AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Human Microbiome | Announcement

Draft genome of a human-derived *pks*+ *E. coli* that caused spontaneous disseminated infection in a mouse

Allison M. Weis,[1,2] O'Connor J. Matthews,[3,4] Matthew A. Mulvey,[3,4] June L. Round[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We present the draft genome of a novel human-derived *Escherichia coli* strain isolated from a healthy control human microbiota that, when put into a mouse, spontaneously disseminated from the gut to the kidneys.

KEYWORDS *E. coli*, gastrointestinal tract, genome, pathogen

Escherichia coli is a common gut commensal as well as a devastating and persistent pathogen. Pathogenic strains of *E. coli* can cause myriad types of infections, including diarrhea, urinary tract infections, bacteremia, and sepsis (1–3). Furthermore, *E. coli* strains that carry the *pks*+ gene cluster, encoding the secreted DNA-damaging toxin colibactin, can promote tumor development and are associated with colorectal cancer (4–6). Here, we present the draft genome of a novel human-derived strain isolated from a mouse after spontaneous dissemination from the gut.

This *E. coli* strain was recovered from the kidneys of an ex-germ-free Swiss Webster mouse, raised in house, which had undergone a human-to-mouse fecal microbial transplant in the Round lab at the University of Utah, IACUC protocol 00001562. The human sample was from the ColoCare study (7). The mouse was discovered near death, and upon necropsy, both kidneys were red and enlarged. Kidney homogenates were plated on LB agar and were incubated aerobically at 37°C overnight. The kidneys contained numerous bacteria, all seemingly the same type, which were then streaked to isolation. Plating on McConkey agar led to brilliant pink hues of the colonies, indicative of *E. coli*. We named the isolate AW001.

DNA was extracted from a pure culture of AW001 grown overnight in LB broth at 37°C in the Mulvey lab using the Qiagen DNeasy Blood & Tissue kit, and libraries were made using Tecan Ultralow V2, both according to the manufacturer protocols. Sequencing was performed by Illumina NovaSEQ6000 at the University of Colorado Anschutz with 151 bp paired-end reads. The total read count was 8,779,520, with an average depth of 257.79. Software default parameters were used, except where noted. Read preprocessing was performed using Trim Galore v0.6.5dev (8). Sequences were assembled *de novo* using Unicycler v.0.4.8 through BV-BRC v3.35.5 using -t 12 –min_fasta_length 300 –keep 2 –no_pilon (9–11). The assembled genome contains 109 contigs with an N50 value of 268,981. The genome was annotated using Genbank PGAP v6.6 (12). CheckM v1.0.5 revealed completeness of 99.97 (13).

The assembled AW001 genome contains 5,008,932 bps, comprising 4,952 genes, 4,868 CDS, and 50.7% GC content. AW001 was predicted to be *E. coli* phylogroup B2 using ClermonTyping (v23.06) (14). By MLST, it was ST "unknown," closely related STs being 12998 and 2831 (MLST-2.0 Server). Using AMRFinderPlus v3.11.26 (15), putative antimicrobial resistance genes were identified including fosfomycin-resistant *glpT* variant (16), multidrug-resistance *marR*, colistin-resistant *pmrB* (17), efflux pump acrF (18), and beta-lactamase *blaEC* (19) (Table 1). This indicates that AW001 is likely resistant to multiple classes of clinically relevant antibiotics.

Editor Vanja Klepac-Ceraj, Department of Biological Sciences, Wellesley College, Wellesley, Massachusetts, USA

Address correspondence to June L. Round, June.round@path.utah.edu.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 12 April 2024
Accepted 2 May 2024
Published 4 June 2024

