterocolitica* phylogroup each sequence type belongs to (see Table S2 in the supplemental material).

TABLE 3 Level diversity across all 171 genome-sequenced strains for each of the MLST regions as determined by START

Gene	Size of fragment (bp)	% GC content	% polymorphic sites	*dN*/*dS* ratio	No. of alleles
aarF	500	44	37.4	0.0049	58
dfp	500	47.8	40.8	0.0599	61
galR	500	49.7	44.8	0.028	70
glnS	500	48.5	38.4	0.0221	68
hemA	500	51.3	39.8	0.0222	65
rfaE	429	54.1	40.3	0.019	60
speA	490	48.9	34.7	0.0232	50
Mean	488.4	49.2	39.5	0.0256	61.7

DISCUSSION

The enteropathogenic *Yersinia* spp. are the third most common cause of bacterial infectious intestinal disease in the developed world (5). Despite this, the isolation and identification of infections with *Y. enterocolitica* and *Y. pseudotuberculosis* are still heavily reliant on classical biochemical techniques that may be open to

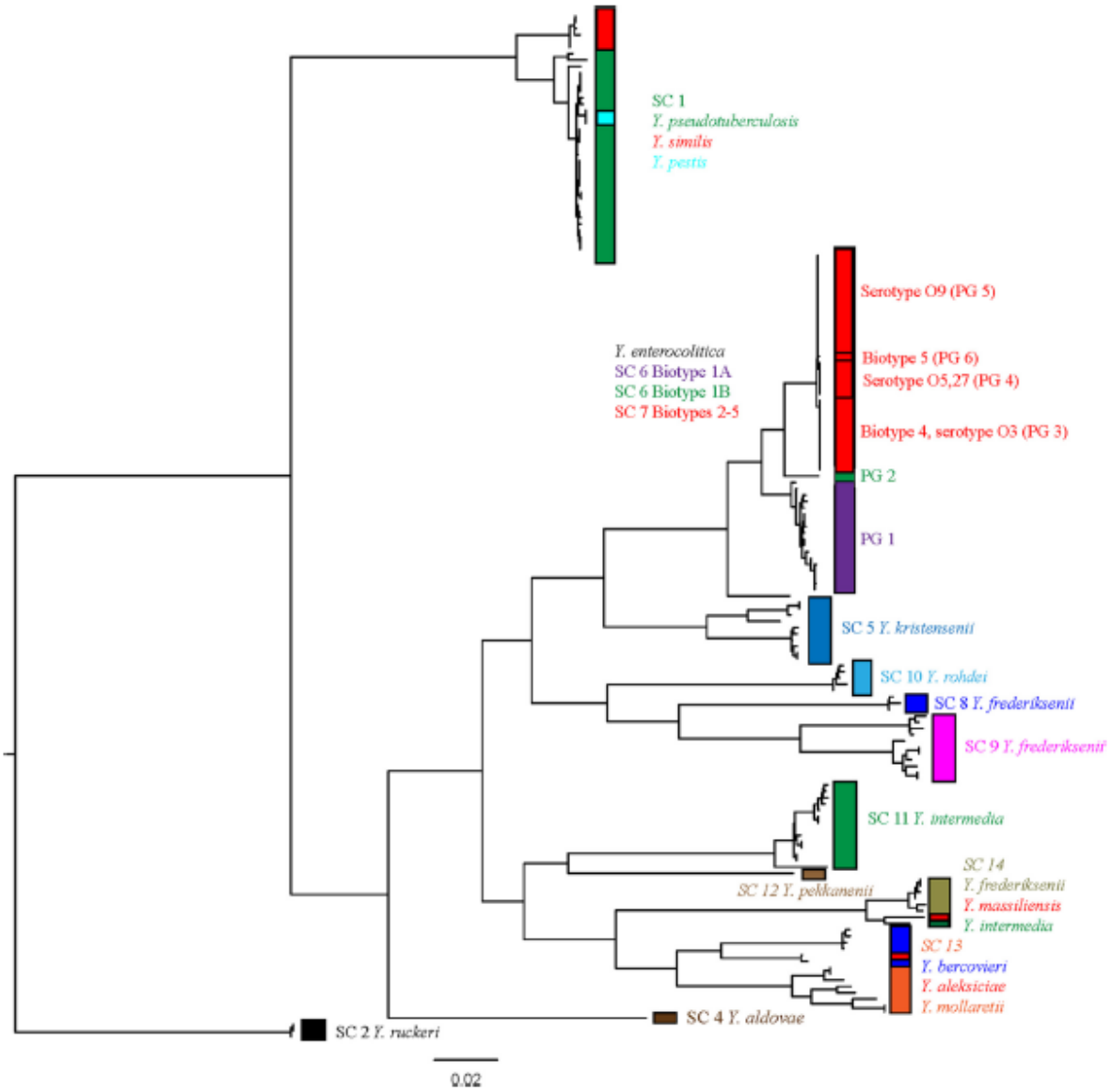


FIG 2 Maximum likelihood phylogeny of concatenated alleles derived from every unique sequence type obtained from 171 genomes from across the genus. The species contained within each sequence type are indicated, and species clusters are labeled as defined in our previous genome study (13), with the MLST tree showing complete concordance with our previous phylogeny.

subjective interpretation to provide a definitive identification (4, 10). This subjective biochemical typing is even more problematic when applied to subtyping of isolates, which is of importance in epidemiological tracking, and in the case of *Y. enterocolitica* may be of clinical importance in distinguishing between the carriage of a nonpathogenic organism, a self-limiting infection with a low-pathogenic strain, or an infection with a more aggressive high-pathogenic strain type. Similarly, nonpathogenic species within the genus may be biochemically typed as atypical *Y. enterocolitica*, leading to misidentification of clinical episodes, administration of unnecessary treatments, and skewed data in environmental and livestock surveys of enteropathogenic *Yersinia* prevalence (26, 27).

Despite the proven levels of resolution offered by molecular typing techniques for bacterial pathogens to overcome such problems, there is no such approved and standardized methodology in place for *Y. enterocolitica*, the most common cause of human gastrointestinal yersiniosis. An MLST scheme does exist for *Y. pseudotuberculosis* but is designed and validated to be used as an epidemiological and population genetics tool solely for that species (15). In this study, we have utilized the comprehensive genus ge-

