nature microbiology

SUPPLEMENTARY INFORMATION

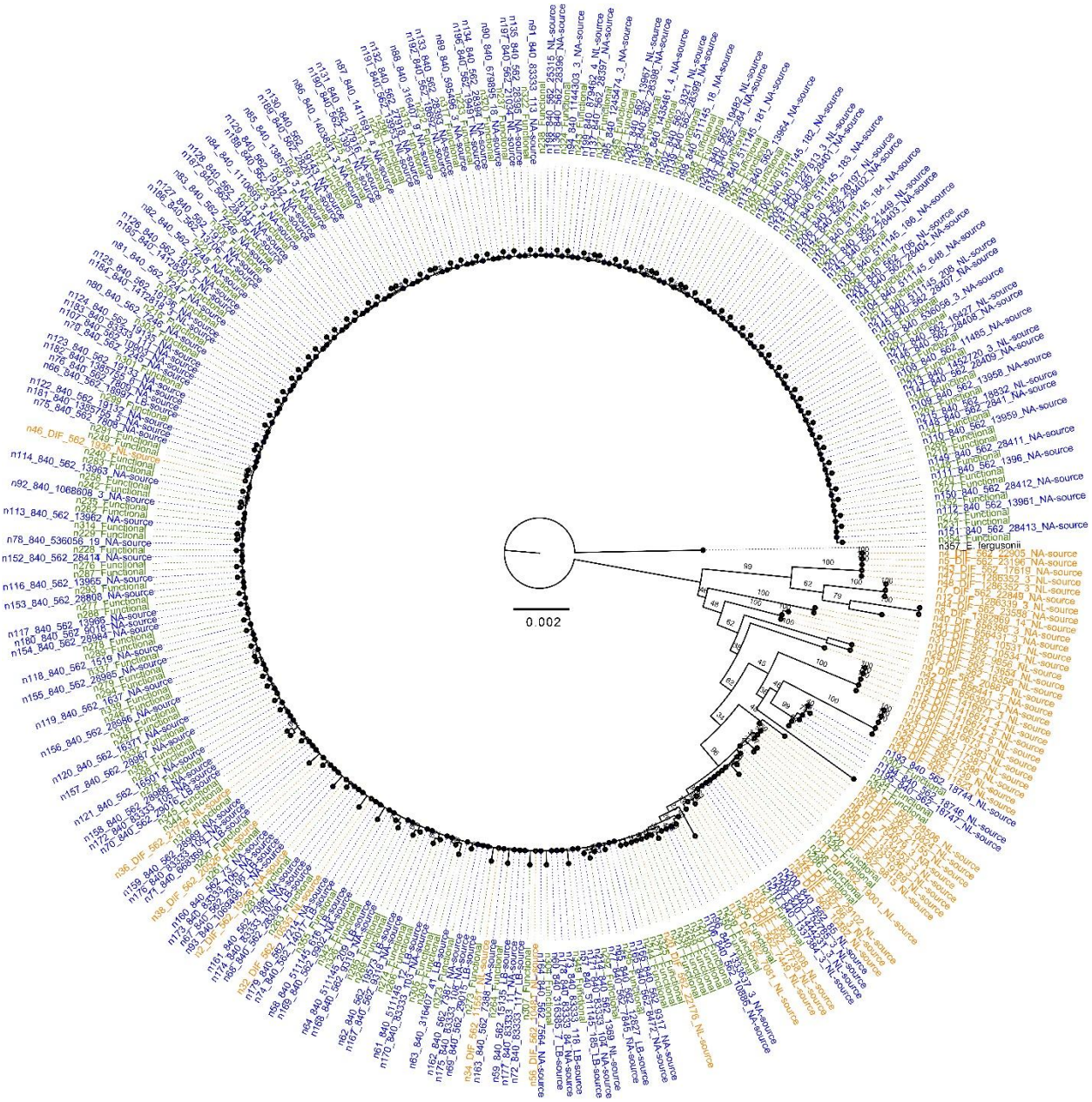
https://doi.org/10.1038/s41564-018-0340-2

In the format provided by the authors and unedited.

