

# Draft Genome Sequence of Uropathogenic *Escherichia coli* Strain J96

Eric A. Klein, Zemer Gitai

Department of Molecular Biology, Princeton University, Princeton, New Jersey, USA

***Escherichia coli* J96 (O4:K6) was isolated from a human pyelonephritis patient. Here, we report the draft genome sequence of *E. coli* J96, which contains virulence genes, including adhesion factors, alpha-hemolysins, and cytotoxic necrotizing factor. J96 infects the kidney and bladder, making it an important tool for studying *E. coli* pathogenesis.**

Received 26 December 2012 Accepted 31 December 2012 Published 21 February 2013

Citation Klein EA, Gitai Z. 2013. Draft genome sequence of uropathogenic *Escherichia coli* strain J96. *Genome Announc.* 1(1):e00245-12. doi:10.1128/genomeA.00245-12.

Copyright © 2013 Klein and Gitai. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Zemer Gitai, [zgitai@princeton.edu](mailto:zgitai@princeton.edu).

Uropathogenic *Escherichia coli* (UPEC) strains are the most common bacteria that cause urinary tract infections. UPEC strains differ primarily by the presence of various virulence factors, often encoded on pathogenicity islands (PAIs). *E. coli* strain J96 was isolated from a human pyelonephritis patient (1). Several studies have characterized J96 virulence factors, including adhesins (*pap*, *prs*, *foc*, and *fim*), alpha-hemolysins (*hly*), and the cytotoxic necrotizing factor (*cnf*) (2, 3). In addition to chromosomal pathogenicity factors, J96 carries an ~115-kb plasmid (pJ96) (4). The variety of virulence genes, together with the ability of J96 to infect bladder and kidney tissues, makes this strain a valuable tool for studying UPEC pathogenesis. A previous work produced a physical-genetic map of the J96 genome (4), as well as the sequences of several of pathogenic factors (5, 6). In this report, the full chromosome and plasmid sequences were subjected to genomic analysis.

Three paired-end libraries were prepared: (i) total DNA (200 bp), (ii) total DNA (500 bp), and (iii) plasmid DNA (200 bp). The genome sequence was determined using shotgun sequencing (read length, 101 bp) on an Illumina HiSeq 2000 (library 1, 116,959,856 reads and 2,300× coverage; library 2, 85,721,146 reads and 1,690× coverage; plasmid, 30,608,372 reads and 26,800× coverage). The reads were quality-filtered and randomly sampled to achieve 100× coverage for the assembly. The plasmid sequence was assembled using Edena v3 (7), resulting in 14 contigs. Plasmid contaminants in the total DNA libraries were removed by aligning reads to the plasmid contigs using Bowtie (8). The chromosomal reads were assembled in two phases. *De novo* assembly was done with Edena v3 (7), Assembly by Short Sequences (ABYSS) v1.3.1 (9), and Velvet v1.2 (10). Reads were also aligned to UPEC strains UTI89 (accession no. NC\_007946), CFT073 (accession no. NC\_004431) (11), and EC536 (accession no. NC\_008253). The *de novo* assemblies and alignment-based contigs were merged using Gap4 (12) and scaffolded with SSAKE-based Scaffolding of Pre-Assembled Contigs after Extension (SSPACE) v2.0 (13), and gaps were PCR-amplified and sequenced by Sanger sequencing. The assembly of chromosomal sequences resulted in 262 contigs, which were annotated via the NCBI Prokaryotic Genome Annotation Pipeline (PGAAP) (14). The J96 chromosome is 5.51 Mb and its G+C content is 51.1%. pJ96 is 112 kb and its G+C content is 50.9%. The chromosome and plas-

mid contain 5,259 and 140 protein-coding sequences, respectively. The chromosome also contains 28 rRNA and 105 tRNA genes.

Phylogenetic analysis based on average nucleotide index (ANI) values (15) among *E. coli* strains showed that J96 is most similar to other UPEC strains. At least three PAIs were identified: PAI-I (*pap*), PAI-II (*hly*, *prs*, *cat-1* [chloramphenicol resistance], and *fim*), and PAI-III (*foc*). Additionally, the genome encodes numerous transposases and insertion sequences, highlighting the importance of genomic exchange in creating the J96 pathogenic phenotype.

Alignment of pJ96 to other UPEC plasmids revealed high similarity to pUTI89 (accession no. NC\_007941), though pJ96 lacked several hypothetical proteins (P018-020 and P084-085) and contained 6 single nucleotide polymorphisms (SNPs).

**Nucleotide sequence accession numbers.** *E. coli* strain J96 can be obtained from ATCC (no. 700336). This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [ALIN00000000](https://www.ncbi.nlm.nih.gov/nuccore/ALIN00000000). The version described in this paper is the first version, [ALIN02000000](https://www.ncbi.nlm.nih.gov/nuccore/ALIN02000000).

## ACKNOWLEDGMENTS

We thank Jessica Buckles, Donna Storton, Wei Wang, and Lance Parsons in the genomic sequencing core facility at Princeton University for their help in obtaining the Illumina reads.

This work was supported by an NRSA postdoctoral fellowship (no. F32GM906842) from the National Institute of General Medical Sciences to E.A.K. and a National Institutes of Health grant no. DP2OD004389 to Z.G.

## REFERENCES

1. Hull RA, Gill RE, Hsu P, Minshew BH, Falkow S. 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. *Infect. Immun.* 33:933–938.
2. Blum G, Falbo V, Caprioli A, Hacker J. 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. *FEMS Microbiol. Lett.* 126:189–195.
3. Swenson DL, Bukanov NO, Berg DE, Welch RA. 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. *Infect. Immun.* 64:3736–3743.
4. Melkerson-Watson LJ, Rode CK, Zhang L, Foxman B, Bloch CA. 2000. Integrated genomic map from uropathogenic *Escherichia coli* J96. *Infect. Immun.* 68:5933–5942.

5. Felmlee T, Pellett S, Welch RA. 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. *J. Bacteriol.* 163:94–105.
6. Marklund BI, Tennent JM, Garcia E, Hamers A, Båga M, Lindberg F, Gaastra W, Normark S. 1992. Horizontal gene transfer of the *Escherichia coli* pap and prs pili operons as a mechanism for the development of tissue-specific adhesive properties. *Mol. Microbiol.* 6:2225–2242.
7. Hernandez D, François P, Farinelli L, Osterås M, Schrenzel J. 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res.* 18:802–809.
8. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
9. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19:1117–1123.
10. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
11. Welch RA, Burland V, Plunkett G 3rd, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HL, Donnenberg MS, Blattner FR. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 99:17020–17024.
12. Bonfield JK, Smith Kf, Staden R. 1995. A new DNA sequence assembly program. *Nucleic Acids Res.* 23:4992–4999.
13. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27: 578–579.
14. Daraselia N, Dernovoy D, Tian Y, Borodovsky M, Tatusov R, Tatusova T. 2003. Reannotation of *Shewanella oneidensis* genome. *Omics* 7:171–175.
15. Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106: 19126–19131.