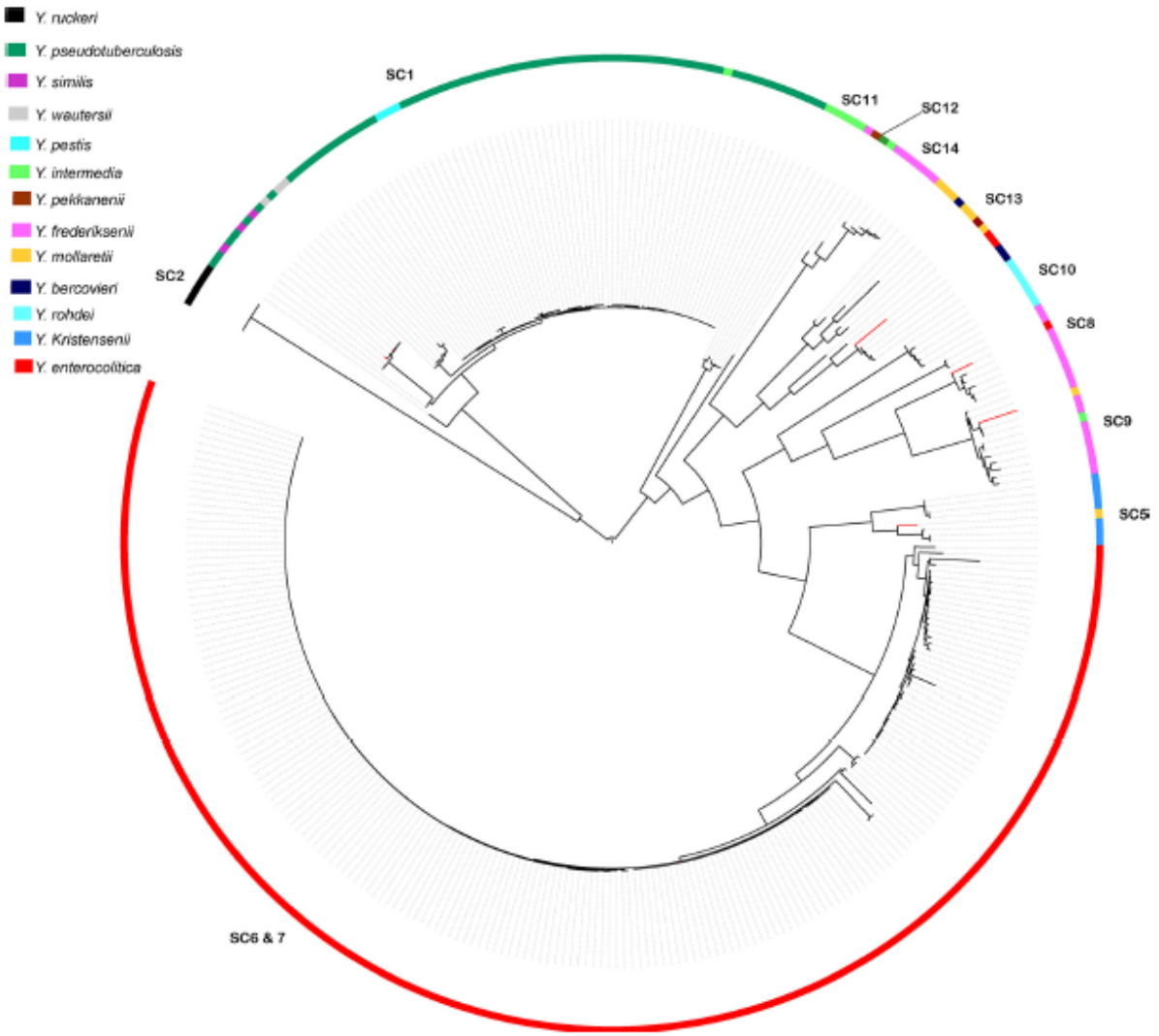


FIG 3 Maximum likelihood phylogeny of concatenated alleles derived from 385 reference laboratory strains of *Yersinia*. The species clusters are designated on each clade as SC, with the biochemically determined species of each strain denoted by the described color coding. Strains whose species cluster differed from biochemical typing results are denoted by red branches on the phylogeny.

nome sequence data set previously produced by our group (13) to inform the design of an MLST-based scheme that can rapidly and reproducibly assign any strain to a defined species cluster and any *Y. enterocolitica* to a defined phylogroup.

Previous attempts have been made to create MLST typing tools for *Y. enterocolitica*. The first scheme (16) was a 5-locus scheme incorporating 16S that was developed to allow phylogenetic inferences within the genus *Yersinia*. However, when the phylogeny published in that pregenomics era study is compared to our definitive phylogeny recently published (13), it is clear that the 5-locus phylogeny is inaccurate with *Y. enterocolitica* deeply embedded within environmental species (16). As such, determining the species using this scheme on an unknown isolate would not offer sufficiently robust resolution for reference laboratory adoption. A conventional 7-locus scheme was developed from a semirandom selection of housekeeping genes to investigate subgrouping within the nonpathogenic biotype 1A *Y. enterocolitica* isolates (17). While the loci in this scheme are among the 84 genes conserved across the genus, *in silico* analysis suggests that the primers designed may not be optimal across the genus due to base mismatches at the primer sites and as such would not be suitable for the purposes of identifying *Yersinia* isolates to the species level. Most recently, a scheme was developed to differentiate the three human-pathogenic species of the *Yersinia* genus using a 7-locus MLST scheme (18). This scheme accurately subtyped *Y. enterocolitica* into distinct subtypes, including serotype-specific clades within the low-pathogenic strains, as observed both in our scheme and in our genomic phylogeny (13). However, this scheme also

