Data in Brief 36 (2021) 107065

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Epigenetic regulation of gene expression in Shiga toxin-producing *Escherichia coli*: Transcriptomic data

Michelle Qiu Carter[a,*], Bin Hu[b], Patrick S.G. Chain[b]

[a] *U.S. Department of Agriculture, Agricultural Research Service, Western Regional Research Center, Produce Safety and Microbiology Research Unit, Albany, CA, USA*
[b] *Biosecurity and Public Health Group, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM, USA*

ARTICLE INFO

Article history:
Received 25 February 2021
Revised 9 April 2021
Accepted 13 April 2021
Available online 20 April 2021

Keywords:
Shiga toxin-producing *Escherichia coli* (STEC)
DNA adenine methylase (Dam)
Restriction and modification (R-M) systems
Epigenetic regulation
RNA-Seq transcriptomics

ABSTRACT

Shiga toxin-producing *Escherichia coli* (STEC) strain RM13514 is a clinical isolate linked to the 2010 romaine lettuce-associated outbreak in the U.S. The genes encoding a type II restriction and modification system, PstI R-M, are located in a prophage genome that is also encoding Shiga toxin. In-frame deletion of the PstI R-M genes or *dam*, encoding a DNA adenine methylase, in strain RM13514 were generated, resulting in two mutant strains MQC599 and MQC602, respectively. The mutant strain MQC599 exhibited a similar growth rate as the wild-type (WT) strain RM13514 when grown in Luria-Bertani (LB) broth at 37 °C. In contrast, the growth of mutant strain MQC602 was significantly slower than either RM13514 or MQC599. Genes transcriptionally regulated by the PstI R-M system or by Dam were examined by the RNA-Seq based comparative transcriptomics. The total RNA was extracted from cells of each strain grown in LB broth at exponential and stationary phases. Three biological replicates were collected for each strain. After removal of ribosomal RNA, the mRNAs were converted to cDNAs followed by Illumina sequence library construction. For strains RM13514 and MQC599, six libraires were generated for each, three from the cells in the exponential growth phase and three from the cells in the stationary phase. For

DOI of original article: 10.1016/j.fm.2020.103722
* Corresponding author.
E-mail address: michelle.carter@usda.gov (M.Q. Carter).

https://doi.org/10.1016/j.dib.2021.107065

strain MQC602, three additional libraries were constructed from the cells in the early exponential growth phase. The resulting 21 libraries were combined in equal amounts and sequenced on an Illumina HighSeq 2000 instrument with the Paired End 100 bp (PE100) read format, generating a total of 45.83 Gbp sequence reads. This set of sequence data is available in the NCBI SRA database under the BioProject accession number PRJNA684587. This set of transcriptomic data provides information on methylation-mediated epigenetic regulation in STEC, an important foodborne pathogen that is frequently associated with large epidemic outbreaks and can cause life-threatening disease in humans [1]. This set of data will be useful for researchers who are interested in physiology and pathogenicity of foodborne pathogens or in the fundamental mechanisms of epigenetic regulation in bacteria.

Specifications Table

Subject	Microbiology: Bacteriology
Specific subject area	Molecular microbiology; Shiga toxin-producing *Escherichia coli*
Type of data	Tables, Figure, and DNA sequences
How data were acquired	Bacterial culture, RNA extraction, Reverse transcription, and Illumina sequencing
Data format	Raw and analyzed
Parameters for data collection	Bacterial cells grown in Luria-Bertani broth at 37 °C in exponential and stationary phases were harvested for RNA-Seq based transcriptomic analyses.
Description of data collection	The total RNA was extracted using a Qiagen RNeasy mini kit and quantified using a Nanodrop. The RNA quality was monitored with an Agilent 2100 Bioanalyzer (Agilent Technologies). After removal of ribosomal RNA using a RiboZero rRNA depletion kit (Epicentre Biotechnologies), the samples were converted to cDNA prior to Illumina sequence library construction using a ScriptSeq™ v2RNA-Seq Library Preparation Kit (Epicentre Biotechnologies). The resulting libraries were combined equally (100 pM each) and submitted to Genomic Core at Michigan State University for sequencing. The combined library was sequenced on an Illumina HiSeq2000 instrument with a read format of PE100.
Data source location	The data were collected by The Research Technology Support Facility (RTSF) Genomics Core at Michigan State University (East Lansing, MI, USA) and deposited in NCBI Sequence Read Archive (SRA) Database
Data accessibility	Primary Data: The SRA database under the BioProject accession number PRJNA684587
Related research article	Carter MQ, Pham A, Huynh S, Parker CT, Miller A, He X, Hu B, Chain PSG. 2021. DNA adenine methylase, not the PstI restriction-modification system, regulates virulence gene expression in Shiga toxin-producing *Escherichia coli*. Food Microbiol 96:103722 (DOI: 10.1016/j.fm.2020.103722).

