

Article

Comparative Genomics of *Escherichia coli* Serogroups 64474, O179, O188 and *Shigella boydii* O16

Edwin Omar Desales-Decaro ¹, Graciela Castro-Escarpulli ¹ , Andres Saldaña-Padilla ¹ , Alejandro Cravioto ² , Hugo G. Castelán-Sánchez ^{3,*}  and Armando Navarro-Ocaña ^{4,*} 

¹ Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Mexico City 11340, Mexico

² Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; dracravioto@hotmail.com

³ Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 5C1, Canada

⁴ Public Health Department, Faculty of Medicine, Universidad Nacional Autónoma de México (UNAM), Avenida Universidad 3000, Ciudad Universitaria, Mexico City 04510, Mexico

* Correspondence: hcastelans@gmail.com (H.G.C.-S.); arnava@unam.mx (A.N.-O.); Tel.: +52-5544547025 (A.N.-O.)

Abstract

Shigella spp., and *Escherichia coli* exhibit notable genomic and phenotypic similarities, including serologically and genetically related somatic antigens. For example, the relationship among pathogenic strains *E. coli* 64474, O179, O188, and *S. boydii* O16 suggests a shared clonal origin. To evaluate their genomic proximity, a comparative genomics study was conducted using whole-genome sequencing. Comparative genomics involved *rfb* gene cluster regions and whole-genome comparisons. Phylogenomic inferences were performed using the virtual genome fingerprint (VGF) method with bootstrap support. The results revealed a high degree of genomic similarity and a close evolutionary relationship among *E. coli* strains, which also demonstrated genetic associations with clinically relevant pathotypes through the presence of virulence genes. Furthermore, serogroups 64474, O188, and *S. boydii* O16 exhibited close genetic relationships, suggesting that serotype 64474 could represent a novel serogroup, although its similarity to O188 indicates the influence of divergent factors. These findings support the hypothesis that these *E. coli* strains originated from a common clonal lineage, enhancing our understanding of serogroup diversity and the evolutionary dynamics within enteric pathogens.

Keywords: comparative genomics; *Escherichia coli*; *Shigella* spp.; serogroup O-antigen; phylogenomics; pangenome analysis; clonality; virulence genes



Academic Editor: Bogdan Nowicki

Received: 28 March 2026

Revised: 15 April 2026

Accepted: 23 April 2026

Published: 24 April 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Escherichia coli is a facultative anaerobic bacterium that includes both commensal and pathogenic strains; these pathogens are capable of causing a wide variety of diseases in different animal species Liu, 2020 [1]. Eight pathotypes that cause human diseases have been described, six of which are intestinal pathogens: enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC). The other two cause extraintestinal infections and are therefore called ExPEC; uropathogenic *E. coli* (UPEC) and meningitis-associated *E. coli* (MNEC) (Kaper, 2004; Pokharel, 2023) [2,3]. *Shigella* was recognized in 1890 as *Bacillus dysenteriae*, a facultative anaerobe intracellular pathogen

that causes bacillary dysentery or shigellosis (Lan, 2002; Stenhouse, 2023) [4,5]. *E. coli* and *Shigella* spp., are closely related species that were part of the same genus (then called *Bacillus*) until 1950, when *Shigella* was assigned to its own genus and classified into four species: *S. boydii*, *S. dysenteriae*, *S. flexneri*, and *S. sonnei*, also known as subgroups A, B, C, and D, respectively (Stenhouse, 2023; Pupo, 2000) [5,6]. Since this taxonomic separation, classification systems for each genus have been developed independently, with somatic (O) antigen serotyping established to identify *Shigella* species, as they lack surface (K) and flagellar (H) antigens. In contrast, *E. coli* classification is mainly genotypic, although serotyping is also used, based on distinctions between O, K and H antigens [1,7]. Most genes involved in O-antigen biosynthesis are grouped in the *rfb* cluster [8], which is typically located between the *galF* and *gnd* genes in both *E. coli* and *Shigella* [1].

Strain identification of *E. coli* and *Shigella* spp., can be performed using biochemical and molecular tools; however, differentiation depends on the level of genetic relatedness and the purpose of the study. A high degree of genotypic and phenotypic similarity may complicate differentiation. It is common to find *E. coli* and *Shigella* spp. strains with similar O-antigen structures, which may lead to a misdiagnosis, particularly in infections causing bacillary dysentery—commonly associated with both *Shigella* spp., and EIEC due to their genetic similarity [9]. This is evident in the strains studied here: *E. coli* 64474, O179, O188, and *S. boydii* type 16 (*S. boydii* O16), which share epitopes, particularly *E. coli* 64474, O188 and *S. boydii* O16. These strains express highly similar O-antigen phenotypes and share some *rfb* genes, such as *wzx* (O-antigen flippase) and *wzy* (O-antigen polymerase) [10,11].

It is worth noting that *E. coli* 64474 strains have undefined O-antigen and were isolated from fecal specimens of pediatric patients with diarrhea in different countries [10]. Meanwhile, *E. coli* O188 has been associated with the antibiotic resistance genes *mcr-9.1*, *bla_{VIM-1}*, *bla_{KPC-3}*, which confer resistance to colistin and beta-lactams, respectively. These findings are relevant given the high agglutination similarity of these serotypes with the O-antigen of *S. boydii* O16, which may also harbor or acquire these genes [12]. Lastly, *E. coli* O179 is a Shiga toxin-producing strain (STEC) [13], and although it shows lower agglutination levels, it remains relevant to this study.

The aim of this study is to explore the genomic relatedness and evolutionary relationships among these strains through comparative genomics and phylogenomic analysis, to assess the existence of clonality.

2. Materials and Methods

2.1. Strains

E. coli 64474, O179:H8 (E43478) and *S. boydii* O16 (G1219) strains were obtained from a previous study by the laboratory [10] and *E. coli* O188:H10 was obtained from the Statens Serum Institut (Copenhagen, Denmark).

2.2. Serotyping

The *E. coli* 64474, O179, *E. coli* O188 and *S. boydii* O16 strains were serotyped by agglutination assays [14] using 96-well microtiter plates with rabbit antisera (SERUNAM) obtained against 188 somatic antigens and 53 flagellar antigens for *E. coli*, and against 45 somatic antigens for *Shigella* species. Rabbit serum against the *E. coli* O188 strain was also prepared, and the anti-*E. coli* 64474, O179, and *S. boydii* O16 sera were previously obtained from the study referenced above.

2.3. Absorption Assays

Rabbit sera prepared against *E. coli* 64474, O179, O188 and *S. boydii* O16 strains were absorbed with homologous and heterologous antigens according to the method described by Ewing [15].

2.4. DNA Extraction and Sequencing

Genomic libraries were prepared using the Illumina TruSeq DNA Nano protocol, with an average insert size of approximately 550 bp. Library quality was assessed using a High Sensitivity Bioanalyzer chip. Sequencing was performed in a paired-end 2 × 300 bp format. The i7 index sequences assigned to the four libraries were ATCACG, CGATGT, TTAGGC, and TGACCA, respectively. In addition, A-tailing was performed prior to adapter ligation. The Index 1 (i7) adapter sequence was (A)GATCGGAAGAGCACACGTCTGAACTCCAGTCAC[i7]ATCTCGTATGCCGTCTTCTGCTTG, where the base in parentheses corresponds to the additional A incorporated during A-tailing. For adapter trimming, the following sequences were used: Read 1, AGATCGGAAGAGCACACGTCTGAACTCCAGTCA; Read 2, AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT. Genomic sequencing was performed by the Genomic Services Laboratory of the Center for Research and Advanced Studies of the IPN (CINVESTAV) Irapuato, Mexico.

