

RESEARCH ARTICLE

Genomic epidemiology and carbon metabolism of *Escherichia coli* serogroup O145 reflect contrasting phylogenies

Rose M. Collis^{1,2}, Patrick J. Biggs^{2,3,4}, Anne C. Midwinter², A. Springer Browne², David A. Wilkinson^{2,4}, Hamid Irshad⁵, Nigel P. French^{2,4}, Gale Brightwell^{1,4}, Adrian L. Cookson^{1,2*}

1 AgResearch Ltd, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand, **2** Molecular Epidemiology and Veterinary Public Health Laboratory (^mEpiLab), Infectious Disease Research Centre, School of Veterinary Science, Massey University, Palmerston North, New Zealand, **3** School of Fundamental Sciences, Massey University, Palmerston North, New Zealand, **4** New Zealand Food Safety Science and Research Centre, Massey University, Palmerston North, New Zealand, **5** Animal Health Programme, National Agricultural Research Centre, Islamabad, Pakistan

* Adrian.Cookson@agresearch.co.nz



OPEN ACCESS

Citation: Collis RM, Biggs PJ, Midwinter AC, Browne AS, Wilkinson DA, Irshad H, et al. (2020) Genomic epidemiology and carbon metabolism of *Escherichia coli* serogroup O145 reflect contrasting phylogenies. PLoS ONE 15(6): e0235066. <https://doi.org/10.1371/journal.pone.0235066>

Editor: Baochuan Lin, Defense Threat Reduction Agency, UNITED STATES

Received: January 30, 2020

Accepted: June 8, 2020

Published: June 25, 2020

Copyright: © 2020 Collis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The accession numbers for *E. coli* serogroup O145 strains whose genome sequenced in this study are listed in [S4 Table](#) and are deposited with NCBI under the BioProject number PRJNA435641.

Funding: This research was supported by the AgResearch Food Provenance and Assurance Strategic Science Investment (SSI) Fund programme awarded to GB, the Massey University School of Veterinary Science post-graduate research fund Massey University awarded to RMC

Abstract

Shiga toxin-producing *Escherichia coli* (STEC) are a leading cause of foodborne outbreaks of human disease, but they reside harmlessly as an asymptomatic commensal in the ruminant gut. STEC serogroup O145 are difficult to isolate as routine diagnostic methods are unable to distinguish non-O157 serogroups due to their heterogeneous metabolic characteristics, resulting in under-reporting which is likely to conceal their true prevalence. In light of these deficiencies, the purpose of this study was a twofold approach to investigate enhanced STEC O145 diagnostic culture-based methods: firstly, to use a genomic epidemiology approach to understand the genetic diversity and population structure of serogroup O145 at both a local (New Zealand) ($n = 47$) and global scale ($n = 75$) and, secondly, to identify metabolic characteristics that will help the development of a differential media for this serogroup. Analysis of a subset of *E. coli* serogroup O145 strains demonstrated considerable diversity in carbon utilisation, which varied in association with *eae* subtype and sequence type. Several carbon substrates, such as D-serine and D-malic acid, were utilised by the majority of serogroup O145 strains, which, when coupled with current molecular and culture-based methods, could aid in the identification of presumptive *E. coli* serogroup O145 isolates. These carbon substrates warrant subsequent testing with additional serogroup O145 strains and non-O145 strains. Serogroup O145 strains displayed extensive genetic heterogeneity that was correlated with sequence type and *eae* subtype, suggesting these genetic markers are good indicators for distinct *E. coli* phylogenetic lineages. Pangenome analysis identified a core of 3,036 genes and an open pangenome of >14,000 genes, which is consistent with the identification of distinct phylogenetic lineages. Overall, this study highlighted the phenotypic and genotypic heterogeneity within *E. coli* serogroup O145, suggesting that the development of a differential media targeting this serogroup will be challenging.

and the Palmerston North Medical Research Foundation awarded to ALC. The funder AgResearch Ltd provided support in the form of salaries for the authors ALC and GB, but AgResearch Ltd and other Funders did not have any additional role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: Two of the authors (ALC and GB) are employed by a commercial company, AgResearch Ltd. This does not alter our adherence to PLOS ONE policies on sharing data and materials. The remaining authors have declared that no competing interests exist.

Introduction

Shiga toxin-producing *Escherichia coli* (STEC) are zoonotic pathogens residing harmlessly in the gut of bovine reservoirs, but capable of causing human disease with a broad range of symptoms; from diarrhoea to life-threatening haemolytic uraemic syndrome (HUS) [1, 2]. STEC can be shed in large numbers in faeces excreted by ruminants [3, 4], particularly calves [5], and are an important source of both foodborne and environmentally acquired STEC infections through direct contact with faeces or faecally-contaminated environments. Most human infections are associated with sporadic outbreaks where risk factors include contact with cattle, animal manure, recreational waters [6] or consumption of contaminated food [7]. STEC have been identified as the causative pathogenic agent in disease outbreaks associated with a wide variety of contaminated food products such as romaine lettuce [8], ice-cream [9], and hamburger patties [10]. In an attempt to manage food-related risk, seven serogroups (O26, O45, O103, O111, O121, O145 and O157) collectively described as the "Top 7" have been declared adulterants of ground beef in the United States of America (USA) [11, 12] impacting food safety regulations and international trade. A cross-sectional study investigating the prevalence of STEC in young calves (2–21 days of age) throughout New Zealand (NZ) identified STEC O145 as the most prevalent serogroup (43%) at the dairy farm level compared with the other "Top 7" serogroups [13]. These prevalence data indicate that, as a zoonotic pathogen, *E. coli* serogroup O145 represents both a risk to public health and a regulatory issue for NZ's meat export industry.

STEC express Shiga toxins encoded by the *stx1* and *stx2* genes within lambdoid bacteriophage [14] maintained in a lysogenic state [15]. Stx toxin production is a component of STEC pathogenesis in humans that occurs during bacterial adhesion and intestinal colonisation, leading to impaired intestinal epithelial cell barrier function and diarrhoea [16]. Systemic dissemination of Stx toxin through the cardiovascular system may also lead to HUS and other sequelae [16]. Other important virulence factors for STEC pathogenicity include enterohaemolysin, a plasmid-associated pore-forming RTX toxin encoded by the *ehxA* gene [17, 18] and an outer membrane adhesin, intimin, encoded by the *eae* gene located within the Locus of Enterocyte Effacement (LEE) pathogenicity island [19]. Intimin and other LEE-encoded type III secretion system components and effector proteins mediate the formation of attaching and effacing lesions [19], which are actin pedestals characterised by microvilli effacement and bacterial attachment to the intestinal epithelial cells [19]. The C-terminal end of intimin has a highly variable amino acid sequence thought to be associated with contrasting host tissue tropisms [20, 21] to the extent that the *eae* gene has been differentiated into at least 28 different subtypes [22]. Some STEC serotypes are characterised by a single *eae* subtype such as O157:H7 (γ), O26:H11 (β), O103:H2 (ϵ), O111:H8 (θ) and O145:H28 (γ) [23], however, multiple *eae* subtypes may be associated with other serogroups [19, 24]. The LEE pathogenicity island is inserted in the *E. coli* genome near tRNA genes such as *selC*, *pheV* and *pheU* [19] and is found in enteropathogenic *E. coli* (EPEC), that lack *stx* genes, in addition to the STEC pathotype [25]. Importantly, many STEC do not possess the *eae* and *ehxA* molecular markers; for example, a large foodborne outbreak was caused by a hybrid STEC/Enterotoxigenic *E. coli* O104:H4 strain which was *stx2*-positive and negative for both *eae* and *ehxA* [26, 27], suggesting that all STEC should be treated as pathogenic, regardless of specific O-serogroups [28]. These distinct diarrhoeagenic *E. coli* pathotypes such as STEC and EPEC are often identified according to the presence or absence of specific virulence factors, such as the *stx* and *eae* genes, but display significant genetic heterogeneity and readily acquire new genetic material via horizontal gene transfer (HGT) [25].

Current culture-based detection methods for non-O157 STEC do not provide sufficient discrimination between serogroups due to the lack of differential characteristics between non-

O157 STEC serogroups in comparison to non-pathogenic *E. coli* [29, 30]. A variety of selective media currently available have been developed for the detection and isolation of STEC utilising carbohydrate fermentation patterns to detect specific serogroups based on colony colour [9, 30]. Such media containing carbon substrates have recently been proposed to differentiate the serogroups O26, O103, O111, O145 and O157 [31], in conjunction with other previously developed methods [32]. However, the efficacy of such media utilising these substrates has not been fully validated [31], and this is likely to be a key factor associated with highly variable isolation rates of non-O157 STEC serogroups between studies [29, 33–35] and their probable under-reporting [36]. Despite the extra efforts required with culture-based techniques for the isolation of non-O157 serogroups such as O145, in comparison to rapid molecular methods [37, 38], the identification and isolation of individual bacterial strains provides subsequent opportunities for further epidemiological and clinical analysis [30, 35, 38]. Other studies have analysed data from large panels of serogroup-specific STEC strains [39], such as environmental and clinical isolates [40], but these isolates are often associated with a distinct geographical area [40].

Whole genome sequencing (WGS) provides the ability for high-resolution genetic typing analysis that can be used in epidemiological investigations whilst simultaneously providing information on an isolate's gene content. Previous studies to provide STEC serogroup phylogenies have been limited through the analysis of WGS data from a limited number of serogroup-specific strains [41–43], or from datasets biased towards human isolates [44], causing wide-ranging serogroup-specific diversity to be overlooked.

Therefore, the purpose of this study was firstly, to take a broad-ranging approach to understand the genetic diversity and population structure of serogroup O145 at both a local (NZ) and a global scale using genomic epidemiology methods and, secondly, to identify characteristic metabolic traits associated with serogroup O145 which could prove beneficial in the development of culture-dependent tests for this serogroup.

Materials and methods

E. coli serogroup O145 strains

In this study, 53 *E. coli* serogroup O145 strains (S1 Table) were whole genome sequenced from NZ (n = 47), Norway (n = 4), Australia (n = 1) (provided by Roy Robbins-Browne, University of Melbourne) and USA (n = 1). These were isolated from bovine (n = 36), environmental (n = 6) and human clinical sources (n = 11). The serogroup of these O145 strains (S1 Table) was confirmed using an O145 serogroup-specific PCR, with the primers (5' -GCGGGTGTTCGCCGTTCTGT-3') and (5' -ACGGCATTCGCTGCGAGTT-3') [29] and subsequently with analysis of WGS data. Whole genome sequence data from an additional 69 overseas isolates (human clinical cases: n = 36; bovine: n = 12; food: n = 3; ground beef: n = 1; intact beef: n = 1; ground pork: n = 2; intact pork: n = 3; swine: n = 3; wolf: n = 2; and unknown: n = 6) were included in the comparative genome analysis (S2 Table) to provide a global context. Most of the 122 isolates analysed in the global study were from New Zealand (n = 47) and the USA (n = 47) with the remainder from the UK (n = 15), Norway (n = 4), Canada (n = 2) and one each from Australia, Denmark, Germany, Italy and Uruguay. Two isolates were of unknown geographic origin.

DNA extraction, library preparation and whole genome sequencing

Previously described DNA extraction and library preparation methods [13] were used to prepare the *E. coli* serogroup O145 isolates for WGS. Aliquots of the first four library preparations and an aliquot of the pooled libraries underwent a quality control check (Bioanalyzer 2100

[Agilent Genomics, Santa Clara, CA, USA]) at New Zealand Genomics Limited (NZGL, Massey Genome Service, Massey University, Palmerston North, New Zealand). WGS was performed by NZGL (University of Otago, Dunedin, New Zealand) using an Illumina HiSeq paired-end v4 platform (2 x 125 bp).

Genome quality control, assembly and annotation

The raw sequencing reads were evaluated using quality control software (QCtool) [45]. The sequences were *de novo* assembled using SPAdes v3.9.1 [46]. The quality assessment tool QUAST was used to assess and compare the quality of the genome assemblies [47], which were annotated using Prokka (v1.12-beta) [48]. Genome assembly statistics are displayed in the supplementary information (S1 Fig).

Downloading publicly available serogroup O145 raw sequence data

Serogroup O145 strains were identified from NCBI [49], EnteroBase [50] and published papers (S2 Table). Only whole genome sequences in which the raw read sequence data was available were further analysed using the same analysis pipeline (namely quality assessment, assembly and genome analysis). Publicly available whole genome sequences were excluded from the analysis if any discrepancies indicative of potential contamination such as genome size (<4 Mb or >6Mb) or GC content (<48% or >51%) were identified during the quality assessment, or if an over-representation of unassigned/ambiguous nucleotides (Ns) in the reads was identified using FastQC, or if the identity of the *wzx* and *wzy* genes of the O-antigen biosynthesis gene cluster could not be confirmed as homologous to those from serogroup O145.

