

# Draft Genome Sequences of Three European Laboratory Derivatives from Enterohemorrhagic *Escherichia coli* O157:H7 Strain EDL933, Including Two Plasmids

Lea Fellner,<sup>a</sup> Christopher Huptas,<sup>a</sup> Svenja Simon,<sup>b</sup> Anna Mühlig,<sup>a</sup> Siegfried Scherer,<sup>a</sup> Klaus Neuhaus<sup>a</sup>

Lehrstuhl für Mikrobielle Ökologie, Wissenschaftszentrum Weihenstephan, Technische Universität München, Freising, Germany<sup>a</sup>; Lehrstuhl für Datenanalyse und Visualisierung, Fachbereich Informatik und Informationswissenschaft, Universität Konstanz, Konstanz, Germany<sup>b</sup>

***Escherichia coli* O157:H7 EDL933, isolated in 1982 in the United States, was the first enterohemorrhagic *E. coli* (EHEC) strain sequenced. Unfortunately, European labs can no longer receive the original strain. We checked three European EDL933 derivatives and found major genetic deviations (deletions, inversions) in two strains. All EDL933 strains contain the cryptic EHEC-plasmid, not reported before.**

Received 29 September 2015 Accepted 24 February 2016 Published 7 April 2016

**Citation** Fellner L, Huptas C, Simon S, Mühlig A, Scherer S, Neuhaus K. 2016. Draft genome sequences of three European laboratory derivatives from enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933, including two plasmids. *Genome Announc* 4(2):e01331-15. doi:10.1128/genomeA.01331-15.

**Copyright** © 2016 Fellner et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Siegfried Scherer, [siegfried.scherer@wzw.tum.de](mailto:siegfried.scherer@wzw.tum.de).

Enterohemorrhagic *Escherichia coli* (EHEC) O157:H7 strain EDL933 (ATCC 43895), isolated in 1982, was the first EHEC strain to be sequenced in 2001 (1). In the same year, the genome of EHEC strain Sakai was published (2). The large 92-kb-plasmid pO157 was published in 1998 (3). Because the first EDL933 sequence was inferior, it was recently resequenced (4). Despite its frequent use in research, neither the strain nor its DNA are available to European researchers due to export restrictions of the United States. Unfortunately, EHEC tends to undergo genomic rearrangements (5), which often remain unnoticed.

We sequenced and compared three European derivatives of EDL933. The first derivative, CIP 106327 (Collection de l'Institut Pasteur, Paris, France), was obtained in 2003, freeze-dried after a single sub-cultivation, and deposited in our Weihenstephan strain collection as WS4202. The next derivative WS4435 (= BFEL E135) from R. Pichner (Max-Rubner Institut, Kulmbach, Germany) should be a direct derivative of ATCC 43895. The third derivative, WS4678, was donated by H. Schmidt (Universität Hohenheim, Germany), who received it from L. Beutin (Bundesinstitut für Risikobewertung, Berlin, Germany).

DNA was isolated using CTAB (6) and fragmented using a Covaris E220. Adapter sequences were added using the TruSeq DNA sample preparation kit (Illumina). Libraries were sequenced on a MiSeq station after preparation using the Illumina MiSeq reagent kit v2 (read length 2 × 150) according to the manual.

DNA of WS4202 was additionally sequenced on a Pacific Bioscience PacBio RS II (GATC Biotech AG, Konstanz, Germany) for ring closure. Reads from the Illumina-based sequencing were mapped to the original EHEC genome and plasmid sequence (GenBank accession numbers NC\_002655 and NC\_007414) using Bowtie2 version 2.0.5 (7), considering reads with a mapping quality ≥99% and a base quality ≥99%. In WS4202 and WS4678 we found a mega-inversion of about 1.4 Mbp and larger regions, including OI#48, were missing. After further analysis using the primer system of Bielaszewska et al. (5), OI#48 appeared to be deleted in WS4202 and WS4678. Only the genome of WS4435 appeared to be comparable to the original strain, i.e., similar to the sequence published by Latif et al. (4).

Genomic rearrangements of EHEC have been observed before (5, 8, 9). In the similar strain Sakai, deletions in various prophages were found (10, 11). Chances are good that derivatives possess rearranged genomes and, hence, changed phenotypes (5). This should be kept in mind before choosing strains to conduct experiments.

