Communication

Whole-Genome Sequencing Analysis of a *stx*-Negative *Escherichia coli* O63:H6 Isolate Associated with Hemolytic Uremic Syndrome

Tae Yeul Kim [1,†] [ID], **Tae-Min La** [2,†], **Taesoo Kim** [2], **Sun Ae Yun** [3], **Sang-Won Lee** [2,*], **Hee Jae Huh** [1,*] [ID] and **Nam Yong Lee** [1]

1   Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea; voltaire0925@gmail.com (T.Y.K.); micro.lee@samsung.com (N.Y.L.)
2   College of Veterinary Medicine, Konkuk University, Seoul 05029, Korea; white_tm@naver.com (T.-M.L.); taesoo95811@naver.com (T.K.)
3   Center for Clinical Medicine, Samsung Biomedical Research Institute, Samsung Medical Center, Seoul 06351, Korea; sm10040730@naver.com
*   Correspondence: odssey@konkuk.ac.kr (S.-W.L.); pmhhj77@gmail.com (H.J.H.); Tel.: +82-2-450-0445 (S.-W.L.); +82-2-3410-1836 (H.J.H.); Fax: +82-2-3437-1941 (S.-W.L.); +82-2-3410-2719 (H.J.H.)
†   These authors contributed equally to this work.

Abstract: Shiga toxin-encoding genes (*stx*) of enterohemorrhagic *Escherichia coli* (EHEC) can be lost during infection or in vitro cultivation, and in clinical practice, it is difficult to distinguish EHEC that have lost *stx* (EHEC-LST) from enteropathogenic *E. coli* (EPEC), as both are *stx*-negative and *eae*-positive. In this study, we performed whole-genome sequencing (WGS) of a *stx*-negative, *eae*-positive *E. coli* O63:H6 isolate from a child with hemolytic uremic syndrome and compared its genome with those of nine *E. coli* O63:H6 strains in public databases. Virulence gene profiles were analyzed and core-genome multilocus sequence typing (cgMLST) was conducted. The virulence gene profile of our isolate was consistent with EHEC, except for the absence of *stx*, and the isolate clustered with seven EHEC strains but was distant from two EPEC strains in cgMLST. In *genome alignment*, our isolate exhibited a high nucleotide identity with EHEC strain 377323_2f but displayed a gap corresponding to the *stx*-harboring prophage sequence. Overall, our isolate was genetically closely related to EHEC strains, consistent with this being an EHEC-LST strain. As EHEC-LST may be misdiagnosed as EPEC in routine laboratories, comparative genomic analysis using WGS can be useful to determine whether *stx*-negative and *eae*-positive isolates are EHEC-LST or EPEC.

Keywords: *E. coli* O63:H6; STEC; EHEC; *stx* loss; hemolytic uremic syndrome; whole-genome sequencing

1. Introduction

Shiga toxin-producing *Escherichia coli* (STEC) can cause human illnesses ranging from uncomplicated diarrhea to life-threatening complications such as hemolytic uremic syndrome (HUS). More than 200 STEC serotypes are known to be associated with human illnesses, among which O157:H7 is considered the most common serotype [1]; however, in recent years, sporadic cases and outbreaks caused by non-O157 STEC serotypes have been increasingly reported [1]. The serogroup/serotype distribution of non-O157 STEC strains varies from country to country. In the United States, six serogroups (O26, O103, O111, O121, O45, and O145) account for the majority of non-O157 STEC cases [2]. In Germany, the Netherlands, and Belgium, a substantial proportion of non-O157 STEC strains belong to the serotype O63:H6 [3–6], but this serotype has been rarely reported elsewhere.

Shiga toxin (Stx) is the primary virulence factor of STEC involved in the pathogenesis of HUS, and the genes encoding Stx (*stx*) are located in bacteriophage genomes integrated

into the bacterial chromosome [7]. In addition to *stx*, a subset of STEC strains, termed entero-hemorrhagic *E. coli* (EHEC), possess the locus of enterocyte effacement (LEE) pathogenicity island containing *eae*, which encodes intimin, a protein required for intimate bacterial attachment to the host intestinal epithelium [7]. Similar to EHEC, enteropathogenic *E. coli* (EPEC) carries the LEE pathogenicity island, but in contrast to EHEC, EPEC lacks *stx* and is not associated with HUS [7,8]. EPEC strains are divided into typical EPEC and atypical EPEC based on the presence of the EPEC adherence factor plasmid carrying the bundle-forming pilus (*bfp*) operon, a cluster of 14 genes that includes *bfpA*; typical EPEC strains are *stx*-negative, *eae*-positive, and *bfp*-positive, whereas atypical EPEC strains are *stx*-negative, *eae*-positive, and *bfp*-negative [7,9]. However, EHEC strains can lose *stx*-harboring prophages during infection or in vitro cultivation [10–12], and in clinical practice, it is difficult to distinguish EHEC that have lost *stx* (EHEC-LST) from EPEC, as both are *stx*-negative and *eae*-positive.

