AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences of 12 Shiga Toxin-Producing *Escherichia coli* Strains Isolated from Dairy Cattle in Portugal

Gregor Fiedler,[a] Andressa Ballem,[b] Erik Brinks,[a] Carina Almeida,[b] Charles M. A. P. Franz,[a] Hugo Oliveira[c]

[a]Department of Microbiology and Biotechnology, Max Rubner-Institut, Federal Research Institute of Nutrition and Food, Kiel, Germany
[b]INIAV, IP—National Institute for Agrarian and Veterinary Research, Vairão, Vila do Conde, Portugal
[c]Centre of Biological Engineering, University of Minho, Braga, Portugal

ABSTRACT Shiga toxin-producing *Escherichia coli* (STEC) is a foodborne pathogen transmitted from animals to humans through contaminated food. Cattle are the main reservoir of STEC, but their genetic diversity is still poorly characterized, especially regarding strains isolated in Portugal. We therefore present the draft genomic sequences of 12 STEC strains isolated from cattle in the north of Portugal.

Shiga toxin-producing *Escherichia coli* (STEC) is able to produce Shiga toxin as its major virulence factor, which is encoded by prophage regions (1). These prophages act as mobile elements and are therefore a driving force in the dissemination of Shiga toxin genes among STEC, other *E. coli* pathotypes, and members of *Enterobacterales* (2). STEC uses the gut of ruminant animals as a natural reservoir and is generally transmitted to humans through the consumption of contaminated foods. From the 500 STEC serotypes known to infect humans, only a few serotypes are responsible for the majority of foodborne cases (3).

In Portugal, the genetic diversity of STEC strains present in cattle has never been evaluated. To increase our knowledge of STEC population epidemiology, the repertoire of STEC genome sequences, and ecology, whole-genome sequencing was performed on 12 strains isolated from dairy cows on a specific farm located in Guilhabreu, Vila do Conde, Portugal.

The genome characteristics and virulence profiles are presented in Table 1. STEC strains were isolated from feces by enrichment with modified tryptone soy broth medium (supplemented with 20 mg liter^{-1} of novobiocin) and finally cultivated on both MacConkey sorbitol and tryptone bile X-glucuronide agar. Next, typical *E. coli* colonies were screened for the presence of *stx* genes by multiplex quantitative PCR (qPCR) according to International Organization for Standardization technical specification 13136:2012. Apart from strains E7N6P4C8A, E7N6P4C8C, and E7N6P4C8F, which are clones isolated from the same animal, the remaining strains originated from different cows/heifers. Strains were grown overnight in brain heart infusion broth at 37°C, and genomic DNA was extracted using the ZR fungal/bacterial DNA miniprep kit (Zymo Research, Freiburg, Germany). The DNA was quantified using a Qubit 3 fluorometer (Invitrogen, Darmstadt, Germany). The sequencing library was prepared with a TruSeq Nano DNA prep kit (Illumina, San Diego, CA, USA) and run on an Illumina MiSeq instrument with 2 × 251-bp paired ends using the MiSeq reagent kit v2 (Illumina). The raw data were trimmed with BBDuk (trimming adapters and low-quality reads and discarding reads shorted than 50 bp) (4), normalized (target coverage, 50×; minimum depth, 6×), and assembled with the Geneious Prime 2020 (Biomatters Ltd., New Zealand) *de novo* assembler with the medium/low-sensitivity settings (5). Read quality and assembly quality were assessed with FastQC v0.11.5 (6) and the Geneious Prime assembler, respectively. Contigs shorter than 500 bp or with coverage less than 8×

Citation Fiedler G, Ballem A, Brinks E, Almeida C, Franz CMAP, Oliveira H. 2020. Draft genome sequences of 12 Shiga toxin-producing *Escherichia coli* strains isolated from dairy cattle in Portugal. Microbiol Resour Announc 9:e00790-20. https://doi.org/10.1128/MRA.00790-20.

Editor David Rasko, University of Maryland School of Medicine

Address correspondence to Hugo Oliveira, hugooliveira@deb.uminho.pt.