Genetic characterisation of *E. coli* serogroup O145 strains

Assembled genomes were batch uploaded to the Center for Genomic Epidemiology (CGE) server [51] for identification of serotype (O and H antigens; threshold of 85% identity (ID) and a minimum gene fractional length of 60%) [52], species [53], *E. coli* associated virulence factors (n = 76; threshold of 90% ID and a minimum gene fractional length of 60%) [54], plasmids (threshold of 95% ID and a minimum gene fractional length of 60%) [55], antibiotic resistance genes (threshold of 90% ID and a minimum gene fractional length of 60%) [56] and multi-locus sequence typing (MLST) [57]. The *stx* variants were determined by VirulenceFinder, and the *eae* subtype was determined by identifying the best nucleotide match using BLASTN [58]. The EPEC-associated bundle forming pilus subunit *bfpA* [59] was detected using Geneious v8.1 [60]. *In silico* analysis of ribosomal multi-locus sequence types (rMLST) [61] was generated from single nucleotide polymorphisms identified in 51 genes encoding the ribosome protein subunits (*rps*, *rpm* and *rpl*). The *in silico* rMLST analysis was visualised using neighbour-joining methods in SplitsTree v4.14.4 [62] and edited using the Interactive Tree of Life (iTOL) webserver [63].

The presence or absence of 37 virulence genes identified using VirulenceFinder [54], which differed between strains, were used to make a Neighbour-joining tree using the Jaccard index and converted to the Newick file format using R 3.6.0 [64] and the packages 'vegan' [65] and 'ctc' [66]. The tree was edited using the iTOL webserver [63] and isolate metadata was included for *eae* subtype, sequence type (ST) and isolation source.

The LEE pathogenicity island integration sites were identified using either the location of LEE-encoded genes including a prophage integrase adjacent to a potential tRNA (*selC*, *pheU* or *pheV*) integration site, or contigs were assembled to a reference genome and the likely tRNA integration predicted based on the mapped contigs and gene synteny using Geneious v8.1 [60]. The reference genomes used were STEC O145:H28 RM13514 (NZ_CP006027) or

STEC O145:H28 RM12761 (NZ_CP007133) where the LEE is integrated at tRNA *selC* [43], STEC O26:H11 11368 (AP010953) [67] where the LEE is integrated at tRNA *pheU* or STEC O103:H2 12009 (AP010958) where the LEE is integrated at tRNA *pheV* [67].

Comparative genomics

Single nucleotide polymorphisms (SNPs) were identified in the paired-end sequencing reads using Snippy v3.0 [68] and STEC O145:H28 RM12761 (NZ_CP007133), associated with a foodborne STEC outbreak in Belgium, was used as the reference genome [43]. This isolate has three contigs (a chromosome and two large plasmids) and has a virulence profile similar to several of the O145 isolates in this study (*stx*-positive, *eae* subtype γ). At the time of this study, there were no publicly available genome sequences from STEC O145 isolated in NZ. Randomised Accelerated Maximum Likelihood (RAxML) Next-Generation [69] maximum-likelihood trees were generated of the core SNP alignment using a general time-reversible model and random seed to perform 20 tree searches using ten random and ten parsimony-based starting trees. The best-scoring maximum-likelihood tree was viewed in iTOL [63]. Roary [70] was used to identify the pangenome and the core and accessory genes in the *E. coli* serogroup O145 strains.

Accession numbers

The accession numbers for *E. coli* serogroup O145 strains whole genome sequenced in this study are listed in S3 Table and are deposited with NCBI under the BioProject number PRJNA435641.

Biolog phenotypic microarray assays

The Omnilog phenotypic microarray system (Biolog Inc, Hayward, California, USA) was used to investigate the metabolic characteristics of serogroup O145 strains. Serogroup O145 strains to be examined were selected using random sampling, stratified by the variables: *eae* subtype, ST, the geographic origin of isolation and a *stx*-positive or *stx*-negative genotype (Table 1). The plates were prepared as previously described [31], except the colonies were re-suspended at a light transmittance of 42%. Half of the isolates ($n = 14$) were tested in replicate (analysed on separate days) and two in duplicate (analysed on the same day) for the PM1 MicroPlatesTM, and four biological replicates were tested for the PM2A MicroPlatesTM. The raw Omnilog data was analysed using R 3.3.1 [64] and the packages 'opm' [71] and 'gplots' [72]. To compare carbon substrate utilisation between the strains, the end-point values per serogroup O145 strain ($n = 28$) for each carbon substrate on the phenotypic microarray plates ($n = 95$) was recorded and used to produce a cluster dendrogram using hierarchical clustering, with height indicating the distance between pairs.

Results and discussion

Population structure and genome composition of *E. coli* serogroup O145 strains

A total of 122 *E. coli* serogroup O145 strains were analysed, 53 sequenced in this study (S1 Table) and 69 publicly available genome sequences (S2 Table). Using the Achtman MLST system [73], a total of 14 distinct STs were identified for the 122 isolates (Fig 1). The predominant type was ST32 (83 of 122, 68.0%), followed by ST137 (8 of 122, 6.6%), ST17 and ST722 (7 of 122, 5.73% each), ST342 (5 of 122, 4.1%), ST10, ST48 and ST1877 (2 of 122, 1.64% each) and ST20, ST35, ST526, ST6529, ST7413 and ST unknown (1 of 122, 0.82% each). ST32 was also

Table 1. Serogroup O145 strains analysed using the Omnilog phenotypic microarray system.

Strain	MicroPlates™	Serotype ^a	Source	Origin	Sequence type	Virulence profile	<i>eae</i> subtype
Trh7	PM1, PM2A ^b	O145:H40	Human	Norway	ST-10	<i>eae</i>	β
TW07865	PM1 ^b , PM2A	O145:H28	Human	Germany	ST-137	<i>stx2, eae, ehxA</i>	γ
AGR718	PM1 ^c	O145:H46	Bovine	Manawatu, New Zealand	ST-137	<i>eae, ehxA</i>	γ
ERL020412	PM1 ^c	O145:H28	Human	New Zealand	ST-137	<i>eae, ehxA</i>	γ
16ER0267A	PM1 ^b , PM2A ^b	O145:H2	Human	Auckland, New Zealand	ST-17	<i>stx1, eae, ehxA</i>	ε
16ER0517A	PM1 ^b , PM2A	O145:H2	Human	Auckland, New Zealand	ST-17	<i>stx1, eae, ehxA</i>	ε
116B	PM1, PM2A	O145:H2	Bovine	Taranaki, New Zealand	ST-17	<i>eae, ehxA</i>	ε
188B	PM1	O145:H2	Bovine	Taranaki, New Zealand	ST-17	<i>eae, ehxA</i>	ε
267P	PM1 ^b , PM2A	O145:H2	Bovine	Taranaki, New Zealand	ST-17	<i>eae, ehxA</i>	ε
54B	PM1 ^b	O145:H2	Bovine	Taranaki, New Zealand	ST-17	<i>eae, ehxA</i>	ε
13ER3103A	PM1, PM2A	O145:H28	Human	Auckland, New Zealand	ST-32	<i>stx2, eae, ehxA</i>	γ
VC1281m	PM1 ^b , PM2A	O145:H28	Bovine	Canterbury, New Zealand	ST-32	<i>eae, ehxA</i>	γ
VC847m	PM1	O145:H28	Bovine	Manawatu-Wellington, New Zealand	ST-32	<i>eae, ehxA</i>	γ
13ER4824	PM1	O145:H28	Bovine	New Zealand	ST-32	<i>stx2, eae, ehxA</i>	γ
13ER5640	PM1, PM2A	O145:H28	Bovine	New Zealand	ST-32	<i>stx2, eae, ehxA</i>	γ
14ER2392	PM1, PM2A	O145:H28	Bovine	New Zealand	ST-32	<i>stx2, eae, ehxA</i>	γ
H12ESR01231	PM1 ^b , PM2A	O145:H28	Bovine	New Zealand	ST-32	<i>eae, ehxA</i>	γ
H12ESR01387	PM1, PM2A	O145:H28	Bovine	New Zealand	ST-32	<i>stx2, eae, ehxA</i>	γ
H12ESR03525	PM1, PM2A	O145:H28	Bovine	New Zealand	ST-32	<i>stx2, eae, ehxA</i>	γ
VC308m	PM1 ^b	O145:H28	Bovine	Northland, New Zealand	ST-32	<i>eae, ehxA</i>	γ
Trh30	PM1, PM2A	O145:H28	Human	Norway	ST-32	<i>eae, ehxA</i>	γ
VC1413m	PM1 ^b , PM2A ^b	O145:H28	Bovine	Southland, New Zealand	ST-32	<i>stx2, eae, ehxA</i>	γ
VC1506m	PM1, PM2A	O145:H28	Bovine	Southland, New Zealand	ST-32	<i>eae, ehxA</i>	γ
F5J	PM1	O145:H2	Environmental	Waikato, New Zealand	ST-32	<i>eae, ehxA</i>	γ
P2B1	PM1 ^b , PM2A	O145:H28	Environmental	Waikato, New Zealand	ST-32	<i>eae, ehxA</i>	γ
Trh42	PM1 ^b , PM2A	O145:H34	Human	Norway	ST-35	<i>eae</i>	ι
13ER6723A	PM1 ^b , PM2A	O145:H34	Human	Auckland, New Zealand	ST-722	<i>stx2, eae</i>	ι
R249-1	PM1 ^b , PM2A ^b	O145:H34	Human	Australia	ST-722	<i>eae</i>	ι

a: O antigen: H antigen

b: MicroPlates™ were completed in replicate (analysed on separate days)

c: MicroPlates™ were completed in duplicate (analysed on the same day)

<https://doi.org/10.1371/journal.pone.0235066.t001>

found to be predominant (230 of 239, 96.2%) in a study of O145:H28 strains [44]. Each ST was also associated with a specific *eae* subtype, as highlighted in the neighbour-joining tree generated from SNPs identified in 51 genes encoding the ribosome protein subunits (Fig 1).

A summary of the genome composition for the *E. coli* serogroup O145 strains (n = 122) is shown in Fig 2, indicating genome length (bp), coding sequence counts (CDS) and GC content (%). For all three parameters, clustering occurred according to *eae* subtype. The shortest genome lengths were associated with *eae*-negative, *eae* subtype ι and α2 strains (4,640,737–5,010,707 bp). Similarly, the *eae*-negative and three *eae* subtype ι strains also had the lowest CDS counts (3,687–3,946). The CDS counts for the remaining *eae* subtype ι strains (n = 6), *eae* subtype α2, β, ε and λ strains ranged from 4,087 to 5,485. The *eae* subtype ε strains had both the longest genome length (5,400,785–5,494,427 bp) and highest CDS counts (5,311–5,434). The *eae*-negative strains had a slightly higher GC content (50.81%), however, the GC content of all serogroup O145 strains was relatively similar (50.22–50.81%).

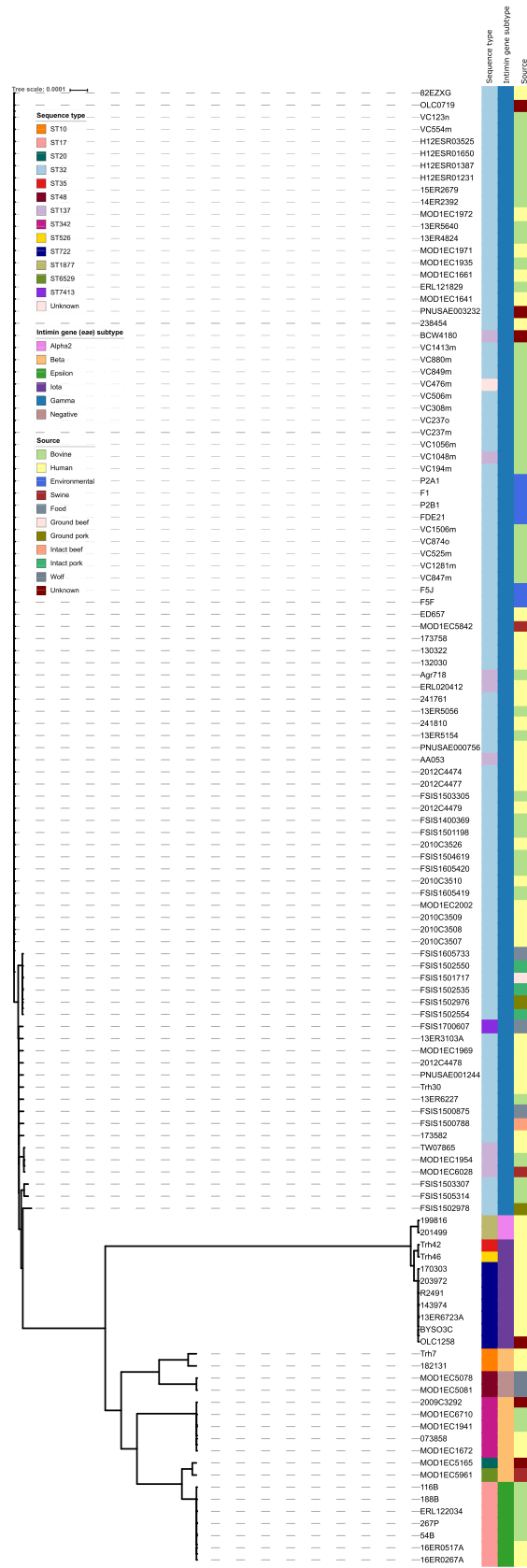


Fig 1. Neighbour-joining phylogeny constructed using *in silico* ribosomal multi-locus sequence typing. A Neighbour-joining tree of ribosomal multi-locus sequence types (rMLST) generated from single nucleotide polymorphisms identified in 51 genes encoding the ribosome protein subunits (*rps*, *rpm* and *rpl*). The *in silico* rMLST analysis was visualised using neighbour-joining methods in SplitsTree and edited using the iTOL (Interactive Tree of Life) webserver. Isolate metadata is included for sequence type, *eae* subtype and isolation source, as indicated by the colour keys.

