



Escherichia coli NGF-1, a Genetically Tractable, Efficiently Colonizing Murine Gut Isolate

 Marika Ziesack,^{a,b} Michiel A. P. Karrenbelt,^{a,b} Johannes Bues,^{a,b} Elena Schaefer,^{a,b} Pamela Silver,^{a,b} Jeffrey Way^{a,b}

^aWyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, Massachusetts, USA

^bDepartment of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA

ABSTRACT The genome of the murine commensal strain *Escherichia coli* NGF-1 contains a 5.03-Mbp chromosome and plasmids of 40.2 kbp and 8.56 kbp. NGF-1 efficiently colonizes the mouse gut and is genetically tractable. The genome sequence reported here facilitates genetic engineering and research in mouse models of healthy and diseased intestine.

The gut microbiome plays a key role in health and disease (1). *Escherichia coli* NGF-1 is of particular interest because it was isolated from a healthy BALB/c mouse from Charles River Labs, colonizes mice efficiently, and can be engineered with complex genetic circuits (2–6) (Table 1). Kotula and colleagues placed a tetracycline-inducible trigger element and a memory element in *E. coli* K-12 and *E. coli* NGF-1 and found essentially identical responses to tetracycline treatment *in vitro* and in the mouse gut (2). Riglar et al. showed that engineered NGF-1 was stable in C57BL/6 mice for over 6 months (3). In the work of Certain et al., *E. coli* NGF-1 survived near a surgical implant, allowing study of persistent infection (4). Kim et al. constructed a communication system which could be observed in the mouse gut, and Ziesack et al. introduced NGF-1 as a member of an engineered consortium into gnotobiotic mice (5, 6).

To obtain the NGF-1 sequence, a glycerol stock was used to inoculate an LB agar plate for the isolation of single colonies. A single colony was then used to inoculate an overnight culture in LB broth (37°C, with shaking at 220 rpm). Genomic DNA was extracted using the Qiagen DNeasy blood and tissue kit and quantified using a Life Technologies Quant-iT PicoGreen double-stranded DNA (dsDNA) assay kit. The DNA was sheared on a Covaris S2000 machine, and a library was prepared using an Illumina TruSeq kit. Sequencing was done using the Illumina MiSeq reagent kit v2 (2 × 250 bp), and quality filtering, trimming, and filtering of adapter sequences were performed using FastQC with standard settings (7). The resulting 6.6 million paired-end reads with an average length of 207 bp were assembled *de novo* with SPAdes version 3.7.1 (8), using the “careful” option to minimize mismatches in the final contigs. The 64 resulting contigs were filtered for contaminants via a BLAST search against several *E. coli* genomes using Projector2 (9) and Ragout (10) with standard parameters, resulting in 51 contigs at an average read coverage of 22×. Contig ends were joined by two methods. First, some contig ends were identified that had ends with identical segments that fell below the alignment threshold of the joining software. These joinings were validated by alignment of the joined sequences with sequences of other *E. coli* strains. Second, in cases such as those where identical rRNA genes prevented inference of continuity between sequences on either side of the repeated element, we hypothesized associations based on other *E. coli* sequences and then confirmed the association by PCR using unique sequence flanking primers and observation of a DNA fragment of predicted size. One case of sequence ambiguity was attributed to an inverting-phase variation-type element.

Received 29 October 2018 Accepted 9 November 2018 Published 6 December 2018

Citation Ziesack M, Karrenbelt MAP, Bues J, Schaefer E, Silver P, Way J. 2018. *Escherichia coli* NGF-1, a genetically tractable, efficiently colonizing murine gut isolate. Microbiol Resour Announc 7:e01416-18. <https://doi.org/10.1128/MRA.01416-18>.

Editor Frank J. Stewart, Georgia Institute of Technology

Copyright © 2018 Ziesack et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Jeffrey Way, jeff.way@wyss.harvard.edu.

M.Z., M.A.P.K., and J.B. contributed equally to this work.

TABLE 1 Applications of NGF-1 using artificial genetic circuits

Application	Engineering ^a	Reference
Gut sensor (molecule)	ATC-inducible trigger element; memory element	2
Gut sensor (pathogen)	Tetrathionate-inducible trigger element; memory element	3
Chronic infection sensor	ATC-inducible trigger element; memory element	4
Interspecies quorum sensing	ATC-inducible signaling element; Lux-triggered memory element	5
Metabolite cross-feeding consortium member in the gut	Triple KO of amino acid biosynthetic pathways; methionine overproduction through antimetabolite selection	6

^a ATC, anhydrotetracycline; KO, knockout; Lux, luciferase.