In a recent genome update (4), the cryptic EHEC-plasmid has gone unnoticed again for EDL933. We assembled unmapped reads using ABySS v1.3.3 (12). A *de novo* contig exactly matched pOSAK1 (3.3 kbp, NC\_002127) (2). However, the original pO157 of EDL933 was found to be incomplete, 636 bp are missing. The

TABLE 1 Genome accession numbers

Strain	Accession no.	Local ID	No. of contigs	Sequence length (bp)
WS4202	<a href="#">CP012802</a>	genome-WS4202	1	5,332,063
WS4202	<a href="#">CP012803</a>	pO157-WS4202	1	92,739
WS4202	<a href="#">CP012804</a>	pOSAK1-WS4202	1	3,306
WS4435	<a href="#">LKAK00000000</a>	Whole genome shotgun	143	
WS4678	<a href="#">LKAL00000000</a>	Whole genome shotgun	195	

combined ratio (i.e., copy number) of chromosome:pO157:pOSAK1 for all three European strains is approximately 1:2:30.

**Nucleotide sequence accession numbers.** The genome sequences were deposited in GenBank (see Table 1). The versions for WS4202 and WS4678 described in this paper are the first versions; for WS4435 it is the second version.

## ACKNOWLEDGMENTS

We thank R. Pichner (Max-Rubner Institut, Kulmbach) and H. Schmidt (Universität Hohenheim) for donating the strains.

## FUNDING INFORMATION

This work, including the efforts of Lea Fellner, Christopher Huptas, Siegfried Scherer, and Klaus Neuhaus, was funded by Deutsche Forschungsgemeinschaft (DFG) (SCHE316/3-1,2,3). This work, including the efforts of Svenja Simon, was funded by Deutsche Forschungsgemeinschaft (DFG) (KE740/13-1,2,3).

## REFERENCES

1. Perna NT, Plunkett G III, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Pósfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533. <http://dx.doi.org/10.1038/35054089>.
2. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 8:11–22. <http://dx.doi.org/10.1093/dnares/8.1.11>.
3. Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR. 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res* 26:4196–4204. <http://dx.doi.org/10.1093/nar/26.18.4196>.
4. Latif H, Li HJ, Charusanti P, Pálsson BØ, Aziz RK. 2014. A gapless, unambiguous genome sequence of the enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933. *Genome Announc* 2(4):e00821-00814. <http://dx.doi.org/10.1128/genomeA.00821-14>.
5. Bielaszewska M, Middendorf B, Tarr PI, Zhang W, Prager R, Aldick T, Dobrindt U, Karch H, Mellmann A. 2011. Chromosomal instability in enterohaemorrhagic *Escherichia coli* O157:H7: impact on adherence, tellurite resistance and colony phenotype. *Mol Microbiol* 79:1024–1044. <http://dx.doi.org/10.1111/j.1365-2958.2010.07499.x>.
6. Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8:4321–4325. <http://dx.doi.org/10.1093/nar/8.19.4321>.
7. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <http://dx.doi.org/10.1038/nmeth.1923>.
8. Iguchi A, Iyoda S, Terajima J, Watanabe H, Osawa R. 2006. Spontaneous recombination between homologous prophage regions causes large-scale inversions within the *Escherichia coli* O157:H7 chromosome. *Gene* 372:199–207. <http://dx.doi.org/10.1016/j.gene.2006.01.005>.
9. Yang Z, Kim J, Zhang C, Zhang M, Nietfeldt J, Southward CM, Surette MG, Kachman SD, Benson AK. 2009. Genomic instability in regions adjacent to a highly conserved *pch* prophage in *Escherichia coli* O157:H7 generates diversity in expression patterns of the LEE pathogenicity island. *J Bacteriol* 191:3553–3568. <http://dx.doi.org/10.1128/JB.01738-08>.
10. Chen C, Lewis CR, Goswami K, Roberts EL, DebRoy C, Dudley EG. 2013. Identification and characterization of spontaneous deletions within the Sp11-Sp12 prophage region of *Escherichia coli* O157:H7 Sakai. *Appl Environ Microbiol* 79:1934–1941. <http://dx.doi.org/10.1128/AEM.03682-12>.
11. Yoshii N, Ogura Y, Hayashi T, Ajiro T, Sameshima T, Nakazawa M, Kusumoto M, Iwata T, Akiba M. 2009. Pulsed-field gel electrophoresis profile changes resulting from spontaneous chromosomal deletions in enterohemorrhagic *Escherichia coli* O157:H7 during passage in cattle. *Appl Environ Microbiol* 75:5719–5726. <http://dx.doi.org/10.1128/AEM.00558-09>.
12. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol İ. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res* 19:1117–1123. <http://dx.doi.org/10.1101/gr.089532.108>.