2.5. Genome Assembly

The four genomes of *E. coli* 64474, O179, O188, and *S. boydii* O16 strains were sequenced using paired-end reads on an Illumina HiSeq 1500 platform. Raw reads (R1/R2) were evaluated with FastQC v0.12.1 and summarized with MultiQC v1.17. Reads were trimmed using Trimmomatic v0.39 to remove adapters/primers (ILLUMINACLIP), trim low-quality bases at read ends (LEADING:30, TRAILING:30), apply sliding-window trimming (SLIDINGWINDOW:4:15), and discard reads shorter than 75 bp (MINLEN:75). Post-trimming quality was re-assessed with FastQC v0.12.1. Genome assembly was performed with SPAdes v3.15.5 [16] using the “careful” and “cov-cutoff” options. Annotation was carried out with Prokka v1.14.6 [17], specifying genus, species, kingdom, and the selected database.

2.6. Core Genome and Pan-Genome Analysis

Comparative genomic analysis was performed using Roary v3.13.0 to identify core and accessory genes across all genomes included in this study. A total of *n* complete genome assemblies representing *Escherichia*, *Shigella*, and *Salmonella enterica* lineages were retrieved from the NCBI RefSeq database (accession numbers listed in Supplementary Table S1) to provide phylogenetic context. These publicly available reference genomes were downloaded in FASTA format and combined with four newly sequenced isolates generated in this study.

Protein clustering was performed using BLASTp with a minimum sequence identity threshold of 95%. The core gene alignment generated by Roary v3.13.0 was used to reconstruct a maximum-likelihood phylogeny using IQ-TREE2, with the best-fit substitution model selected according to the Bayesian Information Criterion (BIC). Branch support was assessed using the SH-like approximate likelihood ratio test (SH-aLRT) and ultrafast bootstrap approximation (UFBoot) with 1000 replicates. *Salmonella enterica* subsp. *enterica* serovar Enteritidis strain LA5_775 was used as the outgroup for tree rooting.

2.7. Virulence Genes

Specific virulence genes were identified by manually inspecting the annotation files generated with Prokka [17]. This process was guided by previously reported associations in

the literature. In particular, the study by Navarro-García et al. [18] was used as a reference for the presence and characterization of virulence factors, including genes related to the Type VI secretion system [19,20]. The genes of interest included *stx2a*, *stx2b*, *sigA*, *pic*, and *iha*.

3. Results

3.1. Serotyping and Antigenic Cross-Reactions

Serotyping of the *E. coli* strains with 188 anti-O *E. coli* and 45 anti-O *Shigella* sera, showed positive agglutination reactions with *E. coli* 64474, O179, O188 and *S. boydii* O16 rabbit antisera prepared against the homologous O antigens.

3.2. Absorption Assays

To evaluate the presence of common epitopes among *E. coli* 64474, O179, O188, and *S. boydii* O16 strains, serum samples prepared against *E. coli* 64474, O179, O188, and *S. boydii* O16, were absorbed with homologous and heterologous antigens. The agglutination reactions were determined using the OX188 antiserum against *E. coli* O179, 64474, OX188 and *S. boydii* O16 antigens, the registered titers were 1:100, 1:800, 1:1600 and 1:400 respectively (Table 1). Absorption of OX188 antiserum with *E. coli* O179 antigen removed the reaction against the O179 antigen only. In contrast, when OX188 antiserum was absorbed with either *S. boydii* O16 or *E. coli* 64474 antigens, agglutination was completely removed for *E. coli* O179, *S. boydii* 16 and *E. coli* 64474 antigens, which means that OX188 shares a somatic antigen with *E. coli* 64474 and *S. boydii* O16.

Table 1. Agglutination Titers of Absorbed and Unabsorbed *E. coli* O179, 64474:H32, *S. boydii* O16 and OX188 Sera.

Antigen	Titers of Unabsorbed Sera			Titers of Antisera Absorbed with Boiled Cultures						
	<i>E. coli</i> 64474:H32	<i>S. boydii</i> O16	<i>E. coli</i> OX188:H10	<i>E. coli</i> 64474:H32 Absorbed with		<i>S. boydii</i> O16 Absorbed with		<i>E. coli</i> OX188:H10 Absorbed with		
				<i>E. coli</i> O179	<i>S. boydii</i> O16	<i>E. coli</i> O179	<i>E. coli</i> 64474	<i>E. coli</i> O179	<i>S. boydii</i> O16	<i>E. coli</i> 64474
<i>E. coli</i> O179	1:800	1:800	1:100	-	-	-	-	-	-	-
<i>E. coli</i> 64474:H32	1:1600	1:3200	1:800	1:1600	-	1:1600	-	1:400	-	-
<i>E. coli</i> OX188:H10	1:800	1:1600	1:1600	1:800	-	1:800	-	1:800	-	-
<i>S. boydii</i> 16	1:3200	1:6400	1:400	1:800	-	1:3200	-	1:400	-	-

E. coli 64474, *E. coli* O179:H8 (E43478) and *S. boydii* O16 (G1219) strains were obtained from a previous study by the laboratory [10] and *E. coli* O188:H10 was obtained from the Statens Serum Institut (Copenhagen, Denmark).

Regarding to the *E. coli* 64474 strain, we previously reported that it shares a common O antigen with *S. boydii* O16 [10]. In that study, we proposed that *E. coli* 64474 serotypes belong to a new pathogenic serogroup showing a somatic antigen identical to that of *S. boydii* O16. The absorption assays of anti-*E. coli* 64474, anti-*S. boydii* O16, and anti-*E. coli* OX188 sera with 64474 and *S. boydii* O16 antigens showed that the agglutination reactions were completely removed against these antigens (Table 1). The results suggest that the *E. coli* OX188 strain shares a common O antigen with *E. coli* 64474 and *S. boydii* O16, and shares an antigenic fraction with O179, since the absorption assay of the antiserum of *E. coli* OX188 with O179 antigen removed only the reaction against the O179 antigen.

3.3. Comparative Genomics Analysis of *rfb*

The results identified three distinct regions within the *rfb* cluster when comparing of *E. coli* O179 and *E. coli* 64474 (Figure 1). The first region, which contains the *galF* gene, showed similarity levels above 96%. The second region, located in the central part of the cluster, showed scattered short segments with similarity over 80%; however, these are not displayed in the figure due to their limited length. The third region extends from the glycosyltransferase gene upstream of *maC* to the *wzz* at the end of the cluster and shows a similarity greater than 70%. Three additional loci corresponding to *S. boydii* O16, *E. coli* 64474, and O188 shared similarity levels greater than 85%. In O188, the flippase gene matched the *wzx* gene, and it was arranged in the same order as in *S. boydii* O16 and *E. coli* 64474.