<https://doi.org/10.1371/journal.pone.0235066.g001>

Genetic characterisation of virulence factors and antimicrobial resistance

The virulence profiles of the serogroup O145 isolates ($n = 122$) (S4 Table) broadly cluster according to both *eae* subtype and ST (Fig 3). The *eae* subtype γ ($n = 93$) strains cluster together, with some variation according to ST and at the strain level. WGS data analysis of the *eae* subtype γ strains indicates the presence of between 13 and 22 of the 37 virulence factors, as shown in Fig 3. Notably, the plasmid-associated virulence factor *etpD* was present in the eight ST137 strains and absent in the remaining ST32 and ST7413 ($n = 85$) *eae* subtype γ strains.

The presence or absence of 37 virulence genes identified using VirulenceFinder, which differed between strains, were used to make a Neighbour-joining tree using the Jaccard index and converted to Newick format using R 3.6.0 and the packages 'vegan' and 'ctc', respectively. The tree was edited using the iTOL webserver and isolate metadata was included for *eae* subtype, sequence type (ST) and isolation source.

The *eae* subtype $\alpha 2$ ($n = 2$), ι ($n = 9$), *eae*-negative ($n = 2$), β ($n = 9$) and ϵ ($n = 7$) strains broadly form separate clusters, with similarity within each cluster correlating with ST. The *eae* subtype ϵ strains ($n = 7$, ST17) carry between 17 and 19 virulence factors, and the *eae* subtype β strains carry between 8 and 19 virulence factors. In comparison to other serogroup O145 strains, the $\alpha 2$, ι , β (ST10) and *eae*-negative strains carry fewer virulence factors; *eae* subtype $\alpha 2$, ι , β (ST10) and *eae*-negative strains carry 7, 8 to 11, 8 and 9, and 2 and 3 of the 37 virulence factors listed in Fig 3, respectively. The low number of virulence factors carried by the *eae*-negative and *eae* subtype $\alpha 2$, ι and β (ST10) strains may be partially due to the absence of plasmid-acquired virulence factors, such as *etpD*, *ehxA* and *katP* (Fig 3). Five strains (ST722, *eae* subtype ι) are defined as typical EPEC due to the presence of the LEE pathogenicity island and the bundle forming pilus biosynthesis operon [74].

Of the 122 isolates, 65 were *stx*-positive (53.3%) including *stx* variants *stx1a* (12 of 65, 18.5%), *stx2a* (35 of 65, 53.8%), *stx2c* (2 of 65, 3.1%), *stx2d* (4 of 65, 6.2%) and *stx2f* (4 of 65, 6.2%). Eight isolates were both *stx1* and *stx2* positive (*stx1a*, *stx2a*, 1 of 65, 1.5%; *stx1a*, *stx2d*, 7 of 65, 10.7%). The remaining 57 isolates were *stx*-negative. Similar *stx* variants were detected in an analysis of 239 O145:H28 strains [44]; however, *stx2f* was not detected and was only detected in four O145:H34 strains in this study. It has been suggested that different *stx* variants may have varying levels of virulence, for example, *stx2a* has been associated with an increased risk of developing HUS [75, 76], in comparison *stx1* variants were associated with a lower risk [75]. STEC possessing the *stx2f* toxin, first described in pigeons [77], have been described as an emerging pathogen [78]. Preliminary epidemiological data suggested infections caused by *stx2f*-positive STEC were associated with mild clinical disease [79], however, one case of HUS caused by a STEC strain possessing *stx2f*, and *eae* positive and *ehxA* negative, has been reported [80]. Three of the four *stx2f* isolates were from human clinical cases, with the source of the remaining isolate being unknown.

Plasmids were detected in 113 out of 122 strains (S5 Table), with distinct nucleotide matches ($\geq 95\%$ identity and $\geq 60\%$ coverage) of plasmid incompatibility factors indicative of separate plasmids. A single plasmid was identified in 89 O145 strains, two plasmids in 18, three plasmids in five and a single O145 strain was identified with four plasmids. The most commonly detected plasmid incompatibility factor was IncFIB (AP001918) which was

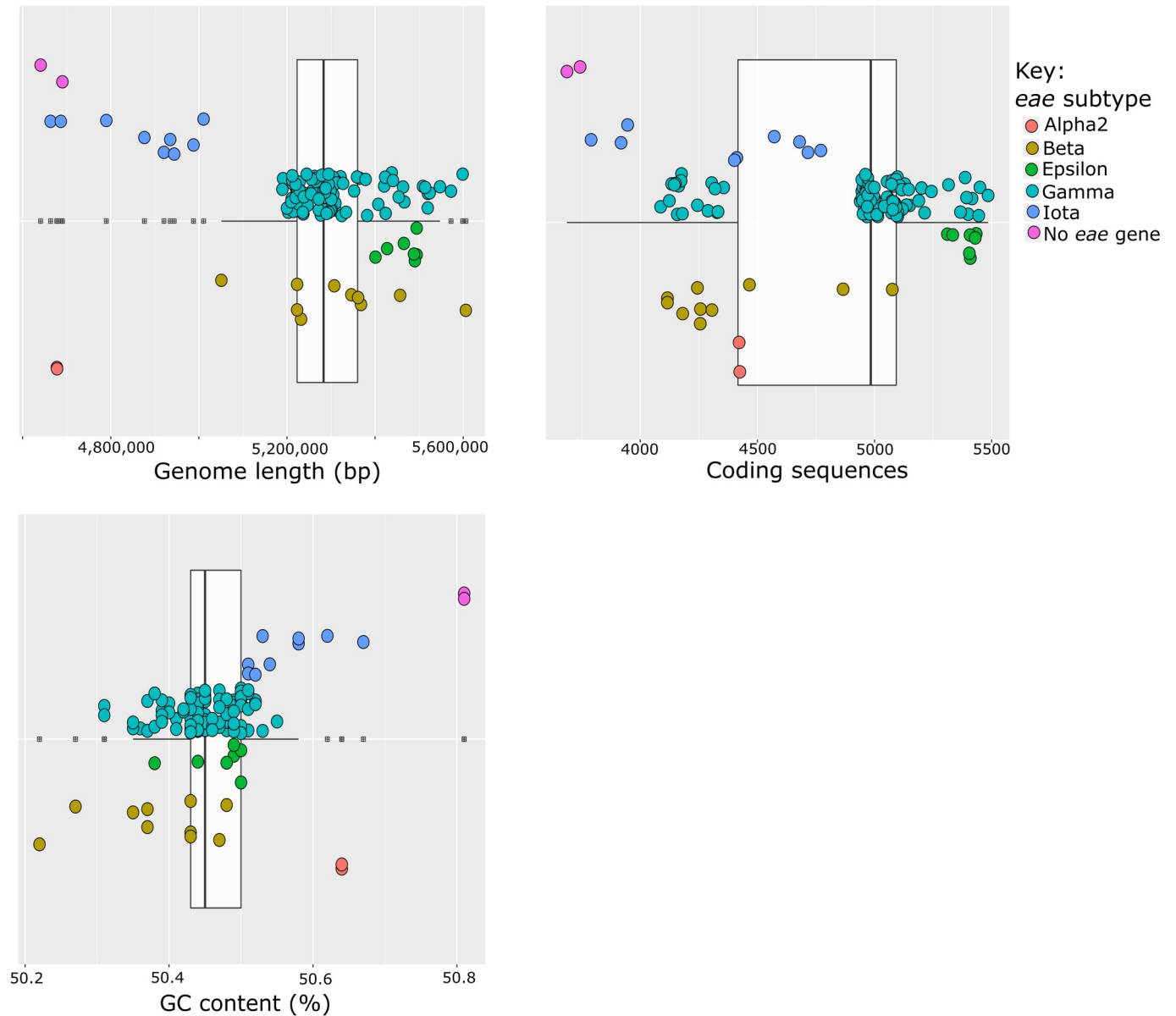


Fig 2. Box and whisker plots indicating the genome composition of *E. coli* serogroup O145 strains (n = 122). The box and whisker plots indicate the genome length (bp), number of coding sequences and GC content (%) for the serogroup O145 strains (n = 122). The box spans the interquartile range, with the lower and upper quartiles indicated by the ends of the box (from left to right) and the median by the vertical line inside the box. Each data point is shown on the plots and has been grouped according *eae* subtype in the y axes, as indicated by the figure key.

<https://doi.org/10.1371/journal.pone.0235066.g002>

detected in 96.5% (109 out of 113) of the strains. Interestingly, the IncFIB and IncB/O/K/Z_3 plasmids were found to be highly conserved within a population of 239 O145:H28 strains [44]. The ubiquity of the plasmid incompatibility factor IncFIB detected in this study may suggest that this plasmid is conserved within the O145 serogroup. Plasmid negative strains (n = 9), belonged to *eae* subtype $\alpha 2$ (n = 2), ι (n = 3), and γ (n = 4).

The LEE pathogenicity island integration sites were identified in 73 out of 120 serogroup O145 strains and are listed in S6 Table. A tRNA *pheV* integration site was identified for *eae* subtype ϵ (n = 7), β (n = 2) and γ (n = 2) strains, *pheU* for the *eae* subtype β (n = 1) strain and *selC* for the *eae* subtype $\alpha 2$ (n = 2), β (n = 1), ι (n = 7), and γ (n = 49) strains. Although the *selC*

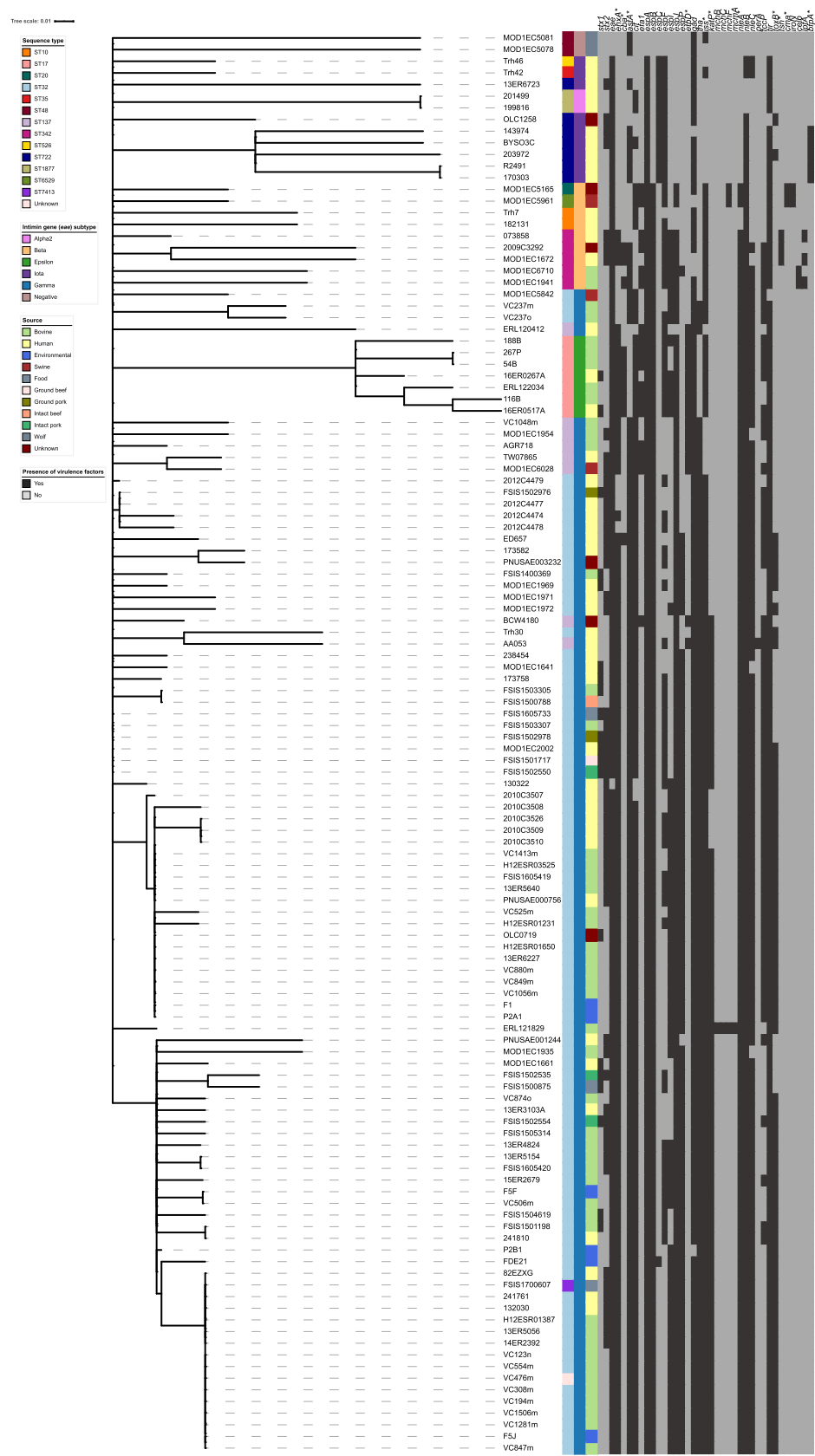


Fig 3. Neighbour-joining tree constructed using the presence or absence data from 37 virulence genes identified by the center for genomic epidemiology VirulenceFinder webserver.

