genomeA
Journals.ASM.org

Draft Genome Sequences of Nine Enteropathogenic *Escherichia coli* Strains from Kenya

Tracy H. Hazen,[a] Michael S. Humphrys,[a] John Benjamin Ochieng,[b] Michele Parsons,[c] Cheryl A. Bopp,[c] Ciara E. O'Reilly,[c] Eric Mintz,[c] David A. Rasko[a,d]

Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA[a]; Kenya Medical Research Institute/Centers for Disease Control and Prevention, Kisumu, Kenya[b]; Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA[c]; Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA[d]

We report here the draft genome sequences of nine enteropathogenic *Escherichia coli* (EPEC) strains isolated from children in Kenya who died during hospitalization with diarrhea. Each of the isolates possess the EPEC adherence factor (EAF) plasmid encoding the bundle-forming pilus, which is characteristic of EPEC. These isolates represent diverse serogroups and EPEC phylogenomic lineages.

Received 22 May 2014 **Accepted** 27 May 2014 **Published** 12 June 2014

Citation Hazen TH, Humphrys MS, Ochieng JB, Parsons M, Bopp CA, O'Reilly CE, Mintz E, Rasko DA. 2014. Draft genome sequences of nine enteropathogenic *Escherichia coli* strains from Kenya. Genome Announc. 2(3):e00582-14. doi:10.1128/genomeA.00582-14.

Address correspondence to David A. Rasko, drasko@som.umaryland.edu.

Attaching and effacing *Escherichia coli* (AEEC) is characterized by the presence of the locus of enterocyte effacement (LEE) pathogenicity region that encodes a type III secretion system involved in pathogenesis (1–3). AEEC is further identified as either LEE-positive Shiga toxin-producing *E. coli* (STEC) or enteropathogenic *E. coli* (EPEC). EPEC strains contain the LEE region but not the Shiga toxin-encoding phage, and they are considered typical EPEC when they possess the bundle-forming pilus (BFP) encoded by the EPEC adherence factor (EAF) plasmid (1, 4, 5). In contrast, LEE-positive *stx*-negative *bfp*-negative *E. coli* strains are considered atypical EPEC (4). The nine *E. coli* isolates described here are LEE positive, *stx* negative, and *bfp* positive and are therefore classified as typical EPEC based on their virulence factor content.

The genomic DNA for sequencing was isolated from an overnight culture of each isolate using the Sigma GenElute kit (Sigma-Aldrich). Genome sequencing was performed at the University of Maryland School of Medicine, Institute for Genome Sciences, at the Genome Resource Center. The genome sequences were generated using paired-end libraries with 300-bp inserts on the Illumina HiSeq 2000. The Illumina reads generated for each isolate were assembled *de novo* using the Velvet assembly program (6), with *k*-mer values determined using VelvetOptimiser version 2.1.4 (http://bioinformatics.net.au/software.velvetoptimiser.shtml). The genome assemblies contained an average of 181 contigs per isolate (range, 117 to 283), with an average G+C content of 50.4% (range, 50.3 to 50.6%) and an average genome size of 5.2 Mb (range, 5.0 to 5.4 Mb).

These nine isolates were isolated from children in Kenya who died during hospitalization with diarrhea (7). These EPEC isolates have diverse serogroups (O126, O119, O111, O53, O127, O55, and O114). Phylogenomic analysis of the whole genomes demonstrated that these isolates occur in two *E. coli* phylogroups (B1 and B2) (8, 9) and occupy four EPEC phylogenomic lineages. Among the four EPEC isolates in phylogroup B2 was isolate S6995 (O127), which was identified in the EPEC1 phylogenomic lineage, with the most widely studied prototype EPEC isolate E2348/69 (10). Also from phylogroup B2 was EPEC isolate S6400 (O119), which was identified in the recently described EPEC4 phylogenomic lineage (11). The remaining two EPEC isolates identified in phylogroup B2, S6966 (O53) and S7438 (O55), were present in unassigned phylogenomic lineages. The five EPEC isolates identified in phylogroup B1, S7005 (O126), S6685 (O114), S5274 (O111), S7380 (serogroup unknown), and S6662 (O111), were all present in the EPEC2 phylogenomic lineage, which contains the EPEC prototype isolate B171 (12).

The genome sequencing of these nine EPEC isolates demonstrates that the EPEC isolates associated with diarrheal illness of children in Kenya exhibit considerable genomic diversity that was not previously appreciated.

Nucleotide sequence accession numbers. The genome assemblies have been deposited at DDBJ/EMBL/GenBank with accession no. JICI00000000, JICH00000000, JICG00000000, JICF00000000, JICE00000000, JICD00000000, JICC00000000, JICB00000000, and JICA00000000, for isolates S6966, S6995, S7005, S7438, S6400, S7380, S6662, S5274, and S6685, respectively.

ACKNOWLEDGMENTS

This project was funded in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract 1U19AI090873, and by funds from the State of Maryland.

REFERENCES

1. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. **2**:123–140. http://dx.doi.org/10.1038/nrmicro818.
2. **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. http://dx.doi.org/10.1073/pnas.92.5.1664.
3. **McDaniel TK, Kaper JB.** 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol. Microbiol. **23:**399–407. http://dx.doi.org/10.1046/j.1365-2958.1997.2311591.x.
4. **Nataro JP, Kaper JB.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142–201.
5. **Donnenberg MS, Kaper JB.** 1992. Enteropathogenic *Escherichia coli*. Infect. Immun. **60:**3953–3961.
6. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
7. **O'Reilly CE, Jaron P, Ochieng B, Nyaguara A, Tate JE, Parsons MB, Bopp CA, Williams KA, Vinje J, Blanton E, Wannemuehler KA, Vulule J, Laserson KF, Breiman RF, Feikin DR, Widdowson MA, Mintz E.** 2012. Risk factors for death among children less than 5 years old hospitalized with diarrhea in rural western Kenya, 2005–2007: a cohort study. PLoS Med. **9:**e1001256. http://dx.doi.org/10.1371/journal.pmed.1001256.
8. **Tenaillon O, Skurnik D, Picard B, Denamur E.** 2010. The population genetics of commensal *Escherichia coli*. Nat. Rev. Microbiol. **8:**207–217. http://dx.doi.org/10.1038/nrmicro2298.
9. **Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbonnelle E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S.** 2008. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. BMC Genomics **9:**560. http://dx.doi.org/10.1186/1471-2164-9-560.
10. **Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G, Hayashi T, Parkhill J, Frankel G.** 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J. Bacteriol. **191:**347–354. http://dx.doi.org/10.1128/JB.01238-08.
11. **Hazen TH, Sahl JW, Fraser CM, Donnenberg MS, Scheutz F, Rasko DA.** 2013. Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **110:**12810–12815. http://dx.doi.org/10.1073/pnas.1306836110.
12. **Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J.** 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J. Bacteriol. **190:**6881–6893. http://dx.doi.org/10.1128/JB.00619-08.