Draft Whole-Genome Sequences of 10 Atypical Enteropathogenic *Escherichia coli* Strains Isolated in Brazil

Fernanda F. Santos,[a] Emily J. Richardson,[b] Jack A. Bryant,[b] Denise Yamamoto,[a*] (ID) Waldir P. Elias,[c] (ID) Tânia A. T. Gomes,[a] Ian R. Henderson[b]

[a]Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, Brazil
[b]Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom
[c]Instituto Butantan, Laboratório de Bacteriologia, São Paulo, Brazil

ABSTRACT The number of diarrhea cases caused by atypical enteropathogenic *Escherichia coli* (aEPEC) has been increasing worldwide. Here, we report the draft whole-genome sequences of 10 aEPEC strains isolated in Brazil. These sequences will provide an important source for future studies concerning aEPEC pathogenicity and genetic markers of potentially virulent strains.

Globally, diarrheal diseases are the second leading cause of death and the leading cause of malnutrition in children under 5 years old (1, 2). One of the most significant etiological agents of moderate-to-severe diarrhea is enteropathogenic *Escherichia coli* (EPEC), which is one of the diarrheagenic *E. coli* pathotypes. EPEC is subdivided into typical EPEC (tEPEC) and atypical EPEC (aEPEC) based on the presence of the EPEC adherence factor (EAF) plasmid in the former group and its absence in the latter group (3, 4).

The main pathogenic mechanism of both tEPEC and aEPEC is the formation of attaching and effacing (AE) lesions in the intestinal epithelium, in which intimate adherence between bacteria and host cells occurs (5–8). The genes required for the establishment of AE lesions are located on the locus of enterocyte effacement (LEE) island, which is approximately 35 kb long (9).

The number of cases of diarrhea caused by aEPEC has increased even in industrialized countries. This information reveals the need for a thorough knowledge of the genetic traits of aEPEC organisms and their mechanisms of pathogenicity in order to develop appropriate control strategies (4, 10, 11).

The 10 aEPEC strains we sequenced were isolated from feces of children under 5 years old during epidemiological studies carried out in São Paulo, Brazil (1985 to 1986 and 1989 to 1990) by the Laboratory of Microbiology of the Universidade Federal de São Paulo (UNIFESP) and in Salvador (2003 to 2004) by the Laboratory of Bacteriology of the Butantan Institute (12, 13). All the strains were stored at −80°C in lysogeny broth (LB) and 15% glycerol and were routinely cultured in LB for 18 to 20 h at 37°C.

DNA extraction was performed using the QIAamp DNA minikit (Qiagen) according to the manufacturer's instructions. DNA libraries from each aEPEC genome were prepared using the Nextera XT library prep kit (Illumina, San Diego, CA) and sequenced at MicrobesNG (University of Birmingham, UK) on the Illumina HiSeq 2500 platform using 2- × 250-bp paired-end reads, achieving between 34 and 157× depth (Table 1). Trimmomatic version 0.30 (14) was used to adapt and quality trim the reads with a sliding window quality cutoff of Q15. The *de novo* assembly and annotation of the genomes were performed using the software SPAdes version 3.9.1 (with the – careful flag) (15) and Prokka version 1.12 (16), respectively. The genome assembly metric, calculated using QUAST, and the number of

Address correspondence to Fernanda F. Santos, ff.santos@unifesp.br, or Tânia A. T. Gomes, tatg.amaral@unifesp.br.

* Present address: Denise Yamamoto, Universidade de Santo Amaro, Campus I, São Paulo, Brazil.

TABLE 1 Genome assembly metrics and number of CDS of 10 aEPEC genomes

Strain	Mean coverage (×)	No. of contigs	Largest contig (bp)	Total length (bp)	GC content (%)	N_{50} (bp)	N_{75} (bp)	L_{50} (contigs)	L_{75} (contigs)	No. of CDS	GenBank accession No.
1331-2	141.587	148	428,643	5,187,915	50.61	112,592	63,355	14	29	4,998	QYVD00000000
2012-1	78.7493	230	247,360	5,607,608	50.43	97,093	41,042	21	43	5,482	QYVC00000000
2531-13	157.673	265	242,288	5,055,689	50.59	73,162	36,314	21	45	4,843	QYVB00000000
3522-6	85.7236	144	317,272	4,730,622	50.57	97,356	51,903	15	32	4,494	QYVA00000000
3881-3	109.173	73	359,010	4,701,486	50.72	241,425	130,938	9	15	4,442	QYUZ00000000
3991-1	34.3279	239	196,577	4,812,795	50.79	73,367	32,807	22	46	4,569	QYUY00000000
4581-2	87.0432	78	406,594	4,696,360	50.41	136,790	95,011	11	20	4,382	QYYF00000000
4632-3	93.7625	174	288,111	4,855,364	50.6	90,810	36,963	16	38	4,617	QYUX00000000
BA2103	85.4008	252	285,340	5,412,725	50.48	102,510	48,630	18	36	5,231	QYUW00000000
BA4095	43.121	71	510,385	4,797,996	50.61	216,769	124,397	8	14	4,568	QYUV00000000

coding sequences (CDS) annotated in each aEPEC genome are shown in Table 1. Unless specified, all software was used with default settings.

