



Draft Genome Sequences of 27 Northern Maine Clinical Isolates

 Larry Feinstein,^{a*} Kendra Batchelder,^a Lydia Tilley,^{a,b} Grace Stafford^c

^aUniversity of Maine at Presque Isle, Presque Isle, Maine, USA

^bMount Desert Island Biological Laboratory, Murawala Lab, Mount Desert Island, Maine, USA

^cThe Jackson Laboratory, Computational Sciences, Mount Desert Island, Maine, USA

ABSTRACT We report the draft genome sequences of 27 common pathogens collected from a northern Maine hospital in 2017. These were sequenced in order to determine temporal and biogeographical patterns of antibiotic gene distribution. A total of 908 antibiotic resistance genes, 848 insertion sequence elements, and 57 plasmids were identified.

Antibiotic resistance genes (ARGs) contribute to a significant public health issue, with ~2.8 million antibiotic-resistant infections responsible for over 35,000 deaths annually in the United States (1). A total of 27 clinical isolates from the most common pathogens collected from the Aroostook Medical Center in Presque Isle, Maine, during winter 2017 were sequenced in order to determine the extent of ARG distribution and the mobile genetic elements that transfer them between bacteria. Institutional review board approval or ethics committee review was not required for this study. The isolate genera consisted of *Klebsiella* (8 isolates), *Escherichia* (6 isolates), *Pseudomonas* (4 isolates), *Enterobacter* (4 isolates), *Staphylococcus* (4 isolates), and *Stenotrophomonas* (1 isolate, originally identified as *Pseudomonas*). Table 1 provides additional information about each isolate sequenced.

Microbes were cultured in tryptic soy medium, and all incubations were 18 h at 37°C. We received strains on slants, and cultures were transferred to broth that was used to create streak plates. Pure cultures were obtained from isolated colonies on streak plates. Genomic DNA was extracted with an UltraClean kit (MoBio, Carlsbad, CA) and quantified with NanoDrop spectrophotometry (Fisher Scientific, Waltham, MA). Isolate cultures were stored in 25% glycerol at –80°C. The Jackson Laboratory (Bar Harbor, ME) constructed whole-genome libraries using the KAPA HyperPrep kit (Kapa Biosystems, Wilmington, MA), which sheared DNA into 150- to 350-bp fragments. Libraries were pooled and sequenced on a NextSeq sequencer (Illumina, San Diego, CA). Sequences were 2 × 150 bp with 100× mean coverage per sample.

All software utilized default parameters unless otherwise specified. FastQC 0.11.3 (2) was used for sequence quality control, and sequences were trimmed with Trimmomatic 0.33 (3). The Burrows-Wheeler Aligner MEM (BWA-MEM) algorithm (4) was used to align the sequences to reference genomes obtained from the RefSeq NCBI database (5). Contigs were assembled with SPAdes 3.10.0 (6), and NCBI's Prokaryotic Genome Annotation Pipeline (7) was used for annotation. Standard features were called using GLIMMER v3.02 (8, 9) and Prodigal (10), and a scan was conducted for the following additional feature types: rRNA, tRNA, selenoproteins, pyrrolysoproteins, repeat regions, and CRISPR. The average (99.87%) and the lowest (99.65%) completeness scores were determined with CheckM (11). We used Spine (12) to identify core genomes for each species. Each isolate genome was compared to its core genome with AGenT (12) to identify a core and accessory genome for individual isolates.

Accessory and core genomes were uploaded to the Comprehensive Antibiotic Resistance Database (CARD) (13) to identify any ARGs present in the sequences. A total

Citation Feinstein L, Batchelder K, Tilley L, Stafford G. 2021. Draft genome sequences of 27 northern Maine clinical isolates. *Microbiol Resour Announc* 10:e01379-20. <https://doi.org/10.1128/MRA.01379-20>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2021 Feinstein et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Larry Feinstein, larry.feinstein@maine.edu.

* Present address: Larry Feinstein, 181 Main Street, Presque Isle, Maine, USA.