Recently, we obtained a *stx*-negative, *eae*-positive *E. coli* O63:H6 isolate from a 3-year-old child diagnosed with HUS based on the presence of non-immune hemolytic anemia, thrombocytopenia, and acute kidney injury and performed whole-genome sequencing (WGS) to determine whether this isolate was EHEC-LST or EPEC.

2. Materials and Methods

A stool sample was collected 7 days after the onset of gastrointestinal symptoms (abdominal pain and vomiting) and sent for culture. Species identification was carried out using the Vitek 2 system (bioMérieux, Marcy l'Etoile, France). Serotyping was performed by agglutination test using O and H antisera (Denka Seiken, Tokyo, Japan). The presence of *stx1*, *stx2*, *stx2f*, and *eae* in both stool sample and *E. coli* isolate was tested by PCR using the primers described in Supplementary Table S1.

We performed WGS for the molecular characterization of the isolate. Genomic DNA was extracted using the MagAttract HMW DNA kit (Qiagen, Hilden, Germany). To obtain the complete genome of our isolate, long read sequencing was performed using MinION (Oxford Nanopore Technologies, Oxford, UK). To overcome the high error rate of Nanopore long read sequencing, short read sequencing was also performed using HiSeq X Ten (Illumina, San Diego, CA, USA), and the de novo hybrid assembly of Illumina short reads and Nanopore long reads was performed using Unicycler v0.4.7 [13]. The assembled genome was annotated and visualized using Geneious Prime (Biomatters, Auckland, New Zealand) [14], and then the sequence was compared with the genome sequences of nine *E. coli* O63:H6 strains found in public databases: GenBank 377323_2f; EnteroBase 179620, 77AZUC, 8X4S73, SSI-AC286, SSI-AC307, OLC2140, 190561, and 191861. We identified virulence genes using VirulenceFinder 2.0 [15] and performed in silico multilocus sequence typing (MLST) using MLST 2.0 [16]. Core-genome MLST (cgMLST) was performed using the EnteroBase scheme [17], and a minimum spanning tree was constructed using GrapeTree [18]. Pairwise genome alignment was performed using Mauve v2.4.0 [19] and BLAST Ring Image Generator (BRIG) [20].

3. Results

Colorless colonies grew on a sorbitol-MacConkey agar plate, and the isolate (strain EC12) was identified as *E. coli* O63:H6. Both the stool sample and strain EC12 isolated from the sample tested negative for *stx1*, *stx2*, and *stx2f* but positive for *eae*. Illumina HiSeq X Ten generated a total of 13,227,764 reads, yielding 403× coverage of the genome. Nanopore MinION generated 324,092 reads, with a mean read length of 6922 bases and a read length N50 of 23,635, resulting in 467× coverage of the genome. The genome sequence of strain EC12 was assembled into three circular contigs: the 4,951,794 bp chromosome; a 163,194 bp IncFIB/FII-type plasmid named pEC12_1; and a 58,414 bp IncI2-type plasmid named pEC12_2 (Figure 1). The chromosome was predicted to contain 4752 protein-coding genes, 22 rRNA genes, and 96 tRNA genes. In the pEC12_1, 201 protein-coding genes and one tRNA gene were predicted. In the pEC12_2, 76 protein-coding genes were

