



Article

Comparative Genomics of *Escherichia coli* Serotype O55:H7 Using Complete Closed Genomes

Margaret D. Weinroth[†] and James L. Bono^{*†}

U.S. Meat Animal Research Center, U.S. Department of Agriculture, Agricultural Research Service, Clay Center, NE 68933, USA; maggie.weinroth@usda.gov

* Correspondence: jim.bono@usda.gov; Tel.: +1-402-762-4363

† Current Address: U.S. National Poultry Research Center, U.S. Department of Agriculture, Agricultural Research Service, Athens, GA 30605, USA.

Abstract: *Escherichia coli* O55:H7 is a human foodborne pathogen and is recognized as the progenitor strain of *E. coli* O157:H7. While this strain is important from a food safety and genomic evolution standpoint, much of the genomic diversity of *E. coli* O55:H7 has been demonstrated using draft genomes. Here, we combine the four publicly available *E. coli* O55:H7 closed genomes with six newly sequenced closed genomes to provide context to this strain's genomic diversity. We found significant diversity within the 10 *E. coli* O55:H7 strains that belonged to three different sequence types. The prophage content was about 10% of the genome, with three prophages common to all strains and seven unique to one strain. Overall, there were 492 insertion sequences identified within the six new sequence strains, with each strain on average containing 75 insertions (range 55 to 114). A total of 31 plasmids were identified between all isolates (range 1 to 6), with one plasmid (pO55) having an identical phylogenetic tree as the chromosome. The release and comparison of these closed genomes provides new insight into *E. coli* O55:H7 diversity and its ability to cause disease in humans.

Keywords: keyword; NGS; whole genome sequencing; genomics; microbial evolution; *Escherichia coli*; O55:H7



Citation: Weinroth, M.D.; Bono, J.L. Comparative Genomics of *Escherichia coli* Serotype O55:H7 Using Complete Closed Genomes. *Microorganisms* **2022**, *10*, 1545. <https://doi.org/10.3390/microorganisms10081545>

Academic Editors: Andrew Hesketh and Grigoris Amoutzias

Received: 31 May 2022

Accepted: 23 July 2022

Published: 30 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Escherichia coli O55:H7 is a human foodborne pathogen with an unknown reservoir, although humans are thought to be the primary host [1]. O55:H7 is usually thought of as an enteropathogenic *Escherichia coli* (EPEC), but strains can also carry a Shiga toxin gene that classifies the strain as a Shiga toxin-containing *Escherichia coli* (STEC). Foodborne outbreaks from O55:H7 can have a wide range of clinical outcomes. An EPEC O55:NM outbreak in Japan resulted in cases of diarrhea [2], while a STEC O55:H7 outbreak in England had a more severe outcome, with some patients developing hemolytic uremic syndrome [3]. From an evolutionary standpoint, *E. coli* O55:H7 has been proposed as the progenitor of Shiga toxin-containing *E. coli* O157:H7 (STEC O157:H7) [4–11]. Briefly, one model of evolution [5] describes an *E. coli* O55:H7 with a pathogenicity island referred to as the Locus of Enterocyte Effacement (LEE) acquiring *stx2* followed by the O157 O-antigen biosynthesis cluster mediating its antigen switch from *E. coli* O55:H7 to STEC O157:H7. From there, the evolutionary isolate lost its ability to ferment sorbitol and gained *stx1*, while also losing the ability to express beta-glucuronidase activity, resulting in the typical STEC O157:H7. More recently, the model was changed to show that O55:H7 strains are placed into two groups that descend from the same common ancestor as STEC O157 strains, but are not part of the stepwise evolution [6].

Whole genome sequencing of bacterial genomes has become the gold standard in both research and foodborne outbreak settings [12–15]. The ability to compare genomes at nucleotide resolution allows for the evaluation of single nucleotide polymorphisms (SNP) within genes of interest. Complex genomes with inversions and repeated regions such as

mobile elements and prophage [16–18] cause problems in genome assemblies similarly to STEC O157:H7 based on short-read sequencing technologies. However, this problem can be addressed by long-read sequencing technologies. Integrating long-read sequencing as scaffolding with short-reads to polish the assembly and to find smaller plasmids produces a higher quality genome than either technology alone [18].

Here, we use this hybrid approach of sequencing to generate complete closed genomes from six *E. coli* O55:H7 and compare them to all other publicly available, complete closed *E. coli* O55:H7 genomes. Previously, only two O55:H7 genomes were closed and used to explore the genetic diversity of O55:H7 strains [6,9]. The six newly sequenced O55:H7 strains in this study were part of a study using multilocus enzyme electrophoresis to determine the genetic relationship between *E. coli* strains causing enteric disease [19] and were also used to understand the stepwise evolution of STEC O157:H7 [4–6,8]. These genomes provide a higher resolution of O55:H7 genomic diversity, including genome architecture, than previously described.

2. Materials and Methods

Isolate selection. A total of 10 *E. coli* O55:H7 strains were used for this study. Six were previously used to describe the phylogeny of O55:H7 and evolution *E. coli* O157:H7 strains [4,18]. The complete closed genomes from four O55:H7 strains were publicly available, with two genomes published [5,8] and two not published (Supplemental Table S1).

Wet Lab work and sequencing. All previously non-publicly available genomes that underwent sequencing were subjected to both long-read (Pacific Biosciences, Menlo Park, CA, USA) and short-read (Illumina, Inc., San Diego, CA, USA) sequencing. DNA extraction using a Genomic tip 100/g (Qiagen, Inc., Valencia, CA, USA) and size selected PacBio libraries were generated using the SMRTbell Template Prep Kit 1.0, as previously described [20]. The library was bound with polymerase P6, followed by sequencing on a Pacific Biosciences (Pacific Biosciences) RS II sequencing platform with chemistry C4 and the 360-min data collection protocol.

For Illumina library preparation, DNA from the same extraction as the long-read libraries was sheared using a microTube AFA Fiber Pre-Slit Snap-Cap 6 × 16 mm (Corvaris, Woburn, MA, USA). Libraries were created using the TruSeq DNA CPR-Free HT Library Preparation kit (Illumina, Inc., San Diego, CA, USA) and quantitated using the KAPA Library Quantification Kit (F. Hoffmann-La Roche Ltd., Basel, Switzerland) prior to pooling. The pooled libraries were run on an Illumina MiSeq with the MiSeq Reagent Kit v3 (600 cycles), resulting in 300 bp paired end read lengths.

Bioinformatics. After sequencing, raw long reads were assembled using HGAP3 [21] in SMRT analysis Version 8.0, and the resulting contigs were imported into Geneious (2020.1.2; Biomatters, Ltd., Auckland, New Zealand). If present, overlapping sequences on the ends of the contigs were removed from the 5' and 3' ends to generate circularized chromosomes and plasmids. The closed chromosomes were reoriented to start with the putative origin of replication using Ori-Finder 2 [22]. The closed chromosomes and plasmids were polished using the Illumina libraries via Pilon [23]. Finally, Illumina reads were mapped to the chromosome and known plasmids using Geneious, and unused reads were de novo assembled (also in Geneious) for small plasmid identification. All genomes and plasmids were annotated with the NCBI Prokaryotic Genome Annotation Pipeline [24]. Parsnp (1.2) [25] was used to generate the core chromosome phylogeny. The SNP alignment file was extracted from the Parsnp output .ggr file using HarvestTools [25]. We obtained branch supports with the ultrafast bootstrap [26] implemented in the IQ-TREE software [27] with ascertainment bias correction. The pangenome of all chromosome was visualized using Gview Server [28]. Multiple alignments were visualized using Mauve (1.1.3) [29]. EasyFig (2.2) was used to visualize plasmid similarities [30]. Sequence type, virulence genes, and plasmid types were identified using MLST 2.0, VirulenceFinder 2.0, and PlasmidFinder 2.1, respectively (<http://www.genomicepidemiology.org> last accessed on 17 June 2022). Insertion sequences were identified through NCBI annotation and prophages were identified

with PHASTER [31]. Both were visualized with the heatmap3 package (version 1.1.9) [32] in R (version 4.0.2). Chromosome and pO55 phylogenetic trees were compared using Phylo (<https://phylo.io> accessed on 12 May 2022) [33]. Graphical feature format (GFF) files were produced in Prokka (v1.14.6) [34] and used with Roary (3.13.0) [35] to generate set specific unique genes.

3. Results

3.1. Genomic Overview

Ten *E. coli* O55:H7 strains were included in the analysis: six that were subjected to long-read sequencing for the first time and four that had been previously subjected to long-read sequencing and were publicly available. Individual strain information is provided in Supplemental Table S1. The chromosomes ranged in size from 5196 kb to 5571 kb (average 5340 kb), with strains having an average of 3.1 plasmids (range 1 to 6; size average 38.8 kb, range 3.3 to 99.7 kb). The average number of protein-coding sequences (CDS) within strains ranged from 4912 to 5452 (average 5120). A phylogenomic tree of the chromosomes with 8680 core chromosomal-derived SNPs showed several groupings of O55:H7 with diverse branches (Figure 1A) using Parsnp [24]. Bootstrapping using UFboot and SH-aLRT in IQ-TREE supported the internal branches, but the branches between DEC5B and USDA5905 and USDA5905 and DEC5E were not supported.