Value of the Data

- The data provide information on methylation-mediated epigenetic regulation in Shiga toxin-producing *Escherichia coli*, an important foodborne pathogen that is frequently associated with large epidemic outbreaks and can cause life-threatening disease in humans.
- The data expand our knowledge about the function of Dam, a well conserved DNA methylase in bacteria.

- The data expand our knowledge about the type II Restriction-Modification systems and their roles in bacterial physiology and pathogenicity.
- The data are useful for researchers who are interested in physiology and pathogenicity of foodborne pathogens as well as epigenetic regulation in general.
- The data contribute to a better understanding of the impact of horizontal gene transfer on bacterial physiology and pathogenicity, given that the PstI R-M genes are associated with genetic mobile elements and have potential to be transferred to other bacteria.

1. Data Description

Strain RM13514 is a clinical isolate linked to the 2010 romaine lettuce-associated outbreak in the U.S. [2]. The genome of RM13514 consists of a chromosome and two plasmids, one carrying the key virulence genes of enterohemorrhagic *E. coli* and the other one conferring multidrug resistance to RM13514 [3]. The genes encoding the PstI R-M system (locus IDs, ECRM13514_RS15860 and ECRM13514_RS15865) in strain RM13514 are located in the genome of the Stx2-prophage, approximately 10,000 bp upstream of the stx_{2a} genes. The gene encoding DNA adenine methylase (*dam*) in strain RM13514 (locus ID, ECRM13514_RS21645) is located between the genes encoding the protein DamX (locus ID, ECRM13514_RS21650) and the ribulose-phosphate 3-epimerase (locus ID, ECRM13514_RS21640). The detailed information about the wild-type (WT) strain RM13514 and the two isogenic mutant strains are described in Table 1. When grown in LB broth at 37 °C, strains RM13514 and MQC599 exhibited a similar growth rate with a generation time of 24.2 min and 23.6 min, respectively (Fig. 1). However, the growth of strain MQC602 was significantly slower than that of WT strain RM13514 or mutant strain MQC599 (One-way ANOVA followed by Tukey's multiple comparisons test, adjusted $P < 0.0001$), with a generation time of 31.9 min (Fig. 1). At the time when both RM13514 and MQC599 entered mid-exponential growth phase, strain MQC602 only reached early exponential growth phase. Following additional 30 min incubation at 37 °C, strain MQC602 entered mid-exponential growth phase; thus, for MQC602, cells in both early- and mid-exponential growth phases were harvested for RNA extraction. A total of 22 sequencing libraries were constructed, of which, 21 were constructed from the rRNA-depleted samples and one served as a negative control (no RNA was added in the first step of library construction) (Table 2). The concentrations of the libraries ranged from 0.6 ng/µl to 5.2 ng/µl, and the average size of the libraries ranged from 350 bp to 458 bp. The Illumina sequencing run yielded a total of 249,756,850 raw clusters with 91.8% of the clusters passing filter (PF). The percent of bases having quality scores equal to and above Q30 for read 1 (R1) and read 2 (R2) were 90.0 and 86.7, respectively. The biological samples, sequence quality scores, and sequence yields for the 22 libraries are described in Table 2. The raw sequence reads are available in NCBI Sequence Read Archive (SRA) Database under the BioProject accession number PRJNA684587. This BioProject includes three SRA experiments. The experiment RM13514_RNA_seq contains a total of 10.2G bases, including the reads from sequence samples Lib1-3 (strain RM13514 in the exponential growth phase) and Lib13-15 (strain RM13514 in the stationary phase); the experiment MQC599_RNA_seq contains a total of 9.7G bases, including the reads from sequence samples Lib4-Lib6 (strain MQC599 in the exponential growth phase) and Lib16-Lib18 (strain MQC599 in the stationary phase); the experiment MQC602_RNA_seq contains

Table 1
Strains used in this study.

