Genome Sequence of Enterotoxigenic *Escherichia coli* Strain FMU073332

Zeus Saldaña-Ahuactzi,[a] Ariadnna Cruz-Córdova,[a] Gerardo E. Rodea,[a]
Helena Porta,[b] Armando Navarro-Ocaña,[c] Carlos Eslava-Campos,[c]
Miguel A. Cevallos,[d] Juan Xicohtencatl-Cortes[a]

Laboratorio de Investigación en Bacteriología Intestinal, Hospital Infantil de México Federico Gómez, Ciudad de México, México[a]; Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México[b]; Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad Universitaria, Ciudad de México, México[c]; Centro de Ciencias Genómicas, Programa de Genómica Evolutiva, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México[d]

ABSTRACT   Enterotoxigenic *Escherichia coli* (ETEC) is an important cause of bacterial diarrheal illness, affecting practically every population worldwide, and was estimated to cause 120,800 deaths in 2010. Here, we report the genome sequence of ETEC strain FMU073332, isolated from a 25-month-old girl from Tlaltizapán, Morelos, México.

Address correspondence to Miguel A. Cevallos, mac@ccg.unam.mx, or Juan Xicohtencatl-Cortes, juanxico@yahoo.com.

Enterotoxigenic *Escherichia coli* (ETEC) is a type of pathogenic bacteria that causes traveler's diarrhea and affect populations worldwide, with children under 2 years old being the most affected (1, 2). The mortality caused by ETEC in all age groups in 2010 was estimated to be 120,800 deaths (3). ETEC strain FMU073332 serotype O6:H16 is a clinical isolate, recovered in 1987 from the stool of a 25-month-old girl from Tlaltizapán, Morelos, México (4). This strain has been used in previous studies to characterize the CS21 colonization factor in the adherence of ETEC to HT-29 cells and self-aggregation (5). We used the genomic sequences of ETEC strain FMU073332 to extract gene sequences and perform multilocus sequence typing (MLST). According to the PubMLST (Achtman) and Pasteur systems, the strain belongs to the sequence type 4 (ST4) and ST215 clonal groups, respectively (6, 7). This strain carries the classic virulence genes *eltA*, *eltB*, *sta*2, *cstH*, and *lngA* and the nonclassic *etpA* and *etpB* genes.

A high-quality draft of the genome of this strain was determined on a PacBio RSII platform. A large-insert (20 to 25 kb) library was constructed and sequenced, using one single-molecule real-time (SMRT) cell with a P6 polymerase and C4 chemistry combination (P6-C4) with a 180-min movie. The statistical analysis obtained from the genome assembly gave 43,603 reads with an N_{50} length of 11,395 bp. The *de novo* assembly was made using an RS Hierarchical Genome Assembly Process (HGAP) protocol version 3 in SMRT Analysis version 2.3 (Pacific Biosciences) (8). To improve regions of low coverage, a library with a 2 × 100-bp paired-end configuration was sequenced on an Illumina HiSeq platform. This platform yielded 11,282,773 reads. A hybrid assembly was constructed with the Illumina HiSeq reads and PacBio RII subreads, using SPAdes 3.5.0 (9). The plasmids and chromosome were circularized, when possible, using apc, a perl script available at https://github.com/jfass/apc. Functional annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline.

The hybrid assembly consisted of five contigs: the largest comprises the complete ETEC strain FMU073332 circular chromosome (4,718,719 bp). The other four contigs represent the four plasmids of this strain: pEcoFMU073332a (5,538 bp), pEcoFMU073332b (47,563 bp), pEcoFMU073332c (113,343 bp), and pEcoFMU073332d (137,665 bp). The sequences of plasmids pEcoFMU073332b and pEcoFMU073332d are not complete; however, an electrophoretic plasmid profile analysis indicated that the number and sizes of the

observed plasmids match with the sequence sizes of plasmids pEcoFMU073332b and pEcoFMU073332d from genome assembly.

The chromosome has seven rRNA operons, 91 tRNAs, and 4,936 coding sequences (CDSs). Some genes relevant for virulence are located in plasmids, for example, genes encoding subunits A and B of the heat-labile enterotoxin (*eltA* and *eltB*); the secreted autotransporter serine protease (*eatA*); and the mayor fimbrial subunit (*cstH*) and the heat-stable enterotoxin ST (*sta*2) located in plasmid pEcoFMU073332d. In contrast, gene encoding the CS21 pilus structural subunit (*lngA*) is located in plasmid pEcoFMU073332b. The high-quality draft genome of ETEC Mexican strain FMU073332 and its four plasmids will aid in precise genetic manipulation and therefore further improve the study of ETEC virulence.

Accession number(s). The complete sequences of the chromosome and the four plasmids of ETEC strain FMU073332 have been submitted to GenBank under the accession numbers CP017844 to CP017848.

REFERENCES

1. Qadri F, Svennerholm AM, Faruque AS, Sack RB. 2005. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev 18:465–483. https://doi.org/10.1128/CMR.18.3.465-483.2005.
2. Svennerholm AM, Tobias J. 2008. Vaccines against enterotoxigenic *Escherichia coli*. Expert Rev Vaccines 7:795–804. https://doi.org/10.1586/14760584.7.6.795.
3. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, et al. 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. Lancet 380:2095–2128. https://doi.org/10.1016/S0140-6736(12)61728-0.
4. Cravioto A, Reyes RE, Trujillo F, Uribe F, Navarro A, De La Roca JM, Hernández JM, Pérez G, Vázquez V. 1990. Risk of diarrhea during the first year of life associated with initial and subsequent colonization by specific enteropathogens. Am J Epidemiol 131:886–904.
5. Cruz-Córdova A, Espinosa-Mazariego K, Ochoa SA, Saldaña Z, Rodea GE, Cázares-Domínguez V, Rodríguez-Ramírez V, Eslava-Campos CA, Navarro-Ocaña A, Arrellano-Galindo J, Hernández-Castro R, Gómez-Duarte OG, Qadri F, Xicohtencatl-Cortes J. 2014. CS21 positive multidrug-resistant ETEC clinical isolates from children with diarrhea are associated with self-aggregation, and adherence. Front Microbiol 5:709. https://doi.org/10.3389/fmicb.2014.00709.
6. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 60:1136–1151. https://doi.org/10.1111/j.1365-2958.2006.05172.x.
7. Clermont O, Gordon D, Denamur E. 2015. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. Microbiology 161:980–988. https://doi.org/10.1099/mic.0.000063.
8. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
9. 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.