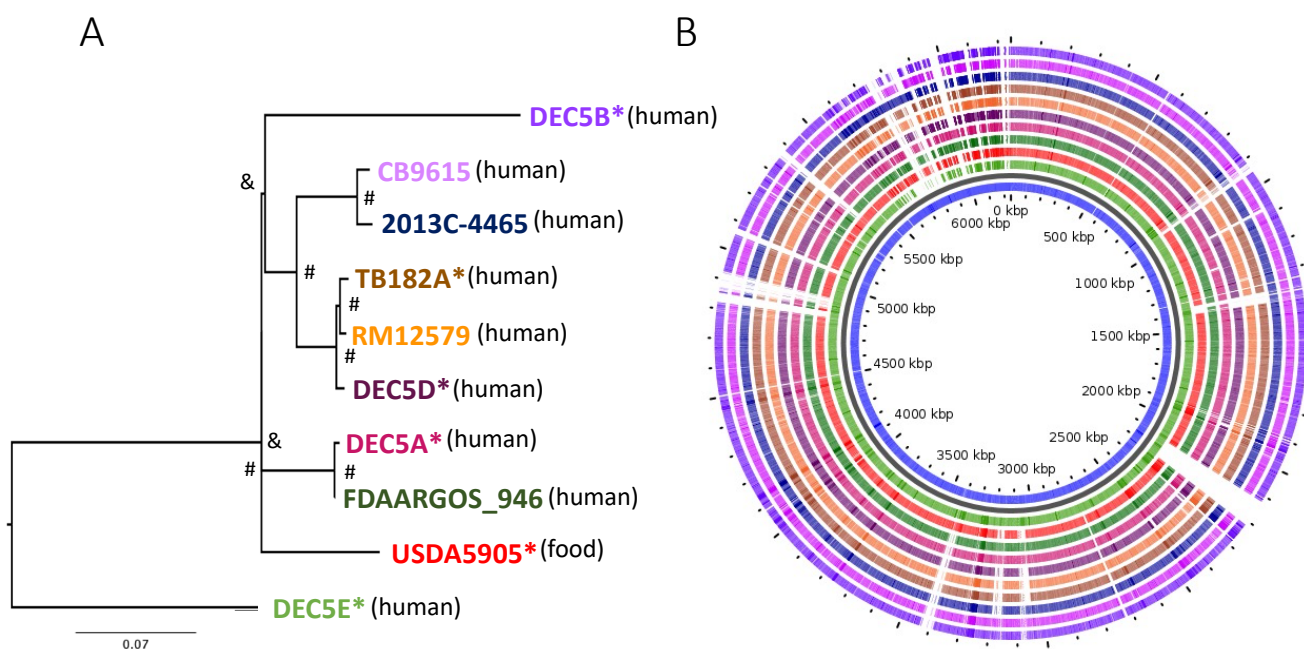


Figure 1. (A) Phylogenetic tree of the core regions from the chromosomes of *E. coli* O55:H7 strains, with newly sequenced chromosomes denoted with an asterisk (*). The tree was constructed using Parsnp with 6860 core chromosome-derived SNPs. Bootstrapping values were assigned using UFboot and SH-aLRT, with values above 95% represented by a hashtag (#) and values below represented by an ampersand (&). The source of the isolate is in parenthesis next to the strain name. (B) The pangenome of the chromosomes from *E. coli* O55:H7 strains using GView. The colored rings match the color of the strains at the end of the branches of the phylogenetic tree in (A) with blank areas representing a section of chromosome not found within a strain. The inner blue circle is the complete pangenome of the chromosomes. Strain DEC5E was used as the seed genome to build the pangenome.

Figure 1B illustrates the diversity of genomic content within the serotype. Within all O55 isolates, the core genome was made up of 4273 genes. There were 193 genes only missing from one of the chromosomes and 1151 genes unique to only one isolate, with many of the unique genes belonging to strain DEC5E (inner most green ring). Multi locus sequence typing (MLST) examines the allelic profile of seven housekeeping genes to

determine the relatedness of strains to each other [36]. Three different sequence types (ST) representing two different clonal complexes were identified from the 10 strains (Figure 2). DEC5E belonged to ST-61 while DEC5A and FDAARGOS-946 belonged to ST 7444. The remaining strains belong to ST 355, and all O55:H7 strains fall within a large group called ST11 clonal complex (CC ST11) (https://pubmlst.org/bigssdb?db=pubmlst_ecoli_achtman_seqdef&set_id=4&page=downloadProfiles&scheme_id=4 accessed 20 April 2022). CC ST11 also contains STEC O157:H7 strains, as would be expected, as it is the descendant of O55:H7 [37]. The main allelic difference between O55:H7 and STEC O157:H7 strains was in the *adk* gene, where O55:H7 had allele 29 while STEC O157:H7 had allele 12.

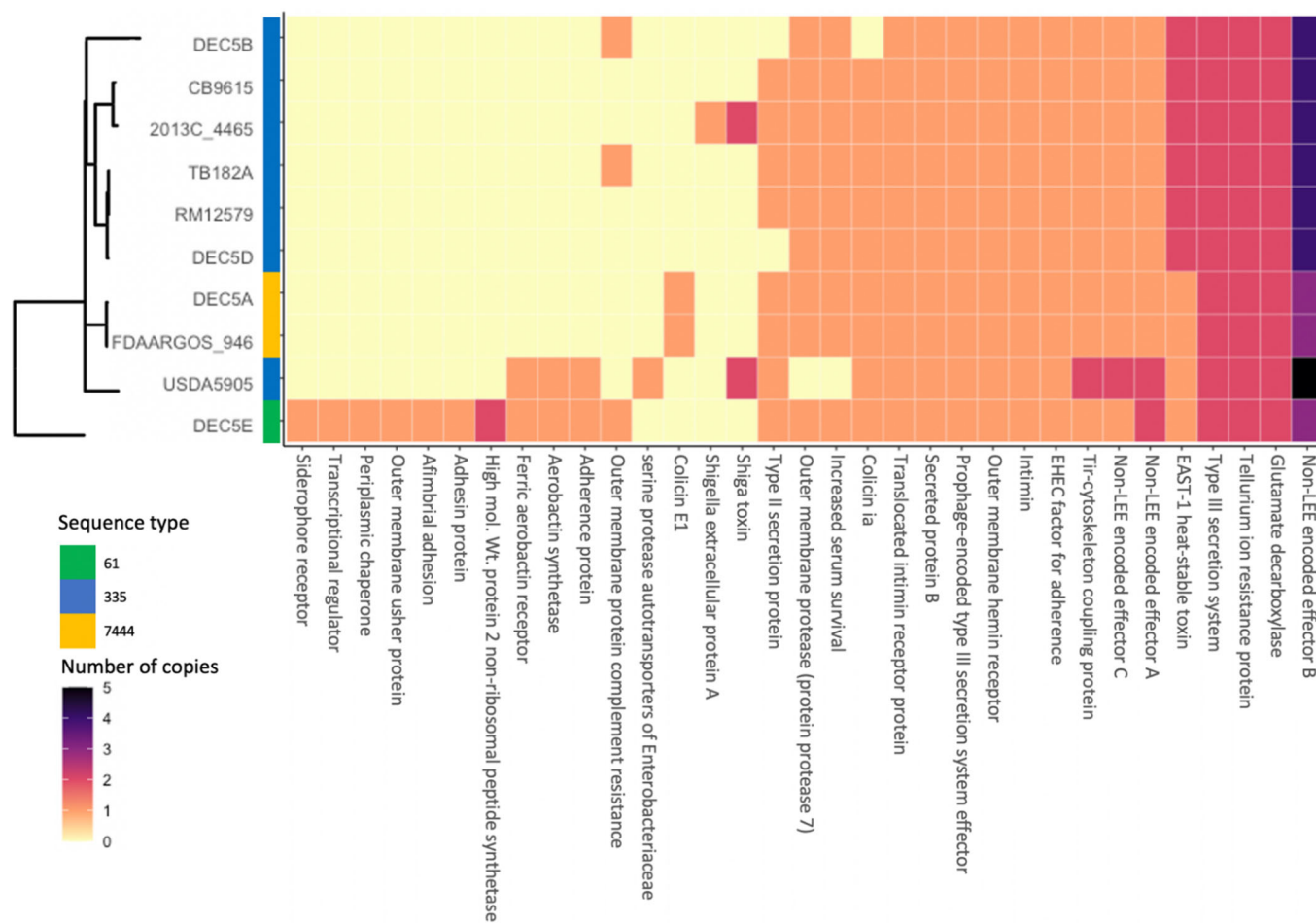


Figure 2. Comparison of the phylogeny of O55:H7 strain with sequence type and virulence gene content. The phylogenomic tree, strain names, and sequence type are on the *y* axis of the heatmap while the virulence factors are on the *x* axis. The number of copies of a virulence factor is represented by a different color. The phylogenomic tree is the same as Figure 1.

Based on the tree, DEC5E was proposed to be the earliest ancestor of the O55 isolates, with DEC5B being the most recent, and agrees with the phylogenetic analysis described by Kyle et al., 2012 [6]. When overall chromosome architecture was considered, there was a high degree of synteny (except for DEC5E which showed an inversion near the replication terminus) (Supplemental Figure S1). These large-chromosomal rearrangements (LCRs) are seen in STEC O157:H7 strains and are usually flanked by prophages that share a homologous region. RNA transcriptional profiling and phenotyping of specific structural variants indicated that important virulence phenotypes such as Shiga-toxin production, type-3 secretion, and motility can be affected by them [38]. LCRs have been seen in the chromosomes of other bacteria, including *Campylobacter*, *Staphylococcus*, and *Salmonella* [39–41]. In *Campylobacter*, the orientation of LCRs can provide resistance to

phage infection [39]. Further research will be required to identify if the LCRs in O55:H7 strain DEC5E are related to any of its phenotypes.