Received 15 July 2020
Accepted 13 August 2020
Published 17 September 2020

TABLE 1 Sequencing features of STEC strains

Strain	SRA accession no.	GenBank accession no.	N_{50} (bp)	No. of contigs	No. of raw reads	Genome coverage (×)	Genome size (bp)	No. of CDS[a]	GC content (%)	Serotype	No. of prophages	Shiga toxin gene type(s)	Other virulence genes	gMLST[b]
E7V3P1C1	SRR12187013	JACBWP000000000	113,570	82	3,095,354	62	5,459,339	5,376	49.5	O1:H10	14	stx_1	*astA, ehxA, espI, hra, katP, lpfA, sta1, terC, traT*	1258
E7V4P1C10	SRR12187014	JACBWO000000000	147,484	78	3,120,412	63	5,458,642	5,362	49.4	O1:H10	12	stx_1	*astA, ehxA, espI, espP, gad, hra, katP, lpfA, sta1, terC, traT*	1258
E7V12P1C4G	SRR12187005	JACBWN000000000	126,836	105	3,366,256	64	5,352,924	5,449	49.9	O182:H25	25	stx_1	*astA, eae, ehxA, espA, espF, espJ, espP, etpD, gad, iss, lpfA, nleA, nleB, ompT, tccP, terC, tir, traT*	300
E7N3P7C8A	SRR12187006	JACBWM000000000	130,890	121	2,751,882	62	5,547,926	5,582	51.1	O150:H2	25	stx_1, stx_2	*astA, eae, ehxA, espA, espF, espI, espJ, espP, gad, iha, iucC, iutA, lpfA, nleA, nleB, ompT, tccP, terC, tir, traT*	306
E7N6P4C8A	SRR12187007	JACBWL000000000	210,932	76	3,759,806	60	5,302,937	5,130	51.3	O29:H12	14	stx_1, stx_2	*cea, celb, cia, ehxA, epeA, gad, hra, iha, lpfA, mchB, mchC, mchF, ompT, terC*	155
E7N6P4C8C	SRR12187008	JACBWK000000000	242,239	87	2,707,852	55	5,306,582	5,133	51.0	O29:H12	15	stx_1, stx_2	*cea, celb, cia, ehxA, epeA, gad, hra, iha, lpfA, mchB, mchC, mchF, ompT, terC, terC*	155
E7N6P4C8F	SRR12187009	JACBWJ000000000	160,056	79	3,083,908	59	5,302,542	5,126	51.0	O29:H12	15	stx_1, stx_2	*cea, celb, cia, ehxA, epeA, gad, hra, iha, lpfA, mchB, mchC, mchF, ompT, terC*	155
E7N8P4C1	SRR12187010	JACBWI000000000	238,307	83	2,368,706	51	5,271,944	5,113	51.4	O113:H21	13	stx_2	*cdtB, cea, celb, cia, cib, ehxA, epeA, espP, gad, hra, iha, iss, lpfA, ompT, subA, terC, traT*	223
E7N12P2C4	SRR12187011	JACBWH000000000	127,918	144	2,996,158	52	5,811,737	5,937	51.1	O26:H11	22	stx_1	*astA, celb, cif, eae, efa1, ehxA, espA, espB, espF, espJ, espP, fyuA, gad, iha, irp2, iss, iucC, iutA, katP, lpfA, nleA, nleB, nleC, ompT, tccP, terC, tir, toxB, traT*	21
E7N15P4C10	SRR12187012	JACBWG000000000	155,331	72	1,999,722	53	5,182,995	4,968	51.2	O116:H21	11	stx_2	*cba, cdtB, cea, celb, cia, ehxA, epeA, espP, gad, hra, iha, lpfA, ompT, subA, terC, traT*	Unknown
E7N16P4C9	SRR12187015	JACBWF000000000	142,475	87	2,079,114	48	5,328,386	5,173	51.1	O29:H12	17	stx_1, stx_2	*cea, celb, cia, ehxA, epeA, gad, hra, iha, lpfA, mchB, mchC, mchF, ompT, terC*	155
E7N18P5C8G	SRR12187016	JACBWE000000000	299,977	44	2,479,862	54	5,084,573	4,977	51.9	ONT:H28	10	stx_1, stx_2	*cea, cia, cib, cvaC, ehxA, epeA, espP, gad, hra, iha, iss, lpfA, ompT, subA, terC, traT*	156

[a] CDS, DNA coding sequences.
[b] gMLST, genomic multilocus sequence typing.

were removed. The assembly N_{50} values ranged from 111.973 to 299.997 kb. The assembled contigs were screened using VirulenceFinder (7), SerotypeFinder (8), and PHASTER (9) with default parameters for *in silico* typing of *stx* genes and identification of prophages. PHASTER retrieves and classifies a given prophage as incomplete, questionable, or intact on the basis of its genome completeness or potential viability. Draft genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline v4.12 (10). The genomic data show a diversity of STEC serotypes and virulence genes, with nine different serotypes detected on a single farm.

Data availability. The assembled genome and sequencing reads were deposited under the BioProject accession number PRJNA643688. The GenBank and Sequence Read Archive (SRA) accession numbers are presented in Table 1.

ACKNOWLEDGMENTS

This study was supported by project PhageSTEC PTDC/CVT-CVT/29628/2017 (POCI-01-0145-FEDER-029628) funded by FEDER through COMPETE2020 (Programa Operacional Competitividade e Internacionalização) and by National Funds through FCT (Fundação para a Ciência e a Tecnologia). This study was also supported by FCT grant POCI-01-0247-FEDER-033679 and strategic funding of unit UIDB/04469/2020 and the BioTecNorte operation (NORTE-01-0145-FEDER-000004), funded by the European Regional Development Fund under the scope of Norte2020–Programa Operacional Regional do Norte.

REFERENCES

1. Kruger A, Lucchesi PMA. 2015. Shiga toxins and stx phages: highly diverse entities. Microbiology 161:451–462. https://doi.org/10.1099/mic.0.000003.
2. Khalil RKS, Skinner C, Patfield S, He X. 2016. Phage-mediated Shiga toxin (Stx) horizontal gene transfer and expression in non-Shiga toxigenic Enterobacter and Escherichia coli strains. Pathog Dis 74:ftw037. https://doi.org/10.1093/femspd/ftw037.
3. Gould LH, Mody RK, Ong KL, Clogher P, Cronquist AB, Garman KN, Lathrop S, Medus C, Spina NL, Webb TH, White PL, Wymore K, Gierke RE, Mahon BE, Griffin PM, Emerging Infections Program FoodNet Working Group. 2013. Increased recognition of non-O157 Shiga toxin-producing Escherichia coli infections in the United States during 2000–2010: epidemiologic features and comparison with E. coli O157 infections. Foodborne Pathog Dis 10:453–460. https://doi.org/10.1089/fpd.2012.1401.
4. Bushnell B. 2014. BBMap. https://sourceforge.net/projects/bbmap/.
5. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199.
6. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
7. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol 52:1501–1510. https://doi.org/10.1128/JCM.03617-13.
8. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. J Clin Microbiol 53:2410–2426. https://doi.org/10.1128/JCM.00008-15.
9. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
10. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.