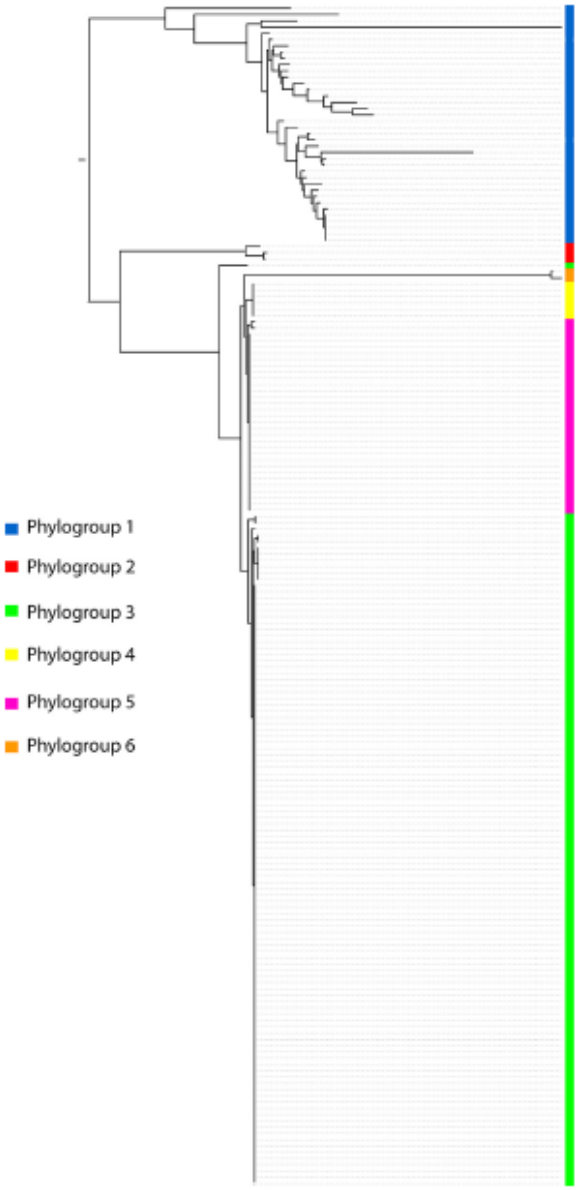


FIG 4 Maximum likelihood phylogeny of concatenated alleles derived from 188 *Y. enterocolitica* strains. The phylogroups are indicated by the described color coding.

uses primers that *in silico* analysis suggests would not anneal to sequences from some species. Additionally, neither of the latter two schemes has been set up with a database and protocols to allow its wide-scale adoption for reference typing.

In conclusion, we present a model and novel design strategy for molecular typing tools based on genome sequence data across an entire genus containing human-pathogenic species. By using these data, we can design a simple MLST-based scheme that provides the power of resolution of whole-genome sequencing to quickly and accurately identify isolates to the species level and also subtype strains of *Y. enterocolitica*. While next-generation sequencing is becoming commonplace in a small number of public health laboratories, there are still many front-line clinical microbiology laboratories that are not yet in a position to employ benchtop sequencing due to the cost or bioinformatics resources. Our scheme provides a blueprint for the efficient design of simple molecular-based tools that provide an equal level of resolution for typing, although obviously not for SNP-based molecular epidemiological investigations. We encourage the public health microbiology community to adopt our scheme and further validate it as a universal typing tool for the entire *Yersinia* genus and as a subtyping and population genetics tool for the important human pathogen *Y. enterocolitica*.