3.2. Virulence Genes

Thirty-three virulence genes (VGs) were identified in the 10 O55:H7 strains by VirulenceFinder (Figure 2). While eighteen VGs were found in more than 80% of the strains, nine VGs were only found in one strain, five VGs in two strains and three VGs in one strain (Figure 2). DEC5E contained the most VGs with thirty, and with seven of them being unique to this strain. The least number of VGs was eighteen, and was shared by five strains: DEC5B, CB9615, TB182A, RM12579, and DEC5D. DEC5E and USDA5905 shared three VGs that encode for the ferric aerobactin receptor, aerobactin synthetase, and an adherence protein. Aerobactin is used to acquire iron and enhance biofilm formation, oxidative stress resistance, and virulence in *Yersinia pseudotuberculosis* [42]. DEC5E also carried an additional siderophore receptor that binds iron, two more adhesion proteins, and a complement resistance protein. The combination of shared and strain-specific virulence genes indicated that this strain might be more infective than the other O55:H7 strains. This hypothesis would need to be tested to determine its validity, which is beyond the scope of this paper. In addition to the aerobactin genes and adherence genes shared with DEC5E, USDA5905 contained the *stx2_d* Shiga toxin variant. Shiga toxin is a virulence factor that was originally identified in *Shigella dysenteriae* but has been found in other bacteria, including several *E. coli* serogroups [43]. Stx2d is one of the most potent Shiga toxin variants along with Stx2a. DEC5A and FDAARGOS-946 belong to the same phylogenomic clade and share one VG found only in this group. Colicin E1, a protein that can puncture the bacterial cell wall, causing cell death, was found on a 6-kb plasmid unique to these strains. As previously described, USDA5905 contained the Stx2d variant, while strain 2013C-4465 had the Stx1a variant, the least potent of the two. Colicin 1A was found in 9 of the 10 strains. DEC5B lacked this gene because it did not carry the pO55 plasmid. All 10 strains contained proteins encoded by VGs that are shared with STEC O157:H7, a progenitor of O55:H7. These include the translocated intimin receptor, intimin, EHEC factor for adherence, non-LEE encoded effectors, tellurium resistance, and glutamate decarboxylase. All strains possessed multiple copies of the *nleB* gene, which encodes for the non-LEE encoded effector B protein, in their genomes. USDA5905 has five copies of this gene, while DEC5A, FDAARGOS-946, and DEC5E each had three. The remaining strains had four copies. These results differ from those of the STEC O157:H7 strain Sakai, which only contains two *nleB* genes. The *nleB*-encoded protein is thought to inhibit proinflammatory signaling and necroptosis [44] and has been identified as one of three proteins essential for effective colonization [45–47]. Of the 33 VGs identified, 14 proteins encoded by VGs were identified in all strains from human cases, while 19 were found in single or multiple strains, but not in all strains. The VGs not identified in all the strains could increase the infectivity or severity of the disease depending upon the combination. Another factor to consider when trying to determine virulence is the immune status of the host. Humans with a compromised immune system might be more susceptible to a pathogen with few virulence factors.

3.3. Prophages

PHASTER can predict if a prophage is complete, incomplete, or questionable based on prophage similarity using BLAST+ [31]. However, prophages predicted by PHASTER with the same name may have different genetic structures. The comparison of prophage content between O55:H7 strains reported the predicted classification, and is not based on prophage structure similarity. Chromosomes were determined to be comprised of 10.0% prophage content (range 7.6 to 12.4%) and encoded between 10 and 17 phages (average 14.2) (Figure 3).

Accession Number	Prophage Name	<i>E. coli</i> O55:H7 Strain									
		DEC5B	CB9615	2013C-4465	TB182A	RM12579	DEC5D	DEC5A	FDAARGOS	USDA5905	DEC5E
NC_004813	Enterobacteria phage BP-4795	2	2	1	1	2	1	1	2	3	3
NC_042057	Enterobacteria phage DE3	2	1	3	3	3	2	3	2	2	2
NC_001416	Enterobacteria phage lambda	2	2	2	2	2	2	1	1	2	1
NC_005344	Shigella phage Sf6	2	1	0	1	1	1	1	1	1	1
NC_019716	Enterobacteria phage mEp460	1	1	0	1	0	2	1	1	1	1
NC_009514	Enterobacteria phage cdtI	0	2	1	1	1	0	1	0	1	0
NC_031940	Salmonella phage 118970_sal3	0	2	1	0	1	0	1	1	1	0
NC_028699	Salmonella phage SEN34	1	1	1	0	0	0	1	1	0	2
NC_031927	Synechococcus phage S-CAM7	1	0	0	1	0	1	1	0	1	1
NC_026014	Enterobacteria phage P88	1	1	0	0	0	0	1	1	0	1
NC_010392	Phage Gifsy-1	0	1	1	0	0	0	0	0	1	1
NC_026927	Synechococcus phage ACG-2014f	0	1	1	0	1	0	0	1	0	0
NC_049342	Escherichia phage 500465-1	0	0	2	0	1	0	0	1	0	0
NC_011357	Stx2-converting phage 1717	0	0	0	0	0	1	1	0	1	0
NC_049924	Stx2-converting phage Stx2a_F451	0	0	1	0	1	0	0	1	0	0
NC_001609	Enterobacteria phage P4	0	0	0	0	0	0	1	1	0	0
NC_011356	Enterobacteria phage YZ-2008	0	0	0	0	1	0	0	0	1	0
NC_049919	Escherichia phage SH2026Stx1	0	1	1	0	0	0	0	0	0	0
NC_021073	Vibrio phage henriette 12B8	0	0	0	0	0	0	1	1	0	0
NC_019711	Escherichia phage HK629	0	1	0	0	0	0	0	0	0	0
NC_021190	Enterobacteria phage phi80	0	0	0	0	0	0	0	1	0	0
NC_049341	Escherichia phage 503458	0	0	1	0	0	0	0	0	0	0
NC_041935	Escherichia phage PA28	0	0	0	0	1	0	0	0	0	0
NC_028943	Escherichia phage pro483	0	0	0	0	0	0	0	0	1	0
NC_021857	Shigella phage Sfil	0	0	1	0	0	0	0	0	0	0
NC_004745	Yersinia phage L-413C	0	0	0	0	0	0	0	0	0	1

Figure 3. Heatmap of the number of prophages presents in *E. coli* O55:H7 chromosomes with the number of phages also included in the cell. The prophage accession number and name are on the *y* axis, while the strain names are on the *x* axis. The number of copies of a prophage is represented by a different color.

Three prophages were present in all genomes (sometimes in multiple copies): Enterobacteria phage BP-4795 (NC_004813), Enterobacteria phage DE3 (NC_042057), and Enterobacteria phage lambda (NC_001416). Enterobacteria phage BP-4795 has been characterized to contain two IS629 elements and a type III secretion system effector, NleA₄₇₉₅ [48]. There were seven phages unique to one isolate, while other phages were only found in strains grouped by phylogeny. For example, *Escherichia* phage SH2026Stx1 was found in strain CB9615 and 2013C-4465, while phages Enterobacteria phage p4 and *Vibrio* phage Henriette 12B8 were found in strain DEC5A and FDAARGOS-946 (Figure 3).

Two strains, USDA5905 and 2013C-4465, had prophages that encode the Shiga toxin genes, *stx2_d* and *stx1_a*, respectively. *stx2_d*-encoding prophages were extracted from nine closed genomes deposited in NCBI and were compared with the *stx2_d*-encoding prophage from USDA5905 (Supplemental Figure S2A). The *stx2_d*-encoding prophages could be placed into one of five groups. Prophages shared regions of homology within each serogroup but shared little homology with different serogroups. Six of the prophages, from strains USDA5905, NCCP15955, STEC 313, STEC 316, STEC 367, and STEC 1025, were integrated into *yecE*, a gene with an unknown function [49,50]. Integration of a *stx2_d*-encoding prophage at *yecE* is not unique and has been observed in STEC serotype O26:H11 [51]. The *yecE* gene is one of the six sites that *stx*-encoding prophages are known to integrate. Another integration site is *yehV*, a gene encoding a transcriptional regulator [48,52]. The *stx2_d* prophage in RM10410, an *E. coli* serotype O111:H4, is integrated into this site. The integration site of *stx2_d* prophages in strains M7424, M00057, and M11957 was at a site not previously recognized as a site of integration for *stx*-encoding prophages. The prophage was integrated between the *zinT* and *mtfA* genes. This integration site was occupied in other O55:H7 and STEC O157 strains with a non-*stx*-encoding prophage (Supplemental Figure S3). In STEC O157:H7 strain NE1169-1, a tellurite resistance protein was in place of the *stx2_d* gene, while O55:H7 strain USDA5905 had a hypothetical protein and DUF1327 domain-containing protein in this site. A phylogenetic tree of the *stx2_d* genes identified

four different variants (Supplemental Figure S2B). The *stx2_d* gene from USDA5905 was one identified variant, while the *stx2_d* genes RM10410 and NCCP 15955 formed a second variant. The third *stx2_d* variant was found in strains M7424, M00057, and M11957, with the fourth *stx2_d* variant in strains STEC313, STEC316, STEC367, and STEC1025. The grouping of the *stx2_d* phylogenetic tree reflects the same grouping in the *stx2_d* prophage alignment, demonstrating the specificity of the *stx2_d* gene to the prophage that carries it.

