AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of an *Escherichia coli* Strain Isolated from Laboratory Mouse Stool for Use as a Chassis for Transgene Delivery to the Murine Microbiome

Nicole Siguenza,[a] Baylee J. Russell,[a] R. Alexander Richter,[a] Amir Zarrinpar[a,b,c]

[a]Division of Gastroenterology, University of California, San Diego, La Jolla, California, USA
[b]VA Health Sciences San Diego, La Jolla, California, USA
[c]Center for Microbiome Innovation, University of California, San Diego, La Jolla, California, USA

Nicole Siguenza and Baylee J. Russell contributed equally to this work. Author order was determined randomly.

ABSTRACT Tools to explore functional changes in the microbiome are limited. Here, we report the complete genome sequence of a strain of *Escherichia coli* that was isolated from murine stool. This sequence will provide essential information to further develop this tool, and similar tools, to explore the complex murine microbiome.

As demonstrated recently, a *Escherichia coli* strain derived from a conventional murine microbiome can be engineered to produce a function of interest and reintroduced to the microbiome to induce physiological change (1). This provides an essential tool to probe the impact of genes of interest on the microbiome and the murine host (2–4). Here, we provide the complete genome of the mouse-derived *E. coli* strain described in the cited study (1).

To isolate this bacterium, stool was collected from a C57BL/6 male mouse that had been acquired from the Jackson Laboratory (Bar Harbor, ME) and was suspended in sterile deionized water. The stool was homogenized for 2 min at 3,500 rpm in a Mini-Beadbeater-24 (Biospec, Bartlesville, OK). The homogenized stool sample was plated on MacConkey agar containing lactose. Potential *E. coli* colonies were selected based on colony shape and the ability to ferment lactose, as indicated by the MacConkey agar. The colonies were struck out for isolation twice, and one isolate was used for further identification and use (termed EcAZ-1 in the cited study [1]). The isolate was stored at −80°C in 25% glycerol. Further culturing was performed in LB and/or Super Optimal Broth (SOB) medium.

Genomic DNA was extracted using the Invitrogen PureLink genomic DNA minikit (Thermo Fisher Scientific, Carlsbad, CA), following the manufacturer's protocol. The concentration and quality of DNA were determined using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The DNA was neither sheared nor size selected. A high-molecular-weight library was generated using the DNA template preparation kit v3.0 (Pacific Biosciences [PacBio], Menlo Park, CA). The DNA was sequenced on a PacBio GS2 system and base called using basecaller v1 (PacBio), which resulted in 623,841,140 bases in 70,098 reads. Additionally, a library was generated using TruSeq HT barcodes (Illumina, San Diego, CA), and 2 × 151-bp paired-end reads were sequenced on an Illumina MiSeq system and base called with bcl2fastq v1.8.4 (Illumina), which resulted in 8,672,400 bases in 28,908 read pairs. Reads were cleaned and adapter trimmed using fastp v0.23.2 (5). The PacBio reads were unfiltered, ranging in length from 35 bp to 43,696 bp, with an N_{50} value of 18,790 bp. PacBio reads were assembled using Unicycler v0.5.0 (6), which circularizes and rotates the assembled chromosomes using *dnaA* or *recA* as the circularization gene. The assembly was then

Editor Vanja Klepac-Ceraj, Wellesley College

Address correspondence to Amir Zarrinpar, azarrinpar@ucsd.edu.

The authors declare a conflict of interest. A.Z. is a cofounder, the acting chief medical officer, and an equity-holder for Endure Biotherapeutics.

Received 21 October 2022
Accepted 12 February 2023

polished through five rounds of mapping of the Illumina reads to the assembly using BWA v0.7.17 (7), followed by individual base call correction with Pilon v1.24 (8); no correction for multiple hits were needed to be performed.

This process resulted in three closed molecules, namely, a primary chromosome of 5,011,906 bases (G+C content of 51%), one plasmid of 42,565 bases (G+C content of 41%), and another plasmid of 8,571 bases (G+C content of 47%). The assembly was then annotated using the Prokka v14.6 pipeline (9), with *E. coli* as the assigned species and with Pfam-A v34 (10) and UniProt/Swiss-Prot (downloaded 16 June 2021) (11) as additional annotation sources. Default parameters were used for all software unless otherwise noted.

Data availability. The complete, annotated genome sequence of EcAZ-1 has been deposited at the European Bioinformatics Institute under the accession numbers OX341604.1 (chromosome), OX341605.1 (plasmid 1), and OX341606.1 (plasmid 2). The SRA accession numbers are ERR10187966 (Illumina sequencing) and ERR10187964 (PacBio sequencing).

ACKNOWLEDGMENTS

N.S. is supported by a Biolegend Fellowship and the National Science Foundation Graduate Research Fellowship Program (grant 1000340660). A.Z. is supported by a VA Merit BLR&D Award (grant I01 BX005707) and NIH grants K08 DK102902, R03 DK114536, R01 HL148801, R01 EB030134, and U01 CA265719. All authors receive institutional support from NIH grants P30 DK120515, P30 DK063491, P30 CA014195, P50 AA011999, and UL1 TR001442. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the U.S. Government. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

A.Z. is a cofounder, the acting chief medical officer, and an equity-holder for Endure Biotherapeutics.

REFERENCES

1. Russell BJ, Brown SD, Siguenza N, Mai I, Saran AR, Lingaraju A, Maissy ES, Dantas Machado AC, Pinto AFM, Sanchez C, Rossitto L-A, Miyamoto Y, Richter RA, Ho SB, Eckmann L, Hasty J, Gonzalez DJ, Saghatelian A, Knight R, Zarrinpar A. 2022. Intestinal transgene delivery with native *E. coli* chassis allows persistent physiological changes. Cell 185:3263–3277.e15. https://doi.org/10.1016/j.cell.2022.06.050.
2. Walker RL, Owen RL. 1990. Intestinal barriers to bacteria and their toxins. Annu Rev Med 41:393–400. https://doi.org/10.1146/annurev.me.41.020190.002141.
3. Claesen J, Fischbach MA. 2015. Synthetic microbes as drug delivery systems. ACS Synth Biol 4:358–364. https://doi.org/10.1021/sb500258b.
4. Pedrolli DB, Ribeiro NV, Squizato PN, de Jesus VN, Cozetto DA, Tuma RB, Gracindo A, Cesar MB, Freire PJC, da Costa AFM, Lins MRCR, Correa GG, Cerri MO, Team AQA Unesp at iGEM 2017. 2019. Engineering microbial living therapeutics: the synthetic biology toolbox. Trends Biotechnol 37:100–115. https://doi.org/10.1016/j.tibtech.2018.09.005.
5. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
6. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
7. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997. https://doi.org/10.48550/ARXIV.1303.3997.
8. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
9. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
10. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432. https://doi.org/10.1093/nar/gky995.
11. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. 2007. UniProtKB/Swiss-Prot, p 89–112. *In* Edwards D (ed), Plant bioinformatics. Humana Press, Totowa, NJ.