

Draft Genome Sequence of *Escherichia coli* Strain Nissle 1917 (Serovar O6:K5:H1)

Brady F. Cress,^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 announce the availability of the 5.023-Mbp high-quality draft assembly of the *Escherichia coli* strain Nissle 1917 (serovar O6:K5:H1) genome. Short genomic segments from this important probiotic strain have been available in public databases, but the full genome sequence has remained inaccessible. Thus, high-coverage, whole genome sequencing of *E. coli* Nissle 1917 is presented herein. Reannotation and metabolic reconstruction will enable comparative genomics analysis and model-guided predictions of genetic manipulations leading to increased production of the K5 capsular polysaccharide known as *N*-acetyl heparosan, a precursor to the anticoagulant pharmaceutical heparin.

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

Citation Cress BF, Linhardt RJ, Koffas MAG. 2013. Draft genome sequence of *Escherichia coli* strain Nissle 1917 (serovar O6:K5:H1). *Genome Announc.* 1(2):e00047-13. doi:10.1128/genomeA.00047-13.

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.

The commensal *Escherichia coli* strain Nissle 1917 is one of the oldest, most well-characterized probiotic agents and has shown promising results in treatment of various intestinal diseases and disorders (1) since its isolation from the feces of a World War I soldier exhibiting unique intestinal fortitude compared to his cohort, all of whom developed infectious diarrhea (2). Assorted genomics studies have been performed on Nissle 1917, including tRNA screening, genomic island sequencing, DNA-DNA hybridization (3), and even low-coverage genomic shotgun sequencing (4). Until now, however, the whole genome sequence has been inaccessible. Serotyping of *E. coli* Nissle 1917 has identified the presence of a K5 antigen, which is known to be composed of *N*-acetyl heparosan (a precursor to the anticoagulant pharmaceutical heparin), a group 2 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 (5). Under certain growth conditions, *E. coli* Nissle 1917 produces significantly more CPS than *E. coli* K5 (data not shown), making the organism attractive as a production strain for bioengineered heparin.

Escherichia coli Nissle 1917 was cultured from a Mutaflor tablet (Ardeypharm, Herdecke, Germany) and plated on Luria-Bertani (LB) agar medium; a single colony was picked and grown in LB medium at 37°C overnight, and the genomic DNA was purified with an Invitrogen PureLink genomic DNA mini kit. Prior to whole-genome sequencing, successful PCR amplification of the genes *kfiA*, *kfiB*, *kfiC*, and *kfiD* in region 2 of the capsular polysaccharide biosynthetic gene cluster confirmed probable isolation of Nissle 1917. The genome was sequenced using the Illumina HiSeq 2000 sequencing system, which produced 110 M paired-end reads of 101 bp with an insert size of 400 bp. Approximately 28 M random reads were assembled with Velvet v1.2.07 (6) at an optimal hash length of 91. The final genome assembly has 51-fold coverage

and contains 125 supercontigs composed of 143 contigs (>200 bp in length) with a total size of 5,023,325 bp, an N₅₀ contig length of 253,628 nucleotides, and a mean G+C content of 50.5%. 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 (7) using Glimmer3 as a gene caller (8), which predicted 4,846 coding sequences (CDSs) with an average length of 900 bp (3,739 CDSs have functional predictions), 80 tRNA-encoding genes, and 9 rRNA-encoding genes. RAST was also used to construct a draft metabolic model (9) containing 1,179 genes, corresponding to 1,388 reactions with 1,089 metabolites (including 4 gap-filling reactions and an artificial biomass reaction). Comparison of metabolic reconstructions will uncover differential carbohydrate and polysaccharide biosynthetic pathways between Nissle 1917 and K5 while yielding insight on growth conditions leading to maximum CPS production. A comparative genomics analysis currently ongoing in our lab between Nissle 1917 and related uropathogenic *E. coli* (UPEC) strains producing similar capsular polysaccharides will also guide understanding of the role of glycosaminoglycan-like capsules in pathogenesis.

Nucleotide sequence accession numbers. The annotated draft genome sequence was deposited in DDBJ/EMBL/GenBank under accession no. [CAPM000000000](https://www.ncbi.nlm.nih.gov/nuccore/CAPM000000000). The version described in this paper is the first version, [CAPM010000000](https://www.ncbi.nlm.nih.gov/nuccore/CAPM010000000).

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. Sonnenborn U, Schulze J. 2009. The non-pathogenic *Escherichia coli* strain Nissle 1917—features of a versatile probiotic. *Microb. Ecol. Health Dis.* 21:122–158.
2. Nissle A. 1918. Die antagonistische Behandlung chronischer Darmstörungen mit Colibakterien. *Med. Klin.* 2:29–33.
3. Grozdanov L, Raasch C, Schulze J, Sonnenborn U, Gottschalk G, Hacker J, Dobrindt U. 2004. Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J. Bacteriol.* 186:5432–5441.
4. Sun J, Gunzer F, Westendorf AM, Buer J, Scharfe M, Jarek M, Gösling F, Blöcker H, Zeng A-P. 2005. Genomic peculiarity of coding sequences and metabolic potential of probiotic *Escherichia coli* strain Nissle 1917 inferred from raw genome data. *J. Biotechnol.* 117:147–161.
5. 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.
6. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
7. 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.
8. Delcher AL, Bratke Ka, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* 23:673–679.
9. 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.