TABLE 1 Key AMR and virulence factors identified in the AW001 *E. coli* genome

Gene(s)	Protein ID	Type	Protein name and function
acrF	WP_001273251.1	AMR	AcrF multidrug efflux RND transporter permease
blaEC	WP_001556381.1	AMR	BlaEC family class C beta-lactamase
glpT	WP_000948732.1	AMR	GlpT fosfomycin resistant
marR	WP_000799375.1	AMR	MarR multidrug resistant
pmrB	WP_001052123.1	AMR	PmrB colistin resistant
ariR	WP_000888771.1	Stress	Biofilm/acid-resistance regulator AriR
emrE	WP_001070440.1	Stress	EmrE multidrug efflux SMR transporter
astA	WP_000989438.1	Virulence	EAST1 heat-stable enterotoxin
cbtA	WP_000854814.1	Virulence	CbtA type IV toxin-antitoxin system
ccdA	WP_000125566.1	Virulence	CcdA type II toxin-antitoxin system CcdA
chuA	WP_000089583.1	Virulence	ChuA outer membrane hemin receptor
ClbA	WP_001217110.1	Virulence	Colibactin synthesis proteins (*pks*)
entH	WP_000637953.1	Virulence	EntH proofreading thioesterase
fdeC	WP_000092543.1	Virulence	FdeC inverse autotransporter adhesin
fimH	WP_000832236.1	Virulence	SfaH fimbrial protein subunit
fliP	WP_334615852.1	Virulence	FliP flagellar type III secretion system
gspA	WP_000107592.1	Virulence	GspA,C,D type II secretion system
gspG	WP_001087296.1	Virulence	GspG type II secretion system major pseudopilin
hcp	WP_000458845.1	Virulence	Hcp family type VI secretion system effector
hecB	WP_334616364.1	Virulence	HecB family hemolysin secretion/activation protein
hhA	WP_001333231.1	Virulence	Hha hemolysin expression modulator
ibsE	WP_001387082.1	Virulence	Ibs family toxin type I toxin-antitoxin system
Irp1,2	WP_000369530.1	Virulence	Yersiniabactin siderophore
iss	WP_001298464.1	Virulence	Iss increased serum survival lipoprotein
ldrD	WP_001295224.1	Virulence	Ldr family protein type I toxin-antitoxin system
mchB	WP_001375214.1	Virulence	H47 microcin
mchF	WP_001518504.1	Virulence	MchF microcin H47 export transporter peptidase
neuC	WP_000723250.1	Virulence	Polysialic acid biosynthesis protein P7
ompA	WP_001518466.1	Virulence	Outer membrane protein
paeA	WP_000935036.1	Virulence	Hemolysin family protein
sitC	WP_001101732.1	Virulence	MntB manganese transport membrane protein
sslE	WP_001034565.1	Virulence	SslE lipoprotein metalloprotease
TssE	WP_000106967.1	Virulence	TssE type VI secretion system baseplate subunit
tssJ	WP_000484008.1	Virulence	TssJ type VI secretion system lipoprotein
vgrG	WP_001350146.1	Virulence	VgrG type VI secretion system tip protein
ybtE	WP_001518699.1	Virulence	2,3-dihydroxybenzoate-AMP ligase
ybtP	WP_001327262.1	Virulence	YbtP yersiniabactin ABC transporter
ybtQ	WP_001295637.1	Virulence	YbtQ yersiniabactin ABC transporter
yqfA	WP_000250274.1	Virulence	Hemolysin III family protein

AW001 was analyzed for putative fitness and virulence factors using AMRFinderPlus, BLAST+ (v2.14.0+), and targeted searching (20). Like *E. coli* strain NC101 (21), AW001 carries the *clb (pks)* gene cluster that produces colibactin. AW001 contains genes involved in iron acquisition, adhesion, capsule biosynthesis, and mucin degradation (Table 1). AW001 also encodes toxin-antitoxin systems; enterotoxin AstA; antimicrobial microcin H47; type II, III, and VI secretion systems; and hemolysin biosynthesis genes (Table 1).

ACKNOWLEDGMENTS

Thanks to the Round and Mulvey labs for help with isolation, identification, and sequencing of this isolate.

This work was funded by R01AT011423-03, a W. M. Keck Award, by a Burrough's Welcome grant to J.L.R., and by R01GM134331 and Department of Defense award W81XWH-22-1-0800 (SC210103) to M.A.M. A.M.W. was funded by NIH NCI NRSA F32CA243501, and O.J.M. was supported by NIH Genetics T32 training, grant no. GM007464.

AUTHOR AFFILIATIONS

[1]Department of Pathology, Division of Microbiology and Immunology, University of Utah School of Medicine, Salt Lake City, Utah, USA
[2]Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, USA
[3]School of Biological Sciences, University of Utah, Salt Lake City, Utah, USA
[4]Henry Eyring Center for Cell & Genome Science, University of Utah, Salt Lake City, Utah, USA

AUTHOR ORCIDs

Allison M. Weis http://orcid.org/0000-0001-9991-1661
June L. Round http://orcid.org/0000-0002-7158-9874

FUNDING

Funder	Grant(s)	Author(s)
HHS | NIH | National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)	1R01DK124317	June L. Round
HHS | NIH | National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)	R01AT011423	June L. Round
W. M. Keck Foundation (WMKF)		June L. Round

DATA AVAILABILITY

Data are in GenBank under accession JAYWIW000000000. Raw reads are under SRR28249370.