predicted. Consistent with the PCR results indicating that it was *stx*-negative but *eae*-positive, the chromosome of strain EC12 contained the LEE pathogenicity island but lacked *stx*-harboring prophages. Plasmid pEC12_1 contained the *bfp* operon. The virulence genes identified in the genome of strain EC12 and other *E. coli* O63:H6 strains are compared in Table 1. Seven *stx*-positive, *eae*-positive strains (377323_2f, 179620, 77AZUC, 8X4S73, SSI_AC286, SSI_AC307, and OLC2140) and two *stx*-negative, *eae*-positive, *bfp*-negative strains (190561 and 191861) were classified as EHEC and atypical EPEC, respectively. Strain EC12 had the virulence gene profile consistent with EHEC, except for *stx*. In silico MLST showed that strain EC12 and EHEC strains belong to sequence type 583 (ST583), whereas atypical EPEC strains belong to ST122. The minimum spanning tree based on cgMLST results revealed that strain EC12 was closely related to EHEC strains but distinct from atypical EPEC strains (Figure 2). In genome alignment using Mauve and BRIG, strain EC12 exhibited a high nucleotide identity with EHEC strain 377323_2f but displayed a gap of around 42.5 kbp corresponding to the *stx*-harboring prophage sequence (Figure 3A,B). Furthermore, plasmid pEC12_1 contained the whole sequence of p377323_2f, a plasmid of strain 377323_2f, in Mauve alignment (Figure 3C). The complete genome sequences of strain EC12 have been deposited in GenBank under accession no. CP070229 (chromosome), CP070230 (pEC12_1), and CP070231 (pEC12_2).
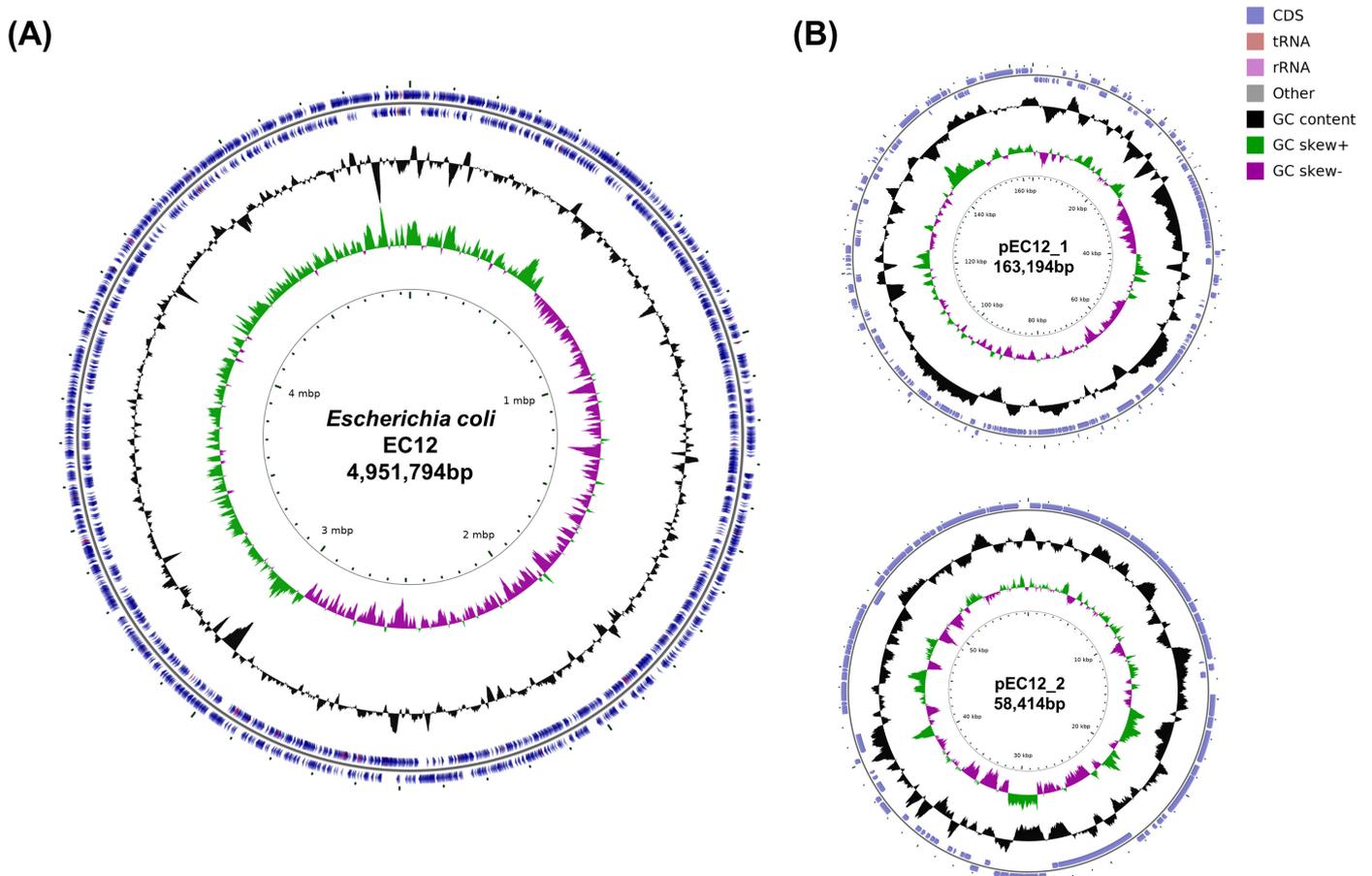