Strain	Serotype	Shiga toxin gene	Characteristics	References
RM13514	O145:H28	stx_{2a}	Human isolate linked to the 2010 U.S. romaine lettuce-associated outbreak	[3]
MQC599	O145:H28	stx_{2a}	The PstI R-M deletion mutant of strain RM13514	[1]
MQC602	O145:H28	stx_{2a}	The *dam* deletion mutant of strain RM13514	[1]

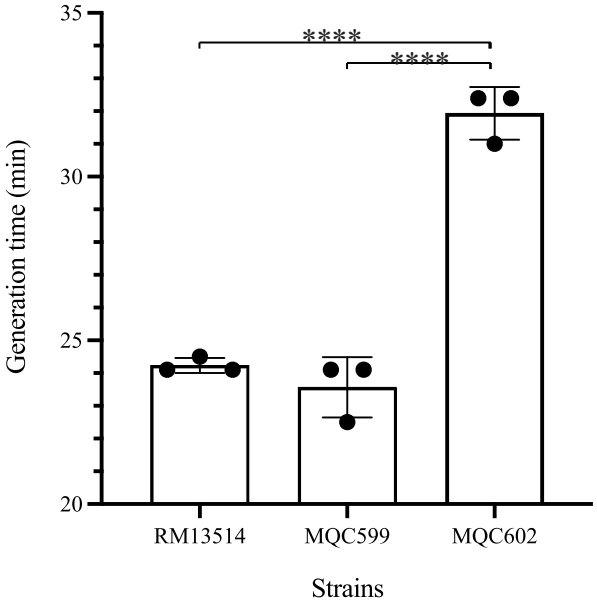

Fig. 1. Generation time of STEC O145:H28 WT and mutant strains incubated in LB broth at 37 °C. The generation time was calculated using the population sizes of each strain at 2 h and 3 h following the inoculation. The value presented here is the mean ± s.d. from three independent experiments. The differences in generation time among the STEC strains are indicated by the *P*-value of the One-way ANOVA followed by Turkey's multiple comparisons test (****$P \leq 0.0001$).

a total of 13.8G bases, including the reads from sequence samples Lib7-Lib12 (strain MQC602 in the exponential growth phase) and Lib19-Lib21 (strain MQC602 in the stationary phase). Genes that are transcriptionally regulated by the PstI R-M systems and Dam are detailed in the related research article [1].

2. Experimental Design, Materials and Methods

2.1. Bacterial strains and growth media

Bacterial strains were grown routinely in Luria-Bertani (LB) broth (10 g tryptone, 5 g yeast extract, and 5 g NaCl per liter). The complete genome of RM13514 is available in GenBank under accession numbers NZ_CP006027.1, NZ_CP006028.1, and NZ_CP006029.1.

2.2. Construction of deletion mutants

In-frame deletion was carried out using the Lambda Red-mediated gene replacement method [4]. Briefly, the Km cassette was amplified from plasmid pKD4 (GenBank accession number: AY048743) using the gene knockout primers described previously [1]. The PCR products were agarose gel purified and transformed by electroporation into the strain RM13514 carrying the helper plasmid pKD46 (GenBank accession number: AY048746). The Km^R transformants were selected on LB agar plates containing Km. The deletion mutants were first verified by colony PCR using primers flanking the target genes described previously [1]. Mutants were sub-cultured at 37 °C and tested for Amp sensitivity, indicating loss of pKD46. The Km cassette was eliminated by transforming the Flp-expressing plasmid pCP20 [4] into the Km^R deletion mutants. The

Table 2
Illumina sequence libraries and sequence reads.

