

# Draft Genome Sequence of *Escherichia coli* Strain ATCC 23502 (Serovar O5:K4:H4)

Brady F. Cress,<sup>a</sup> Zachary R. Greene,<sup>a</sup> Robert J. Linhardt,<sup>a,b,c</sup> Mattheos A. G. Koffas<sup>a</sup>

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

**We report the 4.682-Mbp high-quality draft assembly of the *Escherichia coli* strain ATCC 23502 (serovar O5:K4:H4, also known as NCDC U1-41) genome. This uropathogenic strain, commonly referred to as *E. coli* K4, produces a glycosaminoglycan-like capsular polysaccharide with a backbone similar in structure to unsulfated chondroitin, a precursor to the nutraceutically and potentially pharmaceutically valuable compound chondroitin sulfate. Metabolic reconstruction of this genome will enable prediction of genetic engineering strategies leading to increased chondroitin production.**

Received 22 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 23502, serovar O5:K4:H4. *Genome Announc.* 1(2):e00046-13. doi:10.1128/genomeA.00046-13.

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* strain ATCC 23502 is a uropathogenic strain that produces a group 2 capsular polysaccharide known as K4 capsular polysaccharide (K4 CPS) (1). The K4 capsular polysaccharide consists of a repeating [ $\rightarrow$ 4]  $\beta$ -D-glucuronic acid (GlcA) (1  $\rightarrow$  3) *N*-acetyl- $\beta$ -D-galactosamine (GalNAc) (1 $\rightarrow$ )]<sub>n</sub> disaccharide backbone with bisecting  $\beta$ -fructofuranose subunits linked to C3 of GlcA (2); the polysaccharide resembles unsulfated chondroitin sulfate with the exception of the terminal fructose residues present in K4 CPS. Group 2 capsules are of interest to the pharmaceutical industry as precursors in non-animal-sourced glycosaminoglycan (GAG) production, a safer alternative to animal tissue extraction routes that are susceptible to mammalian virus and prion contamination. Chemical or enzymatic sulfation of defructosylated K4 CPS could produce a range of chondroitin sulfate polysaccharides that are already used in the treatment of osteoarthritis and that have shown promise as skin substitutes, antivirals, and vaccines against maternal malaria (3). Genome-scale reconstruction of *E. coli* strain ATCC 23502 metabolism will allow for the prediction of gene deletions and overexpressions capable of increasing chondroitin production.

Genomic DNA was purified from *E. coli* strain ATCC 2352 with an Invitrogen PureLink Genomic DNA mini kit. The genome was sequenced using the Illumina HiSeq 2000 sequencing system, which produced 65 M paired-end reads of 101 bp with an insert size of 400 bp. Approximately 28 M random reads were assembled using Velvet v1.2.07 (4) at an optimal hash length of 93. The final genome assembly has approximately 43-fold coverage and contains 144 supercontigs composed of 160 contigs (>200 bp in length) with a total size of 4,682,525 bp, an N<sub>50</sub> contig length of 105,137 nucleotides, and a mean G+C content of 50.6%. All assembly data were deposited in the EMBL nucleotide sequence database.

The draft genome was annotated by the RAST (Rapid Annotation using Subsystem Technology) server (5) using Glimmer3 as a

gene caller (6), which predicted 4,560 coding sequences (CDSs) with an average length of 894 bp (3,541 CDSs have functional predictions), 83 tRNA-encoding genes, and 25 rRNA-encoding genes. RAST was also used to construct a draft metabolic model (7) containing 1,141 genes, corresponding to 1,389 reactions with 1,098 metabolites (including 4 gap-filling reactions and an artificial biomass reaction). Further analysis of the genome will provide useful information to characterize the entire mechanism of K4 CPS biosynthesis, transport, and extracellular attachment. Reconstruction of this strain's unique metabolic network will guide metabolic engineering modeling efforts aimed at increasing chondroitin production through targeted genetic manipulations. A comparative genomics analysis between *E. coli* strain ATCC 23502 and other strains producing glycosaminoglycan-like CPS is under way.

**Nucleotide sequence accession numbers.** The annotated draft genome sequence was deposited in DDBJ/EMBL/GenBank under accession no. [CPAL00000000](https://ncbi.nlm.nih.gov/nucl/CPAL00000000). The version described in this paper is the first version, [CPAL01000000](https://ncbi.nlm.nih.gov/nucl/CPAL01000000).

## 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

- Whitfield C. 2006. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* 75:39–68.
- Rodriguez ML, Jann B, Jann K. 1988. Structure and serological characteristics of the capsular K4 antigen of *Escherichia coli* O5:K4:H4, a fructose-containing polysaccharide with a chondroitin backbone. *Eur. J. Biochem.* 177:117–124.

3. Schiraldi C, Cimini D, De Rosa M. 2010. Production of chondroitin sulfate and chondroitin. *Appl. Microbiol. Biotechnol.* **87**:1209–1220.
4. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**:821–829.
5. 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.
6. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* **23**: 673–679.
7. 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.