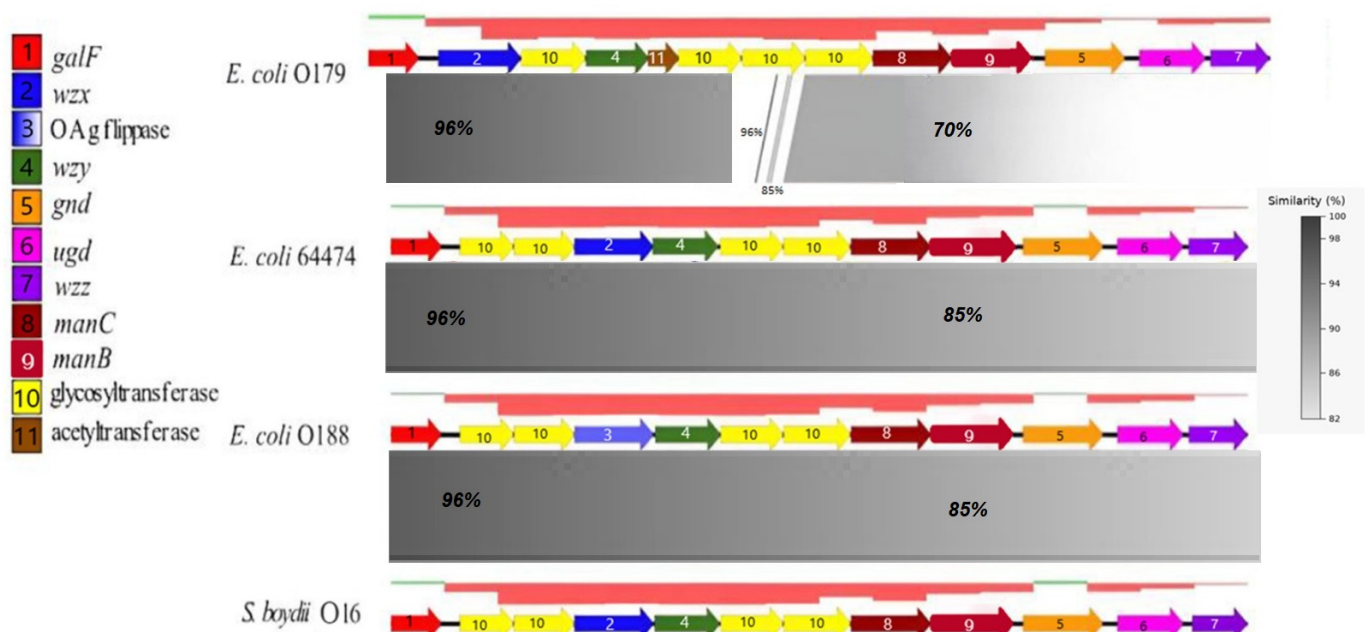


Figure 1. Comparative analysis of the *rfb* gene clusters in the four strains and their GC content. Each gene alignment is represented by a colored arrow as indicated in the legend on the left side. The percentage of similarity scale between aligned regions is shown in the lower right corner, while GC content is depicted above each gene cluster (green: positive values; red: negative values). Alignments were performed using BLAST v. 2.14.1 from Easyfig v. 2.2.2.

3.4. The Comparative Genomics Analysis

Comparison of the whole genome of *E. coli* O104:H4 strain 2011C-3493 with the study strains revealed regions with sequence similarity above 80%. Regions with 90% and 100% similarity are highlighted using distinct color labels. Additionally, GC content and GC skew are represented graphically (Figure 2).

3.5. Pangenome Structure and Core Genome Phylogeny

Analysis of the pangenome of *E. coli* 64474, O179, O188, and *S. boydii* O16 was performed using Roary v3.13.0 with a 95% BLASTp identity threshold. A total of 5199 gene clusters were identified across the four genomes. Of these, 3583 (68.9%) were classified as core genes, defined as those present in ≥ 99 –100% of the genomes analyzed (i.e., all four strains), while the remaining 1616 (31.1%) constituted the accessory genome.