The *stx1_a*-encoding prophage was extracted from six closed STEC O157:H7 genomes deposited in NCBI and was compared to the *stx1_a*-encoding prophage from strain 2013C-4465 (Supplemental Figure S2C). The 2013C-4465 *stx1_a*-encoding prophage was integrated into *argW*, which encodes the transfer RNA, tRNA-Arg [53,54], while the six *stx1_a*-encoding prophages from STEC O157:H7 were integrated into the *yehV* site. When comparing the prophage structure, the STEC O157:H7 prophages were placed into one of two groups while the 2013C-4465 prophage structure was grouped by itself. This demonstrated that prophage-encoding *stx1_a* genes are significantly different structurally but contain *stx1_a* genes with identical sequence. This agrees with the presence of the *stx1_a*-encoding prophage in STEC O157:H7 strain in UK [55]. Prophage-encoding *stx1_a* did not share a high level of similarity across lineages, geographical regions, or time, but shared similarity at the gene level. The two STEC O157:H7 *stx1_a*-encoding prophage groups were associated with a polymorphism in the *tir* gene that associates with a strain's ability to cause disease in humans [53]. This same grouping can be seen in the phylogenetic tree of the *stx1_a* genes (Supplemental Figure S2D). The *stx1_a* gene from 2013C-4465 was identical to the *stx1_a* genes from STEC O157:H7 *tir* 255T allele that associates with human disease. In STEC O157:H7, 98% of clinical isolates from humans have the *tir* 255T allele, while 2% have the *tir* 255A allele [53]. The strains carrying the *tir* 255A allele are proposed to have diverged from strains carrying *tir* 255T early in the evolution of STEC O157:H7. This suggests a possible lineage specificity of *stx1_a*-encoding prophages in O55:H7 and STEC O157:H7 strains depending upon the *tir* 255 allele variant.

3.4. Insertion Sequences

Insertion sequence elements (IS elements) were specifically evaluated between a subset of isolates (those with common annotation in Genbank) shown to play a role in *E. coli* O157:H7 diversification in the population, specifically IS629 [56]. Across the six isolates studied, 450 IS elements were identified: 349 within chromosomes and 101 within plasmids (Figure 4).

On average, within the chromosome of O55:H7, there were 55 (ranging from 39 to 102) IS elements, while the plasmids had an average of 16 IS elements (ranging from 10 to 35). Eight IS elements were unique to one isolate, and eight were found in at least one copy within all isolates (six on the chromosome and one on the plasmids). The IS3 family transposase was found in all isolates on both the chromosome and plasmids in the highest copy number. One other IS element, IS4 family transposase, was found within at least one plasmid in all strains.

IS elements are drivers of evolution and diversity in bacteria. In STEC O157:H7, there are multiple reports of the Shiga toxin gene being inactivated by IS elements. In the strain DEC5B, there are three genes encoding for *citC*, fimbrial biogenesis outer membrane usher protein, and autotransporter outer membrane beta-barrel domain-containing protein that have IS elements inserted into them, presumably inactivating these genes. Two of these genes have an inserted IS3 family transposase, while one has an IS3-like element ISEc31 family transposase. *citC* is part of an operon that is responsible for citrate fermentation and would presumably not function in DEC5B. Recently, a novel overlapping, open reading frame was identified internal to the *citC* gene. The novel protein, Nog1, is thought to provide a growth advantage in the presence of MgCl₂ and is transcribed about 14-fold higher in cow dung compared to Luria broth [57]. Since these genes are intact in STEC O157, the proposed progeny of DEC5B, the IS elements were either excised or STEC O157 descended from another closely related strain missing the inserted IS elements.

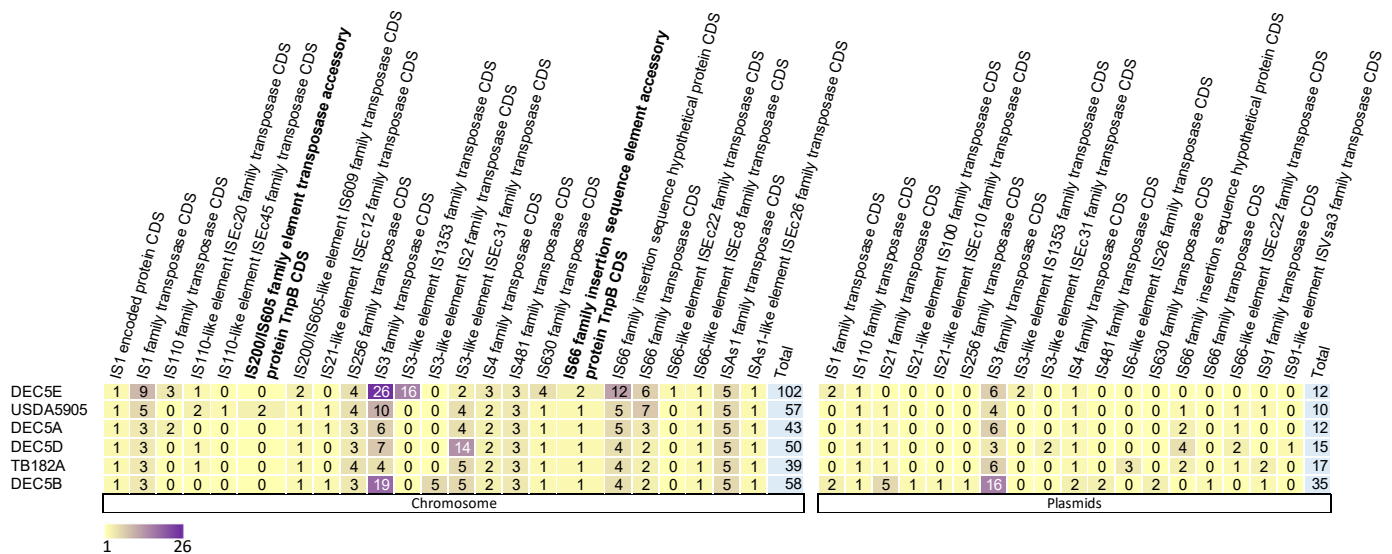


Figure 4. Heat map of insertion sequence (IS) elements in the chromosome and plasmids of six *E. coli* O55:H7. Plasmids are not broken down by individual plasmids; instead, all plasmids present in an isolate are aggregated into one. The strain names are on the y axis with the IS elements on the x axis. The number of copies of an IS element is represented by a different color.

3.5. Plasmids

Plasmid carriage was diverse, with 31 plasmids found between all isolates that grouped into six common backbones. Five of the strains carried an antibiotic resistance cassette, including two strains that had multiple cassettes. All plasmids were placed into one of six groups except plasmids pUSDA5905-1 and p12579-1 that appeared to be similar bacteriophages. All plasmids except pUSDA5905-1 and pUSDA5905-5 were circularized by trimming overlapping 5' and 3' end to make a closed circle.

Plasmid pDEC5B-3 was highly related to plasmids pUSDA5905-3, P12579-4, TB182A-3, and pDEC5D-2 (Supplemental Figure S4A). DEC5B-3 was 3.3 kb longer than the other related plasmids and had an additional *sul2* gene, while DEC5D-2 had an IS91-like element in place of the *aph(3')*-Ib and *aph(6)*-Id antibiotic resistance genes. Interestingly, pTB182A-3 and pTB182A-5 share the same antibiotic resistance cassette, *sul2*, *aph(3')*-Ib, and *aph(6)*-Id, indicating again the potential for this strain to have increased resistance to these antibiotics. These plasmids contain a plasmid replication initiation gene whose protein product recruits and positions an active helicase at the plasmid replication origin [58].

Plasmids pTB182A-2, pDEC5B-2, pDEC5A-3, and pFDAARGOS-3 are between 5.4 kb and 6.8 kb in size and share just over 3 kb of homology to each other (Supplemental Figure S4B). Plasmids pDEC5B-2 and pTB182A-2 are the same size and only differ by nine bases. The genes found in the related regions were *mobC* and *mbeA*, whose encoded proteins are multifunctional and promote conjugal plasmid mobilization [59,60]. These plasmids are missing the *mobB* gene, which is part of the mobilization operon. If present, these strains could use the conjugation system (*tra* operon) on an IncF plasmid in the cell to be horizontally transferred to other strains. Plasmids pDEC5B-2 and pTB182A-2 possess a nickel transport and two hypothetical genes in the non-conserved region, while pDEC5A-3 and pFDAARGOS-3 have a colicin 10 operon.

pDEC5A-3, pFDAARGOS-2, and pUSDA5905-5 are small plasmids with a roughly 1.7-kb homologous region that contains the gene *nikA*. The Nika protein from the plasmid R64 combines with NikB to form a relaxation complex at the *oriT* region of the plasmid and prepares the plasmid to be replicated [61]. The relaxation complex and replication may not function in these plasmids because there is no annotated *nikB* gene or other genes involved in plasmid replication. (Supplemental Figure S4C). pDEC5A-3 and pFDAARGOS-2 are identical and contain an ATP-binding protein in the non-conserved region, while plasmid

pUSDA5905-5 has an RNA-directed DNA polymerase. A group of plasmids was only found in strains USDA5905 and RM12579.

Plasmids p12579-3, p12579-5, and pUSDA5905-4 have homology with a conserved region of approximately 3.0 kb that contains the complete plasmid mobilization operon *mobABC*. The proteins encoded by this operon are required for formation of the relaxosome for mobilizing the small plasmid (Supplemental Figure S4D). Like the previously described plasmids, these plasmids are in strains that contain an IncF plasmid that contains the *tra* operon for transferring mobile elements between strains and is compatible with mobilizing plasmids with *mobABC* [62]. This indicates the potential of these small plasmids to be horizontally transferred to other strains. The IncF plasmid p12579-3 has three hypothetical genes, a transposase, and a beta-lactamase TEM antibiotic resistance gene in the non-conserved region, while p12579-5 only has two hypothetical genes. pUSDA5905-6 only has a hypothetical gene in the non-conserved region, while pUSDA5905-4 has a site-specific methyl transferase and restriction endonuclease genes.

pDEC5B-5 was most similar to plasmids pDEC5E-2 and TB182A-5 (Supplemental Figure S4E). These plasmids range in size from 70 kb to 99 kb, with the defining feature being the *tra* conjugation transfer systems. While this region appears to be conserved, the non-conserved regions are quite different. Plasmid pDEC5B-5 has 16 IS elements and a translesion error-prone DNA polymerase (*umuC* and *umuD*), while pTB182A-5 contains antibiotic resistance genes TEM-1, *sul2*, *aph(3'')*-Ib, *aph(6)*-Id, and *dfrA8*. pDEC5E-2 contains the previously described antibiotic resistance region and a mercury resistance operon. These plasmids belong to the IncFII plasmid incompatibility group.

