AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Draft Genome Sequences of Five Novel Polyketide Synthetase-Containing Mouse *Escherichia coli* Strains

Anthony Mannion, Zeli Shen, Yan Feng, Alexis Garcia,* James G. Fox

Division of Comparative Medicine, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

* Present address: Alexis Garcia, Molecular Sciences Research Center, University of Puerto Rico, San Juan, Puerto Rico, USA.

We report herein the draft genomes of five novel *Escherichia coli* strains isolated from surveillance and experimental mice housed at MIT and the Whitehead Institute and describe their genomic characteristics in context with the polyketide synthetase (PKS)-containing pathogenic *E. coli* strains NC101, IHE3034, and A192PP.

Received 12 August 2016 **Accepted** 18 August 2016 **Published** 6 October 2016

Citation Mannion A, Shen Z, Feng Y, Garcia A, Fox JG. 2016. Draft genome sequences of five novel polyketide synthetase-containing mouse *Escherichia coli* strains. Genome Announc 4(5):e01082-16. doi:10.1128/genomeA.01082-16.

Copyright © 2016 Mannion et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to James G. Fox, jgfox@mit.edu.

Escherichia coli is typically considered a harmless commensal bacteria colonizing the mammalian intestine. However, infection by numerous strains from the B2 phylogroup can cause septicemia/bacteremia, urinary tract infection, and meningitis and are increasingly being linked to inflammatory bowel disease and colorectal tumors in humans (1–3). These strains commonly harbor an ~54-kb polyketide synthetase (PKS) pathogenicity island that produces a genotoxin called colibactin (4, 5). Recently, colibactin expression by the mouse commensal *E. coli* strain NC101 was shown to potentiate colitis-associated colorectal cancer in IL-10$^{-/-}$ mice (2). Furthermore, aside from causing double-strand DNA breaks, colibactin facilitates intestinal colonization and systemic translocation that causes septicemia and meningitis (6). Thus, the virulence and oncogenic modalities of colibactin-producing *E. coli* strains in experimental mice may represent a potential confounder of results, especially in models of inflammatory disease and cancer.

In this report, we announce the draft genomes of five novel B2-phylotype *E. coli* strains isolated from the feces of two asymptomatic surveillance mice, the uterine fluid of a transgenic mouse, and from the blood of two experimental immunocompromised mice with sepsis housed under specific-pathogen-free conditions. We previously found all these strains except A4 were PCR-positive for the PKS genes clbA and clbQ and caused megalocytosis and γ-H2AX-detectable DNA damage in HeLa cells *in vitro*, suggesting colibactin activity by these organisms ([7]; our unpublished data). We briefly describe the features of these novel *E. coli* genomes in context to the previously sequenced genomes of NC101 and the PKS-containing septicemia- and meningitis-causing strains IHE3034 and A192PP (4, 6).

Genomes were sequenced using PacBio RSII and resulting sequencing data were assembled into two to five contigs using RS_HGAP_Assembly.3 from the SMRT Portal 2.3. Contigs were annotated using the RAST tool kit (RASTtk) (8). Novel strains have genomic sizes, G+C content, protein-coding sequences, and RNAs comparable to those of the genomes of NC101 (GenBank: NZ_AEFA00000000.1), IHE3034 (GenBank: NC_017628.1), and A192PP (GenBank: NZ_CVOH00000000.1) (Table 1). All novel genomes except A4 contained complete PKS islands, with all genes having ≥99% sequence identity and coverage as well as identical synteny to those from NC101, IHE3034, and A192PP.

PATRIC's proteome comparison service was used to determine the number of homologous genes between NC101, IHE3034, and A192PP as references versus the novel *E. coli* strains (9). NC101 and the novel PKS-containing strains shared >4,780 proteins (>97%) as homologs, whereas between NC101 and A4, 4,732 proteins (~89%) were homologous. All novel *E. coli* strain shared considerably less homologs to IHE3034 (4,386 to 4,525 proteins; ~87 to 90%) and A192PP (4,472 to 4,624; ~80 to 83%).