REFERENCES

1. Mulvey MA. 2002. Adhesion and entry of uropathogenic *Escherichia coli*. Cell Microbiol 4:257–271. https://doi.org/10.1046/j.1462-5822.2002.00193.x
2. Lewis AJ, Richards AC, Mulvey MA. 2016. Invasion of host cells and tissues by uropathogenic bacteria. Microbiol Spectr 4. https://doi.org/10.1128/microbiolspec.UTI-0026-2016
3. Kaper JB, Nataro JP, Mobley HLT. 2004. Pathogenic *Escherichia coli.* Nat Rev Microbiol 2:123–140. https://doi.org/10.1038/nrmicro818
4. Dejea CM, Fathi P, Craig JM, Boleij A, Taddese R, Geis AL, Wu X, DeStefano Shields CE, Hechenbleikner EM, Huso DL, Anders RA, Giardiello FM, Wick EC, Wang H, Wu S, Pardoll DM, Housseau F, Sears CL. 2018. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science 359:592–597. https://doi.org/10.1126/science.aah3648
5. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, Gurjao C, Manders F, Dalmasso G, Stege PB, et al. 2020. Mutational signature in colorectal cancer caused by genotoxic *pks+ E. coli*. Nature 580:269–273. https://doi.org/10.1038/s41586-020-2080-8
6. Wilson MR, Jiang Y, Villalta PW, Stornetta A, Boudreau PD, Carrá A, Brennan CA, Chun E, Ngo L, Samson LD, Engelward BP, Garrett WS, Balbo S, Balskus EP. 2019. The human gut bacterial genotoxin colibactin alkylates DNA. Science 363:eaar7785. https://doi.org/10.1126/science.aar7785
7. Ulrich CM, Gigic B, Böhm J, Ose J, Viskochil R, Schneider M, Colditz GA, Figueiredo JC, Grady WM, Li CI, Shibata D, Siegel EM, Toriola AT, Ulrich A. 2019. The ColoCare study: a paradigm of transdisciplinary science in colorectal cancer outcomes. Cancer Epidemiol Biomarkers Prev 28:591–601. https://doi.org/10.1158/1055-9965.EPI-18-0773
8. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
9. 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
10. 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
11. Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, Nordberg EK, Scott M, Schulman JR, Snyder EE, Sullivan DE, Wang C, Warren A, Williams KP, Xue T, Seung Yoo H, Zhang C, Zhang Y, Will R, Kenyon RW, Sobral BW. 2011. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. Infect Immun 79:4286–4298. https://doi.org/10.1128/IAI.00207-11
12. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569
13. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from

isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114
14. Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. 2018. ClermonTyping: an easy-to-use and accurate *in silico* method for *Escherichia* genus strain phylotyping. Microb Genom 4:e000192. https://doi.org/10.1099/mgen.0.000192
15. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M, Pettengill JB, Prasad AB, Tillman GE, Tyson GH, Klimke W. 2021. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep 11:12728. https://doi.org/10.1038/s41598-021-91456-0
16. Mowlaboccus S, Daley DA, Birdsall J, Gottlieb T, Merlino J, Nimmo GR, George N, Korman T, Streitberg R, Robson J, Peachey G, Collignon P, Bradbury S, Rogers BA, Coombs GW. 2021. Molecular characterization of fosfomycin-resistant *Escherichia coli* urinary tract infection isolates from Australia. Clin Microbiol Infect 27:1360–1361. https://doi.org/10.1016/j.cmi.2021.04.029
17. Phan M-D, Nhu NTK, Achard MES, Forde BM, Hong KW, Chong TM, Yin W-F, Chan K-G, West NP, Walker MJ, Paterson DL, Beatson SA, Schembri MA. 2017. Modifications in the pmrB gene are the primary mechanism for the development of chromosomally encoded resistance to polymyxins in uropathogenic *Escherichia coli*. J Antimicrob Chemother 72:2729–2736. https://doi.org/10.1093/jac/dkx204
18. Gaurav A, Bakht P, Saini M, Pandey S, Pathania R. 2023. Role of bacterial efflux pumps in antibiotic resistance, virulence, and strategies to discover novel efflux pump inhibitors. Microbiology (Reading) 169:001333. https://doi.org/10.1099/mic.0.001333
19. Schmidt J, Zdarska V, Kolar M, Mlynarcik P. 2023. Analysis of BlaEC family class C beta-lactamase. FEMS Microbiol Lett 370:fnad097. https://doi.org/10.1093/femsle/fnad097
20. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421
21. 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 18:777–786. https://doi.org/10.1016/j.micinf.2016.07.005