A 57-kb to 69-kb plasmid (pO55) was found in all strains except for DEC5B (Figure 5). This plasmid belongs to the IncFIB incompatibility group and shares similar regions with the pO157:H7 plasmid in STEC O157:H7 [9]. However, unlike pO157, whose plasmids are almost identical except for a few IS elements, there were four variants of pO55. The differences were due to deletions, insertions or a replacement of a region with other genes. The most common genes shared by pO55 and pO157 were for conjugation, a type II secretion system, a colicin, and a type III secretion system effector, *nleA*. Interestingly, the plasmid pDEC5D-3 was missing the type II secretion system and colicin genes, while the plasmid pDEC5E-3 also contains the antibiotic resistance genes *aadA1*, *qacEDelta1*, and *sul1*, along with an operon for mercury resistance. This 16kb resistance region was also found on the pDEC5E-2, another plasmid in DEC5E, indicating the potential of increased resistance to these antibiotics. pDEC5E-2 also contains a type A-1 chloramphenicol resistance gene. The pO55 plasmid appears to evolve at the same rate as the chromosome. When the phylogenetic trees of the chromosome and pO55 were compared, the structure of the subtrees matched at its best corresponding node. (Supplemental Figure S5). This result is similar to STEC O157:H7, the proposed descendant of O55:H7. Nyong et al. showed a stable evolutionary relationship between the host chromosome and pO157 plasmid [20]. Plasmids, being mobile elements, are thought to be transient in bacteria. However, there are now examples in two *E. coli* serotypes where plasmids have taken on the role of extrachromosomal elements and are stably maintained in the population.

Strain DEC5B was the lone O55:H7 isolate that did not contain a pO55 plasmid. Plasmid pDEC5B-4 belongs to the IncFII incompatibility group and was the closest plasmid in DEC5B that resembled pO55 or pO157. (Supplement Figure S6). The homology between pDEC5B-4 and pO157 includes several hypothetical genes and conjugation genes (*tra*) (Supplement Figure S6A). DEC5B-4 shows the most similarity to pDEC5B-5, another plasmid in DEC5B, pDEC5E-2, and TB182A-2. This region was about 11 kb in length, with most of the similarity with the *tra* genes (Supplement Figure S6B). Comparison of pDEC5B-4 to pO55 shows homology with the same hypothetical genes as pO157, but to a different region of the *tra* operon (Supplement Figure S6C). Interestingly, DEC5B was the strain most closely related to STEC O157:H7 strains [6] and had the same sequence type as five other O55 strains. However, it did not contain a plasmid that was similar to either pO55 or pO157. This suggests that at some time this strain lost the pO55 plasmid

and acquired plasmid pDEC5B-4, or this strain belongs to a lineage of O55:H7 strains that never acquired plasmid pO55. While strain DEC5B is currently the closest known ancestor to STEC O157:H7, there is a yet to be discovered strain from which the STEC O157:H7 lineage descended.

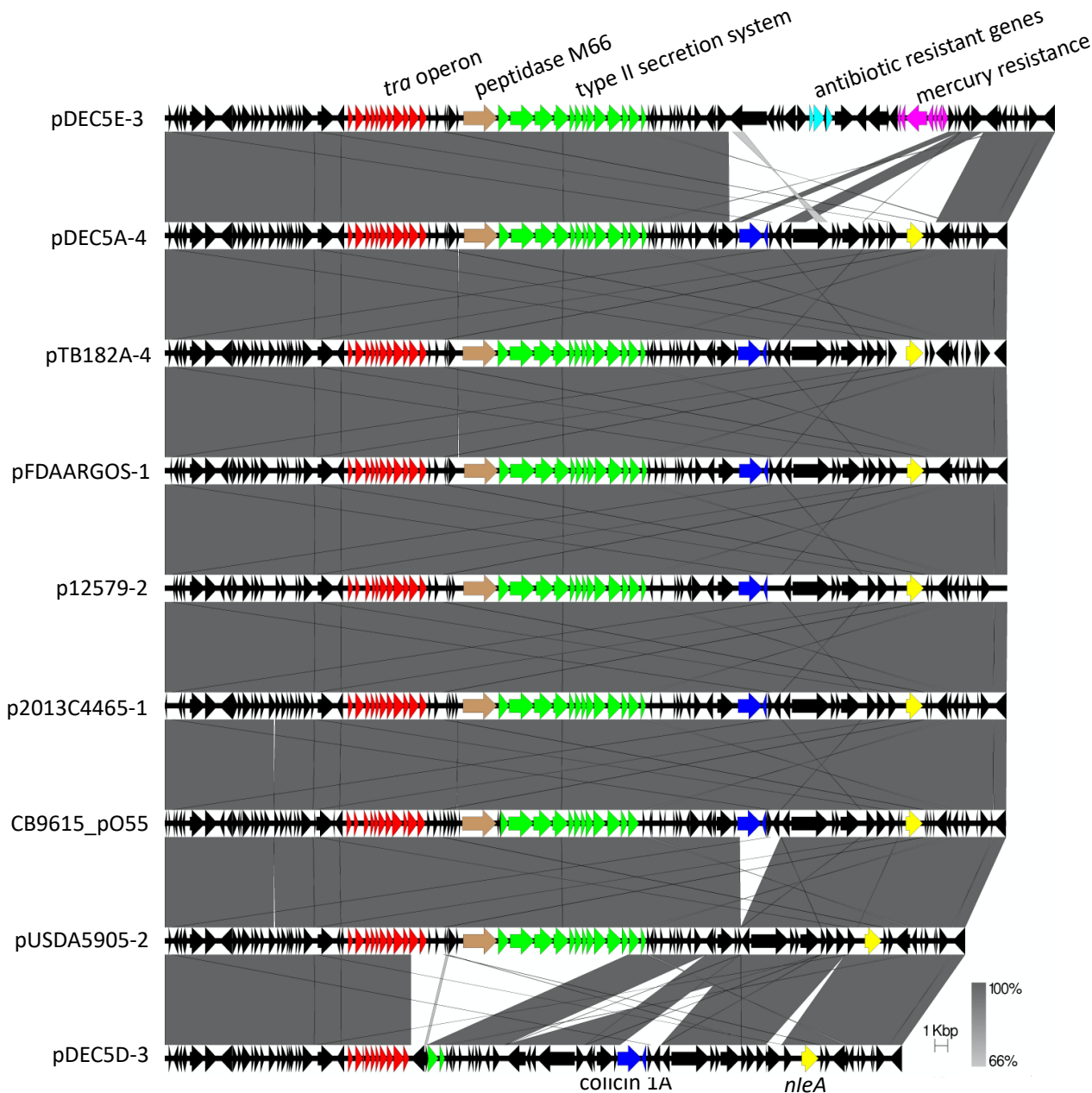


Figure 5. Genetic map showing the O55:H7 57-kb to 69-kb plasmid present in 9 of the 10 strains. Each colored arrow represents a different gene or group of genes labeled above or below the genetic map. The black arrows are other genes on the plasmids. The shades of gray between the genetic map represents the percent similarity of the plasmids to each other. The scale bar represents plasmid size in kilobases.

3.6. Comparison of Genes in O55:H7 Complete Genomes

Using Roary, the gene content of the 10 O55:H7 strains was compared based on phylogenetic tree placement (Figure 1). Three strains, DEC5E, USDA5905, and DEC5B, were the lone occupants at branch tips and were compared as individuals. The rest of the strains could be placed into one of three groups: group 1—DEC5A and FDAARGOS_946, group 2—DEC5D, RM12579 and TB182A, and group 3—CB9615 and 2013C-4465 (Supplemental

Table S4). DEC5E had the most unique genes when compared to the other strains ($n = 503$), with 73 unique to the plasmids and 430 unique to the chromosome. Group 2 had the least unique genes ($n = 42$), with all of them on the chromosome. Most of the unique genes in these comparisons were hypothetical proteins or phage-related and were classified as belonging to gene families (Supplemental Table S4).

The Roary output classified duplicated genes caused by a missense mutation that disrupted the open reading frame. Some of the mutated genes provided essential functions for the bacterium (Supplemental Table S4). For instance, in DEC5E, a base substitution in the *eutA* gene created a truncated protein. The *eutA* gene is part of the operon for ethanolamine utilization. Ethanolamine catabolism is associated with bacterial pathogenicity in *S. Typhimurium*. Transcript expression studies link the increase in *eut* expression with the activity of global regulators including CsrA and Fis [63]. The DEC5E strain was isolated from a human specimen, indicating that in this situation a complete *eutA* gene was not needed for causing human disease. DEC5E does contain seven virulence factors that were not found in the other O55:H7 strains and four that were shared with two or three other strains, so the possibility exists that these factors were able to overcome the loss of *eutA* and cause disease in this individual.

The chromosome of USDA5905 contained a missense mutation in the *rssB* gene that encodes for the response regulator RssB. RssB acts as a proteolytic recognition targeting factor for RpoS, a stationary phase sigma factor that controls many genes involved in helping cells deal with the stresses from being in a stationary growth phase. RpoS is regulated at the transcriptional and translational levels, but RssB regulates RpoS by specifically targeting the protein for degradation by ClpXP [64]. *rssB* mutants express high levels of RpoS and impaired osmotic regulation and stationary phase response [65]. In *S. typhimurium*, inactivation of *movA*, the homolog of *rssB* in *E. coli*, caused a growth defect resulting in small colonies and attenuated virulence [66]. While no colony morphology difference was noted between USDA5905 and the other O55:H7 strain, additional studies would be needed to determine if this strain has similar phenotypes to the *S. typhimurium movA* mutant.