Received 1 December 2020

Accepted 14 January 2021

Published 4 February 2021

TABLE 1 Data about each isolate sequenced

Species	Strain	Infection location	Genome size (bp)	GC content (%)	No. of reads	No. of contigs	N_{50} (bp)	Assembly accession no.	Raw read accession no.
<i>Escherichia coli</i>	ECPF15	Urinary tract	4,928,364	50.52	12,575,734	124	119,476	JACJHQC000000000	SRX9264872
<i>E. coli</i>	SE6-1	Urinary tract	5,052,494	50.75	12,901,582	106	142,608	JACJHP000000000	SRX9264873
<i>E. coli</i>	HB37	Urinary tract	5,124,920	50.52	11,305,112	90	221,887	JACJHO000000000	SRX9264883
<i>E. coli</i>	ECPF5	Urinary tract	5,131,891	50.41	11,746,384	122	334,530	JACJHN000000000	SRX9264892
<i>E. coli</i>	144	Urinary tract	4,962,945	50.62	11,293,956	127	189,818	JACJHM000000000	SRX9264893
<i>E. coli</i>	WP3-S18-ESBL-09	Urinary tract	5,057,591	50.62	11,986,542	119	153,491	JACJHL000000000	SRX9264894
<i>Enterobacter hormaechei</i>	RHBSTW-00564	Urinary tract	4,813,154	55.48	12,252,196	84	246,389	JACJHK000000000	SRX9264895
<i>E. hormaechei</i>	BW	Urinary tract	4,663,914	55.45	7,378,720	27	508,613	JACJHI000000000	SRX9264897
<i>E. hormaechei</i>	RHBSTW-00564	Respiratory	4,765,656	55.43	11,114,030	81	182,399	JACJHH000000000	SRX9264898
<i>Enterobacter ludwigii</i>	L140	Urinary tract	5,039,619	54.33	10,019,726	22	546,290	JACJHJ000000000	SRX9264896
<i>Klebsiella aerogenes</i>	HNHf1	Urinary tract	5,328,186	54.95	11,193,060	35	325,541	JACJHG000000000	SRX9264874
<i>K. aerogenes</i>	HNHf1	Urinary tract	5,287,865	54.97	10,112,974	45	320,838	JACJHF000000000	SRX9264875
<i>Klebsiella pneumoniae</i>	90-17	Respiratory	5,405,495	57.27	11,438,618	59	361,870	JACJHE000000000	SRX9264876
<i>K. pneumoniae</i>	MS14393	Urinary tract	5,286,397	57.48	9,976,828	46	589,446	JACJHD000000000	SRX9264877
<i>K. pneumoniae</i>	E16KP0288	Urinary tract	5,584,709	57.23	12,749,122	113	174,417	JACJHC000000000	SRX9264878
<i>K. pneumoniae</i>	SMKP03	Urinary tract	5,595,242	57.17	11,964,910	93	237,479	JACJHB000000000	SRX9264879
<i>K. pneumoniae</i>	Kp8701	Urinary tract	5,165,081	57.56	12,430,324	45	357,399	JACJHA000000000	SRX9264880
<i>Klebsiella variicola</i>	LEMb11	Urinary tract	5,481,031	57.44	12,292,892	28	576,709	JACJGZ000000000	SRX9264881
<i>Pseudomonas aeruginosa</i>	PA0750	Urinary tract	6,775,059	66.19	10,683,770	62	374,669	JACJGY000000000	SRX9264882
<i>P. aeruginosa</i>	PA0750	Respiratory	6,392,362	66.43	12,503,800	54	373,308	JACJGW000000000	SRX9264884
<i>P. aeruginosa</i>	PA0750	Wound lesion	6,601,920	66.17	11,790,784	108	376,348	JACJGV000000000	SRX9264885
<i>P. aeruginosa</i>	PA0750	Respiratory	6,832,764	66.06	1,923,108	77	258,204	JACJGU000000000	SRX9264886
<i>Staphylococcus aureus</i>	14505	Urinary tract	2,741,077	32.75	10,826,960	30	243,925	JACJGT000000000	SRX9264887
<i>S. aureus</i>	14505	Urinary tract	2,767,079	32.67	12,991,802	55	135,119	JACJGS000000000	SRX9264888
<i>S. aureus</i>	14505	Skin lesion	2,826,084	32.60	11,951,566	27	345,651	JACJGR000000000	SRX9264889
<i>S. aureus</i>	MJ015	Blood	2,736,988	32.74	13,200,774	32	290,137	JACJGQ000000000	SRX9264890
<i>Stenotrophomonas maltophilia</i>	NCTC10498	Urinary tract	4,577,239	66.42	13,956,724	44	202,957	JACJGP000000000	SRX9264891

of 908 ARGs were identified. The interleaved raw read files were analyzed using the plasmidSPAdes pipeline in SPAdes 3.12.0 (14) to identify plasmids. The contigs given by plasmidSPAdes were uploaded to CARD to identify ARGs. A total of 57 plasmids were identified, with 10 having at least one ARG. The assembled contigs were given to ISEScan (15) to identify insertion sequence elements. The nucleotide open reading frames were annotated using blastn (16). Gene functions were obtained with Uniprot (17). These data are being utilized to quantify temporal and biogeographical ARG distribution patterns in the northeastern United States.

Data availability. The raw reads and assembled contigs are available at NCBI BioProject number [PRJNA649713](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA649713).

ACKNOWLEDGMENTS

We gratefully acknowledge Tim McNamee at the Aroostook Medical Center for supplying us with clinical isolates.

This research was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant numbers 5P20GM103423-16 and P20GM103423 and an ASM summer fellowship.

REFERENCES

- Centers for Disease Control and Prevention. 2019. More people in the United States dying from antibiotic-resistant infections than previously estimated. Centers for Disease Control and Prevention, Atlanta, GA. <https://www.cdc.gov/media/releases/2019/p1113-antibiotic-resistant.html>.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997v2. <https://arxiv.org/abs/1303.3997v2>.
- Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>.
- Nurk S. 2013. Assembling genomes and mini-metagenomes from highly chimeric reads. *In* Deng M, Jiang R, Sun F, Zhang X (ed), *Research in computational molecular biology*. RECOMB 2013. Lecture notes in computer science, vol 7821. Springer, Berlin, Germany. https://doi.org/10.1007/978-3-642-37195-0_13.
- 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>.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636–4641. <https://doi.org/10.1093/nar/27.23.4636>.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <https://doi.org/10.1093/bioinformatics/btm009>.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholz P, Tyson GW. 2015. 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>.
- Ozer EA, Allen JP, Hauser AR. 2014. Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics* 15:737. <https://doi.org/10.1186/1471-2164-15-737>.
- Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistance surveillance with the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* 48:D517–D525. <https://doi.org/10.1093/nar/gkz935>.
- Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32:3380–3387. <https://doi.org/10.1093/bioinformatics/btw493>.
- Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33:3340–3347. <https://doi.org/10.1093/bioinformatics/btx433>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–D515. <https://doi.org/10.1093/nar/gky1049>.