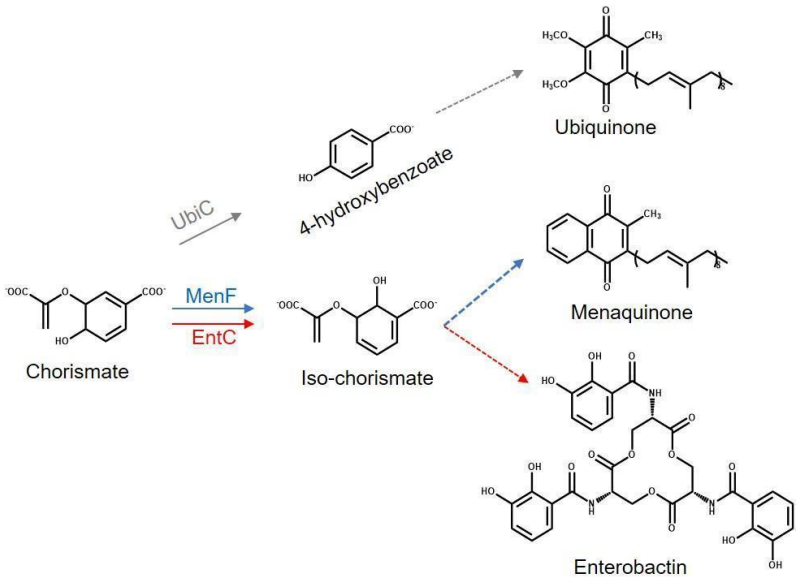
Pseudogene repair driven by selection pressure applied in experimental evolution

Amitesh Anand[1], Connor A. Olson[1], Laurence Yang[1], Anand V. Sastry[1], Edward Catoiu[1], Kumari Sonal Choudhary[1], Patrick V. Phaneuf[1], Troy E. Sandberg[1], Sibei Xu[1], Ying Hefner[1], Richard Szubin[1], Adam M. Feist[1,2] and Bernhard O. Palsson[1,2]*

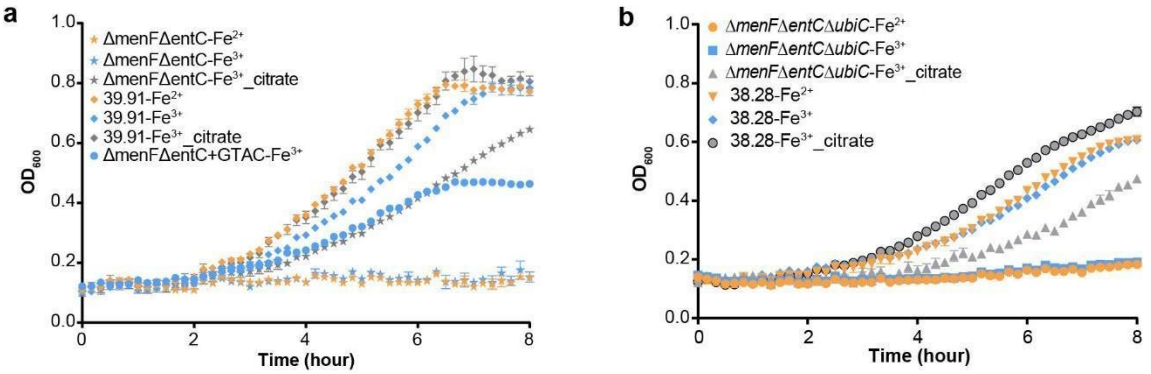
[1]Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. [2]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. *e-mail: palsson@ucsd.edu



