Draft Genome Sequences of the Diarrheagenic _Escherichia coli_ Collection

Tracy H. Hazen, Jason W. Sahl, Julia C. Redman, Carolyn R. Morris, Sean C. Daugherty, Marcus C. Chibucos, Naomi A. Sengamalay, Claire M. Fraser-Liggett, Hans Steinsland, Thomas S. Whittam, Beth Whittam, Shannon D. Manning, and David A. Rasko

Department of Microbiology and Immunology and Institute for Genome Sciences; University of Maryland School of Medicine, Baltimore, Maryland, USA; University of Maryland School of Medicine, Department of Medicine, Baltimore, Maryland, USA; Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA; and Centre for International Health and Department of Biomedicine, University of Bergen, Bergen, Norway

We report the draft genome sequences of the collection referred to as the Escherichia coli DECA collection, which was assembled to contain representative isolates of the 15 most common diarrheagenic clones in humans (http://shigatox.net/new/). These genomes represent a valuable resource to the community of researchers who examine these enteric pathogens.

The most comprehensive diarrheal studies indicate that there are greater than 110 million cases of diarrhea in children under 5 each year (3) and approximately 2 million people die each year as a direct result of diarrheal disease; a large proportion of these are children. While many assume this a problem only for the developing world, the NIDDK indicates that the rate of diarrhea among the U.S. population is 100% per year; i.e., each person in the United States contracts diarrhea at least once each year (http://www.niddk.nih.gov/). The primary bacterial pathogens that contribute to diarrheal disease are _Escherichia coli_ and _Shigella_ species. Recent food-borne outbreaks attributable to both _E. coli_ and _Shigella_ illustrate that these diarrheal pathogens also constitute a significant public health problem in the developed world (1, 8, 11). While genome sequencing is entering a phase where rapid sequencing will become part of the normal clinical diagnostic paradigm, it has been demonstrated that adequate and reliable reference genomes are required for useful comparative studies (6, 9, 14).

The collection of isolates in this announcement represent the dominant clonal types of diarrheagenic _E. coli_ and have been used in innumerous studies to highlight the diversity among _E. coli_ isolates. Each of these isolates has been examined using multilocus sequencing typing schema (10), which was confirmed with each of the draft genome sequences generated in this project. The generation of these genomes allows the direct comparison of housekeeping gene typing schema with the large-scale genome phylogeny methods that are evolving (2, 4, 8, 12).

Genomic DNA was isolated from an overnight culture using the Sigma GenElute kit (Sigma-Aldrich) and was sequenced at the Institute for Genome Sciences, Genome Resource Center (http://www.igs.umaryland.edu/). The genome data have been deposited in GenBank with accession numbers AIEV00000000 to AIEZ00000000, AIFA00000000 to AIFZ00000000, AIGA00000000 to AIGZ00000000, and AIHA00000000 to AIHS00000000. Please see http://gscid.igms.umd.edu/wp.php?wp=emerging_diarrheal_pathogens to match the GenBank accession number to the isolate of interest.

ACKNOWLEDGMENT

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 number HHSN27220090009C.

REFERENCES


Received 17 March 2012 Accepted 27 March 2012
Address correspondence to David A. Rasko, drasko@som.umaryland.edu.
* Present address: Translational Genomics Research Institute, Flagstaff, Arizona, USA.
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JB.00426-12