Throughout group 1, there were 127 unique genes, with 8 found on plasmids and 121 on the chromosome. This group of strains possessed a truncated *uidB* gene created by a missense mutation at bp 1168 (G-T) that resulted in a protein that is 68 amino acids shorter than the other strains. UidB is a proton-dependent transporter for α - and β -glucuronides, transporting them from the environment into the cell where they are cleaved to yield glucuronate [67]. Glucuronate is then available as a source of carbon for growth. In most STEC O157:H7 cells, the UidA protein is truncated by an insertion in the *uidA* gene [68]. The result of this inactivated protein provides one of the key phenotypic diagnostic features of this serotype. The inability of UidA to cleave β -glucuronides substrates that produce a color or fluorescence, along with the inability to ferment sorbitol, is used in different methods to identify potential STEC O157:H7 isolates. Group 1 strains appear to share the same UidA phenotype as STEC O157:H7, as the truncated UidB protein would not be able to import α - and β -glucuronides into the cells allowing UidA to cleave them. The loss of the ability to use α - and β -glucuronides as a carbon source does not appear to prevent O55:H7 group 1 strains and STEC O157:H7 strains from causing disease in humans.

Groups 2 and 3 did not have any unique gene in their plasmids and had a total of 43 and 46 unique genes, respectively, on the chromosome. For Group 2 strains, sulfoquinovose isomerase, a gene in the sulfoquinovose degradation I system, had a thymidine inserted at bp 482, creating a truncated protein. The insertion of a base may be caused by a sequencing error. However, this insertion was present in all three strains of this group which were sequenced independently in three different labs. The strain sequenced from our group was verified by mapping short-read sequences to the insertion site. Sulfoquinovose (SQ) is one of the most abundant organic sulfur compounds in nature and is found in many plants. Sulfoquinovose isomerase is the first enzyme in the Embden-Meyerhof-Parnas pathway, and converts sulfoquinovose to sulfofructose [69]. The end products of the pathway are dihydroxyacetone phosphate, which provides energy for growth, and 2,3-

dihydroxypropane sulfonate, which is exported from the cell. In culture, *E. coli* strain K-12 can use SQ as a sole source of carbon and energy. It is proposed that this pathway provides a source of bacterial energy in the intestinal tracts of humans and animals where available sources of metabolites are limited [70]. The inactivation of this pathway in Group 2 strains would suggest that these strains may not grow as vigorously as strains with a complete pathway.

In Group 3, the *pgaB* had a single bp deletion at bp 778 of the gene. This region of the gene has a stretch of seven adenosines in the full-length gene but only six in the two strains from Group 3. Strains CB9615 and 2013C-4465 were sequenced by other groups, so the deletion could not be verified. However, this deletion was not seen in the other strains in the study and were independently sequenced by two groups, so we assume the deletion is real. The *pgaABCD* operon affects biofilm development by promoting abiotic surface binding and intercellular adhesion by synthesizing and exporting poly-N-acetyl glucosamine (PNAG). All genes in this operon are required for optimal biofilm formation [71]. In uropathogenic *E. coli*, the *pgaABCD* operon is required for fitness in a mouse model of bacteremia and urinary tract infection and promotes biofilm formation in uropathogenic *E. coli* [72]. The inability to make PNAG in strains CB9615 and 2013C-4465 leaves them at a disadvantage during competition with other bacteria for resources in the environment and potentially reduces their ability to cause disease in humans.

Lastly, DEC5B had 288 unique genes, with 149 unique to the plasmids and 139 unique to the chromosome. In DEC5B, the *malE* gene contained a missense mutation at bp 551 (G-A) that created a truncated MalE protein. MalE is a maltose/maltodextrin ABC transporter substrate-binding protein responsible for delivering maltose or maltodextrin to the transport complex for internalization into the cell. Maltose is then cleaved by an amylase to release two glucose molecules for glycolysis. Maltose has been shown to be important for colonization of pathogenic and commensal *E. coli* strain in the intestines of mice [73]. Without the ability to utilize maltose, DEC5B might be at a disadvantage when colonizing the intestines of humans and thus reducing their ability to cause disease.

Nine of the ten strains used in this study were isolated from humans, while USDA5905 was isolated from food (meat). Interestingly, there were truncated genes in each of the human strains that were involved in metabolism, colonization, or the ability to cause disease in humans. This indicates that the other virulence factors in the strains were able to overcome the loss of function from these inactivated genes. However, we do not know much about the immune or health status of the humans that were infected with these strains. Additional research would be needed to understand how the combination of virulence factors and other genomic diversity of these strains affect their ability to cause disease.

4. Conclusions

This study leveraged the complete closed genomes from 10 *E. coli* O55:H7 strains to look at the diversity of these human pathogenic strains. The strains were assigned to three clades. Differentiation of clades could be attributed to sequence type and virulence gene profile but not to the chromosome architecture, which was similar except for one strain. Prophage and insertion sequence content did not associate with clade assignment or sequence type. Two strains contained a prophage that encoded for two different Shiga toxin genes, *stx1_a* and *stx2_d*. The O55:H7 prophage containing *stx1_a* had a different gene structure than those from STEC O157:H7, but the *stx1* gene was identical to a STEC O157:H7 *stx1* gene. The O55:H7 *stx2_d*-containing prophage and the *stx2_d* gene were not similar to *stx2_d*-containing prophages or *stx2_d* genes from other serotypes. The O55:H7 strains contained many plasmids that did not associate with clade or sequence type. These plasmids could be classified into eight groups with one group closely related to prophages. There were several strains that contained plasmids with multiple copies of the same antibiotic resistance cassettes and mercury resistance operons, indicating the potential of these strains to be resistant to a higher concentration of these antimicrobials. Plasmid pO55 was found in nine of the ten strains. This plasmid showed high homology across the length of the plasmid

except in two strains with an insert, with antibiotic and heavy-metal resistance in one strain and a deletion of the type II secretions system in the other strain. The phylogenetic tree from the core genome of this plasmid had the same branching pattern as the chromosome, indicating that pO55 has a stable evolutionary relationship with the chromosome. Finally, missense mutations in genes related to metabolism, colonization, and virulence factors were identified that associated with clade assignment. Despite the diverse genome, these O55:H7 strains were still able to cause disease in humans.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms10081545/s1>, Figure S1: Mauve alignment of all *E. coli* O55:H7 chromosomes, Figure S2: Comparison of *stx*-encoding prophages in O55:H7 to related prophages and *stx* genes, Figure S3: Integration site of *stx2_d* prophage in O17/O44:H18 compared to the prophages integrated into the same site in O55:H7 and STEC O157:H7, Figure S4: Additional plasmid identified in O55:H7 strains, Figure S5: Comparison of pDEC5B-4 to pO157, pO55, and additional pO55 plasmids, Figure S6: Comparison of the phylogenetic trees from the chromosome and pO55 plasmids from O55:H7. Table S1: Strains used in the study with associated metadata, Table S2: The prophage content of each strain as determined by PHASTER, Table S3: The insertion element content of *E. coli* O55:H7 strains as determined by NCBI annotation, Table S4: The unique gene content of *E. coli* O55:H7 strains as determined by Roary.

Author Contributions: Conceptualization, J.L.B.; formal analysis, M.D.W. and J.L.B.; investigation, M.D.W. and J.L.B.; data curation, M.D.W. and J.L.B.; writing—original draft preparation, M.D.W.; writing—review and editing, J.L.B. and M.D.W.; supervision, J.L.B.; project administration, J.L.B.; funding acquisition, J.L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the USDA Agricultural Research Service project number 3040-42000-017-00D.

Data Availability Statement: All newly sequenced fastq files and assemblies have been submitted to the National Center for Biotechnology Information (NCBI) under BioProject PRJNA528413.

Acknowledgments: We would like to thank Sandy Fryda-Bradley and Kerry Brader for technical assistance; Bob Lee, Kelsey McClure, and Kristen Kuhn for sequencing support; Aaron Rogge and Randy Bradley for IT support; and Jody Gallagher for secretarial assistance. The use of product and company names is necessary to accurately report the methods and results; however, the United States Department of Agriculture (USDA) neither guarantees nor warrants the standard of the products, and the use of names by the USDA implies no approval of the product to the exclusion of others that may also be suitable. The USDA is an equal opportunity provider and employer.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Spears, K.J.; Roe, A.J.; Gally, D.L. A comparison of enteropathogenic and enterohaemorrhagic *Escherichia coli* pathogenesis. *FEMS Microbiol. Lett.* **2006**, *255*, 187–202. [[CrossRef](#)]
2. Yatsuyanagi, J.; Saito, S.; Sato, H.; Miyajima, Y.; Amano, K.-I.; Enomoto, K. Characterization of enteropathogenic and enteroaggregative *Escherichia coli* Isolated from Diarrheal Outbreaks. *J. Clin. Microbiol.* **2002**, *40*, 294–297. [[CrossRef](#)] [[PubMed](#)]
3. McFarland, N.; Bundle, N.; Jenkins, C.; Godbole, G.; Mikhail, A.; Dallman, T.; O'Connor, C.; McCarthy, N.; O'Connell, E.; Treacy, J.; et al. Recurrent seasonal outbreak of an emerging serotype of Shiga toxin-producing *Escherichia coli* (STEC O55:H7 Stx2a) in the south west of England, July 2014 to September 2015. *Euro Surveill.* **2017**, *22*, 30610. [[CrossRef](#)] [[PubMed](#)]
4. Feng, P.C.H.; Monday, S.R.; Lacher, D.W.; Allison, L.; Siitonen, A.; Keys, C.; Eklund, M.; Nagano, H.; Karch, H.; Keen, J.; et al. Genetic Diversity among Clonal Lineages within *Escherichia coli* O157:H7 Stepwise Evolutionary Model. *Emerg. Infect. Dis.* **2007**, *13*, 1701–1706. [[CrossRef](#)] [[PubMed](#)]
5. Feng, P.; Lampel, K.A.; Karch, H.; Whittam, T.S. Genotypic and Phenotypic Changes in the Emergence of *Escherichia coli* O157:H7. *J. Infect. Dis.* **1998**, *177*, 1750–1753. [[CrossRef](#)]
6. Kyle, J.L.; Cummings, C.A.; Parker, C.T.; Quiñones, B.; Vatta, P.; Newton, E.; Huynh, S.; Swimley, M.; Degoricija, L.; Barker, M.; et al. *Escherichia coli* Serotype O55:H7 Diversity Supports Parallel Acquisition of Bacteriophage at Shiga Toxin Phage Insertion Sites during Evolution of the O157:H7 Lineage. *J. Bacteriol.* **2012**, *194*, 1885–1896. [[CrossRef](#)]