<https://doi.org/10.1371/journal.pone.0235066.g003>

LEE integration site for *eae* subtype ι strains has not previously been identified, the precise LEE integration site could not be determined for one *eae* subtype ι strain Trh42 but was located near tRNA *leu*. This potential LEE insertion site was also observed for the *eae* subtype β strain 73858. The LEE insertion site could not be determined in 47 strains, likely due to incomplete genome assemblies as a result of using short-read sequencing. The common *stx*-bacteriophage insertion sites for serogroup O145 [81] were analysed to identify whether these sites were occupied or available in *stx*-negative strains. Although some sites were vacant in the majority of *stx*-negative strains, not all insertion sites could be detected. This was possibly due to the genes surrounding these sites being unannotated; the sites being occupied and the insertion site therefore disrupted; or the genome assembly being incomplete. The detection of *stx*-bacteriophage insertion sites in serogroup O145 isolates is problematic due to multiple potential insertion sites, variations in prophage structure and variation between integration sites, including between phage which encode the same Stx subtype [44]. In addition, “Stx₂-like” prophage, which appear to be defective as a result of nonsense mutations in the *stx*₂A subunit or absent *stx*₂A and *stx*₂B genes, have been detected in serogroup O145 strains [42], further complicating the detection of *stx*-bacteriophage insertion sites in this serogroup.

The serogroup O145 genome sequences (n = 122) were examined for antibiotic resistance genes using ResFinder [56]. Twenty-three of 122 strains (18.9%) carried one or more resistance genes with resistance to up to a maximum of five classes of antibiotics being detected (Table 2). These strains were from a variety of sources, belonged to multiple STs and *eae* subtypes and were isolated in the USA (n = 17), UK (n = 2), NZ (n = 1), Germany (n = 1) and Canada (n = 1) with the geographic isolation of one strain being unknown. Notably, all 23 of the resistant strains carried genes conferring aminoglycoside resistance. The variation in the carriage of antibiotic resistance genes in the serogroup O145 genomes may be a result of varying selective pressures that may impact the development and transmission of resistance, such as antimicrobial use in different geographical regions. For example, only one out of 35 serogroup O145 strains isolated from bovine sources in NZ carried an antibiotic resistance gene; which may be reflective of the low antimicrobial use in the dairy industry in NZ [82]. As a result of this variability, it is unlikely that antimicrobial resistance is a property that could be utilised in the development of a media for the differentiation of serogroup O145.

Comparative genomics of the 122 serogroup O145 strains from diverse geographical regions and distinct hosts/sources highlighted the genetic heterogeneity within this serogroup. The strains analysed belonged to 14 different STs, carried between 2 and 22 *E. coli* associated virulence factors and 18.9% (23 out of 122) carried genes known to confer antibiotic resistance (Figs 1 and 3, and Table 2). Genome analysis indicated strains of the same *eae* subtype had a similar genome size and number of CDS counts (Fig 2), consistent with other “Top 7” serogroups [42–44, 67, 83]. However, the *eae*-negative and *eae* subtype α 2 and ι strains had smaller genomes (Fig 2) compared to other serogroup O145 strains. The WGS data from α 2 and ι strains in this study are consistent with the genome parameters of two further atypical (bundle forming pilus negative) EPEC (aEPEC) O145:H34 isolates [84] with identical STs and *eae* subtypes (α 2 and ι) recently analysed (S1 and S2 Tables).

Core and pangenome analysis

The number of conserved and total genes present in serogroup O145 strains (n = 122) is shown in Fig 4. The core is defined as genes present in all strains (100%) strains, the soft-core

Table 2. Detection of genes conferring resistance to certain classes of antibiotics.

Strain	Origin	Source	Sequence type	Antibiotic classes				
				Aminoglycoside	Beta-lactam	Phenicol	Sulphonamide	Tetracycline
182131	UK	Human	10	+	+	-	-	-
MOD1EC5165	Unknown	Unknown	20	+	+	-	+	-
OLC0719	Canada	Unknown	32	+	-	-	+	+
VC874o	NZ	Bovine	32	+	-	-	+	+
173758	UK	Human	32	+	-	-	+	-
FSIS1605420	USA	Bovine	32	+	-	+	+	+
FSIS1605419	USA	Bovine	32	+	-	+	+	+
MOD1EC1935	USA	Bovine	32	+	-	-	+	+
2010C3507	USA	Human	32	+	-	+	+	+
2010C3508	USA	Human	32	+	-	+	+	+
2010C3509	USA	Human	32	+	-	+	+	+
2010C3510	USA	Human	32	+	-	+	+	+
2010C3526	USA	Human	32	+	-	+	+	+
MOD1EC1971	USA	Human	32	+	-	+	+	+
MOD1EC1972	USA	Human	32	+	-	+	+	+
MOD1EC5842	USA	Swine	32	+	-	-	+	+
MOD1EC1954	Germany	Bovine	137	+	+	+	+	+
TW07865	USA	Human	137	+	+	+	+	+
MOD1EC6028	USA	Swine	137	+	+	+	+	+
MOD1EC6710	USA	Bovine	342	+	+	-	+	+
MOD1EC1941	USA	Bovine	342	+	+	-	+	+
2009C3292	USA	Unknown	342	+	-	-	-	-
MOD1EC5961	USA	Swine	6529	+	+	-	+	+

<https://doi.org/10.1371/journal.pone.0235066.t002>

as genes present in between ≥ 115 and ≤ 121 strains ($\geq 95\%$ — $\leq 99\%$) strains, the shell as genes present in ≥ 18 - < 115 strains ($\geq 15\%$ — $< 95\%$) strains and the cloud as genes present in < 18 strains ($> 0\%$ — $< 15\%$) (Fig 5). The pangenome analysis suggested a core gene set of 3,036 genes, a soft-core of 252 genes, a shell of 2,916 genes and a cloud of 14,942 genes (Fig 5).

Pangenome analysis of serogroup O145 strains ($n = 122$) supported the diverse genetic heterogeneity within this serogroup (Fig 5). For a given population, when additional genome sequences are included, an open pangenome will identify un-characterised genes, whereas a closed pangenome will have approached a constant number [85]. The serogroup O145 pangenome of $> 14,000$ genes was open (Fig 4), demonstrating the genetic heterogeneity of this dataset, and is consistent with the identification of distinct phylogenetic lineages. The number of core genes reported for *E. coli* varies among studies and ranges from 1,472 to 5,173 [43, 44, 86–88]. Although pangenomes consisting of $> 13,000$ genes have been reported for *E. coli* [86, 88], pangenome analysis of 325 *E. coli* O26 genome sequences and 239 O145:H28 strains indicated an open pangenome with an accessory genome of only 8,804 genes [88] and 9,342 [44], respectively. These core genome variations are likely to be due to factors such as the number of genomes analysed and the genetic similarity of the strains included for comparison. For example, analysis of two serogroup O145:H28 strains identified a large core gene set of 5,173 as the two strains are likely to be genetically very similar [43] and the core gene set of 239 O145:H28 strains was 3,804 [44]. In addition, different software and identity thresholds can be used to define pangenomes, for example, a study of 53 *E. coli* genomes identified a core genome of 1,472 when reporting gene families rather than individual genes [86]. Therefore, such

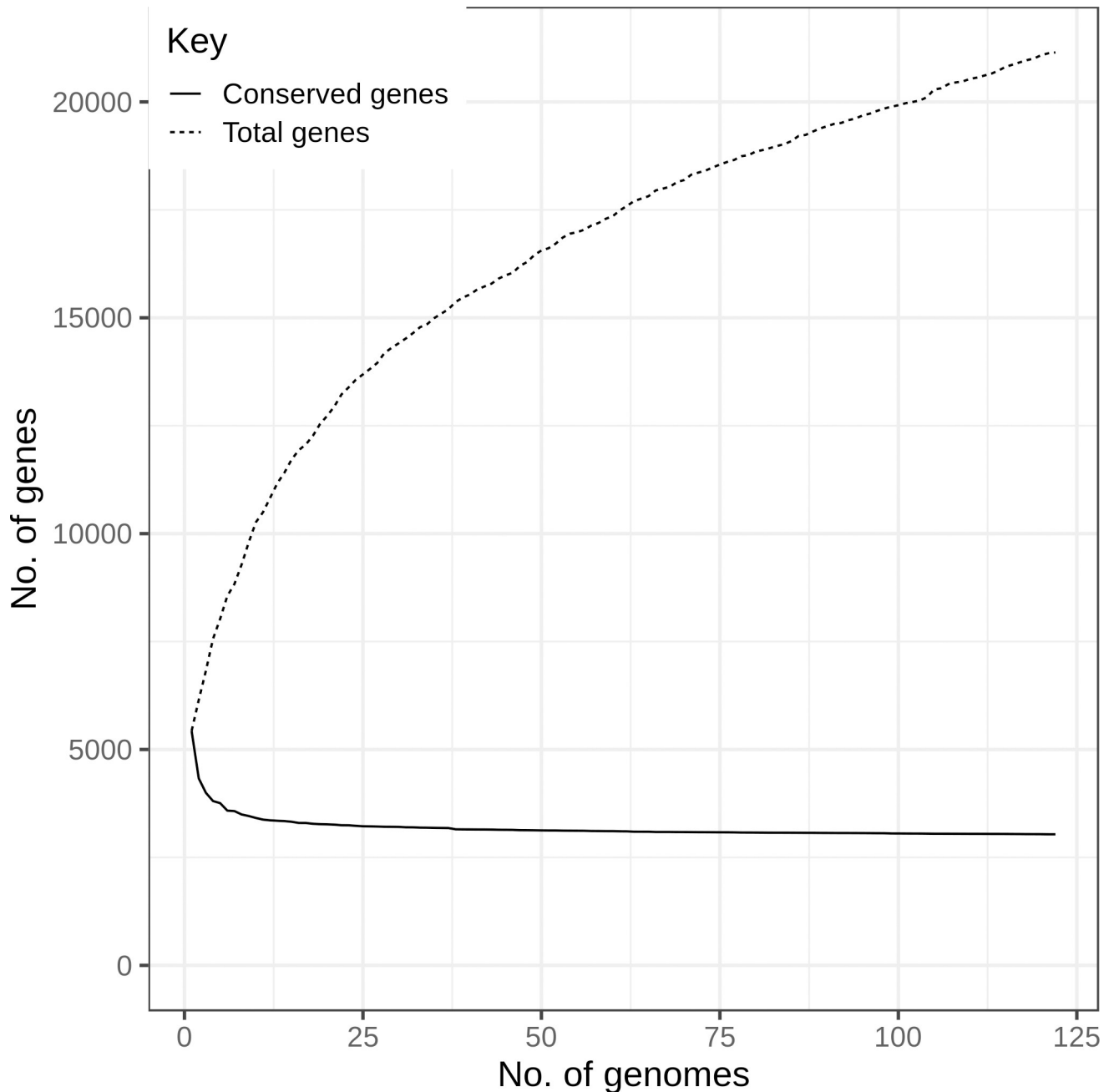


Fig 4. Comparison of the number of conserved and total genes in the serogroup O145 pangenome with increasing number of genomes. This analysis indicates the effect an increasing number of serogroup O145 genomes included in the analysis has on the number of conserved and total genes.

