

Draft Genome Sequence of *Escherichia coli* Strain ATCC 23506 (Serovar O10:K5:H4)

Brady F. Cress,^a Zachary R. Greene,^a Robert J. Linhardt,^{a,b,c} Mattheos A. G. Koffas^a

Department of Chemical and Biological Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York, USA^a; Department of Chemistry, Chemical Biology, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York, USA^b; Department of Biology, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York, USA^c

We report the 5.101-Mbp high-quality draft assembly of the *Escherichia coli* strain ATCC 23506 (serovar O10:K5:H4, also known as NCDC Bi 8337-41) genome. This uropathogenic strain, commonly referred to as *E. coli* K5, produces *N*-acetyl heparosan, a glycosaminoglycan-like capsular polysaccharide and precursor to the anticoagulant pharmaceutical heparin. Metabolic reconstruction of this genome will enable the prediction of gene deletions and overexpressions that lead to increased heparosan production.

Received 21 January 2013 Accepted 23 January 2013 Published 28 February 2013

Citation Cress BF, Greene ZR, Linhardt RJ, Koffas MAG. 2013. Draft genome sequence of *Escherichia coli* strain ATCC 23506 (serovar O10:K5:H4). *Genome Announc.* 1(2):e00049-13. doi:10.1128/genomeA.00049-13.

Copyright © 2013 Cress et al. 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 Mattheos A. G. Koffas, koffam@rpi.edu.

Escherichia coli is the most well-characterized organism commonly utilized in metabolic engineering research. As metabolic engineers seek to further increase the innate production capacities of unique *E. coli* strains, it is valuable to develop genome-scale reconstructions (GSR) of metabolism in order to accurately predict genetic engineering interventions that lead to an improved phenotype. The K5 capsule is composed of *N*-acetyl heparosan, a group II capsular polysaccharide (CPS) consisting of a repeating [\rightarrow 4] β -D-glucuronic acid (GlcA) (1 \rightarrow 4) *N*-acetyl- α -D-glucosamine (GlcNAc) (1 \rightarrow)_n disaccharide unit (1). Although the gene cluster encoding the enzymes required for the biosynthesis of K5 CPS has been characterized elsewhere, annotation of the whole genome sequence will lend further insight into the molecular mechanisms of capsular polysaccharide biosynthesis and transport. The characterization of all genes involved in lipopolysaccharide (LPS) biosynthesis will also enhance understanding of CPS-LPS interactions, while comparative genomic studies between this uropathogenic *E. coli* (UPEC) strain and nonpathogenic strains might identify the virulence factors required for infection of the urinary tract.

Genomic DNA was purified from *E. coli* strain ATCC 23506 with an Invitrogen PureLink Genomic DNA mini kit. The genome was sequenced using the Illumina HiSeq 2000 sequencing system, which produced 104 M paired-end reads of 101 bp, with an insert size of 400 bp. Approximately 28M random reads were assembled with Velvet v1.2.07 (2) at an optimal hash length of 93. The final genome assembly has approximately 38-fold coverage and contains 190 supercontigs composed of 224 contigs (>200 bp in length) with a total size of 5,101,025 bp, an N₅₀ contig length of 129,677 nucleotides, and a mean G+C content of 50.6%. Assembly data were deposited in the EMBL nucleotide sequence database.

The draft genome was annotated by the Rapid Annotations using Subsystems Technology (RAST) server (3) using Glimmer3 as a gene caller (4), which predicted 5,030 coding sequences (CDSs) with an average length of 880 bp (3,815 CDSs have func-

tional predictions), 86 tRNA-encoding genes, and 25 rRNA-encoding genes. RAST was also used to construct a draft metabolic model (5) containing 1,156 genes, corresponding to 1,408 reactions with 1,112 metabolites (including 4 gap-filling reactions and an artificial biomass reaction).

Of particular interest, the sigma factor *rpoF* (gene *fliA*)—required for upregulation of the flagellar regulon—was absent from the genome, along with several other flagellar biosynthetic genes; a motility assay confirmed that uropathogenic *E. coli* strain ATCC 23506 is nonmotile in soft tryptone agar (data not shown), a result consistent with those of a previous investigation of an *E. coli fliA* deletion mutant (6). A detailed comparative genomics study is under way between this strain and other recently sequenced strains that also produce glycosaminoglycan-like capsular polysaccharides of pharmaceutical and nutraceutical relevance. Such analyses will improve the understanding of CPS biosynthesis regulation and the effect of the metabolic landscape on CPS production in pathogenic strains that depend upon the capsule as a “molecular camouflage” for host colonization.

Nucleotide sequence accession numbers. The annotated draft genome sequence was deposited in DDBJ/EMBL/GenBank under the accession no. [CAPK000000000](https://www.ncbi.nlm.nih.gov/nuccore/CAPK000000000). The version described in this paper is the first version, [CAPK010000000](https://www.ncbi.nlm.nih.gov/nuccore/CAPK010000000).

ACKNOWLEDGMENTS

We thank Jenny Xiang and colleagues for sequencing performed at the Genomics Resources Core Facility of Weill Cornell Medical College in New York, NY.

This work was funded by a Rensselaer Polytechnic Institute startup fund. B.F.C. is partially supported by a graduate fellowship from Rensselaer Polytechnic Institute.

REFERENCES

1. Ly M, Wang Z, Laremore TN, Zhang F, Zhong W, Pu D, Zagorevski DV, Dordick JS, Linhardt RJ. 2011. Analysis of *E. coli* K5 capsular polysaccharide heparosan. *Anal. Bioanal. Chem.* 399:737–745.

- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
- DeJongh M, Formsma K, Boillot P, Gould J, Rycenga M, Best A. 2007. Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC Bioinformatics* 8:139.
- Zhao K, Liu M, Burgess RR. 2007. Adaptation in bacterial flagellar and motility systems: from regulon members to ‘foraging’-like behavior in *E. coli*. *Nucleic Acids Res.* 35:4441–4452.