Figure 1. Circular maps of the chromosome (**A**) and two plasmids (**B**) of strain EC12. From the outer circle moving inward, circles 1 and 2 represent protein-coding genes (blue), tRNA genes (reddish brown), and rRNA genes (light purple) on the positive and negative strands; circle 3 shows the GC content (black); circle 4 shows the GC skew (positive GC skew, green; negative GC skew, purple); and circle 5 displays the genomic coordinates.

Table 1. Distribution of virulence genes identified in the genomes of *E. coli* O63:H6 strains.

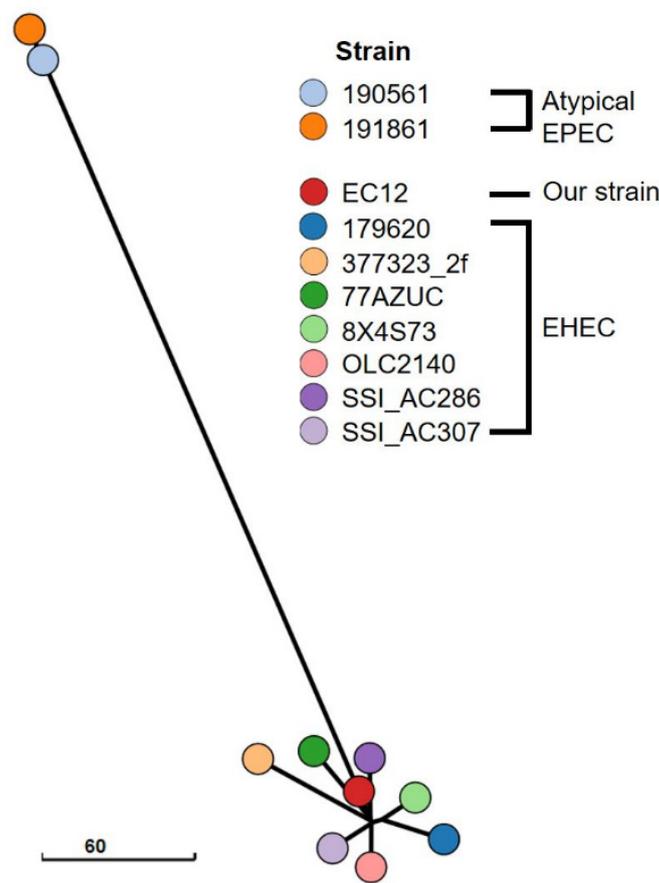| Strain	*astA*	*bfpA*	*chuA*	*cif*	*eae*	*espA*	*espC*	*espF*	*espJ*	*gad*	*ibeA*	*nleB*	*nleC*	*ompT*	*stx*	*tccP*	*terC*	*tir*	*yfcV*
EC12	+	+	+	+	+	+	+	+	+	+	+	−	+	+	−	+	+	+	+
377323_2f	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+ *	+	+	+	+
179620	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+ *	−	+	+	+
77AZUC	+	+	+	+	+	+	+	+	+	−	+	−	+	+	+ *	−	+	+	+
8X4S73	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+ *	+	+	+	+
SSI_AC286	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+ *	−	+	+	+
SSI_AC307	+	+	+	+	+	+	+	+	+	−	+	−	+	+	+ *	−	+	+	+
OLC2140	−	−	+	+	+	+	+	+	+	−	+	−	+	+	+ *	+	+	+	+
190561	−	−	+	+	+	+	+	+	+	+	+	+	−	+	−	−	+	+	+
191861	−	−	+	+	+	+	+	+	+	+	+	+	−	+	−	−	+	+	+

* The *stx* subtype was *stx2f*.