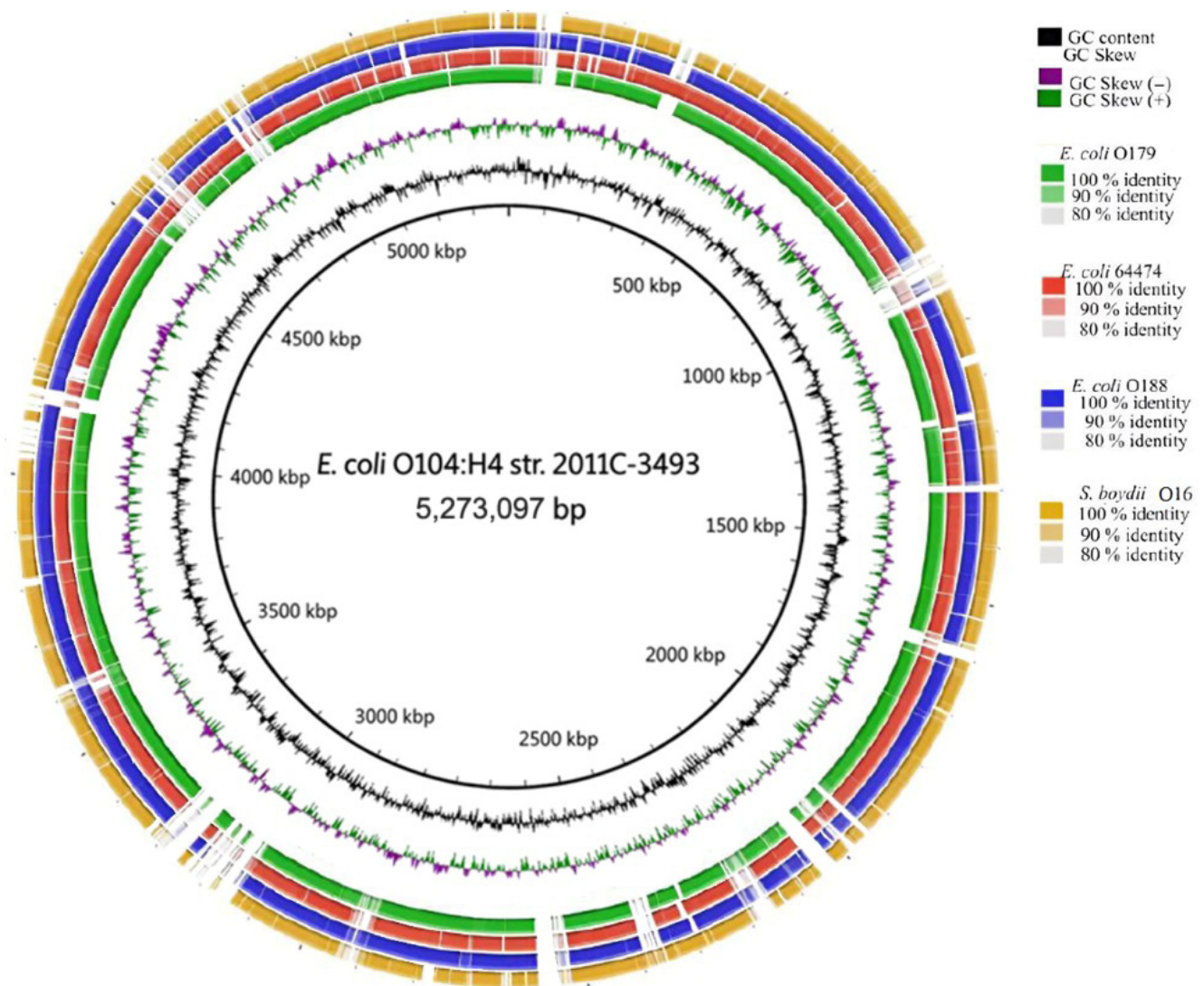


Figure 2. Comparative genomics of strains against a reference genome of *E. coli* O104:H4 strain 2011C-3493 (GCF_000299455.1). The analyses and their corresponding color codes are shown in the upper-right corner. The position along the reference genome is indicated in kilobase pairs (kbp) in the center of the circular map. Alignments were performed using BLAST v. 2.14.1 via BRIG v. 0.95.

Figure 3 illustrates gene presence–absence patterns across the four genomes using an UpSet plot. Most accessory gene clusters are shared by only one or two genomes, indicating substantial genomic variability among strains. Notably, although 3583 gene clusters were classified as core genes, only a subset ($n = 377$) is represented as shared across all genomes in the UpSet plot. This discrepancy reflects the structure of the visualization, which does not display the full set of core genes but rather the subset captured in the plotted presence–absence matrix. Overall, the high proportion of core genes indicates the presence of a conserved genomic backbone among the four strains, supporting their close evolutionary relationship. In contrast, the accessory genome reflects lineage-specific variation, likely driven by horizontal gene transfer, prophage integration, and other mobile genetic elements contributing to genomic diversification.

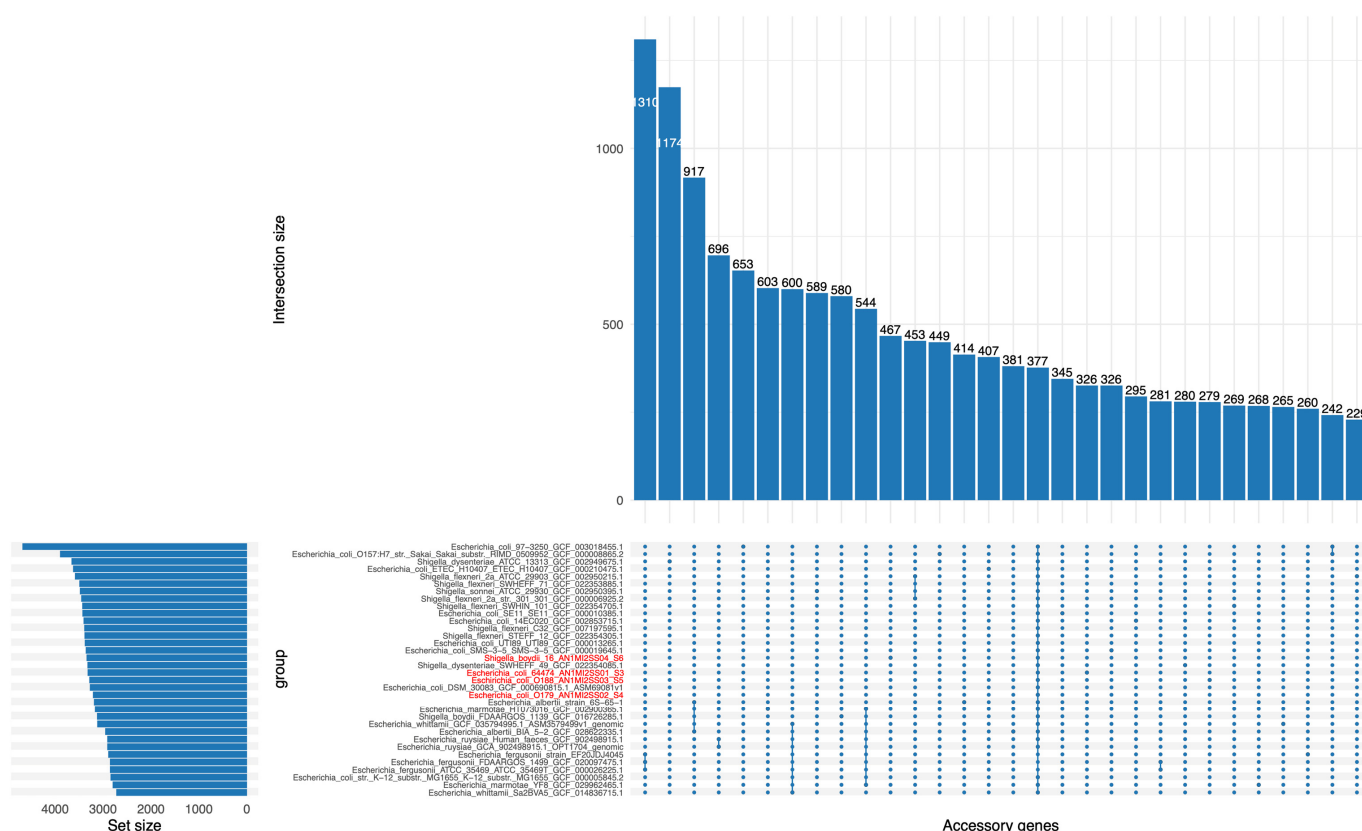


Figure 3. Roary-based pangenome structure and identification of genes present or absent in different genome sets. The UpSet plot shows the occurrence of gene clusters in *E. coli* 64474, O179, O188, and *S. boydii* O16 using Roary software version 3.13.0 with a BLASTp identity threshold of $\geq 95\%$. The top bar chart displays the number of gene clusters found in each combination of genome sets, while the left bar chart shows the total number of genes in each individual genome. The pangenome contains 5199 gene clusters, including 3583 core genes (gene clusters present in ≥ 99 –100% of genome sets, i.e., all four genomes) and 1616 accessory genes. An UpSet plot highlighting a subset of gene clusters found in all four genomes ($n = 377$) is included as a representative portion of the core genome.

The high proportion of shared core genes indicates a strong conserved genomic backbone among the four strains, supporting their close evolutionary relationship. In contrast, the accessory genome comprised genes variably distributed among strains, likely reflecting horizontal gene transfer events, prophage insertions, and other mobile genetic elements that contribute to genomic diversification.

To contextualize these findings within a broader evolutionary framework, a core genome alignment generated by Roary was used to reconstruct a maximum-likelihood phylogeny using IQ-TREE2. The tree was rooted with *Salmonella enterica* subsp. *enterica* serovar Enteritidis strain LA5_775.

