AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequences of 14 *Escherichia coli* O157:H7 Strains Isolated before and during the Time Frame of the 2018 Multistate Outbreak Associated with Romaine Lettuce

Eliot Stanton,[a,b] Sara Wagner,[c] Kelsey R. Florek,[c] Charles W. Kaspar[b]

[a]Department of Bacteriology, University of Wisconsin—Madison, Madison, Wisconsin, USA
[b]Food Research Institute, Madison, Wisconsin, USA
[c]Wisconsin State Laboratory of Hygiene, Madison, Wisconsin, USA

ABSTRACT Several outbreaks of *Escherichia coli* O157:H7 associated with contaminated leafy green vegetables have been documented. Here, we report the draft genome sequences of 14 strains isolated from human patients in the state of Wisconsin during a multistate outbreak in early 2018 that was linked to consumption of romaine lettuce.

Transmission of enterohemorrhagic *Escherichia coli* O157:H7 (EHEC) from the bovine reservoir to humans is usually due to fecal contamination of food or water sources (1). EHEC carries an assortment of genes encoding human virulence factors that may include the Shiga-like toxins Stx1 and Stx2 (2). Processing improvements and regulations have led to a decline in outbreaks linked to ground beef (1). Leafy green vegetables typically consumed raw have become an increasingly prominent vehicle of infection. Consumption of romaine lettuce contaminated with EHEC was implicated in a multistate outbreak between March and June 2018 that resulted in 210 reported infections, 96 hospitalizations, and 5 deaths (3). In the current study, sequence data were used to produce draft whole-genome assemblies of 14 EHEC strains isolated from patients in the state of Wisconsin.

The presence of three major phylogenetic lineages within EHEC has been previously demonstrated (4–7). Six PCR markers are used in the lineage-specific polymorphism assay (LSPA-6) that discriminates between lineages I, II, and I/II. LSPA-6 typing was performed *in silico* using reference nucleotide sequences and BLASTn to distinguish the allele sizes using the highest scoring result. The subtypes stx_{2a} and stx_{2c} were similarly determined *in silico* (8). All isolates belonged to lineage I/II. The majority of strains possessed genes encoding two Stx2 subtypes (Stx2a and Stx2c) and lacked genes encoding Stx1, which is also indicative of lineage I/II. One strain possessed only stx_{2c}, whereas three strains possessed only stx_{2a} (Table 1). Interestingly, EHEC strains isolated from a multistate outbreak associated with spinach in 2006 also belonged to lineage I/II (9).

Clinical isolates were grown on blood agar plate (BAP) medium and Sorbitol-MacConkey agar (SMAC) medium and subsequently checked with Difco and SSI O157 antisera. Biochemical assays consisted of triple sugar iron, motility, and the API 20E fast identification system. The presence/absence of the genes stx_1, stx_2, *eae*, and *ehxA* was determined using PCR amplification. The isolates were stored frozen and briefly thawed; approximately 10 μl was then streaked onto Columbia blood agar (Remel). The plates were incubated overnight at 37°C. Growth from a single colony was streaked onto a new plate and incubated overnight. The growth was transferred to 200 μl of UltraPure molecular-grade (Sigma) water containing 180 μl of lysis buffer from a MagNA Pure LC DNA isolation kit III (Roche). The samples were vortexed until homog-

Citation Stanton E, Wagner S, Florek KR, Kaspar CW. 2020. Genome sequences of 14 *Escherichia coli* O157:H7 strains isolated before and during the time frame of the 2018 multistate outbreak associated with romaine lettuce. Microbiol Resour Announc 9:e00458-20. https://doi.org/10.1128/MRA.00458-20.

Editor David Rasko, University of Maryland School of Medicine

Address correspondence to Charles W. Kaspar, cwkaspar@wisc.edu.

Received 23 April 2020
Accepted 20 June 2020
Published 16 July 2020

TABLE 1 Accession numbers, assembly statistics, LSPA-6 typing, and *stx* genotyping