<https://doi.org/10.1371/journal.pone.0235066.g004>

parameters should be considered when comparing between studies. Pangenome analysis has indicated a significant proportion of the *E. coli* genome as comprised of diverse genes. In O145:H28 strains ($n = 239$), plasmid- and phage-associated genes comprised a large proportion of the pangenome [44]. This highlights the genetic heterogeneity of *E. coli* through HGT, incorporation of phage genetic material and through gene loss or duplication that can lead to genetically diverse populations, even within the same serogroup.

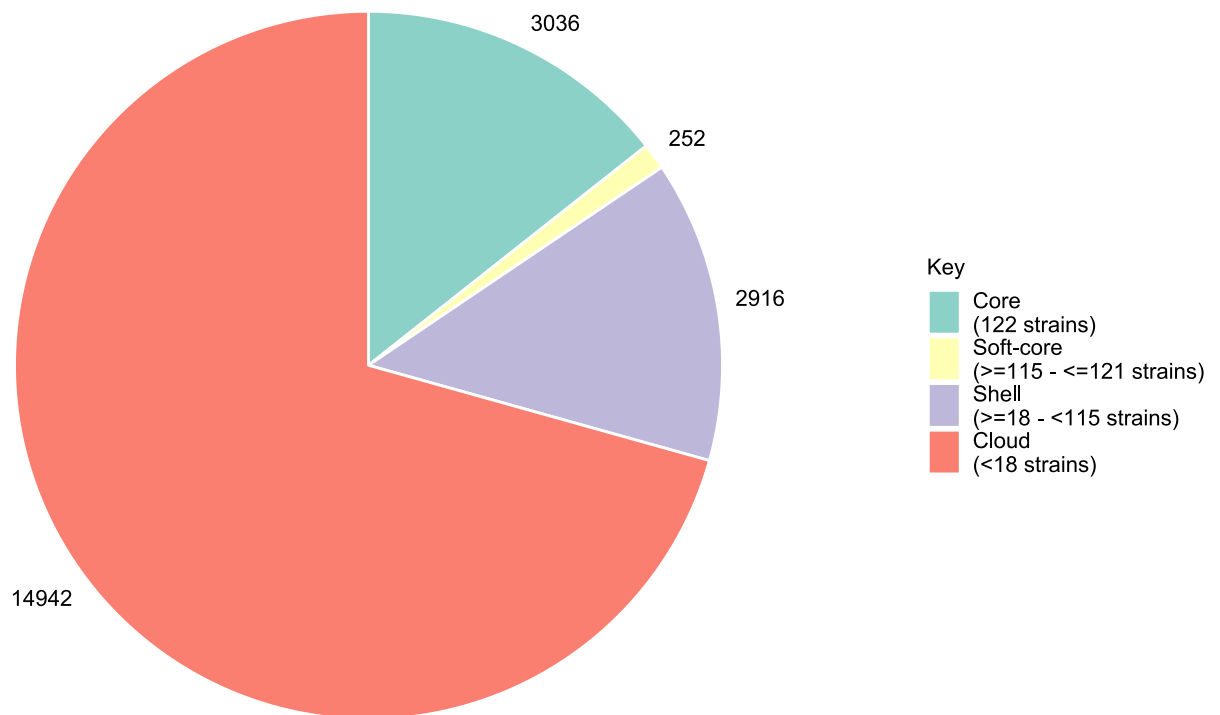


Fig 5. The pan-genome composition of serogroup O145 strains (n = 122). Pan genome composition of serogroup O145 strains (n = 122) showing the core as genes present in all strains (100%), the soft-core as genes present in ≥ 115 and ≤ 121 strains ($\geq 95\%$ — $\leq 99\%$), the shell as genes present in ≥ 18 - < 115 strains ($\geq 15\%$ — $< 95\%$) and the cloud < 18 strains ($> 0\%$ — $< 15\%$).

<https://doi.org/10.1371/journal.pone.0235066.g005>

Core SNP analysis

Core SNP analysis of the 122 serogroup O145 strains (Fig 6) separated the strains into five phylogenetic clades which correlated with both *eae* subtype and ST. Clade 1 consisted of *eae* subtypes ι (n = 9) and $\alpha 2$ (n = 2). Clade 2 consisted of *eae* subtype β strains belonging to ST10 (n = 2) and the two *eae*-negative strains. *eae* subtype β strains belonging to ST342 (n = 5) formed Clade 3, whilst Clade 4 consisted of *eae* subtype ϵ strains (n = 7) and two *eae* subtype β strains belonging to ST20 and ST6529. The largest group, Clade 5, consisted of *eae* subtype γ strains (n = 93) and 139,513 SNPs were identified within the core genome of these strains (n = 122). Genome-wide core SNP analysis of 69 *E. coli* strains across 31 serogroups also identified significant genetic diversity with 86,350 SNPs identified across 1,371 core genes [42]. These results support the hypothesis of the evolution of distinct *E. coli* phylogenetic lineages with different *eae* subtypes, with subsequent mutations and/or HGT resulting in a large amount of genetic heterogeneity.

To resolve the phylogeny of the serogroup O145 strains, a core SNP analysis was performed on the *eae* subtype γ strains (n = 93) (Fig 7) identifying 6,534 SNPs, accounting for only 4.7% of the variation seen in the core genome of serogroup O145 strains (n = 122). This indicates these strains are genetically more similar compared to the other serogroup O145 strains analysed in this study.

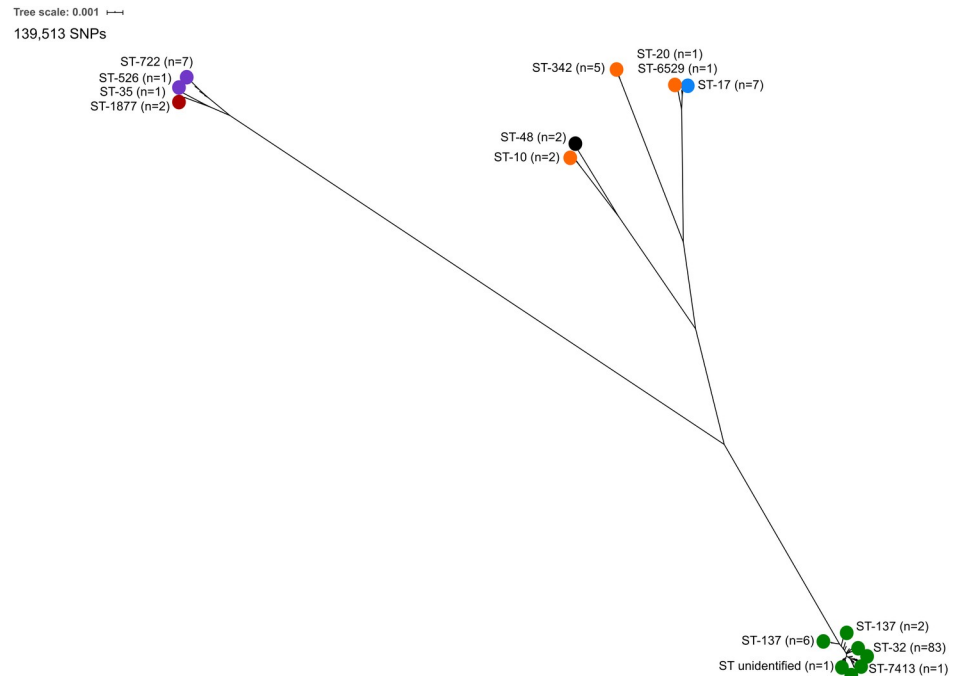


Fig 6. Maximum-likelihood tree of core single nucleotide polymorphism analysis of serogroup O145 strains (n = 122). RAxML Next-Generation maximum-likelihood tree of the core single nucleotide polymorphism (SNP) genome analysis from serogroup O145 genome sequences (n = 122). The tree was generated using 139,513 core SNPs. Metadata is included for *eae* subtype and sequence type, and additional information for each isolate can be found in [S1](#) and [S2](#) Tables.

<https://doi.org/10.1371/journal.pone.0235066.g006>

Utilisation of carbon substrates

A dendrogram was generated according to the clustering of the utilisation of carbon substrates on the PM1 MicroPlates™, and significant metabolic variation was observed between serogroup O145 strains (Fig 8). There was no relationship between the utilisation of specific carbon substrates and whether a strain was *stx*-positive or *stx*-negative. Similarly, strains of human and bovine origin did not cluster together (Fig 8, S2 Fig). Instead, clustering of serogroup O145 strains by carbon utilisation was broadly associated with *eae* subtype and ST, which is consistent with the genomic analysis and highlights both the metabolic and genetic heterogeneity of this serogroup.

The utilisation of carbon substrates and subsequent cluster analysis of metabolic characteristics displayed by serogroup O145 strains (n = 20) on the PM2A MicroPlates™ (S3 Fig) was similar to that seen with the PM1 MicroPlates™. However, far fewer carbon substrates (23.2%; 22 out of 95) were utilised by ≥ 1 *E. coli* strain on the PM2A MicroPlates™. The clustering observed was similar when the replicate (analysed on separate days) and duplicate (analysed on the same day) data was included, however, there was contrasting utilisation of some carbon substrates between replicates and duplicates on the PM1 (11.6%; 11/95) and PM2A (4.2%; 4/95) MicroPlates™ (S2 and S3 Figs). Notably, the utilisation of some substrates, such as D-psi-cose and glucuronamide, was inconsistent between replicates and duplicates for multiple serogroup O145 strains, and due to the inconsistency in utilisation, these substrates are likely to be unsuitable for use in a differential media for serogroup O145.

Analysis of the utilisation of 190 carbon substrates (PM1 MicroPlates™, PM2A MicroPlates™) failed to identify any specific carbon substrates that would be likely to definitively

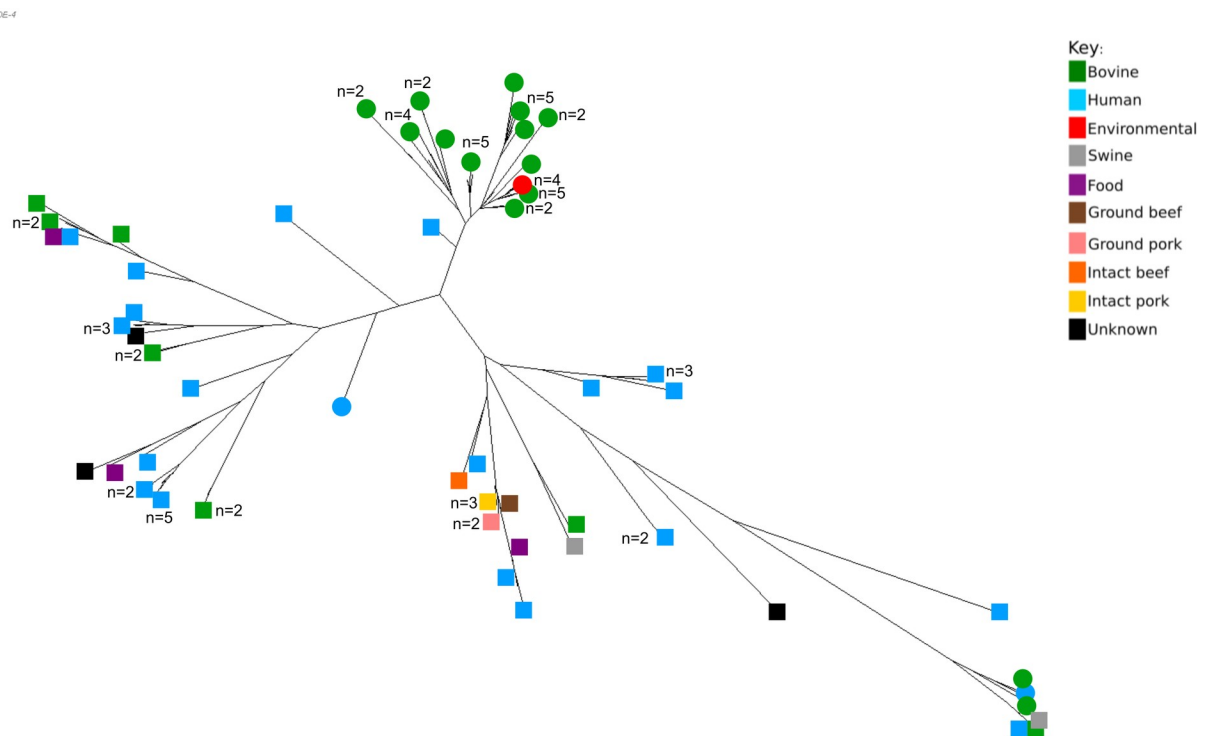


Fig 7. Maximum-likelihood tree of core single nucleotide polymorphism analysis from *eae* subtype γ serogroup O145 strains (n = 93). RAxML Next-Generation maximum-likelihood tree of the core single nucleotide polymorphism (SNP) genome analysis from serogroup O145 *eae* subtype γ genome sequences (n = 93). The tree was generated using 6,534 core SNPs. Circular and square nodes indicate genomes of New Zealand and non-New Zealand origin, respectively. Metadata is included for isolation source, and additional information for each isolate can be found in S1 and S2 Tables.