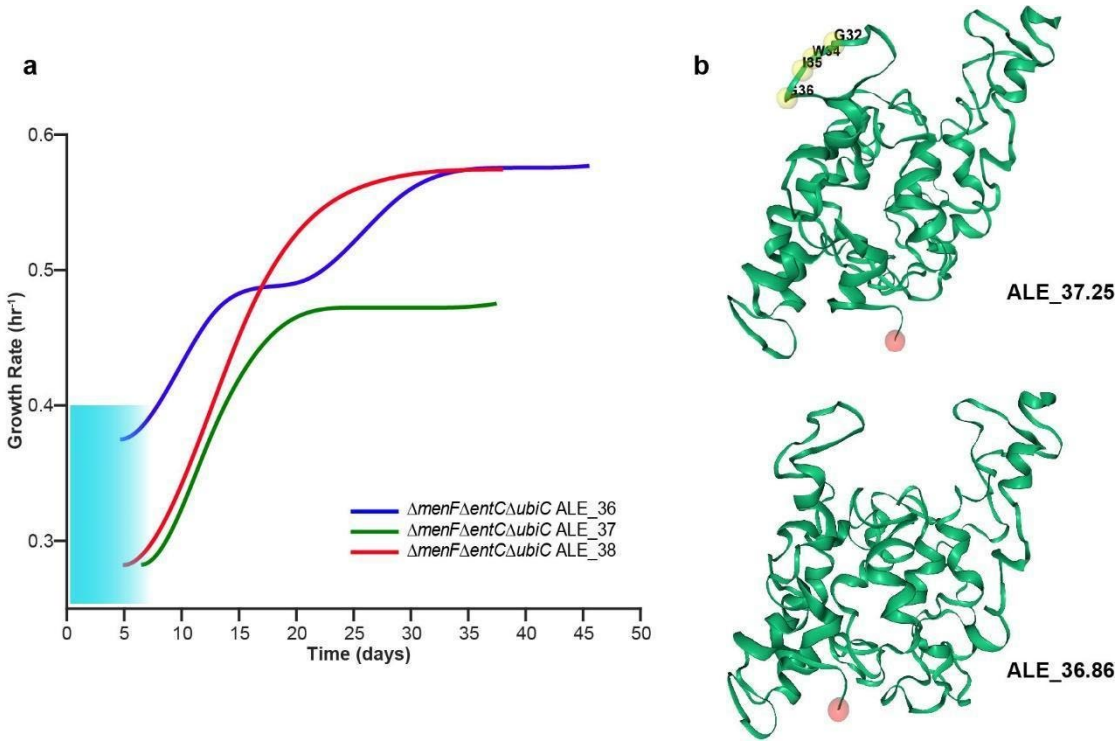
Supplementary Figure 1. *rpoB* based phylogenetic tree of 212 *E. coli* genome IDs wherein *efeU* is fragmented into two ORFs along with selected genome IDs with complete *efeU* (refer to method for selection criteria). '840' represents *efeU* fragmented as 120 & 720 bp (blue), 'DIF' represents other fragmentation types (orange) and 'Functional' refers to complete *efeU* gene (green). Source information: LB (lab related), NL (lab unrelated) and NA (not available).



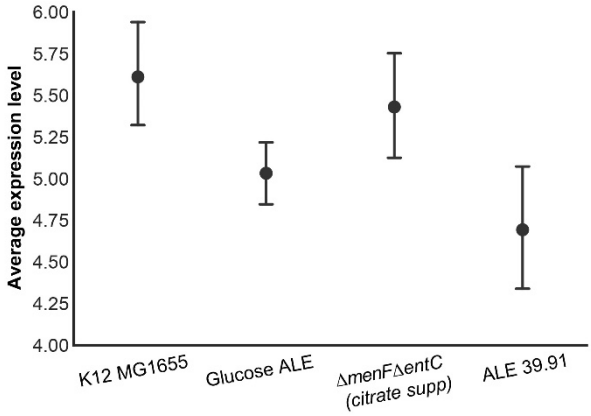
Supplementary Figure 2. Rationale behind the strain design. Chorismate serve as common precursor for the biosynthesis of enterobactin, menaquinone and ubiquinone. MenF and EntC catalyzes its conversion to isochorismate; whereas UbiC results in 4-hydroxybenzoate formation.