The phylogenetic tree reconstruction shows that *E. coli* and *Shigella* species are not differentiated into distinct, separate phylogenetic groups. Instead, they form a single, interspersed clade. This finding is consistent with the current genomic organization of these species and their relatedness. Within this clade, *E. coli* O179 and *E. coli* O188 are grouped together with *E. coli* SE11 and other related isolates. *E. coli* 64474 is also part of a single clade with other *E. coli* strains, indicating no differentiation from established *E. coli* lineages. In contrast, *S. boydii* O16 is grouped with *S. boydii* FDAARGOS_1139 and other related *Shigella* species, including *S. flexneri* and *S. sonnei* (Figure 4). In summary, all these isolates are part of a single clade with known species of the *Escherichia/Shigella* complex.

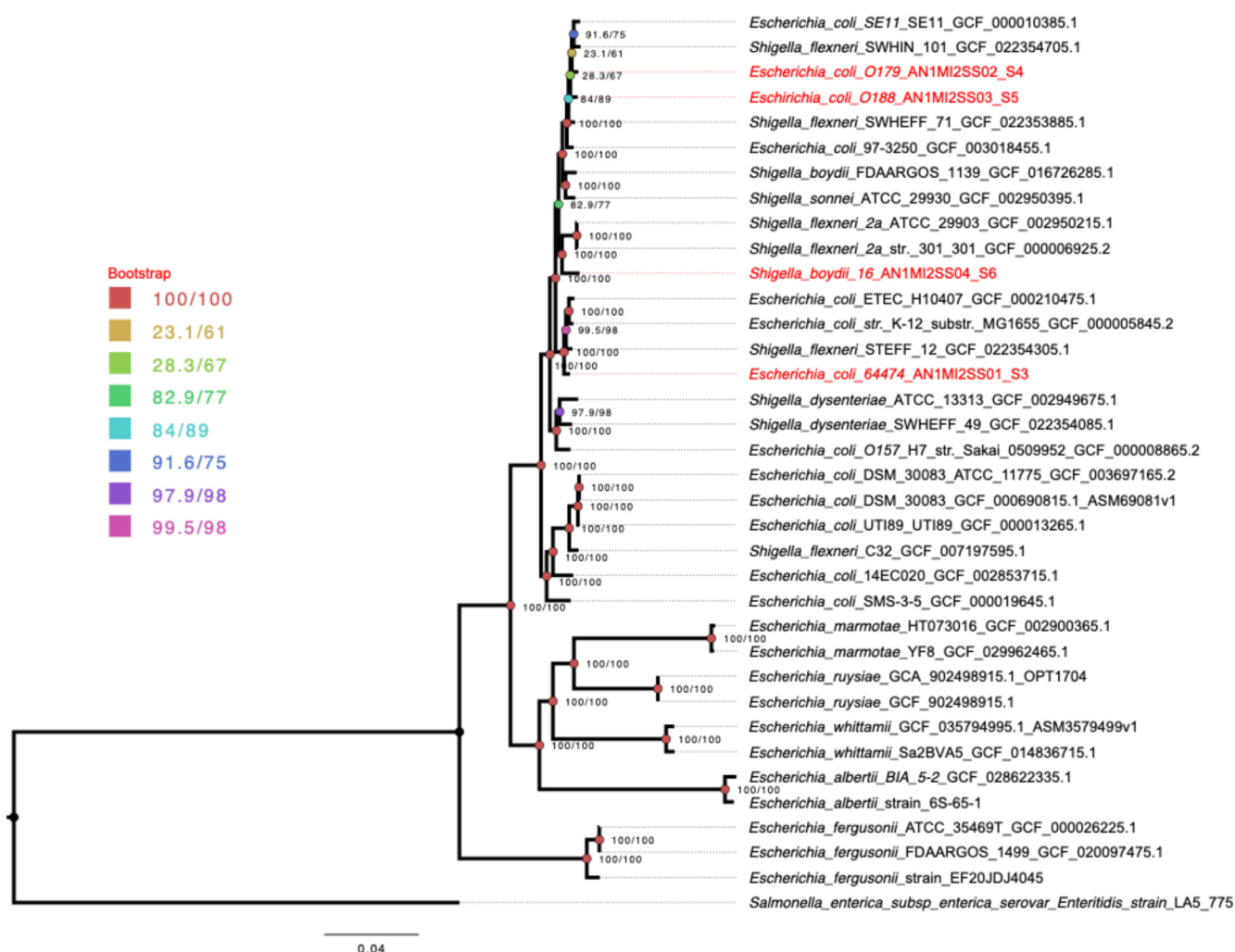


Figure 4. Core genome maximum-likelihood phylogeny inferred from Roary alignment. Phylogenetic tree reconstructed from the core gene alignment generated by Roary v3.13.0 and inferred using IQ-TREE2 under the best-fit substitution model selected according to the Bayesian Information Criterion (BIC). Branch support was assessed using SH-like approximate likelihood ratio test (SH-aLRT) and ultrafast bootstrap approximation (UFBoot) with 1000 replicates, and support values (SH-aLRT/UFBoot) are indicated at the nodes. The tree was rooted with *Salmonella enterica* subsp. *enterica* serovar *Enteritidis* strain LA5_775. Our sequenced strains are highlighted in red, while reference strains are shown in black. The topology illustrates the interspersed distribution of *E. coli* and *Shigella* lineages within a single major clade.

3.6. Virulence Genes

To better contextualize the biological relevance of the analyzed strains, we screened a panel of virulence genes representative of different *E. coli* pathotypes. The resulting profiles showed that the strains harbor markers commonly associated with STEC, UPEC, EAEC, ETEC, and *Shigella*-related virulence schemes, suggesting a notable overlap of pathogenic traits among them. In particular, some strains carried combinations of genes that are typically linked to more than one pathotype, supporting the idea that these isolates may display a hybrid virulence profiles (Table 2). At the strain level, *E. coli* O179 carried *stx2a/stx2b*, *iha*, and *sigA*, consistent with virulence traits associated with STEC and UPEC. In contrast, *E. coli* O188 harbored *sigA*, *pic*, *aggR*, *aatA*, and *aap*, a profile mainly related to EAEC and also shared with virulence schemes reported in STEC. Likewise, *E. coli* 64474 carried *eltA/eltB* and *aggR*, indicating a combination of ETEC- and EAEC-associated

markers. Overall, these findings support the presence of diverse and partially overlapping virulence repertoires among the analyzed strains.

Table 2. Virulence genes identified in the analyzed strains and their association with pathogenic *E. coli* and *Shigella* spp.

Gene	Strain	Association	Function
<i>stx2a stx2b, iha sigA</i>	<i>E. coli</i> O179	STEC, UPEC	Stx: Shiga toxin subunit A Stx2c and Shiga toxin subunit B Stx2a [19,21]. Iha adhesin: Bifunctional siderophore receptor [22]. SigA: serine protease autotransporter toxin [23].
<i>sigA pic aggR aatA and aap</i>	<i>E. coli</i> O188	EAEC, <i>Shigella</i> spp.	SigA: serine protease autotransporter toxin [23]. Pic: protein involved in intestinal colonization, serine protease autotransporter toxin [24]. AggR: transcriptional activator from EAEC of virulence factors. <i>aatA</i> : Facilitate the export of the dispersin Aap protein across the outer membrane. <i>aap</i> (dispersin): It functions in dispersion over the surface of the intestinal mucosa
<i>eltA/eltB, aggR</i>	<i>E. coli</i> 64474	ETEC, EAEC	Lt: Activates intracellular adenylate cyclase [3]. AggR: transcriptional activator from EAEC of virulence factors

EAEC: Enteragggregative *E. coli*; STEC: Shiga toxin-producing *E. coli*; UPEC: Uropathogenic, *E. coli*; Enterotoxigenic *E. coli*: ETEC.