The genome was annotated with the Rapid Annotation using Subsystem Technology (RAST) server (11) and the Pan-Genomes Analysis Pipeline (PGAP) (12), followed by manual curation. NGF-1 contains a 5,026,105-bp chromosome and two plasmids, pNGF1-CROD2 (40,158 bp) and pNGF-coLY (8,556 bp), encoding 5,218, 57, and 10 genes, respectively. The colicin-producing plasmid may help explain the efficient colonizing ability of this strain.

NGF-1 is similar to *E. coli* K-12 and murine *E. coli* strains. Specifically, NGF-1 has 98% nucleotide sequence identity with K-12 and >99% with mouse-derived strains, such as MP1, ATCC 25922, and LF82. NGF-1 is distinct from all known *E. coli* strains, but its genome is a mosaic of genes known from other strains, plus prophage genes. NGF is a niacinamide auxotroph, likely caused by a sense mutation in the *nadC* gene (13).

In sum, *E. coli* NGF-1 is both engineerable and able to colonize the mouse gut and other experimentally relevant environments. Knowledge of its genome sequence should facilitate further studies of gut colonization and may facilitate development of living therapeutics and diagnostics.

Data availability. This whole-genome sequencing project has been deposited in GenBank under the accession number [CP016007](https://doi.org/10.1093/ilar/ilv031). Raw reads are available under the BioProject accession number [PRJNA380756](https://doi.org/10.1093/ilar/ilv031).

ACKNOWLEDGMENTS

This genome sequencing project was funded by the Defense Advanced Research Projects Agency (DARPA grant HR0011-15-C-0094) and by the Wyss Institute for Biologically Inspired Engineering at Harvard University.

We declare no conflict of interest.

REFERENCES

- Bleich A, Fox JG. 2015. The mammalian microbiome and its importance in laboratory animal research. *ILAR J* 56:153–158. <https://doi.org/10.1093/ilar/ilv031>.
- Kotula JW, Kerns SJ, Shakel LA, Siraj L, Collins JJ, Way JC, Silver PA. 2014. Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc Natl Acad Sci U S A* 111:4838–4843. <https://doi.org/10.1073/pnas.1321321111>.
- Riglar DT, Giessen TW, Baym M, Kerns SJ, Niederhuber MJ, Bronson RT, Kotula JW, Gerber GK, Way JC, Silver PA. 2017. Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. *Nat Biotechnol* 35:653–658. <https://doi.org/10.1038/nbt.3879>.
- Certain LK, Way JC, Pezone MJ, Collins JJ. 2017. Using engineered bacteria to characterize infection dynamics and antibiotic effects *in vivo*. *Cell Host Microbe* 22:263–268.e4. <https://doi.org/10.1016/j.chom.2017.08.001>.
- Kim S, Kerns SJ, Ziesack M, Bry L, Gerber GK, Way JC, Silver PA. 2018. Quorum sensing can be repurposed to promote information transfer between bacteria in the mammalian gut. *ACS Synth Biol* 7:2270–2281. <https://doi.org/10.1021/acssynbio.8b00271>.
- Ziesack M, Gibson T, Shumaker AM, Oliver JK, Riglar DT, Giessen TW, DiBenedetto NV, Lall K, Hsu BB, Bry L, Way JC, Silver PA, Gerber GK. 2018. Inducible cooperation in a synthetic gut bacterial consortium introduces population balance and stability. *bioRxiv* <https://doi.org/10.1101/426171>.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Van Hijum SAFT, Zomer AL, Kuipers OP, Kok J. 2005. Projector 2: contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. *Nucleic Acids Res* 33:W560–W566. <https://doi.org/10.1093/nar/gki356>.
- Kolmogorov M, Raney B, Paten B, Pham S. 2014. Ragout—a reference-assisted assembly tool for bacterial genomes. *Bioinformatics* 30:i302–i309. <https://doi.org/10.1093/bioinformatics/btu280>.
- Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
- Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. 2012. PGAP: Pan-Genomes Analysis Pipeline. *Bioinformatics* 28:416–418. <https://doi.org/10.1093/bioinformatics/btr655>.
- Li Z, Bouckaert J, Deboeck F, De Greve H, Hernalsteens J. 2012. Nicotinamide dependence of uropathogenic *Escherichia coli* UT189 and application of *nadB* as a neutral insertion site. *Microbiology* 158:736–745. <https://doi.org/10.1099/mic.0.052043-0>.