Complete Genome Sequence of Escherichia coli BE104, an MC4100 Derivative Lacking the Methionine Reductive Pathway

Brian P. Anton, Richard Morgan, Benjamin Ezraty, Bruno Manta, Frédéric Barras, Mehmet Berkmen

Complete genome sequence of *E. coli* BE104, an MC4100 derivative lacking the methionine reductive pathway

Brian P. Anton[1], Rick Morgan[1], Benjamin Ezraty[2], Bruno Manta[1], Frédéric Barras[2] and Mehmet Berkmen[1*]

[1]New England Biolabs, 240 County Road, Ipswich, MA 01938, USA

[2]Aix-Marseille Université, CNRS, Laboratoire de Chimie Bactérienne, Marseille, France.

[*]Correspondence: Phone: (978) 380-7519; Fax: (978) 921-1350; Email: berkmen@neb.com

Keywords: *Escherichia coli*, MC4100, methionine sulfoxide reductase

Abstract:

Escherichia coli and its derivatives have played a profound role in the development of molecular biology in the last 50 years. With the advent of the genomic age, many *E. coli* strains are being sequenced, adding to our knowledge of the *E. coli* pan genome and its biology. Here we present the complete, annotated genome sequence of an *E. coli* MC4100 mutant strain. This strain has several methionine sulfoxide reductase deletions, making it ideal for studying enzymes that alter the redox state of methionine.

Announcement:

Escherichia coli is the predominant prokaryotic model organism. Its use in molecular biology has led to enzyme discovery, the development of myriad research tools, and the understanding of many biological principles. Not surprisingly, over the past half a century, many mutant *E. coli* strains have been genetically engineered or selected.

There are two popular ancestral strains of *E. coli*, K-12 and B [1]. In general, *E. coli* K-12 is used for DNA manipulation and cloning, while *E. coli* B is mostly used for protein expression. There are many *E. coli* K-12 derivatives currently used in research and production laboratories worldwide, including MC4100, engineered by Malcolm Casadaban [2]. Of the 801 completely sequenced *E. coli* genomes, only 33 are from K-12 derived laboratory strains and of that only 1 is from MC4100. MC4100 has been used extensively since its inception [3], resulting in thousands of publications. Its genome has been previously sequenced [4] and its relationship to other *E. coli* K-12 strains has been analyzed [5].

The strain BE104, a derivative of MC4100 methionine auxotroph mutant lacking five methionine sulfoxide reductases (Msrs), was previously constructed in order to characterize an

enzymatic system (MsrPQ) responsible for repairing proteins containing methionine sulfoxide in the bacterial periplasm [6]. Briefly, BE104 was derived from MC4100 a methionine auxotroph mutant by (1) a series of P1vir crosses to delete all cytoplasmic msrs (*msrA*, *msrB*, *msrC* and *bisC*) by replacement with corresponding alleles from Keio KO strains [7]; (2) selection for suppressor strains that could reduce methionine sulfoxide; and (3) deletion of *msrP*, a consequently discovered periplasmic msr [5].

As BE104 is being used in our research and will be further engineered, we sequenced its genome using the Pacific Biosciences (PacBio) RS II sequencing platform as described previously [8]. Genomic DNA was isolated using the Monarch gDNA kit (New England Biolabs). A SMRTbell library was constructed from 5 µg genomic DNA, sheared to ~10 kb using a gTube (Covaris). The library was sequenced on two SMRT cells using P6-C4 chemistry. The first cell yielded 258 Mb sequence from 22,107 (15%) P1 reads with mean polymerase read length of 11,672 and mean read of insert length of 6096 (180-minute data collection time). A second cell was sequenced to increase coverage, yielding 905 Mb sequence from 63,262 (42%) P1 reads with mean polymerase read length of 14,318 and mean read of insert of 4594 (240-minute data collection time). Sequencing reads were processed and assembled with Pacific Biosciences SMRT analysis v2.3.0 software using the HGAP3 protocol (5 Mb expected genome size, filters set to: minimum subread length 1000, minimum polymerase read length 2000, minimum read quality 0.80) and polished using Quiver (4). The 1.1 Gb of sequence assembled into a single closed circular genome of 4,xxx,xxx bp with mean 200-fold coverage. The assembled sequence was annotated using the NCBI Prokaryotic Genomes Annotation Pipeline (PGAP).

The following expected deletions were all observed: *metB* (*O*-succinyl homoserine lyase) is disrupted by 2 base deletion/frameshift, making this strain auxotrophic for methionine; *msrA* is disrupted by insertion with a spectinomycin cassette; the msr loci *msrB*/*yeaA*, *msrC*/*yebR,* and *msrP*/*yedY* are all deleted*,* as is *bisC* (biotin sulfoxide reductase), an enzyme with weak msr activity on free S-methionine sulfoxide [9]. As expected, *yedV*::IS2 insertion was not observed as *msrP*::kan deletion was transduced using the kanamycin marker from Keio collection (Strain#), co-transducing wt *yedV* allele. Analysis of the genomic sequence revealed several other changes with respect to the parental strain MC4100 in addition to these expected deletions.

This strain should be of general use to the research community studying protein redox states and would be an important addition to the repertoire of sequenced *E. coli* genomes.

Nucleotide sequence accession number. The annotated complete genome sequence is deposited at EMBL, EBI under the following assembly accession numbers:

References:

[1] F.W. Studier, P. Daegelen, R.E. Lenski, S. Maslov, J.F. Kim, Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes, J Mol Biol, 394 (2009) 653-680.

[2] Remembering Malcolm J. Casadaban, J Bacteriol, 192 (2010) 4261-4263.

[3] M.J. Casadaban, Transposition and fusion of the lac genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu, J Mol Biol, 104 (1976) 541-555.

[4] T. Ferenci, Z. Zhou, T. Betteridge, Y. Ren, Y. Liu, L. Feng, P.R. Reeves, L. Wang, Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12, J Bacteriol, 191 (2009) 4025-4029.

[5] J.E. Peters, T.E. Thate, N.L. Craig, Definition of the *Escherichia coli* MC4100 genome by use of a DNA array, J Bacteriol, 185 (2003) 2017-2021.

[6] A. Gennaris, B. Ezraty, C. Henry, R. Agrebi, A. Vergnes, E. Oheix, J. Bos, P. Leverrier, L. Espinosa, J. Szewczyk, D. Vertommen, O. Iranzo, J.F. Collet, F. Barras, Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons, Nature, 528 (2015) 409-412.

[7] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, Mol Syst Biol, 2 (2006) 2006 0008.
[8] C.S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E.E. Eichler, S.W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data, Nat Methods, 10 (2013) 563-569.
[9] B. Ezraty, J. Bos, F. Barras, L. Aussel, Methionine sulfoxide reduction and assimilation in *Escherichia coli*: new role for the biotin sulfoxide reductase BisC, J Bacteriol, 187 (2005) 231-237.