7. Pupo, G.M.; Karaolis, D.K.; Lan, R.; Reeves, P.R. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. *Infect. Immun.* **1997**, *65*, 2685–2692. [[CrossRef](#)]
8. Wick, L.M.; Qi, W.; Lacher, D.W.; Whittam, T.S. Evolution of Genomic Content in the Stepwise Emergence of *Escherichia coli* O157:H7. *J. Bacteriol.* **2005**, *187*, 1783–1791. [[CrossRef](#)]
9. Zhou, Z.; Li, X.; Liu, B.; Beutin, L.; Xu, J.; Ren, Y.; Feng, L.; Lan, R.; Reeves, P.R.; Wang, L. Derivation of *Escherichia coli* O157:H7 from Its O55:H7 Precursor. *PLoS ONE* **2010**, *5*, e8700. [[CrossRef](#)]
10. Leopold, S.R.; Magrini, V.; Holt, N.J.; Shaikh, N.; Mardis, E.R.; Cagno, J.; Ogura, Y.; Iguchi, A.; Hayashi, T.; Mellmann, A.; et al. A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8713–8718. [[CrossRef](#)]
11. Jenke, C.; Leopold, S.R.; Weniger, T.; Rothgänger, J.; Harmsen, D.; Karch, H.; Mellmann, A. Identification of Intermediate in Evolutionary Model of Enterohemorrhagic *Escherichia coli* O157. *Emerg. Infect. Dis.* **2012**, *18*, 582–588. [[CrossRef](#)] [[PubMed](#)]
12. Didelot, X.; Bowden, R.; Wilson, D.J.; Peto, T.E.A.; Crook, D.W. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* **2012**, *13*, 601–612. [[CrossRef](#)] [[PubMed](#)]
13. Tagini, F.; Greub, G. Bacterial genome sequencing in clinical microbiology: A pathogen-oriented review. *Eur. J. Clin. Microbiol. Infect. Dis.* **2017**, *36*, 2007–2020. [[CrossRef](#)] [[PubMed](#)]
14. Jenkins, C.; Dallman, T.J.; Grant, K.A. Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing *Escherichia coli* serogroup O157:H7, England, 2013 to 2017. *Eurosurveillance* **2019**, *24*, 1800346. [[CrossRef](#)]
15. Nadon, C.; Van Walle, I.; Gerner-Smidt, P.; Campos, J.; Chinen, I.; Concepcion-Acevedo, J.; Gilpin, B.; Smith, A.M.; Kam, K.M.; Perez, E.; et al. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Eurosurveillance* **2017**, *22*, 30544. [[CrossRef](#)]
16. Koren, S.; Harhay, G.P.; Smith, T.P.; Bono, J.L.; Harhay, D.M.; Mcvey, S.D.; Radune, D.; Bergman, N.H.; Phillippy, A.M. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol.* **2013**, *14*, R101. [[CrossRef](#)]
17. Koren, S.; Phillippy, A.M. One chromosome, one contig: Complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.* **2015**, *23*, 110–120. [[CrossRef](#)]
18. De Maio, N.; Shaw, L.P.; Hubbard, A.; George, S.; Sanderson, N.D.; Swann, J.; Wick, R.; AbuOun, M.; Stubberfield, E.; Hoosdally, S.J.; et al. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb. Genom.* **2019**, *5*, e000294. [[CrossRef](#)]
19. Whittam, T.S.; Wolfe, M.L.; Wachsmuth, I.K.; Orskov, F.; Orskov, I.; Wilson, R.A. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect. Immun.* **1993**, *61*, 1619–1629. [[CrossRef](#)]
20. Nyong, E.C.; Zaia, S.R.; Allué-Guardia, A.; Rodriguez, A.L.; Irion-Byrd, Z.; Koenig, S.S.K.; Feng, P.; Bono, J.L.; Eppinger, M. Pathogenomes of Atypical Non-shigatoxigenic *Escherichia coli* NSF/SF O157:H7/NM: Comprehensive Phylogenomic Analysis Using Closed Genomes. *Front. Microbiol.* **2020**, *11*, 619. [[CrossRef](#)]
21. Chin, C.-S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E.; et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **2013**, *10*, 563–569. [[CrossRef](#)] [[PubMed](#)]
22. Luo, H.; Zhang, C.-T.; Gao, F. Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. *Front. Microbiol.* **2014**, *5*, 482. [[CrossRef](#)] [[PubMed](#)]
23. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)] [[PubMed](#)]
24. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [[CrossRef](#)]
25. Treangen, T.J.; Ondov, B.D.; Koren, S.; Phillippy, A.M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **2014**, *15*, 524. [[CrossRef](#)]
26. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [[CrossRef](#)]
27. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
28. Petkau, A.; Stuart-Edwards, M.; Stothard, P.; Van Domselaar, G. Interactive microbial genome visualization with GView. *Bioinformatics* **2010**, *26*, 3125–3126. [[CrossRef](#)]
29. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [[CrossRef](#)] [[PubMed](#)]
30. Sullivan, M.J.; Petty, N.K.; Beatson, S.A. Easyfig: A genome comparison visualizer. *Bioinformatics* **2011**, *27*, 1009–1010. [[CrossRef](#)] [[PubMed](#)]
31. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [[CrossRef](#)] [[PubMed](#)]
32. Zhao, S.; Guo, Y.; Sheng, Q.; Shyr, Y. Advanced Heat Map and Clustering Analysis Using Heatmap3. *BioMed Res. Int.* **2014**, *2014*, e986048. [[CrossRef](#)] [[PubMed](#)]