4. Discussion

In a previous study conducted in our laboratory, we showed that *E. coli* 64474 and *S. boydii* O16 strains exhibited similar agglutination reactions with the anti-O sera of these strains. They also share the *wzx* (flippase) and *wzy* (polymerase) genes, which are involved in biosynthesis of the O antigen of *S. boydii* O16 [10,25]. These results suggested that these strains could belong to the same clone. In the present study, we found that the anti-O sera of *E. coli* 64474 and anti-*S. boydii* O16 reacted with the O antigen of *E. coli* O188. To elucidate whether these strains belonged to the same clone due to their similarities at the O antigen level, we performed a comparative genomics and phylogenomic analyses among *E. coli* 64474, *E. coli* O179, *E. coli* O188 and *S. boydii* O16.

There are various types of analyses used to distinguish strains of enterobacteria; however, phylogenomic analyses can differentiate closely related strains because they use a larger amount of genetic information. The phylogenomic analysis revealed a monophyletic clade formed by *E. coli* strains, indicating a close clonal relationship. *E. coli* O188 and O179 were the most closely related, while *E. coli* 64474 was positioned third, contrary to previous serological studies, which showed a closer relationship between serotypes O188 and 64474 and between serotypes O179 and 64474. These discrepancies were supported by the comparative analysis of the *rfb* clusters, where distinct genes were only found in a section of the O179 O-antigen cluster.

4.1. *rfb* Cluster

Discrepancies in the high-similarity patterns between the O-antigen and whole-genome sequences of the strains may result from the high rate of homologous and non-homologous recombination within O-antigen cluster [26]. These recombination events can lead to changes in the entire genes or gene sets within these genomic hotspots [27], which are associated with integration sites [28]. When comparing the *rfb* cluster of 64474 and O179 strains, more differences were observed than when comparing 64474 with O188 and *S.*

boydii O16. Notably, mutations were found in housekeeping genes such as *galF* [29], over 96% similarity, *gnd* [30], and *ugd* [31], these last two genes with over 70% of similarity, which are common within these clusters due to frequent genetic rearrangement. The O179 *rfb* cluster consists of a different gene set compared to the other three *rfb* clusters, explaining the lower serological similarity. Despite these differences, certain loci within the O179 cluster retained high levels of similarity (over 70%), which may indicate a recombination event between O179 and other genomically similar strains [32].

Although *S. boydii* O16 did not exhibit a close clonal relationship with the *E. coli* strains, comparison of their O-antigen clusters revealed high similarity (over 85%) among the *S. boydii* O16, O188 and 64474 clusters. This genomic similarity supports the idea of *S. boydii* O16 is evolutionarily related strain to *E. coli* strains of the study. Furthermore, in addition to O antigen and whole-genome similarities, other studies, such as the structural analysis of O188 have revealed only minor differences from *S. boydii* O16, specifically in the N-acetylated residues of LPS [11]. Some reports also suggest that genome similarities with *E. coli* O188 may indicate that *S. boydii* O16 could act as a gene receptor or donor, potentially acquiring genes that increase virulence or drug resistance [12].

The high similarity among *rfb* clusters of 64474, O188, and *S. boydii* O16 is consistent with the serological findings. Nevertheless, O antigen molecular structures cannot be inferred solely from genome sequences; therefore, these results must be supported by further analyses. Due to the complexity of these polysaccharides, structural characterization using bioinformatics tools combined with NMR spectroscopy may help to determine whether 64474 carries a de novo O antigen [33].

4.2. Comparative Genomics

Although this analysis revealed similarity below 80%; these regions were located at conserved positions across the genomes, showing shared patterns that support the high similarities among the genomes of the study strains. These loci include accessory genes related to phages, mobile elements, or sequences acquired through horizontal gene transfer (HTG), such as genes involved in the *mer* operon (a mercury resistance system) [34], the *sil* operon (which promotes resistance to exogenous silver) [35], and the Type VI Secretion System (or T6SS) [20]. However, most of the genomic elements belong to the core genome (68.9% of genes), as computed from *E. coli* 64474, O179, O188 and *S. boydii* O16 strains, indicating a common evolutionary pattern. When comparing this analysis with broader studies, a report analyzing over 95,525 genome sequences from 14 phylogroups of *E. coli* and *Shigella* spp., revealed a core genome of only 1.96%, highlighting the close evolutionary relationship of the strains studied [36,37].

4.3. Virulence Genes

In addition, virulence genes should be considered when assessing the significance of this relationship. Particularly informative are the similarity levels detected among *E. coli* 64474, *E. coli* O179, *E. coli* O188, *S. boydii* O16, and *E. coli* O104:H4 strain 2011C-3493, this last strain a clinically relevant strain responsible for the 2011 outbreak in Europe, which resulted in 46 deaths, 782 cases of hemolytic uremic syndrome, and 3128 cases of acute gastroenteritis. This strain is genetically related to the EAEC and STEC pathotypes and harbors a broad repertoire of virulence genes. In the analysis, virulence genes recognized as critical for this pathogen's virulence *stx2a*, *stx2b*, *sigA*, *pic*, and *iha* [18] were found in *E. coli* O179 and O188. All these genes, whether present or absent, are associated with horizontal gene transfer (HGT), which may explain the presence of unrelated loci between the strains of the study and *E. coli* O104:H4 strain 2011C-3493 because those are not core genes [38]. This information is relevant because the genome similarity observed between

E. coli O179 and O188 includes shared high-virulence genes with *E. coli* O104:H4 strain 2011C-3493. Moreover, *E. coli* 64474 and *S. boydii* O16 strains, may acquire these genes via genetic recombination, due to their close genetic relationship.

5. Conclusions

E. coli 64474 may possess a de novo O-antigen and be a clone more closely related to *E. coli* O188 and O179. However, additional factors must be considered, such as the isolation date (1985), geographic origin (Mexico), and genome content similarities that could have led to divergence events over time, ultimately affecting phylogenomic interpretations and potentially indicating that this strain descends from an ancestor with the *E. coli* O188 antigen.

Similarity levels of *rfb* clusters should be interpreted alongside other factors, such as genomic plasticity, HGT, homologous recombination [39], and epigenetic effects [40], as well as the presence of virulence and drug resistance genes that may be transferred through HGT mechanisms. These characteristics, frequently observed in *E. coli* and *Shigella* spp., enhance their adaptability and, in some cases, their pathogenicity [41,42].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pathogens15050462/s1>, Table S1: Genomes analyzed in this study and their GenBank accession numbers.