These sequences will provide an important source for futures studies concerning aEPEC pathogenicity and genetic markers of potentially virulent strains. Furthermore, these data will permit comparative studies to be carried out with aEPEC strains isolated in Brazil and others isolated in different countries.

Data availability. The reads used for assembly of the 10 aEPEC genomes were deposited in the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) under the accession number PRJNA490882, and the whole-genome shotgun sequences were deposited in the GenBank database under the accession numbers shown in Table 1. The versions described in this paper are the first versions.

REFERENCES

1. United Nations Children's Fund (UNICEF)/World Health Organization (WHO). 2009. Diarrhoea: why children are still dying and what can be done. whqlibdoc.who.int/publications/2009/9789241598415_eng.pdf.
2. World Health Organization (WHO). 2017. Diarrhoeal disease/fact sheet. http://www.who.int/mediacentre/factsheets/fs330/en/.
3. Kaper JB. Defining EPEC. 1996. Rev Microbiol 27:130–133.
4. Donnenberg MS, Finlay BB. 2013. Combating enteropathogenic Escherichia coli (EPEC) infections: the way forward. Trends Microbiol 21:317–319. https://doi.org/10.1016/j.tim.2013.05.003.
5. Rothbaum R, McAdams AJ, Giannella R, Partin JC. 1982. A clinicopathologic study of enterocyte-adherent Escherichia coli: a cause of protracted diarrhea in infants. Gastroenterology 83:441–454.
6. Moon HW, Whipp SC, Argenzio RA, Levine MM, Giannella RA. 1983. Attaching and effacing activities of rabbit and human enteropathogenic Escherichia coli in pig and rabbit intestines. Infect Immun 41:1340–1351.
7. Hernandes RT, Elias WP, Vieira MAM, Gomes TAT. 2009. An overview of atypical enteropathogenic Escherichia coli. FEMS Microbiol Lett 297:137–149. https://doi.org/10.1111/j.1574-6968.2009.01664.x.
8. Wong AR, Pearson JS, Bright MD, Munera D, Robinson KS, Lee SF, Frankel G, Hartland EL. 2011. Enteropathogenic and enterohaemorrhagic Escherichia coli: even more subversive elements. Mol Microbiol 80:1420–1438. https://doi.org/10.1111/j.1365-2958.2011.07661.x.
9. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc Natl Acad Sci U S A 92:1664–1668. https://doi.org/10.1073/pnas.92.5.1664.
10. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. 2013. Recent advances in understanding enteric pathogenic Escherichia coli. Clin Microbiol Rev 26:822–880. https://doi.org/10.1128/CMR.00022-13.
11. Ingle DJ, Tauschek M, Edwards DJ, Hocking DM, Pickard DJ, Azzopardi KI, Amarasena T, Bennett-Wood V, Pearson JS, Tamboura B, Antonio M, Ochieng JB, Oundo J, Mandomando I, Qureshi S, Ramamurthy T, Hossain A, Kotloff KL, Nataro JP, Dougan G, Levine MM, Robins-Browne RM, Holt KE. 2016. Evolution of atypical enteropathogenic E. coli by repeated acquisition of LEE pathogenicity island variants. Nat Microbiol 1:15010. https://doi.org/10.1038/nmicrobiol.2015.10.
12. Gomes TA, Irino K, Girão DM, Girão VB, Guth BE, Vaz TM, Moreira FC, Chinarelli SH, Vieira MA. 2004. Emerging enteropathogenic Escherichia coli strains? Emerg Infect Dis 10:1851–1855. https://doi.org/10.3201/eid1010.031093.
13. Abe CM, Trabulsi LR, Blanco J, Blanco M, Dahbi G, Blanco JE, Mora A, Franzolin MR, Taddei CR, Martinez MB, Piazza RM, Elias WP. 2009. Virulence features of atypical enteropathogenic Escherichia coli identified by the eae(+) EAF-negative stx(−) genetic profile. Diagn Microbiol Infect Dis 64:357–365. https://doi.org/10.1016/j.diagmicrobio.2009.03.025.
14. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
15. 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.
16. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.