ACKNOWLEDGMENTS

This work was funded by NTU VC studentships awarded to M.H. and S.R. The French Yersinia Reference Laboratory was funded in part by the Institute for Health Surveillance (InVS).

We thank the National Collection of Type Cultures for provision of strains from across the genus for our validation work.

REFERENCES

1. **Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, Kusecek B, Vogler AJ, Wagner DM, Allender CJ, Easterday WR, Chenal-Francisque V, Worsham P, Thomson NR, Parkhill J, Lindler LE, Carniel E, Keim P.** 2004. Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc Natl Acad Sci U S A 101:17837–17842. http://dx.doi.org/10.1073/pnas.0408026101.
2. **Wren BW.** 2003. The *Yersiniae*—a model genus to study the rapid evolution of bacterial pathogens. Nat Rev Microbiol 1:55–64. http://dx.doi.org/10.1038/nrmicro730.
3. **Rosner BM, Stark K, Werber D.** 2010. Epidemiology of reported *Yersinia enterocolitica* infections in Germany, 2001–2008. BMC Public Health 10:337. http://dx.doi.org/10.1186/1471-2458-10-337.
4. **Bottone EJ.** 1999. *Yersinia enterocolitica*: overview and epidemiologic correlates. Microbes Infect 1:323–333. http://dx.doi.org/10.1016/S1286-4579(99)80028-8.
5. **Bottone EJ.** 1997. *Yersinia enterocolitica*: the charisma continues. Clin Microbiol Rev 10:257–276.
6. **Tennant SM, Grant TH, Robins-Browne RM.** 2003. Pathogenicity of *Yersinia enterocolitica* biotype 1A. FEMS Immunol Med Microbiol 38:127–137. http://dx.doi.org/10.1016/S0928-8244(03)00180-9.
7. **Singh I, Virdi JS.** 2004. Production of *Yersinia* stable toxin (YST) and distribution of *yst* genes in biotype 1A strains of *Yersinia enterocolitica*. J Med Microbiol 53:1065–1068. http://dx.doi.org/10.1099/jmm.0.45527-0.
8. **Gierczyński R, Szych J, Rastawicki W, Wardak S, Jagielski M.** 2009. Molecular characterization of human clinical isolates of *Yersinia enterocolitica* bioserotype 1B/O8 in Poland: emergence and dissemination of three highly related clones. J Clin Microbiol 47:1225–1228. http://dx.doi.org/10.1128/JCM.01321-08.
9. **Savin C, Leclercq A, Carniel E.** 2012. Evaluation of a single procedure allowing the isolation of enteropathogenic *Yersinia* along with other bacterial enteropathogens from human stools. PLoS One 7:e41176. http://dx.doi.org/10.1371/journal.pone.0041176.
10. **Wauters G, Kandolo K, Janssens M.** 1987. Revised biogrouping scheme of *Yersinia enterocolitica*. Contrib Microbiol Immunol 9:14–21.
11. **Murros-Kontiainen A, Fredriksson-Ahomaa M, Korkeala H, Johansson P, Rahkila R, Björkroth J.** 2011. *Yersinia nurmii* sp. nov. Int J Syst Evol Microbiol 61:2368–2372. http://dx.doi.org/10.1099/ijs.0.024836-0.
12. **Murros-Kontiainen A, Johansson P, Niskanen T, Fredriksson-Ahomaa M, Korkeala H, Björkroth J.** 2011. *Yersinia pekkanenii* sp. nov. Int J Syst Evol Microbiol 61:2363–2367. http://dx.doi.org/10.1099/ijs.0.019984-0.
13. **Reuter S, Connor T, Barquist L, Walker D, Feltwell T, Harris S, Fookes M, Hall M, Petty N, Fuchs T, Corander J, Dufour M, Ringwood T, Savin C, Bouchier C, Martin L, Miettinen M, Shubin M, Riehm J, Laukkanen-Ninios R, Sihvonen L, Siitonen A, Skurnik M, Falcão J,**

Fukushima H, Scholz H, Prentice M, Wren B, Parkhill J, Carniel E, Achtman M, McNally A, Thomson N. 2014. Parallel independent evolution of pathogenicity within the genus *Yersinia*. Proc Natl Acad Sci U S A 111:6768–6773. http://dx.doi.org/10.1073/pnas.1317161111.

14. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 96:14043–14048. http://dx.doi.org/10.1073/pnas.96.24.14043.
15. Laukkanen-Ninios R, Didelot X, Jolley K, Morelli G, Sangal V, Kristo P, Brehony C, Imori P, Fukushima H, Siitonen A, Tseneva G, Voskressenskaya E, Falcao J, Korkeala H, Maiden M, Mazzoni C, Carniel E, Skurnik M, Achtman M. 2011. Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. Environ Microbiol 13:3114–3127. http://dx.doi.org/10.1111/j.1462-2920.2011.02588.x.
16. Kotetishvili M, Kreger A, Wauters G, Morris JG, Jr, Sulakvelidze A, Stine OC. 2005. Multilocus sequence typing for studying genetic relationships among *Yersinia* species. J Clin Microbiol 43:2674–2684. http://dx.doi.org/10.1128/JCM.43.6.2674-2684.2005.
17. Sihvonen L, Jalkanen K, Huovinen E, Toivonen S, Corander J, Kuusi M, Skurnik M, Siitonen A, Haukka K. 2012. Clinical isolates of *Yersinia enterocolitica* biotype 1A represent two phylogenetic lineages with differing pathogenicity-related properties. BMC Microbiol 12:208. http://dx.doi.org/10.1186/1471-2180-12-208.
18. Duan R, Liang J, Shi G, Cui Z, Hai R, Wang P, Xiao Y, Li K, Qiu H, Gu W, Du X, Jing H, Wang X. 2014. Homology analysis of pathogenic *Yersinia* species *Yersinia enterocolitica*, *Yersinia pseudotuberculosis*, and *Yersinia pestis* based on multilocus sequence typing. J Clin Microbiol 52:20–29. http://dx.doi.org/10.1128/JCM.02185-13.
19. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16:944–945. http://dx.doi.org/10.1093/bioinformatics/16.10.944.
20. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
21. Jolley KS, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. http://dx.doi.org/10.1186/1471-2105-11-595.
22. Jolley KA, Feil EJ, Chan MS, Maiden MC. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230–1231. http://dx.doi.org/10.1093/bioinformatics/17.12.1230.
23. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. http://dx.doi.org/10.1093/molbev/msj030.
24. Francisco AP, Bugalho M, Ramirez M, Carrico JA. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics 10:152. http://dx.doi.org/10.1186/1471-2105-10-152.
25. Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456. http://dx.doi.org/10.1093/bioinformatics/bti191.
26. Milnes A, Stewart I, Clifton-Hadley F, Davies R, Newell D, Sayers A, Cheasty T, Cassar C, Ridley A, Cook A, Evans S, Teale C, Smith R, McNally A, Toszeghy M, Futter R, Kay A, Paiba G. 2008. Intestinal carriage of verocytotoxigenic *Escherichia coli* O157, *Salmonella*, thermophilic *Campylobacter* and *Yersinia enterocolitica*, in cattle, sheep and pigs at slaughter in Great Britain during 2003. Epidemiol Infect 136:739–751. http://dx.doi.org/10.1017/S0950268807009223.
27. McNally A, Cheasty T, Fearnley C, Dalziel RW, Paiba GA, Manning G, Newell DG. 2004. Comparison of the biotypes of *Yersinia enterocolitica* isolated from pigs, cattle and sheep at slaughter and from humans with yersiniosis in Great Britain during 1999–2000. Lett Appl Microbiol 39:103–108. http://dx.doi.org/10.1111/j.1472-765X.2004.01548.x.