Author Contributions: Conceptualization: A.N.-O., G.C.-E., E.O.D.-D., A.C. and H.G.C.-S.; Methodology, E.O.D.-D. and A.S.-P.; Formal analysis: E.O.D.-D., H.G.C.-S., A.N.-O. and A.C.; Investigation, E.O.D.-D. and A.S.-P.; Statistical analysis: H.G.C.-S.; Writing—original draft preparation, E.O.D.-D., A.N.-O. and H.G.C.-S.; Editing, A.N.-O. and H.G.C.-S. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thank the Escuela Nacional de Ciencias Biológicas of Instituto Politécnico Nacional (Project 20232102-20240986) and the Dirección General de Asuntos del Personal Académico—Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (DGAPA-PAPIIT, Project 1N22035) of Universidad Nacional Autónoma de México for their support. G.C.-E. received support from “Sistema Nacional de Investigadores (SNII)” from SECIHTI, Mexico also received support from “Estímulos al Desempeño de los Investigadores, Comisión de Operación y Fomento de Actividades Académicas (Instituto Politécnico Nacional)”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgments: We would like to thank Gabriel Pérez, Liliana Cortes and Delia Licona (Faculty of Medicine, UNAM) for their technical assistance in the laboratory.

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

References

1. Liu, B.; Furevi, A.; Perepelov, A.V.; Guo, X.; Cao, H.; Wang, Q.; Reeves, P.R.; Knirel, Y.A.; Wang, L.; Widmalm, G. Structure and Genetics of *Escherichia coli* O Antigens. *FEMS Microbiol. Rev.* **2020**, *44*, 655–683. [\[CrossRef\]](#)
2. Kaper, J.B.; Nataro, J.P.; Mobley, H.L.T. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2004**, *2*, 123–140. [\[CrossRef\]](#)
3. Pokharel, P.; Dhakal, S.; Dozois, C.M. The Diversity of *Escherichia coli* Pathotypes and Vaccination Strategies against This Versatile Bacterial Pathogen. *Microorganisms* **2023**, *11*, 344. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Lan, R.; Reeves, P.R. *Escherichia coli* in Disguise: Molecular Origins of *Shigella*. *Microbes Infect.* **2002**, *4*, 1125–1132. [\[CrossRef\]](#) [\[PubMed\]](#)

5. Stenhouse, G.E.; Keddy, K.H.; Bengtsson, R.J.; Hall, N.; Smith, A.M.; Thomas, J.; Iturriza-Gómara, M.; Baker, K.S. The Genomic Epidemiology of Shigellosis in South Africa. *Nat. Commun.* **2023**, *14*, 7715. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Pupo, G.M.; Lan, R.; Reeves, P.R. Multiple Independent Origins of *Shigella* Clones of *Escherichia coli* and Convergent Evolution of Many of Their Characteristics. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10567–10572. [\[CrossRef\]](#)
7. Li, Y.; Cao, B.; Liu, B.; Liu, D.; Gao, Q.; Peng, X.; Wu, J.; Bastin, D.A.; Feng, L.; Wang, L. Molecular Detection of All 34 Distinct O-Antigen Forms of *Shigella*. *J. Med. Microbiol.* **2009**, *58*, 69–81. [\[CrossRef\]](#)
8. Coimbra, R.S.; Grimont, F.; Lenormand, P.; Burguière, P.; Beutin, L.; Grimont, P.A.D. Identification of *Escherichia coli* O-Serogroups by Restriction of the Amplified O-Antigen Gene Cluster (Rfb-RFLP). *Res. Microbiol.* **2000**, *151*, 639–654. [\[CrossRef\]](#)
9. Van Den Beld, M.J.C.; Reubsaet, F.A.G.; Pijnacker, R.; Harpal, A.; Kuiling, S.; Heerkens, E.M.; Hoeve-Bakker, B.J.A.D.; Noomen, R.C.E.A.; Hendriks, A.C.A.; Borst, D.; et al. A Multifactorial Approach for Surveillance of *Shigella* spp. and Enteroinvasive *Escherichia coli* Is Important for Detecting (Inter)National Clusters. *Front. Microbiol.* **2020**, *11*, 564103. [\[CrossRef\]](#)
10. Navarro, A.; Eslava, C.; Perea, L.M.; Inzunza, A.; Delgado, G.; Morales-Espinosa, R.; Cheasty, T.; Cravioto, A. New Enterovirulent *Escherichia coli* Serogroup 64,474 Showing antigenic and Genotypic Relationships to *Shigella boydii* 16. *J. Med. Microbiol.* **2010**, *59*, 453–461. [\[CrossRef\]](#)
11. Furevi, A.; Stähle, J.; Muheim, C.; Gkotzsis, S.; Udekwu, K.I.; Daley, D.O.; Widmalm, G. Structural Analysis of the O-Antigen Polysaccharide from *Escherichia coli* O188. *Carbohydr. Res.* **2020**, *498*, 108051. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Simoni, S.; Mingoia, M.; Brenciani, A.; Carelli, M.; Lleò, M.M.; Malerba, G.; Vignaroli, C. First IncHI2 Plasmid Carrying Mcr-9.1, blaVIM-1, and Double Copies of blaKPC-3 in a Multidrug-Resistant *Escherichia coli* Human Isolate. *mSphere* **2021**, *6*, e0030221. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Beutin, L.; Strauch, E. Identification of Sequence Diversity in the *Escherichia coli* fliC Genes Encoding Flagellar Types H8 and H40 and Its Use in Typing of Shiga Toxin-Producing *E. coli* O8, O22, O111, O174, and O179 Strains. *J. Clin. Microbiol.* **2007**, *45*, 333–339. [\[CrossRef\]](#)
14. Ørskov, F.; Ørskov, I. 2 Serotyping of *Escherichia coli* * * The Terminology Used to Describe the Different Classes of Bacterial Antigens Is Explained in the Preface. However, the Authors Would like to Mention That a Different Convention is Used in Some Laboratories, for Example O:L, K:L, H:7 is Equivalent to O1:K1:H7. In *Methods in Microbiology*; Bergan, T., Ed.; Academic Press: Cambridge, MA, USA, 1984; Volume 14, pp. 43–112.
15. Ewing, W.H. *Edwards and Ewing's Identification of Enterobacteriaceae*; Elsevier: Amsterdam, The Netherlands, 1986.
16. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [\[CrossRef\]](#)
17. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Navarro-Garcia, F.; Ruiz-Perez, F.; Cataldi, Á.; Larzábal, M. Type VI Secretion System in Pathogenic *Escherichia coli*: Structure, Role in Virulence, and Acquisition. *Front. Microbiol.* **2019**, *10*, 1965. [\[CrossRef\]](#)
19. Melton-Celsa, A.R. Shiga Toxin (Stx) Classification, Structure, and Function. *Microbiol. Spectr.* **2014**, *2*, 10–1128. [\[CrossRef\]](#)
20. Nesporova, K.; Steemans, B.; Govers, S.K. Large-Scale Genomic Analysis Reveals the Distribution and Diversity of Type VI Secretion Systems in *Escherichia coli*. *mSystems* **2025**, *10*, e00105-25. [\[CrossRef\]](#)
21. Wang, X.; Yu, D.; Chui, L.; Zhou, T.; Feng, Y.; Cao, Y.; Zhi, S. A Comprehensive Review on Shiga Toxin Subtypes and Their Niche-Related Distribution Characteristics in Shiga-Toxin-Producing *E. coli* and Other Bacterial Hosts. *Microorganisms* **2024**, *12*, 687. [\[CrossRef\]](#)
22. Johnson, J.R.; Jelacic, S.; Schoening, L.M.; Clabots, C.; Shaikh, N.; Mobley, H.L.T.; Tarr, P.I. The IrgA Homologue Adhesin Iha is an *Escherichia coli* Virulence Factor in Murine Urinary Tract Infection. *Infect. Immun.* **2005**, *73*, 965–971. [\[CrossRef\]](#)
23. Al-Hasani, K.; Henderson, I.R.; Sakellaris, H.; Rajakumar, K.; Grant, T.; Nataro, J.P.; Robins-Browne, R.; Adler, B. The sigA Gene Which Is Borne on the shePathogenicity Island of *Shigella flexneri* 2a Encodes an Exported Cytopathic Protease Involved in Intestinal Fluid Accumulation. *Infect. Immun.* **2000**, *68*, 2457–2463. [\[CrossRef\]](#)
24. Flores-Sanchez, F.; Chavez-Dueñas, L.; Sanchez-Villamil, J.; Navarro-Garcia, F. Pic Protein From Enterotoxigenic *E. coli* Induces Different Mechanisms for Its Dual Activity as a Mucus Secretagogue and a Mucinas. *Front. Immunol.* **2020**, *11*, 564953. [\[CrossRef\]](#)
25. Liu, B.; Senchenkova, S.N.; Feng, L.; Perepelov, A.V.; Xu, T.; Shevelev, S.D.; Zhu, Y.; Shashkov, A.S.; Zou, M.; Knirel, Y.A.; et al. Structural and Molecular Characterization of *Shigella boydii* Type 16 O Antigen. *Gene* **2006**, *380*, 46–53. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Didelot, X.; Méric, G.; Falush, D.; Darling, A.E. Impact of Homologous and Non-Homologous Recombination in the Genomic Evolution of *Escherichia coli*. *BMC Genom.* **2012**, *13*, 256. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Iguchi, A.; Iyoda, S.; Kikuchi, T.; Ogura, Y.; Katsura, K.; Ohnishi, M.; Hayashi, T.; Thomson, N.R. A Complete View of the Genetic Diversity of the *Escherichia coli* O-Antigen Biosynthesis Gene Cluster. *DNA Res.* **2015**, *22*, 101–107. [\[CrossRef\]](#)
28. Denamur, E.; Clermont, O.; Bonacorsi, S.; Gordon, D. The Population Genetics of Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2021**, *19*, 37–54. [\[CrossRef\]](#) [\[PubMed\]](#)