<https://doi.org/10.1371/journal.pone.0235066.g007>

differentiate serogroup O145 from other *E. coli*. However, several carbon substrates were identified which are utilised by a large proportion of serogroup O145 strains, which, when coupled with current molecular and culture-based methods, could aid in the identification of presumptive *E. coli* serogroup O145 isolates. For example, D-serine is utilised by *eae* subtype γ (ST32 and ST137; n = 18) and β (ST10; n = 1), and D-malic acid is utilised by *eae* subtypes γ (ST32 and ST137; n = 18) and ϵ (ST17; n = 6) (S2 Fig). These carbon substrates warrant subsequent testing with additional serogroup O145 strains and further non-O145 strains by including them as the main energy source in a minimal medium or a selective enrichment media.

However, there is variation in carbon substrate utilisation within serogroup O145 strains of the same ST and *eae* subtype. This suggests that the ability to metabolise certain substrates has either been lost or gained independently on multiple occasions by entirely separate lineages of serogroup O145. Furthermore, this also suggests that HGT events can lead to phenotypic traits that are not homogenous between members of the same phylogenetic cluster (either by MLST or SNP-based typing). Contrasting phenotypic traits leading to variations in metabolic activity may also arise via point mutations. Other traits, such as virulence factors and antimicrobial resistance, are also highly heterogeneous within the *E. coli* serogroup O145 strains studied, highlighting the limitations of making assumptions about isolates belonging to one serogroup from genetic data with limited phylogenetic resolution, such as the seven-gene MLST schemes. As a result, phenotypes cannot always be assumed from genotype, and therefore both phenotype and genotype testing are required to understand the epidemiological origin and potential virulence-associated consequences.

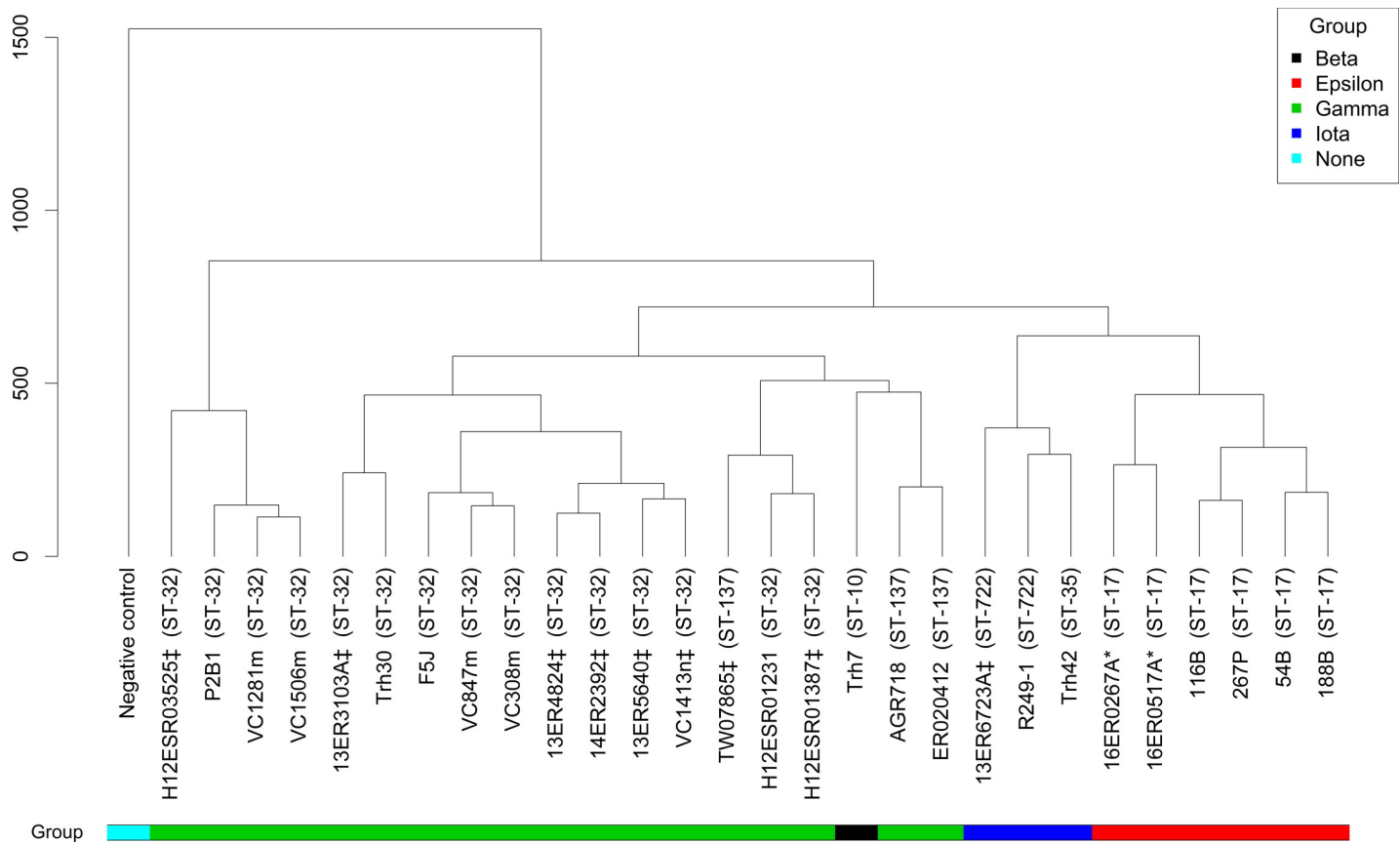


Fig 8. Cluster dendrogram showing the similarities of *E. coli* serogroup O145 strains based on their carbon utilisation profile using PM1 MicroPlate™. The end-point values per serogroup O145 strain ($n = 28$) for each carbon substrate on the phenotypic microarray plates ($n = 95$) was recorded and used to produce a cluster dendrogram using hierarchical clustering, with height indicating the distance between pairs. Metadata is included for sequence type, *eae* subtype and whether the strains were toxigenic or non-toxigenic. Sequence type is shown in brackets, *stx1* positive as * and *stx2* positive as ‡.

<https://doi.org/10.1371/journal.pone.0235066.g008>

Previous studies have examined the carbon utilisation of the “Top 7” *E. coli* serogroups [31, 89], demonstrating the variability in carbon utilisation and how this variability has hindered the development of a differential media for many non-O157 serogroups. However, this is the first study to examine the growth of genetically diverse serogroup O145 strains with a broad range of carbon substrates. In other studies undertaken to compare metabolic capabilities of several STEC serogroups, O145 isolates ($n = 3$) showed little variation in the number of carbon substrates utilised and β -hydroxy-butyric acid was identified as a candidate metabolite for differentiation of O145 from other clinically relevant STEC [31]. However, our study of genetically diverse O145 demonstrated that the utilisation of β -hydroxy-butyric acid was variable (S2 Fig) with some O145 (*eae* subtype β ($n = 1$), ϵ ($n = 4$), ι ($n = 3$) and γ ($n = 1$, ST137)) strains unable to utilise β -hydroxy-butyric acid as the only carbon source. The three O145 strains examined previously [31] displayed similarities in carbon utilisation with O157:H7 strains (*eae* subtype γ), were all O145:H28, and are likely to be *eae* subtype γ , which may account for the limited variation observed in comparison to the heterogeneity seen in our study [31]. In contrast, carbon utilisation of *E. coli* ($n = 153$), *Shigella* ($n = 16$), *Escherichia fergusonii* ($n = 2$), *Escherichia albertii* ($n = 1$) and cryptic *Escherichia* Clade strains ($n = 6$) in another study was shown to be highly variable [90]. The carbon substrate utilisation diversity observed in our study suggests the development of diagnostic media permitting the selective growth and/or

differentiation of all O145 strains based on carbon source utilisation could be difficult. However, if the high prevalence of serogroup O145 *eae* subtype γ strains seen in this study (S1 and S2 Tables) is reflected in their overall zoonotic potential, the development of a medium solely for this *eae* subtype may be beneficial.

Conclusion

In this study, we used comparative genomics and carbon substrate utilisation to understand the genomic epidemiology and metabolic profiles of *E. coli* serogroup O145, respectively. We found considerable genetic heterogeneity within serogroup O145 strains according to the relative abundance of virulence factors, core genome SNPs, and pangenome analysis. ST and *eae* subtype provided an indication of genetic heterogeneity suggesting these parameters are good indicators to separate distinct *E. coli* phylogenetic lineages. The genetic heterogeneity within these strains also provided evidence of a broad virulence continuum; *stx2a*- and *eae*-positive strains are implicated as the cause of severe human disease, both typical and atypical EPEC are associated with mild diarrhoeal disease or asymptomatic carriage, while other serogroup O145 isolated from wolves lacked many STEC-associated virulence factors and appeared to be host-associated and unlikely zoonoses. Carbon substrate utilisation by a subset of *E. coli* serogroup O145 strains demonstrated considerable metabolic variation, which showed a remarkable association with *eae* subtype and ST, consistent with the genomic data. Several carbon substrates, such as D-serine and D-malic acid, were identified which are utilised by many serogroup O145 strains including *eae* subtype γ , the predominant *eae* subtype identified in this study. These carbon substrates, coupled with molecular tests to detect O145-specific *wzx* and *wzy* gene sequences, could provide targets for further investigation in media to assist in the identification of presumptive *E. coli* serogroup O145 strains. Further testing with additional non-O145 isolates is required to test this hypothesis.

Supporting information

S1 Table. Bacterial strains whole genome sequenced in this study.

(DOCX)

S2 Table. Publicly available genome sequences analysed in this study.

(DOCX)

S3 Table. Accession numbers for *E. coli* serogroup O145 stains whole genome sequenced in this study.

(DOCX)

S4 Table. Presence and absence data for serogroup O145 genomes (n = 122) from 37 virulence genes identified by the center for genomic epidemiology VirulenceFinder webserver.

(DOCX)

S5 Table. Plasmid incompatibility factors identified from serogroup O145 whole genome sequence data (n = 122).

(DOCX)

S6 Table. *E. coli* tRNA integration site for the Locus for Enterocyte Effacement (LEE) pathogenicity island for serogroup O145 genomes (n = 120).

(DOCX)

S1 Fig. Box and whisker plots indicating the assembly statistics of *E. coli* serogroup O145 strains (n = 122). The box and whisker plots indicate the number of contigs, largest contig

and N_{50} value for the serogroup O145 strains ($n = 122$). Each data point is shown on the plots and has been colour coded according to *eae* subtype, as indicated by the figure key.

(PNG)

S2 Fig. Heat-map showing *E. coli* serogroup O145 strains carbon utilisation profiles using PM1 MicroPlate™ with replicates and duplicates. Heat-map of PM1 carbon substrate metabolism over a 24-hour incubation period at 37°C by serogroup O145 strains. The end-point utilisation values (Omnilog Units) were grouped into the following three categories: 0–50 representing no utilisation, 51–150 representing moderate utilisation and 151–400 representing extensive utilisation, as indicated by the colour key. Each strain ($n = 28$, $n = 14$ replicates, $n = 2$ duplicates) is indicated on the right and the 95 carbon substrates listed along at the foot of the figure. Metadata is included for *eae* subtype, sequence type, source and whether the strains were toxigenic. *eae* subtype on the left is represented by the colour key, NA is not applicable, sequence type is shown in brackets, isolate source indicated by the colour boxes next to the label name with black, red and blue boxes representing bovine, human and environmental sources, respectively and *stx1* positive as * and *stx2* positive as ‡.

(PNG)

S3 Fig. Heat-map showing *E. coli* serogroup O145 strains carbon utilisation profiles using PM2A MicroPlate™ with replicates. Heat-map of PM2A carbon substrate metabolism over a 24-hour incubation period at 37°C by serogroup O145 strains. The end-point utilisation values (Omnilog Units) were grouped into the following three categories: 0–50 representing no utilisation, 51–150 representing moderate utilisation and 151–400 representing extensive utilisation, as indicated by the colour key. Each strain ($n = 20$ and $n = 4$ replicates) is indicated on the right and the 95 carbon substrates listed along at the foot of the figure. Metadata is included for *eae* subtype, sequence type, source and whether the strains were toxigenic. *eae* subtype on the left is represented by the colour key, NA is not applicable, sequence type is shown in brackets, isolate source indicated by the colour boxes next to the label name with black, red and blue boxes representing bovine, human and environmental sources, respectively *stx1* positive as * and *stx2* positive as ‡.