Figure 2. Minimum spanning tree based on cgMLST profiles of 10 *E. coli* O63:H6 strains included in this study. The branch lengths reflect the number of allelic differences between strains.
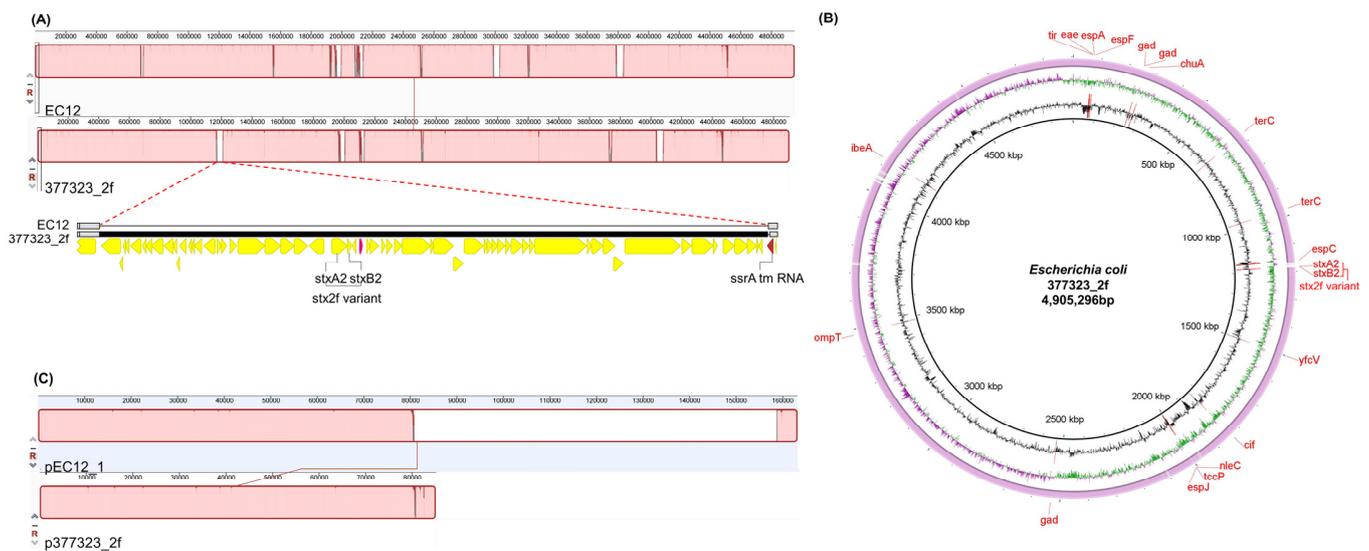
Figure 3. Genome comparison of strains EC12 and 377323_2f. (**A**,**B**) Chromosome comparisons using Mauve (**A**) and BRIG (**B**) demonstrate a gap of around 42.5 kbp in EC12 corresponding to the *stx*-harboring prophage sequence. (**C**) Plasmid alignment using Mauve shows that pEC12_1 contains the whole sequence of p377323_2f. In Mauve, identically colored boxes represent homologous regions shared by the aligned genomes.

4. Discussion

So far, *stx2f* has been the only *stx* subtype found in *E. coli* O63:H6 strains [3–6]. STEC strains carrying *stx2f* are generally associated with mild disease, potentially because these strains carry a relatively small number of virulence genes [5]. However, *stx2f*-positive STEC strains can cause severe disease such as HUS, albeit rarely [6,21]. Although *stx2f*-positive STEC strains are frequently found in pigeons, STEC serotypes associated with human diseases, such as O63:H6, are not found in pigeons, suggesting that the *stx2f*-positive STEC strains responsible for human diseases do not originate from the pigeon reservoir [22].

Infections caused by *stx2f*-positive STEC strains tend to be underreported, as many molecular assays for laboratory diagnosis of STEC infection are not designed to detect *stx2f* [5,6]. The addition of *stx2f* primers to existing PCR assays can be useful for accurate diagnosis of STEC infections, particularly in countries where *stx2f*-carrying STEC strains are frequently encountered.

Of particular note, strain EC12 was positive for *eae* but negative for *stx*. Strain EC12 was initially suspected to be an EHEC-LST strain as it was isolated from an HUS patient. Additionally, as the stool sample of this patient, which was collected 7 days after the onset of gastrointestinal symptoms, was confirmed by PCR to be positive for *eae* but negative for *stx*, strain EC12 is presumed to have lost *stx* in vivo during infection rather than during in vitro culture. The loss of *stx* is a serotype-related phenomenon and has been found in serotypes O26:H11/NM (nonmotile), O103:H2/NM, O145:H28/NM, and O157:H7/NM [10–12]. This study provides the first evidence that *stx* loss can occur in serotype O63:H6.