Strain	Accession no. for:		Read length (bp)	Total no. of reads	Total sequence length (bp)	Genome coverage (%)	No. of contigs	N_{50}	GC content (%)	LSPA-6 genotype for marker:						*stx* genotype		
	Assembly	Sequence reads								FolD-*sfmA*	Z5935	*yhcG*	*rbsB*	*rtcB*	*arp-iclR*	stx_1	stx_{2a}	stx_{2c}
PNUSAE007531	SEKH00000000.1	SRR8499574	250 × 2	1,076,262	5,293,970	98	231	59,778	50.05	2	1	1	1	1	1	−	+	+
PNUSAE010131	SEKI00000000.1	SRR8499573	250 × 2	856,290	5,380,421	78	106	184,501	50.26	2	1	1	1	1	1	−	+	−
PNUSAE010135	SEKJ00000000.1	SRR8499572	250 × 2	868,946	5,358,079	79	104	188,562	50.19	2	1	1	1	1	1	−	+	+
PNUSAE010136	SEKK00000000.1	SRR8499571	250 × 2	711,319	5,434,520	65	117	184,502	50.20	2	1	1	1	1	1	−	+	+
PNUSAE011162	SEKL00000000.1	SRR8499570	250 × 2	1,057,324	5,301,793	96	93	188,562	50.23	2	1	1	1	1	1	−	−	+
PNUSAE011796	SEKM00000000.1	SRR8499569	250 × 2	1,225,680	5,379,475	111	108	188,562	50.23	2	1	1	1	1	1	−	+	+
PNUSAE011802	SEKN00000000.1	SRR8499568	250 × 2	1,164,795	5,377,897	106	103	205,753	50.23	2	1	1	1	1	1	−	+	+
PNUSAE011803	SEKO00000000.1	SRR8499567	250 × 2	905,669	5,384,870	82	104	188,562	50.24	2	1	1	1	1	1	−	+	+
PNUSAE011805	SEKP00000000.1	SRR8499566	250 × 2	712,977	5,386,954	65	108	184,501	50.25	2	1	1	1	1	1	−	+	−
PNUSAE012259	SEKQ00000000.1	SRR8499565	250 × 2	1,079,803	5,373,223	98	95	190,279	50.23	2	1	1	1	1	1	−	+	−
PNUSAE012740	SEKR00000000.1	SRR8499562	250 × 2	1,917,240	5,308,743	174	111	187,686	50.24	2	1	1	1	1	1	−	+	+
PNUSAE013643	SEKS00000000.1	SRR8499561	250 × 2	1,272,417	5,389,036	116	111	184,501	50.25	2	1	1	1	1	1	−	+	+
PNUSAE013919	SEKT00000000.1	SRR8499564	250 × 2	1,008,668	5,254,297	92	125	119,444	50.22	2	1	1	1	1	1	−	+	+
PNUSAE014032	SEKU00000000.1	SRR8499563	250 × 2	672,469	5,293,062	61	97	188,562	50.28	2	1	1	1	1	1	−	+	+

enous, proteinase K was added (20 μl), and the samples were vortexed again (5 to 10 s). The samples were then heated at 65°C for 10 min. Lysates were transferred into the appropriate wells of a MagNA Pure LC sample cartridge and run according to the manufacturer's specifications. DNA from the final eluate was quantified by using a Qubit 2.0 fluorometer (Thermo Fisher Scientific) and a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). Total genomic DNA (gDNA) from each isolate was used to construct a paired-end sequencing library using the Nextera XT DNA library preparation kit (Illumina). Individual libraries were multiplexed and sequenced on the MiSeq platform (Illumina) using the MiSeq reagent kit v2500 cycle (Illumina). Genome assemblies for each strain were produced using SPAdes v3.11.1 in careful mode and utilizing BayesHammer to perform error correction (10). The draft-genome assemblies were subsequently iteratively corrected using Bowtie v1.2.2 to align the paired-end reads with a maximum insert size of 1,000 bp and Pilon v1.22 (11, 12). Contigs less than 1,000 bp long or possessing k-mer coverage less than 20 were excluded from the final assembly. The genomes were annotated in NCBI PGAP v4.7 (13, 14).

Data availability. The sequence data were deposited in GenBank/EMBL/DDBJ under the accession numbers listed in Table 1 and collected under BioProject accession number PRJNA517910.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL. 2005. Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982–2002. Emerg Infect Dis 11:603–609. https://doi.org/10.3201/eid1104.040739.
2. Boerlin P, McEwen SA, Boerlin-Petzold F, Wilson JB, Johnson RP, Gyles CL. 1999. Associations between virulence factors of Shiga toxin-producing Escherichia coli and disease in humans. J Clin Microbiol 37:497–503. https://doi.org/10.1128/JCM.37.3.497-503.1999.
3. CDC. 2018. Multistate outbreak of E. coli O157:H7 infections linked to romaine lettuce (final update). Investigation notice: multistate outbreak of E. coli O157:H7 infections April 2018. https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html.
4. Ziebell K, Steele M, Zhang Y, Benson A, Taboada EN, Laing C, McEwen S, Ciebin B, Johnson R, Gannon V. 2008. Genotypic characterization and prevalence of virulence factors among Canadian Escherichia coli O157:H7 strains. Appl Environ Microbiol 74:4314–4323. https://doi.org/10.1128/AEM.02821-07.
5. Zhang Y, Laing C, Steele M, Ziebell K, Johnson R, Benson AK, Taboada E, Gannon VPJ. 2007. Genome evolution in major Escherichia coli O157:H7 lineages. BMC Genomics 8:121. https://doi.org/10.1186/1471-2164-8-121.
6. Kim J, Nietfeldt J, Benson AK. 1999. Octamer-based genome scanning distinguishes a unique subpopulation of Escherichia coli O157:H7 strains in cattle. Proc Natl Acad Sci U S A 96:13288–13293. https://doi.org/10.1073/pnas.96.23.13288.
7. Laing CR, Buchanan C, Taboada EN, Zhang Y, Karmali MA, Thomas JE, Gannon VPJ. 2009. In silico genomic analyses reveal three distinct lineages of Escherichia coli O157:H7, one of which is associated with hyper-virulence. BMC Genomics 10:287. https://doi.org/10.1186/1471-2164-10-287.
8. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
9. Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA. 2011. Genomic anatomy of Escherichia coli O157:H7 outbreaks. Proc Natl Acad Sci U S A 108:20142–20147. https://doi.org/10.1073/pnas.1107176108.
10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
11. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.
12. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
13. 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.
14. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Oise Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.