

# Complete Genome Sequence of the Engineered *Escherichia coli* SHuffle Strains and Their Wild-Type Parents

Brian P. Anton, Alexey Fomenkov, Elisabeth A. Raleigh, Mehmet Berkmen

New England Biolabs, Ipswich, Massachusetts, USA

SHuffle strains are genetically engineered *Escherichia coli* strains that are capable of oxidizing cysteines within proteins to form disulfide bonds. Here we present the complete genome of both the K-12 and B versions of SHuffle strains along with their parental ancestors. These strains have been of significant use to both the general scientific community and the biotech industry, interested in producing novel disulfide-bonded proteins that were hitherto unable to be expressed in standard *E. coli* expression strains.

Received 17 February 2016 Accepted 18 February 2016 Published 31 March 2016

Citation Anton BP, Fomenkov A, Raleigh EA, Berkmen M. 2016. Complete genome sequence of the engineered *Escherichia coli* SHuffle strains and their wild-type parents. *Genome Announc* 4(2):e00230-16. doi:10.1128/genomeA.00230-16.

Copyright © 2016 Anton 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 Mehmet Berkmen, [berkmen@neb.com](mailto:berkmen@neb.com).

Wild-type *Escherichia coli* strains are incapable of producing disulfide-bonded proteins due to the presence of two reducing pathways, the glutaredoxin and thioredoxin pathways. The reductases in these pathways recognize disulfide-bonded proteins and reduce them to maintain the proteome in its reduced state.

We constructed two *E. coli* strains capable of promoting the production of correctly disulfide-bonded proteins in the cytoplasm by genetically knocking-out *gor* and *trxB* (1) and expressing *dsbC* in the cytoplasm. These strains, SHuffle B and SHuffle K-12 (2), were derived from two commonly used laboratory strains, ER2566 (B background) and DHB4 (K-12 background), respectively. Since their introduction in 2009, numerous studies have used SHuffle strains to express various disulfide-bonded proteins (3), including the recent demonstration of the expression of full-length antibodies in bacterial cytoplasm for the first time (4). Surprisingly, SHuffle strains can also express certain non-disulfide-bonded proteins better than wild-type protein expression strains, indicating a need for a deeper understanding of the biology of these strains. The elucidation of the genomes of these strains should therefore lead to a deeper understanding of the biology of the SHuffle strains.

The four strains were sequenced using the Pacific Biosciences platform on an RSII instrument using P6 chemistry. For each strain, a 10-kb SMRTbell library was prepared from total DNA using the manufacturer's instructions, size-selected (4 to 50 kb) using a BluePippin (Sage Science), and sequenced on 2 SMRT cells with 240-min movies. Assembly was performed using RS\_HGAP\_Assembly followed by manual refinement and re-sequencing using RS\_BridgeMapper. HGAP coverage ranged from roughly 100× to 225×.

The genomic sequence confirmed the genetic deletions of *trxB* and *gor*, but also revealed two unexpected differences: a new allele of the AhpC suppressor and chromosomal duplications of cytoplasmic *dsbC*. Unlike previous studies of *trxB* and

*gor* suppression by AhpC\* in *E. coli* K-12 (5), the naturally selected suppressor of *trxB* and *gor* in *E. coli* B is a triplet codon contraction and not an expansion. The identity of this novel suppressor should result in an expanded understanding of the redox biology. Furthermore, the genetic insertion of the *dsbC* overexpressing plasmid into the lambda attachment site using lambda InCh (6), should result in a single copy insertion. However, the genomic sequences of SHuffle K-12 (C3026) and SHuffle-B (C3029) revealed a triplicate and a duplicate copy of the insertion, respectively. These illegitimate duplication events could be due to the role of cytoplasmic DsbC in alleviating oxidative stress within the SHuffle strains.

These strains have been of significant use to both the general scientific community and the biotech industry in the production of novel disulfide-bonded proteins that are otherwise inexpressible in standard *E. coli* laboratory strains.

**Nucleotide sequence accession numbers.** The annotated complete genome sequences are deposited at EMBL, EBI, under the following assembly accession numbers: NEB T7 express (C2566), [CP014268](https://www.ncbi.nlm.nih.gov/nuclink/CP014268); SHuffle-B (C3029), [CP014269](https://www.ncbi.nlm.nih.gov/nuclink/CP014269); DHB4, [CP014270](https://www.ncbi.nlm.nih.gov/nuclink/CP014270), and its cognate F' episome, [CP014271](https://www.ncbi.nlm.nih.gov/nuclink/CP014271); and SHuffle K-12 (C3026), [CP014272](https://www.ncbi.nlm.nih.gov/nuclink/CP014272), and its cognate F' episome, [CP014273](https://www.ncbi.nlm.nih.gov/nuclink/CP014273).

## ACKNOWLEDGMENT

We thank Tamas Vincze for providing technical support.

## REFERENCES

- Bessette PH, Aslund F, Beckwith J, Georgiou G. 1999. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc Natl Acad Sci USA* 96:13703–13708.
- Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. 2012. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact* 11:56. <http://dx.doi.org/10.1186/1475-2859-11-56>.
- Ke N, Berkmen M. 2014. Production of disulfide-bonded proteins in *Escherichia coli*. *Curr Protoc Mol Biol* 108:1-16.1B2.1. <http://dx.doi.org/10.1002/0471142727.mb1601bs108>.

4. Robinson MP, Ke N, Lobstein J, Peterson C, Szkodny A, Mansell TJ, Tuckey C, Riggs PD, Colussi PA, Noren CJ, Taron CH, DeLisa MP, Berkmen M. 2015. Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nat Commun* 6:8072. <http://dx.doi.org/10.1038/ncomms9072>.
5. Ritz D, Lim J, Reynolds CM, Poole LB, Beckwith J. 2001. Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. *Science* 294:158–160. <http://dx.doi.org/10.1126/science.1063143>.
6. Boyd D, Weiss DS, Chen JC, Beckwith J. 2000. Towards single-copy gene expression systems making gene cloning physiologically relevant: lambda InCh, a simple *Escherichia coli* plasmid-chromosome shuttle system. *J Bacteriol* 182:842–847. <http://dx.doi.org/10.1128/JB.182.3.842-847.2000>.