Sequence sample	Biological sample	Barcode	[a]PF Reads	% of PF Reads	[b]R1 %≥Q30	[c]R2 %≥Q30	R1 Ave Q-Score	R2 Ave Q-Score	Yield (Gbp)
Lib1	RM13514 in exponential growth phase, sample A	ATCACG	13,204,273	5.8%	90.1%	86.3%	34.9	33.8	2.64
Lib2	RM13514 in exponential growth phase, sample B	CGATGT	9,973,866	4.4%	91.3%	86.9%	35.2	34.0	1.99
Lib3	RM13514 in exponential growth phase, sample C	TTAGGC	8,975,219	3.9%	91.0%	86.7%	35.2	33.9	1.80
Lib4	MQC599 in exponential growth phase, sample A	TGACCA	9,509,712	4.1%	90.7%	86.2%	35.1	33.8	1.90
Lib5	MQC599 in exponential growth phase, sample B	ACAGTG	11,335,277	4.9%	90.9%	86.9%	35.1	34.0	2.27
Lib6	MQC599 in exponential growth phase, sample C	GCCAAT	10,667,903	4.7%	91.0%	86.8%	35.2	34.0	2.13
Lib7	MQC602 in exponential growth phase, sample A	CAGATC	9,779,135	4.3%	91.0%	87.7%	35.1	34.2	1.96
Lib8	MQC602 in exponential growth phase, sample B	ACTTGA	13,723,068	6.0%	91.0%	87.7%	35.1	34.2	2.74
Lib9	MQC602 in exponential growth phase, sample C	GATCAG	9,030,685	3.9%	91.0%	87.6%	35.1	34.2	1.81
Lib10	MQC602 in early exponential growth phase, sample A	TAGCTT	9,631,430	4.2%	90.7%	87.3%	35.0	34.1	1.93
Lib11	MQC602 in early exponential growth phase, sample B	GGCTAC	10,443,708	4.6%	90.4%	86.9%	35.0	33.9	2.09
Lib12	MQC602 in early exponential growth phase, sample C	CTTGTA	11,254,839	4.9%	90.7%	87.2%	35.1	34.0	2.25
Lib13	RM13514 in stationary phase, sample A	AGTCAA	11,738,511	5.1%	89.9%	87.6%	34.7	34.2	2.35
Lib14	RM13514 in stationary phase, sample B	AGTTCC	11,967,549	5.2%	90.0%	87.7%	34.7	34.2	2.39
Lib15	RM13514 in stationary phase, sample C	ATGTCA	10,777,017	4.7%	90.4%	87.9%	34.8	34.2	2.16
Lib16	MQC599 in stationary phase, sample A	CCGTCC	11,605,080	5.1%	89.5%	87.1%	34.6	34.0	2.32
Lib17	MQC599 in stationary phase, sample B	GTAGAG	11,203,170	4.9%	89.6%	87.4%	34.6	34.1	2.24
Lib18	MQC599 in stationary phase, sample C	GTGAAA	11,105,581	4.8%	90.1%	87.6%	34.8	34.2	2.22
Lib19	MQC602 in stationary phase, sample A	GTGGCC	9,580,058	4.2%	88.8%	86.4%	34.4	33.7	1.92
Lib20	MQC602 in stationary phase, sample B	CGTACG	10,354,741	4.5%	89.7%	87.8%	34.7	34.2	2.07
Lib21	MQC602 in stationary phase, sample C	GAGTGG	10,274,253	4.5%	89.2%	87.3%	34.5	34.0	2.05
Control	No input cDNA	CACCGG	44,686	0.0%	91.2%	86.4%	35.1	33.6	0.01

[a] PF: passing filter
[b] R1: read 1
[c] R2: read 2.

transformants were selected at 30 °C on LB plates supplemented with carbenicillin, followed by incubation at 43 °C, which resulted in a loss of temperature sensitive plasmid pCP20. The elimination of Km cassette from the targeted loci was confirmed by DNA sequencing.