(PNG)

Acknowledgments

We would like to thank Colleen Ross and Delphine Rapp (AgResearch Limited, Palmerston North, New Zealand), Hugo Strydom, Muriel Dufour and Naveena Karki (Institute of Environmental Science and Research Ltd, Wallaceville, New Zealand), Kaylene Larking (Meat Industry Association, Wellington, New Zealand) and Jan Egil Afset (Norwegian University of Science and Technology and St. Olav's University Hospital, Trondheim, Norway) for generously providing serogroup O145 isolates for use in this study.

Author Contributions

Conceptualization: Patrick J. Biggs, Anne C. Midwinter, A. Springer Browne, Gale Brightwell, Adrian L. Cookson.

Data curation: Rose M. Collis, Patrick J. Biggs, A. Springer Browne, David A. Wilkinson, Adrian L. Cookson.

Formal analysis: Rose M. Collis, Patrick J. Biggs, A. Springer Browne, David A. Wilkinson, Nigel P. French, Adrian L. Cookson.

Funding acquisition: Rose M. Collis, Gale Brightwell, Adrian L. Cookson.

Investigation: Rose M. Collis, Patrick J. Biggs, Anne C. Midwinter, A. Springer Browne, Adrian L. Cookson.

Methodology: Rose M. Collis, Patrick J. Biggs, Anne C. Midwinter, A. Springer Browne, Adrian L. Cookson.

Project administration: Gale Brightwell, Adrian L. Cookson.

Resources: Patrick J. Biggs, A. Springer Browne, Hamid Irshad, Gale Brightwell, Adrian L. Cookson.

Software: Rose M. Collis, Patrick J. Biggs, A. Springer Browne, David A. Wilkinson.

Supervision: Patrick J. Biggs, Anne C. Midwinter, A. Springer Browne, Adrian L. Cookson.

Visualization: Rose M. Collis, Patrick J. Biggs, A. Springer Browne, David A. Wilkinson, Adrian L. Cookson.

Writing – original draft: Rose M. Collis, Patrick J. Biggs, Anne C. Midwinter, A. Springer Browne, David A. Wilkinson, Nigel P. French, Gale Brightwell, Adrian L. Cookson.

Writing – review & editing: Rose M. Collis, Patrick J. Biggs, Anne C. Midwinter, A. Springer Browne, David A. Wilkinson, Hamid Irshad, Nigel P. French, Gale Brightwell, Adrian L. Cookson.

References

1. Griffin PM, Ostroff SM, Tauxe RV, Greene KD, Wells JG, Lewis JH, et al. Illnesses associated with *Escherichia coli* O157:H7 infections: A broad clinical spectrum. *Ann Intern Med*. 1988; 109(9):705–12. <https://doi.org/10.7326/0003-4819-109-9-705> PMID: 3056169
2. Beutin L, Zimmermann S, Gleier K. Human infections with Shiga toxin-producing *Escherichia coli* other than serogroup O157 in Germany. *Emerg Infect Dis*. 1998; 4(4):635–9. <https://doi.org/10.3201/eid0404.980415> PMID: 9866741
3. Dewsbury DM, Renter DG, Shridhar PB, Noll LW, Shi XR, Nagaraja TG, et al. Summer and winter prevalence of Shiga toxin-producing *Escherichia coli* (STEC) O26, O45, O103, O111, O121, O145, and O157 in feces of feedlot cattle. *Foodborne Pathog Dis*. 2015; 12(8):726–32. <https://doi.org/10.1089/fpd.2015.1987> PMID: 26075548
4. Stanford K, Johnson RP, Alexander TW, McAllister TA, Reuter T. Influence of season and feedlot location on prevalence and virulence factors of seven serogroups of *Escherichia coli* in feces of Western-Canadian slaughter cattle. *PLoS One*. 2016; 11(8):18. <https://doi.org/10.1371/journal.pone.0159866> PMID: 27482711
5. Bettelheim KA. Role of non-O157 VTEC. *J Appl Microbiol*. 2000; 88:38–50. <https://doi.org/10.1046/j.1365-2672.2000.00848.x>
6. Jaros P, Cookson AL, Campbell DM, Besser TE, Shringi S, Mackereth GF, et al. A prospective case-control and molecular epidemiological study of human cases of Shiga toxin-producing *Escherichia coli* in New Zealand. *BMC Infect Dis*. 2013; 13:450. <https://doi.org/10.1186/1471-2334-13-450> PMID: 24079470
7. Rivas M, Chinen I, Miliwebsky E, Masana M. Risk factors for Shiga toxin-producing *Escherichia coli* associated human diseases. *Microbiol Spectr*. 2014; 2(5). <https://doi.org/10.1128/microbiolspec.EHEC-0002-2013>
8. Taylor EV, Nguyen TA, Machesky KD, Koch E, Sotir MJ, Bohm SR, et al. Multistate outbreak of *Escherichia coli* O145 infections associated with romaine lettuce consumption, 2010. *J Food Prot*. 2013; 76(6):939–44. <https://doi.org/10.4315/0362-028X.JFP-12-503> PMID: 23726187
9. De Schrijver K, Buvens G, Possé B, Van den Branden D, Oosterlynck O, De Zutter L, et al. Outbreak of verocytotoxin-producing *E. coli* O145 and O26 infections associated with the consumption of ice cream produced at a farm, Belgium, 2007. *Eurosurveillance*. 2008; 13(7):9–10. <https://doi.org/10.2807/ese.13.07.08041-en> PMID: 18445416
10. Bell BP, Goldoft M, Griffin PM, Davis MA, Gordon DC, Tarr PI, et al. A multistate outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers: The

- Washington experience. *J Am Med Assoc.* 1994; 272(17):1349–53. <https://doi.org/10.1001/jama.272.17.1349>
11. U.S. Department of Agriculture FSIS. Risk profile for pathogenic non-O157 Shiga toxin-producing *Escherichia coli* 2012. Available from: http://www.fsis.usda.gov/shared/PDF/Non_O157_STEC_Risk_Profile_May2012.pdf.
 12. U.S. Department of Agriculture FSIS. Shiga toxin-producing *Escherichia coli* in certain raw beef products 2011. Available from: <https://www.gpo.gov/fdsys/pkg/FR-2011-09-20/html/2011-24043.htm>.
 13. Browne AS, Midwinter AC, Withers H, Cookson AL, Biggs PJ, Marshall JC, et al. Molecular epidemiology of Shiga toxin-producing *Escherichia coli* (STEC) on New Zealand dairy farms: Application of a culture-independent assay and whole-genome sequencing. *Appl Environ Microbiol.* 2018; 84(14):16. <https://doi.org/10.1128/aem.00481-18> PMID: 29752274
 14. O'Brien AD, Newland JW, Miller SF, Holmes RK, Smith HW, Formal SB. Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science.* 1984; 226(4675):694–6. <https://doi.org/10.1126/science.6387911> PMID: 6387911
 15. Tozzoli R, Grande L, Michelacci V, Ranieri P, Maugliani A, Caprioli A, et al. Shiga toxin-converting phages and the emergence of new pathogenic *Escherichia coli*: A world in motion. *Front Cell Infect Microbiol.* 2014; 4:80. <https://doi.org/10.3389/fcimb.2014.00080> PMID: 24999453
 16. Paton JC, Paton AW. Pathogenesis and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Clin Microbiol Rev.* 1998; 11(3):450–79. PMID: 9665978
 17. Bielaszewska M, Rüter C, Kunsmann L, Greune L, Bauwens A, Zhang W, et al. Enterohemorrhagic *Escherichia coli* hemolysin employs outer membrane vesicles to target mitochondria and cause endothelial and epithelial apoptosis. *PLoS Pathog.* 2013; 9(12):e1003797–e. <https://doi.org/10.1371/journal.ppat.1003797> PMID: 24348251
 18. Bielaszewska M, Aldick T, Bauwens A, Karch H. Hemolysin of enterohemorrhagic *Escherichia coli*: Structure, transport, biological activity and putative role in virulence. *Int J Med Microbiol.* 2014; 304(5–6):521–9. <https://doi.org/10.1016/j.ijmm.2014.05.005> PMID: 24933303
 19. Jores J, Rumer L, Wieler LH. Impact of the locus of enterocyte effacement pathogenicity island on the evolution of pathogenic *Escherichia coli*. *Int J Med Microbiol.* 2004; 294(2–3):103–13. <https://doi.org/10.1016/j.ijmm.2004.06.024> PMID: 15493820
 20. Phillips AD, Frankel G. Intimin-mediated tissue specificity in enteropathogenic *Escherichia coli* interaction with human intestinal organ cultures. *The Journal of Infectious Diseases.* 2000; 181(4):1496–500. <https://doi.org/10.1086/315404> PMID: 10762584
 21. Girard F, Batisson I, Frankel GM, Harel J, Fairbrother JM. Interaction of enteropathogenic and Shiga toxin-producing *Escherichia coli* and porcine intestinal mucosa: role of intimin and Tir in adherence. *Infect Immun.* 2005; 73(9):6005–16. <https://doi.org/10.1128/IAI.73.9.6005-6016.2005> PMID: 16113321
 22. Lacher DW, Steinsland H, Whittam TS. Allelic subtyping of the intimin locus (*eae*) of pathogenic *Escherichia coli* by fluorescent RFLP. *FEMS Microbiol Lett.* 2006; 261(1):80–7. <https://doi.org/10.1111/j.1574-6968.2006.00328.x> PMID: 16842363
 23. Bibbal D, Loukiadis E, Kerouredan M, de Garam CP, Ferre F, Cartier P, et al. Intimin gene (*eae*) subtype-based real-time PCR strategy for specific detection of Shiga toxin-producing *Escherichia coli* serotypes O157:H7, O26:H11, O103:H2, O111:H8, and O145:H28 in cattle feces. *Appl Environ Microbiol.* 2014; 80(3):1177–84. <https://doi.org/10.1128/AEM.03161-13> PMID: 24296503
 24. Iguchi A, Iyoda S, Ohnishi M, EHEC Study Group. Molecular characterization reveals three distinct clonal groups among clinical Shiga toxin-producing *Escherichia coli* strains of serogroup O103. *J Clin Microbiol.* 2012; 50(9):2894–900. <https://doi.org/10.1128/JCM.00789-12> PMID: 22718945
 25. Robins-Browne RM, Holt KE, Ingle DJ, Hocking DM, Yang J, Tauschek M. Are *Escherichia coli* pathotypes still relevant in the era of whole-genome sequencing? *Front Cell Infect Microbiol.* 2016; 6:141. <https://doi.org/10.3389/fcimb.2016.00141> PMID: 27917373
 26. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, 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–e. <https://doi.org/10.1371/journal.pone.0022751> PMID: 21799941
 27. Bielaszewska M, Mellmann A, Zhang WL, Kock R, Fruth A, Bauwens A, et al. Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis.* 2011; 11(9):671–6. [https://doi.org/10.1016/S1473-3099\(11\)70165-7](https://doi.org/10.1016/S1473-3099(11)70165-7) PMID: 21703928
 28. EFSA BIOHAZ Panel., Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bover-Cid S, Chemaly M, et al. Scientific Opinion on the pathogenicity assessment of Shiga toxin-producing *Escherichia coli* (STEC) and the public health risk posed by contamination of food with STEC. *Efsa J.* 2020; 18(1):105. <https://doi.org/10.2903/j.efsa.2020.5967>