Aside from PKS, the virulence factors enterobactin sidero-

TABLE 1 Genomic characteristics of novel mouse *E. coli* strains

Strain	GenBank accession no.	Origin of mouse isolates	Genome size (bp)	No. of contigs	Fold coverage	G+C content (%)	No. of protein-coding sequences	No. of RNA sequences
E. coli 1409150006 (MIT A2)	MBNU00000000	Feces	5,292,961	3	114.7	50.5	5,203	108
E. coli 1409160003 (MIT A4)	MBNV00000000	Feces	5,066,411	4	166.4	50.5	5,003	114
E. coli 1408270010 (MIT A21)	MBNW00000000	Uterine fluid	5,318,023	5	129.9	50.5	5,280	108
E. coli 1512290008 (Whitehead Institute 0008)	MBNX00000000	Blood	5,289,663	3	147.7	50.5	5,203	108
E. coli 1512290026 (Whitehead Institute 0026)	MBNY00000000	Blood	5,005,444	2	140.6	50.6	4,876	108

Downloaded from http://mra.asm.org/ on September 17, 2019 by guest

phore receptor protein, s-fimbriae minor subunit, glutamate decarboxylase, per-activated serine protease autotransporter enterotoxin, and bor protein precursor were identified in NC101 and the novel PKS-containing strains using VirulenceFinder 1.5 (10). A4 also contained glutamate decarboxylase, per-activated serine protease autotransporter enterotoxin, and bor protein precursor as well as serine protease autotransporters of *Enterobacteriaceae*.

Together, these genomic data suggest that the novel PKS-containing strains are more similar to NC101 than IHE3034 and A192PP. Future studies will investigate the pathogenic and carcinogenic potential of these and other PKS-containing *E. coli* strains in mice under differing experimental protocols.

Accession number(s). The genome sequences have been submitted to GenBank under the accession numbers listed in Table 1.

ACKNOWLEDGMENTS

We thank the BioMicro Center at MIT and the Deep Sequencing Core Facility at the University of Massachusetts Medical School for assistance with PacBio sequencing.

This work was supported by P30ES002109, T32OD010978, and R010D0111141 grants.

FUNDING INFORMATION

This work, including the efforts of James G. Fox, was funded by HHS | National Institutes of Health (NIH) (P30-ES002109, T32-OD010978, and R0-10D0111141).

REFERENCES

1. **Johnson JR, Gajewski A, Lesse AJ, Russo TA**. 2003. Extraintestinal pathogenic Escherichia coli as a cause of invasive nonurinary infections. J Clin Microbiol **41:**5798–5802. http://dx.doi.org/10.1128/JCM.41.12.5798-5802.2003.
2. **Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, Campbell BJ, Abujamel T, Dogan B, Rogers AB, Rhodes JM, Stintzi A, Simpson KW, Hansen JJ, Keku TO, Fodor AA, Jobin C**. 2012. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science **338:**120–123. http://dx.doi.org/10.1126/science.1224820.
3. **Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, Pezet D, Bonnet R**. 2013. High prevalence of mucosa-associated E. coli Producing Cyclomodulin and genotoxin in Colón cancer. PLoS One **8:**e56964. http://dx.doi.org/10.1371/journal.pone.0056964.
4. **Nougayrède JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, Oswald E**. 2006. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. Science **313:**848–851. http://dx.doi.org/10.1126/science.1127059.
5. **Putze J, Hennequin C, Nougayrède JP, Zhang W, Homburg S, Karch H, Bringer MA, Fayolle C, Carniel E, Rabsch W, Oelschlaeger TA, Oswald E, Forestier C, Hacker J, Dobrindt U**. 2009. Genetic structure and distribution of the colibactin genomic island among members of the family *Enterobacteriaceae*. Infect Immun **77:**4696–4703. http://dx.doi.org/10.1128/IAI.00522-09.
6. **McCarthy AJ, Martin P, Cloup E, Stabler RA, Oswald E, Taylor PW**. 2015. The genotoxin Colibactin Is a determinant of virulence in *Escherichia coli* K1 Experimental neonatal systemic infection. Infect Immun **83:**3704–3711. http://dx.doi.org/10.1128/IAI.00716-15.
7. **García A, Mannion A, Feng Y, Madden CM, Bakthavatchalu V, Shen Z, Ge Z, Fox JG**. 2016. Cytotoxic *Escherichia coli* strains encoding colibactin colonize laboratory mice. Microbes Infect, in press. http://dx.doi.org/10.1016/j.micinf.2016.07.005.
8. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R**. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res **42:**D206–D214. http://dx.doi.org/10.1093/nar/gkt1226.
9. **Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJ, Yoo HS, Zhang C, Zhang Y, Sobral BW**. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res **42:**D581–D591. http://dx.doi.org/10.1093/nar/gkt1099.
10. **Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM**. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol **52:**1501–1510. http://dx.doi.org/10.1128/JCM.03617-13.

Downloaded from http://mra.asm.org/ on September 17, 2019 by guest