When using PCR-based assays targeting only *stx* and *eae*, EPEC strains isolated from HUS patients are at risk of being misclassified as EHEC-LST. Whole genome comparison is very useful to determine whether *stx*-negative, *eae*-positive *E. coli* strains isolated from HUS patients are EHEC-LST. A recent WGS study reported that *stx*-negative, *eae*-positive *E. coli* O157:H7/NM strains clustered together with EHEC strains but not with EPEC strains in cgMLST [11]. Furthermore, virulence gene analysis using WGS data revealed that these strains shared a similar virulence gene profile with EHEC strains rather than EPEC strains [11]. Another WGS study demonstrated that a *stx*-negative, *eae*-positive *E. coli* O145:H28 strain had the identical cgMLST and non-*stx* virulence gene profiles with a *stx*-positive, *eae*-positive *E. coli* O145:H28 strain isolated from the same patient [12]. In pairwise genome alignment, the *stx*-negative, *eae*-positive strain was shown to lack the

9 kbp region corresponding to the partial sequence of the *stx*-harboring prophage [12]. In the present study, phylogenetic analysis using cgMLST showed that strain EC12 clustered together with EHEC strains but not with EPEC strains. Furthermore, this strain shared non-*stx* virulence genes with EHEC strains and exhibited a very high nucleotide identity with EHEC strain 377323_2f except for the *stx*-harboring prophage sequence. Taken together, our results indicate that strain EC12 is EHEC-LST, not EPEC. Interestingly, strain EC12 and EHEC strains were *nleB*-negative and *nleC*-positive, whereas EPEC strains were *nleB*-positive and *nleC*-negative. This finding suggests that, in serotype O63:H6, *nleB* and *nleC* may be good molecular markers for differentiating EHEC-LST from EPEC.

In summary, we isolated a *stx*-negative, *eae*-positive *E. coli* O63:H6 strain from a child with HUS. Comparative genomic analysis using WGS data revealed that this strain is genetically more closely related to EHEC than to EPEC, consistent with it being an EHEC-LST strain. Of diagnostic interest, EHEC-LST may be misdiagnosed as EPEC in routine laboratories, and comparative genomic analysis using WGS data can be useful to determine if *stx*-negative and *eae*-positive isolates are EHEC-LST or EPEC.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/diagnostics11101823/s1, Supplementary Table S1: Primers used for detection of *stx1*, *stx2*, *stx2f*, and *eae*.

References

1. Valilis, E.; Ramsey, A.; Sidiq, S.; DuPont, H.L. Non-O157 Shiga toxin-producing *Escherichia coli*—A poorly appreciated enteric pathogen: Systematic review. *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.* **2018**, *76*, 82–87. [CrossRef] [PubMed]
2. Brooks, J.T.; Sowers, E.G.; Wells, J.G.; Greene, K.D.; Griffin, P.M.; Hoekstra, R.M.; Strockbine, N.A. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. *J. Infect. Dis.* **2005**, *192*, 1422–1429. [CrossRef] [PubMed]
3. Prager, R.; Fruth, A.; Siewert, U.; Strutz, U.; Tschäpe, H. *Escherichia coli* encoding Shiga toxin 2f as an emerging human pathogen. *Int. J. Med. Microbiol. IJMM* **2009**, *299*, 343–353. [CrossRef] [PubMed]
4. Buvens, G.; De Gheldre, Y.; Dediste, A.; de Moreau, A.I.; Mascart, G.; Simon, A.; Allemeersch, D.; Scheutz, F.; Lauwers, S.; Piérard, D. Incidence and virulence determinants of verocytotoxin-producing *Escherichia coli* infections in the Brussels-Capital Region, Belgium, in 2008–2010. *J. Clin. Microbiol.* **2012**, *50*, 1336–1345. [CrossRef] [PubMed]
5. Friesema, I.; van der Zwaluw, K.; Schuurman, T.; Kooistra-Smid, M.; Franz, E.; van Duynhoven, Y.; van Pelt, W. Emergence of *Escherichia coli* encoding Shiga toxin 2f in human Shiga toxin-producing *E. coli* (STEC) infections in the Netherlands, January 2008 to December 2011. *Eurosurveillance* **2014**, *19*, 26–32. [CrossRef] [PubMed]
6. De Rauw, K.; Jacobs, S.; Piérard, D. Twenty-seven years of screening for Shiga toxin-producing *Escherichia coli* in a university hospital. Brussels, Belgium, 1987–2014. *PLoS ONE* **2018**, *13*, e0199968. [CrossRef] [PubMed]
7. Kaper, J.B.; Nataro, J.P.; Mobley, H.L. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2004**, *2*, 123–140. [CrossRef]
8. Hartland, E.L.; Leong, J.M. Enteropathogenic and enterohemorrhagic *E. coli*: Ecology, pathogenesis, and evolution. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 15. [CrossRef] [PubMed]
9. Ochoa, T.J.; Contreras, C.A. Enteropathogenic *E. coli* infection (EPEC) in children. *Curr. Opin. Infect. Dis.* **2011**, *24*, 478–483. [CrossRef]