29. Conrad CC, Stanford K, McAllister TA, Thomas J, Reuter T. Further development of sample preparation and detection methods for O157 and the Top 6 non-O157 STEC serogroups in cattle feces. *J Microbiol Methods*. 2014; 105:22–30. <https://doi.org/10.1016/j.mimet.2014.06.020> PMID: 25026274
30. Noll LW, Shridhar PB, Dewsbury DM, Shi XR, Cernicchiaro N, Renter DG, et al. A comparison of culture and PCR based methods to detect six major non-O157 serogroups of Shiga toxin-producing *Escherichia coli* in cattle feces. *PLoS One*. 2015; 10(8):12. <https://doi.org/10.1371/journal.pone.0135446> PMID: 26270482
31. Kerangart S, Cournoyer B, Loukiadis E. C-source metabolic profilings of foodborne Shiga-toxin producing *E. coli* match serogroup differentiations and highlight functional adaptations. *Int J Food Microbiol*. 2017; 266:324–36. <https://doi.org/10.1016/j.ijfoodmicro.2017.10.018> PMID: 29217302
32. Posse B, De Zutter L, Heyndrickx M, Herman L. Quantitative isolation efficiency of O26, O103, O111, O145 and O157 STEC serotypes from artificially contaminated food and cattle faeces samples using a new isolation protocol. *J Appl Microbiol*. 2008; 105(1):227–35. <https://doi.org/10.1111/j.1365-2672.2008.03739.x> PMID: 18284481
33. Gill A, Huszczyński G, Gauthier M, Blais B. Evaluation of eight agar media for the isolation of Shiga toxin-producing *Escherichia coli*. *J Microbiol Methods*. 2014; 96:6–11. <https://doi.org/10.1016/j.mimet.2013.10.022> PMID: 24211606
34. Tillman GE, Wasilenko JL, Simmons M, Lauze TA, Minicozzi J, Oakley BB, et al. Isolation of Shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 from ground beef using modified Rainbow agar and post-immunomagnetic separation acid treatment. *J Food Prot*. 2012; 75(9):1548–54. <https://doi.org/10.4315/0362-028X.JFP-12-110> PMID: 22947460
35. Verhaegen B, De Reu K, Heyndrickx M, De Zutter L. Comparison of six chromogenic agar media for the isolation of a broad variety of non-O157 Shiga toxin-producing *Escherichia coli* (STEC) serogroups. *Int J Environ Res Public Health*. 2015; 12(6):6965–78. <https://doi.org/10.3390/ijerph120606965> PMID: 26090610
36. Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, et al. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. *J Infect Dis*. 2005; 192(8):1422–9. <https://doi.org/10.1086/466536> PMID: 16170761
37. Almeida C, Sousa JM, Rocha R, Cerqueira L, Fanning S, Azevedo NF, et al. Detection of *Escherichia coli* O157 by peptide nucleic acid fluorescence in situ hybridization (PNA-FISH) and comparison to a standard culture method. *Appl Environ Microbiol*. 2013; 79(20):6293–300. <https://doi.org/10.1128/AEM.01009-13> PMID: 23934486
38. Lin A, Sultan O, Lau HK, Wong E, Hartman G, Lauzon CR. O serogroup specific real time PCR assays for the detection and identification of nine clinically relevant non-O157 STECs. *Food Microbiol*. 2011; 28(3):478–83. <https://doi.org/10.1016/j.fm.2010.10.007> PMID: 21356454
39. Haugum K, Johansen J, Gabrielsen C, Brandal LT, Bergh K, Ussery DW, et al. Comparative genomics to delineate pathogenic potential in non-O157 Shiga toxin-producing *Escherichia coli* (STEC) from patients with and without haemolytic uremic syndrome (HUS) in Norway. *PLoS One*. 2014; 9(10). <https://doi.org/10.1371/journal.pone.0111788> PMID: 25360710
40. Carter MQ, Quinones B, He XH, Zhong W, Louie JW, Lee BG, et al. An environmental Shiga toxin-producing *Escherichia coli* O145 clonal population exhibits high-level phenotypic variation that includes virulence traits. *Appl Environ Microbiol*. 2016; 82(4):1090–101. <https://doi.org/10.1128/AEM.03172-15> PMID: 26637597
41. Nyholm O, Halkilahti J, Wiklund G, Okeke U, Paulin L, Auvinen P, et al. Comparative genomics and characterization of hybrid Shiga toxin-producing and enterotoxigenic *Escherichia coli* (STEC/ETEC) strains. *PLoS One*. 2015; 10(8):17. <https://doi.org/10.1371/journal.pone.0135936> PMID: 26313149
42. Lorenz SC, Gonzalez-Escalona N, Kotewicz ML, Fischer M, Kase JA. Genome sequencing and comparative genomics of enterohemorrhagic *Escherichia coli* O145:H25 and O145:H28 reveal distinct evolutionary paths and marked variations in traits associated with virulence & colonization. *BMC Microbiol*. 2017; 17:15. <https://doi.org/10.1186/s12866-017-0930-9> PMID: 28086793
43. Cooper KK, Mandrell RE, Louie JW, Korfach J, Clark TA, Parker CT, et al. Comparative genomics of enterohemorrhagic *Escherichia coli* O145:H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157:H7. *BMC Genomics*. 2014; 15:17. <https://doi.org/10.1186/1471-2164-15-17> PMID: 24410921
44. Nakamura K, Murase K, Sato MP, Toyoda A, Itoh T, Mainil JG, et al. Differential dynamics and impacts of prophages and plasmids on the pangenome and virulence factor repertoires of Shiga toxin-producing *Escherichia coli* O145:H28. *Microb Genom*. 2020. <https://doi.org/10.1099/mgen.0.000323> PMID: 31935184
45. Biggs PJ, Truglio M. QCtool. 2016.

46. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012; 19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021> PMID: 22506599
47. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics.* 2013; 29(8):1072–5. <https://doi.org/10.1093/bioinformatics/btt086> PMID: 23422339
48. Seemann T, Prokka: Rapid prokaryotic genome annotation. *Bioinformatics.* 2014; 30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
49. National Center for Biotechnology Information. National Library of Medicine USA1988. Available from: <https://www.ncbi.nlm.nih.gov/>.
50. Enterobase.warwick.ac.uk. Enterobase. Available from: <http://enterobase.warwick.ac.uk>.
51. Center for Genomic Epidemiology. Overview of services Denmark2011. Available from: <http://www.genomicepidemiology.org/>.
52. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole genome sequencing data. *J Clin Microbiol.* 2015; 53(8):2410–26. <https://doi.org/10.1128/JCM.00008-15> PMID: 25972421
53. Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, et al. Benchmarking of methods for genomic taxonomy. *J Clin Microbiol.* 2014; 52(5):1529–39. <https://doi.org/10.1128/JCM.02981-13> PMID: 24574292
54. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol.* 2014; 52(5):1501–10. <https://doi.org/10.1128/JCM.03617-13> PMID: 24574290
55. Carattoli A, Zankari E, Garcia-Fernandez A, Larsen MV, Lund O, Villa L, et al. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother.* 2014; 58(7):3895–903. <https://doi.org/10.1128/AAC.02412-14> PMID: 24777092
56. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother.* 2012; 67(11):2640–4. <https://doi.org/10.1093/jac/dks261> PMID: 22782487
57. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total genome sequenced bacteria. *J Clin Microbiol.* 2012; 50(4):1355–61. <https://doi.org/10.1128/JCM.06094-11> PMID: 22238442
58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
59. Giron JA, Ho ASY, Schoolnik GK. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. *Science.* 1991; 254(5032):710–3. <https://doi.org/10.1126/science.1683004> PMID: 1683004
60. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012; 28(12):1647–9. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
61. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: Universal characterization of bacteria from domain to strain. *Microbiology.* 2012; 158:1005–15. <https://doi.org/10.1099/mic.0.055459-0> PMID: 22282518
62. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *J Mol Biol.* 2006; 23(2):254–67.
63. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016; 44(W1):242–5. <https://doi.org/10.1093/nar/gkw290> PMID: 27095192
64. R Core Team. R: A language and environment for statistical computing 2016. Available from: www.r-project.org/.
65. Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlenn D, et al. vegan: Community Ecology Package. 2019, R package version 2.5–6. <https://CRAN.R-project.org/package=vegan>.
66. Lucas A, Gautier L. ctc: Cluster and Tree Conversion. 2019, R package version 1.60.0. <http://antoinelucas.free.fr/ctc>.
67. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, et al. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2009; 106(42):17939–44. <https://doi.org/10.1073/pnas.0903585106> PMID: 19815525
68. Seemann T. Snippy. 2016, <https://github.com/tseemann/snippy/blob/master/README.md>.

69. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019; 35(21):4453–5. <https://doi.org/10.1093/bioinformatics/btz305> PMID: 31070718
70. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015; 31(22):3691–3. <https://doi.org/10.1093/bioinformatics/btv421> PMID: 26198102
71. Vaas LAI, Sikorski J, Hofner B, Fiebig A, Buddruhs N, Klenk HP, et al. Opm: An R package for analysing OmniLog (R) phenotype microarray data. *Bioinformatics*. 2013; 29(14):1823–4. <https://doi.org/10.1093/bioinformatics/btt291> PMID: 23740744
72. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WH, Lumley T, et al. gplots: Various R programming tools for plotting data. 2016, <http://CRAN.R-project.org/package=gplots>.
73. Wirth T, Falush D, Lan RT, Colles F, Mensa P, Wieler LH, et al. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol*. 2006; 60(5):1136–51. <https://doi.org/10.1111/j.1365-2958.2006.05172.x> PMID: 16689791
74. Trabulsi LR, Keller R, Gomes TAT. Typical and atypical enteropathogenic *Escherichia coli*. *Emerg Infect Dis*. 2002; 8(5):508–13. <https://doi.org/10.3201/eid0805.010385> PMID: 11996687
75. De Rauw K, Buyl R, Jacquinet S, Piérard D. Risk determinants for the development of typical haemolytic uremic syndrome in Belgium and proposition of a new virulence typing algorithm for Shiga toxin-producing *Escherichia coli*. *Epidemiol Infect*. 2019; 147:e6. <https://doi.org/10.1017/S0950268818002546> PMID: 30182864
76. Brandal LT, Wester AL, Lange H, Løbersli I, Lindstedt B-A, Vold L, et al. Shiga toxin-producing *Escherichia coli* infections in Norway, 1992–2012: Characterization of isolates and identification of risk factors for haemolytic uremic syndrome. *BMC Infect Dis*. 2015; 15:324–. <https://doi.org/10.1186/s12879-015-1017-6> PMID: 26259588
77. Schmidt H, Scheef J, Morabito S, Caprioli A, Wieler LH, Karch H. A new Shiga toxin 2 variant (*Stx2f*) from *Escherichia coli* isolated from pigeons. *Appl Environ Microbiol*. 2000; 66(3):1205–8. <https://doi.org/10.1128/aem.66.3.1205-1208.2000> PMID: 10698793
78. Prager R, Fruth A, Siewert U, Strutz U, Tschape H. *Escherichia coli* encoding Shiga toxin 2f as an emerging human pathogen. *Int J Med Microbiol*. 2009; 299(5):343–53. <https://doi.org/10.1016/j.ijmm.2008.10.008> PMID: 19157976
79. Friesema I, van der Zwaluw K, Schuurman T, Kooistra-Smid M, Franz E, van Duynhoven Y, et al. Emergence of *Escherichia coli* encoding Shiga toxin 2f in human Shiga toxin-producing *E. coli* (STEC) infections in the Netherlands, January 2008 to December 2011. *Eurosurveillance*. 2014; 19(17):26–32. PMID: 24821123
80. De Rauw K, Jacobs S, Piérard D. Twenty-seven years of screening for Shiga toxin-producing *Escherichia coli* in a university hospital. Brussels, Belgium, 1987–2014. *PLoS One*. 2018; 13(7):e0199968. <https://doi.org/10.1371/journal.pone.0199968> PMID: 29965972
81. Kruger A, Lucchesi PMA. Shiga toxins and *stx* phages: Highly diverse entities. *Microbiology*. 2015; 161:451–62. <https://doi.org/10.1099/mic.0.000003> PMID: 25479836
82. Hillerton JE, Irvine CR, Bryan MA, Scott D, Merchant SC. Use of antimicrobials for animals in New Zealand, and in comparison with other countries. *N Z Vet J*. 2017; 65(2):71–7. <https://doi.org/10.1080/00480169.2016.1171736> PMID: 27030313
83. Kwon T, Kim W, Cho S-H. Comparative genomic analysis of Shiga toxin-producing and non-Shiga toxin-producing *Escherichia coli* O157 isolated from outbreaks in Korea. *Gut Pathog*. 2017; 9(1):7. <https://doi.org/10.1186/s13099-017-0156-2> PMID: 28191041
84. Carter MQ, Pham A. Complete genome sequences of two atypical enteropathogenic *Escherichia coli* O145 environmental strains. *Genome Announc*. 2018; 6(19):e00418–18. <https://doi.org/10.1128/genomeA.00418-18> PMID: 29748413
85. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol*. 2008; 11(5):472–7. <https://doi.org/10.1016/j.mib.2008.09.006> PMID: 19086349
86. Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol*. 2010; 60(4):708–20. <https://doi.org/10.1007/s00248-010-9717-3> PMID: 20623278
87. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol*. 2008; 190(20):6881–93. <https://doi.org/10.1128/JB.00619-08> PMID: 18676672
88. Browne AS, Biggs PJ, Wilkinson DA, Cookson AL, Midwinter AC, Bloomfield SJ, et al. Use of genomics to investigate historical importation of Shiga toxin-producing *Escherichia coli* serogroup O26 and non-toxicogenic variants into New Zealand. *Emerg Infect Dis*. 2019; 25(3):489. <https://doi.org/10.3201/eid2503.180899> PMID: 30789138

89. Durso LM, Smith D, Hutkins RW. Measurements of fitness and competition in commensal *Escherichia coli* and *E. coli* O157:H7 strains. *Appl Environ Microbiol.* 2004; 70(11):6466–72. <https://doi.org/10.1128/AEM.70.11.6466-6472.2004> PMID: 15528507
90. Sabarly V, Bouvet O, Glodt J, Clermont O, Skurnik D, Diancourt L, et al. The decoupling between genetic structure and metabolic phenotypes in *Escherichia coli* leads to continuous phenotypic diversity. *J Evol Biol.* 2011; 24(7):1559–71. <https://doi.org/10.1111/j.1420-9101.2011.02287.x> PMID: 21569155