2.3. Transcriptomic study

Overnight cultures from single colonies of each strain grown at 37 °C were inoculated in fresh LB broth at a concentration equivalent to $0.001OD_{600}$ ml^{-1}. Cultures were incubated at 37 °C with shaking (150 rpm). The growth rate of each strain in LB broth at 37 °C was first determined with a Bioscreen C instrument (Labsystems Diagnostitcs). The growth was monitored by recording the optical density of the culture at 600 nm (OD_{600}) every 30 min for 10 hrs. Three biological replicates were run for each strain. The differences in growth rate among the three strains were determined by the *P*-value of the Tukey's multiple comparisons test after a One-way ANOVA test in Prism 9 (GraphPad Software, LLC.). For RNA extraction, an aliquot of 4 ml and 0.8 ml in the mid-exponential and early stationary growth phase, respectively, were removed from each culture as described previously [5,6]. Due to slow growth of strain MQC602, an aliquot of 4 ml in the early exponential growth phases were also sampled. The total RNA was extracted using a Qiagen RNeasy® mini kit (Qiagen, Cat. No. 74104) and quantified using a Nanodrop. The RNA quality was monitored with an Agilent 2100 Bioanalyzer (Agilent Technologies). After removal of ribosomal RNA using a RiboZero rRNA depletion kit (Epicentre Biotechnologies, Cat. No. MRZB12424), the samples were converted to cDNA prior to Illumina sequence library construction using a ScriptSeq v2RNA-Seq Library Preparation Kit (Epicentre Biotechnologies, Cat. No. SSV21124). The quality, quantity, and size distribution of the Illumina libraries were determined using an Agilent 2100 Bioanalyzer. For strains RM13514 and MQC599, six libraires were generated for each, three from the exponential growth phase cells and three from the stationary phase cells. For strain MQC602, three additional libraries were constructed from the cells in early exponential growth phase. The resulting 21 libraries were combined equally (100 pM each) and sequenced using an Illumina HighSeq 2000 instrument with the Paired End 100 bp (PE100) read format at the Genomic Core of Michigan State University (https://rtsf.natsci.msu.edu/genomics/) (Paired-End DNA Sample Prep Kit, Cat. No. PE-102-1001; HiSeq Paired-End Cluster Generation kit, Cat. No. PE-401-1001; HiSeq Sequencing Kit, Cat. No. FC-401-1002). The raw sequence data (primary dataset) are available in the NCBI SRA database under BioProject accession number PRJNA684587.

Ethics Statement

N/A

CRediT Author Statement

Michelle Qiu Carter: Conceptualization, Methodology, Data curtion, Writing - original draft, Writing review & editing, Project administration, Funding acquisition; **Bin Hu:** Data curtion; **Patrick S.G. Chain:** Conceptualization, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Acknowledgments

This work was funded by USDA-ARS CRIS project 2030-42000-050-00D.

References

[1] M.Q. Carter, A. Pham, S. Huynh, C.T. Parker, A. Miller, X. He, B. Hu, P.S.G. Chain, DNA adenine methylase, not the PstI restriction-modification system, regulates virulence gene expression in Shiga toxin-producing *Escherichia coli*, Food Microbiol. 96 (2021) 103722.

[2] E.V. Taylor, T.A. Nguyen, K.D. Machesky, E. Koch, M.J. Sotir, S.R. Bohm, J.P. Folster, R. Bokanyi, A. Kupper, S.A. Bidol, A. Emanuel, K.D. Arends, S.A. Johnson, J. Dunn, S. Stroika, M.K. Patel, I. Williams, Multistate outbreak of *Escherichia coli* O145 infections associated with romaine lettuce consumption, 2010, J. Food Prot. 76 (2013) 939–944.

[3] K.K. Cooper, R.E. Mandrell, J.W. Louie, J. Korlach, T.A. Clark, C.T. Parker, S. Huynh, P.S. Chain, S. Ahmed, M.Q. Carter, Comparative genomics of enterohemorrhagic *Escherichia coli* O145:H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157:H7, BMC Genomics 15 (2014) 17.

[4] K.A. Datsenko, B.L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, Proc. Natl. Acad. Sci. USA 97 (2000) 6640–6645.

[5] M.Q. Carter, C.T. Parker, J.W. Louie, S. Huynh, C.K. Fagerquist, R.E. Mandrell, RcsB contributes to the distinct stress fitness among *Escherichia coli* O157:H7 curli variants of the 1993 hamburger-associated outbreak strains, Appl. Environ. Microbiol. 78 (2012) 7706–7719.

[6] M.Q. Carter, J.W. Louie, S. Huynh, C.T. Parker, Natural *rpoS* mutations contribute to population heterogeneity in *Escherichia coli* O157:H7 strains linked to the 2006 US spinach-associated outbreak, Food Microbiol. 44 (2014) 108–118.