10. Bielaszewska, M.; Köck, R.; Friedrich, A.W.; von Eiff, C.; Zimmerhackl, L.B.; Karch, H.; Mellmann, A. Shiga toxin-mediated hemolytic uremic syndrome: Time to change the diagnostic paradigm? *PLoS ONE* **2007**, *2*, e1024. [CrossRef]

11. Ferdous, M.; Zhou, K.; Mellmann, A.; Morabito, S.; Croughs, P.D.; de Boer, R.F.; Kooistra-Smid, A.M.; Rossen, J.W.; Friedrich, A.W. Is Shiga toxin-negative *Escherichia coli* O157:H7 enteropathogenic or enterohemorrhagic *Escherichia coli*? Comprehensive molecular analysis using whole-genome sequencing. *J. Clin. Microbiol.* **2015**, *53*, 3530–3538. [CrossRef] [PubMed]

12. Senthakumaran, T.; Brandal, L.T.; Lindstedt, B.A.; Jørgensen, S.B.; Charnock, C.; Tunsjø, H.S. Implications of *stx* loss for clinical diagnostics of Shiga toxin-producing *Escherichia coli*. *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.* **2018**, *37*, 2361–2370. [CrossRef] [PubMed]

13. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **2017**, *13*, e1005595. [CrossRef] [PubMed]

14. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]

15. Joensen, K.G.; Scheutz, F.; Lund, O.; Hasman, H.; Kaas, R.S.; Nielsen, E.M.; Aarestrup, F.M. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* **2014**, *52*, 1501–1510. [CrossRef]

16. Larsen, M.V.; Cosentino, S.; Rasmussen, S.; Friis, C.; Hasman, H.; Marvig, R.L.; Jelsbak, L.; Sicheritz-Pontén, T.; Ussery, D.W.; Aarestrup, F.M.; et al. Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* **2012**, *50*, 1355–1361. [CrossRef]

17. Zhou, Z.; Alikhan, N.F.; Mohamed, K.; Fan, Y.; Achtman, M. The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* **2020**, *30*, 138–152. [CrossRef]

18. Zhou, Z.; Alikhan, N.F.; Sergeant, M.J.; Luhmann, N.; Vaz, C.; Francisco, A.P.; Carriço, J.A.; Achtman, M. GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* **2018**, *28*, 1395–1404. [CrossRef]

19. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef]

20. Alikhan, N.F.; Petty, N.K.; Ben Zakour, N.L.; Beatson, S.A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **2011**, *12*, 402. [CrossRef]

21. Friesema, I.H.; Keijzer-Veen, M.G.; Koppejan, M.; Schipper, H.S.; van Griethuysen, A.J.; Heck, M.E.; van Pelt, W. Hemolytic uremic syndrome associated with *Escherichia coli* O8:H19 and Shiga toxin 2f gene. *Emerg. Infect. Dis.* **2015**, *21*, 168–169. [CrossRef] [PubMed]

22. van Hoek, A.; van Veldhuizen, J.N.J.; Friesema, I.; Coipan, C.; Rossen, J.W.A.; Bergval, I.L.; Franz, E. Comparative genomics reveals a lack of evidence for pigeons as a main source of *stx*₂f-carrying *Escherichia coli* causing disease in humans and the common existence of hybrid Shiga toxin-producing and enteropathogenic *E*. coli pathotypes. *BMC Genom.* **2019**, *20*, 271. [CrossRef] [PubMed]