33. Robinson, O.; Dylus, D.; Dessimoz, C. Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. *Mol. Biol. Evol.* **2016**, *33*, 2163–2166. [[CrossRef](#)] [[PubMed](#)]
34. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)] [[PubMed](#)]
35. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)]
36. Wirth, T.; Falush, D.; Lan, R.; Colles, F.; Mensa, P.; Wieler, L.H.; Karch, H.; Reeves, P.R.; Maiden, M.C.J.; Ochman, H.; et al. Sex and virulence in *Escherichia coli*: An evolutionary perspective. *Mol. Microbiol.* **2006**, *60*, 1136–1151. [[CrossRef](#)]
37. Lau, S.H.; Reddy, S.; Cheesbrough, J.; Bolton, F.J.; Willshaw, G.; Cheasty, T.; Fox, A.J.; Upton, M. Major Uropathogenic *Escherichia coli* Strain Isolated in the Northwest of England Identified by Multilocus Sequence Typing. *J. Clin. Microbiol.* **2008**, *46*, 1076–1080. [[CrossRef](#)]
38. Fitzgerald, S.F.; Lupolova, N.; Shaaban, S.; Dallman, T.J.; Greig, D.; Allison, L.; Tongue, S.C.; Evans, J.; Henry, M.K.; McNeilly, T.N.; et al. Genome structural variation in *Escherichia coli* O157:H7. *Microb. Genom.* **2021**, *7*, 000682. [[CrossRef](#)]
39. Scott, A.E.; Timms, A.R.; Connerton, P.L.; Carrillo, C.L.; Radzum, K.A.; Connerton, I.F. Genome Dynamics of *Campylobacter jejuni* in Response to Bacteriophage Predation. *PLoS Pathog.* **2007**, *3*, e119. [[CrossRef](#)] [[PubMed](#)]
40. Guérillot, R.; Kostoulas, X.; Donovan, L.; Li, L.; Carter, G.P.; Hachani, A.; Vandellannoote, K.; Giulieri, S.; Monk, I.R.; Kunitomo, M.; et al. Unstable chromosome rearrangements in *Staphylococcus aureus* cause phenotype switching associated with persistent infections. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 20135–20140. [[CrossRef](#)]
41. Sun, S.; Ke, R.; Hughes, D.; Nilsson, M.; Andersson, D.I. Genome-Wide Detection of Spontaneous Chromosomal Rearrangements in Bacteria. *PLoS ONE* **2012**, *7*, e42639. [[CrossRef](#)] [[PubMed](#)]
42. Li, C.; Pan, D.; Li, M.; Wang, Y.; Song, L.; Yu, D.; Zuo, Y.; Wang, K.; Liu, Y.; Wei, Z.; et al. Aerobactin-Mediated Iron Acquisition Enhances Biofilm Formation, Oxidative Stress Resistance, and Virulence of *Yersinia pseudotuberculosis*. *Front. Microbiol.* **2021**, *12*, 699913. [[CrossRef](#)] [[PubMed](#)]
43. Law, D. Virulence factors of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli*. *J. Appl. Microbiol.* **2000**, *88*, 729–745. [[CrossRef](#)]
44. Slater, S.L.; Frankel, G. Advances and Challenges in Studying Type III Secretion Effectors of Attaching and Effacing Pathogens. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 337. [[CrossRef](#)]
45. Deng, W.; Vallance, B.A.; Li, Y.; Puente, J.L.; Finlay, B.B. *Citrobacter rodentium* translocated intimin receptor (Tir) is an essential virulence factor needed for actin condensation, intestinal colonization and colonic hyperplasia in mice. *Mol. Microbiol.* **2003**, *48*, 95–115. [[CrossRef](#)]
46. Mundy, R.; Petrovska, L.; Smollett, K.; Simpson, N.; Wilson, R.K.; Yu, J.; Tu, X.; Rosenshine, I.; Clare, S.; Dougan, G.; et al. Identification of a novel *Citrobacter rodentium* type III secreted protein, EspI, and roles of this and other secreted proteins in infection. *Infect. Immun.* **2004**, *72*, 2288–2302. [[CrossRef](#)]
47. Kelly, M.; Hart, E.; Mundy, R.; Marchès, O.; Wiles, S.; Badea, L.; Luck, S.; Tauschek, M.; Frankel, G.; Robins-Browne, R.M.; et al. Essential role of the type III secretion system effector NleB in colonization of mice by *Citrobacter rodentium*. *Infect. Immun.* **2006**, *74*, 2328–2337. [[CrossRef](#)] [[PubMed](#)]
48. Creuzburg, K.; Recktenwald, J.; Kuhle, V.; Herold, S.; Hensel, M.; Schmidt, H. The Shiga Toxin 1-Converting Bacteriophage BP-4795 Encodes an NleA-Like Type III Effector Protein. *J. Bacteriol.* **2005**, *187*, 8494–8498. [[CrossRef](#)]
49. De Greve, H.; Qizhi, C.; Deboeck, F.; Hernalsteens, J.-P. The Shiga-toxin VT2-encoding bacteriophage ϕ 297 integrates at a distinct position in the *Escherichia coli* genome. *Biochim. et Biophys. Acta-Gene Struct. Expr.* **2002**, *1579*, 196–202. [[CrossRef](#)]
50. Recktenwald, J.; Schmidt, H. The Nucleotide Sequence of Shiga Toxin (Stx) 2e-Encoding Phage ϕ P27 Is Not Related to Other Stx Phage Genomes, but the Modular Genetic Structure Is Conserved. *Infect. Immun.* **2002**, *70*, 1896–1908. [[CrossRef](#)]
51. Bonanno, L.; Loukiadis, E.; Mariani-Kurkdjian, P.; Oswald, E.; Garnier, L.; Michel, V.; Auvray, F. Diversity of Shiga Toxin-Producing *Escherichia coli* (STEC) O26:H11 Strains Examined via stx Subtypes and Insertion Sites of Stx and EspK Bacteriophages. *Appl. Environ. Microbiol.* **2015**, *81*, 3712–3721. [[CrossRef](#)] [[PubMed](#)]
52. Yokoyama, K.; Makino, K.; Kubota, Y.; Watanabe, M.; Kimura, S.; Yutsudo, C.H.; Kurokawa, K.; Ishii, K.; Hattori, M.; Tatsuno, I.; et al. Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak. *Gene* **2000**, *258*, 127–139. [[CrossRef](#)]
53. Ogura, Y.; Ooka, T.; Asadulghani, Terajima, J.; Nougayrède, J.-P.; Kurokawa, K.; Tashiro, K.; Tobe, T.; Nakayama, K.; Kuhara, S.; et al. Extensive genomic diversity and selective conservation of virulence-determinants in enterohemorrhagic *Escherichia coli* strains of O157 and non-O157 serotypes. *Genome Biol.* **2007**, *8*, R138. [[CrossRef](#)] [[PubMed](#)]
54. Shringi, S.; Schmidt, C.; Katherine, K.; Brayton, K.A.; Hancock, D.D.; Besser, T.E. Carriage of stx2a Differentiates Clinical and Bovine-Biased Strains of *Escherichia coli* O157. *PLoS ONE* **2012**, *7*, e51572. [[CrossRef](#)]
55. Yara, D.A.; Greig, D.R.; Gally, D.L.; Dallman, T.J.; Jenkins, C. Comparison of Shiga toxin-encoding bacteriophages in highly pathogenic strains of Shiga toxin-producing *Escherichia coli* O157:H7 in the UK. *Microb. Genom.* **2020**, *6*, e000334. [[CrossRef](#)]
56. Rump, L.V.; Fischer, M.; Gonzalez-Escalona, N. Prevalence, distribution and evolutionary significance of the IS629 insertion element in the stepwise emergence of *Escherichia coli* O157:H7. *BMC Microbiol.* **2011**, *11*, 133. [[CrossRef](#)]
57. Fellner, L.; Simon, S.; Scherling, C.; Witting, M.; Schober, S.; Polte, C.; Schmitt-Kopplin, P.; Keim, D.A.; Scherer, S.; Neuhaus, K. Evidence for the recent origin of a bacterial protein-coding, overlapping orphan gene by evolutionary overprinting. *BMC Evol. Biol.* **2015**, *15*, 283. [[CrossRef](#)]

58. Jiang, Y.; Pacek, M.; Helinski, D.R.; Konieczny, I.; Toukdarian, A. A multifunctional plasmid-encoded replication initiation protein both recruits and positions an active helicase at the replication origin. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8692–8697. [CrossRef]
59. Zhang, S.; Meyer, R. The relaxosome protein MobC promotes conjugal plasmid mobilization by extending DNA strand separation to the nick site at the origin of transfer. *Mol. Microbiol.* **1997**, *25*, 509–516. [CrossRef]
60. Varsaki, A.; Lucas, M.; Afendra, A.S.; Drainas, C.; de la Cruz, F. Genetic and biochemical characterization of MbeA, the relaxase involved in plasmid ColE1 conjugative mobilization. *Mol. Microbiol.* **2003**, *48*, 481–493. [CrossRef] [PubMed]
61. Furuya, N.; Komano, T. Specific binding of the NikA protein to one arm of 17-base-pair inverted repeat sequences within the oriT region of plasmid R64. *J. Bacteriol.* **1995**, *177*, 46–51. [CrossRef]
62. Smillie, C.; Garcillán-Barcia, M.P.; Francia, M.V.; Rocha, E.P.C.; de la Cruz, F. Mobility of Plasmids. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 434–452. [CrossRef] [PubMed]
63. Garsin, D.A. Ethanolamine Utilization in Bacterial Pathogens: Roles and Regulation. *Nat. Rev. Microbiol.* **2010**, *8*, 290–295. [CrossRef] [PubMed]
64. Zhou, Y.; Gottesman, S.; Hoskins, J.R.; Maurizi, M.R.; Wickner, S. The RssB response regulator directly targets σ^S for degradation by ClpXP. *Genes Dev.* **2001**, *15*, 627–637. [CrossRef] [PubMed]
65. Muffler, A.; Fischer, D.; Altuvia, S.; Storz, G.; Hengge-Aronis, R. The response regulator RssB controls stability of the sigma(S) subunit of RNA polymerase in *Escherichia coli*. *EMBO J.* **1996**, *15*, 1333–1339. [CrossRef] [PubMed]
66. Swords, W.E.; Giddings, A.; Benjamin, W.H. Bacterial phenotypes mediated by *mviA* and their relationship to the mouse virulence of *Salmonella typhimurium*. *Microb. Pathog.* **1997**, *22*, 353–362. [CrossRef] [PubMed]
67. *Escherichia coli* K-12 substr. MG1655 uidB. Available online: <https://biocyc.org/gene?orgid=ECOLI&id=EG11658> (accessed on 30 June 2022).
68. Feng, P.; Lampel, K.A. Genetic analysis of uidA expression in enterohaemorrhagic *Escherichia coli* serotype O157:H7. *Microbiology* **1994**, *140 Pt 8*, 2101–2107. [CrossRef]
69. Felux, A.-K.; Spitteller, D.; Klebensberger, J.; Schleheck, D. Entner–Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E4298–E4305. [CrossRef]
70. Denger, K.; Weiss, M.; Felux, A.-K.; Schneider, A.; Mayer, C.; Spitteller, D.; Huhn, T.; Cook, A.M.; Schleheck, D. Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* **2014**, *507*, 114–117. [CrossRef]
71. Wang, X.; Preston, J.F.; Romeo, T. The pgaABCD Locus of *Escherichia coli* Promotes the Synthesis of a Polysaccharide Adhesin Required for Biofilm Formation. *J. Bacteriol.* **2004**, *186*, 2724–2734. [CrossRef]
72. Subashchandrabose, S.; Smith, S.N.; Spurbeck, R.R.; Kole, M.M.; Mobley, H.L.T. Genome-wide detection of fitness genes in uropathogenic *Escherichia coli* during systemic infection. *PLoS Pathog.* **2013**, *9*, e1003788. [CrossRef] [PubMed]
73. Jones, S.A.; Jorgensen, M.; Chowdhury, F.Z.; Rodgers, R.; Hartline, J.; Leatham, M.P.; Struve, C.; Krogfelt, K.A.; Cohen, P.S.; Conway, T. Glycogen and Maltose Utilization by *Escherichia coli* O157:H7 in the Mouse Intestine. *Infect. Immun.* **2008**, *76*, 2531–2540. [CrossRef] [PubMed]