29. Ebrecht, A.C.; Orlof, A.M.; Sasoni, N.; Figueroa, C.M.; Iglesias, A.A.; Ballicora, M.A. On the Ancestral UDP-Glucose Pyrophosphorylase Activity of GalF from *Escherichia coli*. *Front. Microbiol.* **2015**, *6*, 1253. [[CrossRef](#)]
30. Nelson, K.; Selander, R.K. Intergeneric Transfer and Recombination of the 6-Phosphogluconate Dehydrogenase Gene (Gnd) in Enteric Bacteria. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10227–10231. [[CrossRef](#)] [[PubMed](#)]
31. Drummelsmith, J.; Amor, P.A.; Whitfield, C. Polymorphism, Duplication, and IS1-Mediated Rearrangement in the Chromosomal His-Rfb-Gnd Region of *Escherichia coli* Strains with Group IA and Capsular K Antigens. *J. Bacteriol.* **1997**, *179*, 3232–3238. [[CrossRef](#)]
32. Iguchi, A.; Iyoda, S.; Ohnishi, M. Molecular Characterization Reveals Three Distinct Clonal Groups among Clinical Shiga Toxin-Producing *Escherichia coli* Strains of Serogroup O103. *J. Clin. Microbiol.* **2012**, *50*, 2894–2900. [[CrossRef](#)]
33. Ståhle, J.; Widmalm, G. Lipopolysaccharides of Gram-Negative Bacteria: Biosynthesis and Structural Aspects. *Trends Glycosci. Glycotechnol.* **2019**, *31*, E159–E171. [[CrossRef](#)]
34. Boyd, E.S.; Barkay, T. The Mercury Resistance Operon: From an Origin in a Geothermal Environment to an Efficient Detoxification Machine. *Front. Microbiol.* **2012**, *3*, 349. [[CrossRef](#)]
35. Wang, Q.; Perepelov, A.V.; Wen, L.; Shashkov, A.S.; Wang, X.; Guo, X.; Knirel, Y.A.; Wang, L. Identification of the Two Glycosyltransferase Genes Responsible for the Difference between *Escherichia coli* O107 and O117 O-Antigens. *Glycobiology* **2012**, *22*, 281–287. [[CrossRef](#)]
36. Abram, K.; Udaondo, Z.; Bleker, C.; Wanchai, V.; Wassenaar, T.M.; Robeson, M.S.; Ussery, D.W. Mash-Based Analyses of *Escherichia coli* Genomes Reveal 14 Distinct Phylogroups. *Commun. Biol.* **2021**, *4*, 117. [[CrossRef](#)]
37. Chauhan, S.M.; Ardalani, O.; Hyun, J.C.; Monk, J.M.; Phaneuf, P.V.; Palsson, B.O. Decomposition of the Pangenome Matrix Reveals a Structure in Gene Distribution in the *Escherichia coli* Species. *mSphere* **2024**, *10*, e00532-24. [[CrossRef](#)] [[PubMed](#)]
38. Costa, S.S.; Guimarães, L.C.; Silva, A.; Soares, S.C.; Baraúna, R.A. First Steps in the Analysis of Prokaryotic Pan-Genomes. *Bioinform. Biol. Insights* **2020**, *14*, 1177932220938064. [[CrossRef](#)] [[PubMed](#)]
39. Shikov, A.E.; Savina, I.A.; Nizhnikov, A.A.; Antonets, K.S. Recombination in Bacterial Genomes: Evolutionary Trends. *Toxins* **2023**, *15*, 568. [[CrossRef](#)] [[PubMed](#)]
40. D'Aquila, P.; De Rango, F.; Paparazzo, E.; Passarino, G.; Bellizzi, D. Epigenetic-Based Regulation of Transcriptome in *Escherichia coli* Adaptive Antibiotic Resistance. *Microbiol. Spectr.* **2023**, *11*, e04583-22. [[CrossRef](#)]
41. Touchon, M.; Hoede, C.; Tenaillon, O.; Barbe, V.; Baeriswyl, S.; Bidet, P.; Bingen, E.; Bonacorsi, S.; Bouchier, C.; Bouvet, O.; et al. Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths. *PLoS Genet.* **2009**, *5*, e1000344. [[CrossRef](#)]
42. Touchon, M.; Perrin, A.; De Sousa, J.A.M.; Vangchhia, B.; Burn, S.; O'Brien, C.L.; Denamur, E.; Gordon, D.; Rocha, E.P. Phylogenetic Background and Habitat Drive the Genetic Diversification of *Escherichia coli*. *PLoS Genet.* **2020**, *16*, e1008866. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.