Supplementary Figure 3. Growth curve in minimal media supplemented with one of the following iron sources- (i) ferrous chloride, (ii) ferric chloride, (iii) ferric chloride and sodium citrate. (a) Pre-evolved Δ*menF*Δ*entC* and evolved strain ALE_39.91, (b) pre-evolved Δ*menF*Δ*entC*Δ*ubiC* and evolved strain ALE_38.28. 'Δ*menF*Δ*entC*+GTAC' is the pre-evolved Δ*menF*Δ*entC* strain with the *efeU* frame restoring insertion same as ALE_39.91. Figures present mean, and standard deviation derived from four biologically independent samples.



Supplementary Figure 4. (a) Evolution trajectories of Δ*menF*Δ*entC*Δ*ubiC* replicates. Evolution of Δ*menF*Δ*entC*Δ*ubiC* strain was started with six independent replicates. The shaded area depicts zone of no detectable growth. (b) Overlay of restored protein sequences on the homology model of functional EfeU. Color code: black=deletion, yellow=SNP, green=identical.



Supplementary Figure 5. Expression profile of Fur regulon in WT *E. coli*, glucose adapted *E. coli* along with Δ*menF*Δ*entC* and ALE_39.91 strains. Error bars indicate standard deviation of Fur regulon expression, where center point represents the mean expression (n = 116 genes across seven independent evolutionary endpoints for Glucose ALE, and two biologically independent replicates for all others).

Supplementary Table 1. Growth rate of wild type (WT) and *entC* knockout (Δ*entC*) *E. coli* K12 MG1655 strains

Strain	Growth rate (per hour)	Standard deviation
WT	0.7	0.02
Δ*entC*	0.6	0.01

The data present mean of three biologically independent samples.

Supplementary Table 2. List of mutations observed in evolved strains

Strain	Mutations			
Δ*menF*Δ*entC* ALE_39.91	*efeU* ((+)GTAC)	*putP, efeU* (G→T)	*pyrE, rph* ((A)8→7)	*rpoC* (H419P)
Δ*menF*Δ*entC* ALE_29.82	*glmU* (A322T)	*oxyR* (A204E)	*rph, rph* (Δ82 bp)	*rpoB* (P552L)
Δ*menF*Δ*entC* ALE_30.83	*icd* (H366H, T370T)	*rpoA* (G36V)	*rpsM, rpmJ* (C→G)	*dicA, ydfA* (A→G)
	oxyR (A174E)	*lpoA*(I502M)		
Δ*menF*Δ*entC*Δ*ubiC* ALE_36.86	*efeU* ((TG)3→2)	*prc, proQ, msrC* (Δ1909 bp)	*arcB* (S83L)	*gltD* (C47*)
	rph (Δ29 bp)	*yejM* ((GTGAAAGA)2→3)	*rlmJ* (L224L)	*yneJ* (I226F)
	putP, efeU (T→G)	*lrhA, alaA* ((TGTTA)2→1)		
Δ*menF*Δ*entC*Δ*ubiC* ALE_37.25	*efeU* (+CGAG)	*sdhB* (R9S)	*proQ* (+A)	*lptF* ((TGG)3→2)
	ettA (K120*)	*arcA* (F139Y)		
Δ*menF*Δ*entC*Δ*ubiC* ALE_38.28	*efeU* (+GTAC)	*putP, efeU* (C→A)	*ykfH* (R7P)	*sdhA* (P328T)
	asmA (S326*)	*rpoS* (+CG)	*sspA* (S9L)	

Mutations observed in pre-evolved strains have been removed from the list.

The common mutations related to adaptation to glucose minimal medium have been excluded in the present manuscript (*1*).

Supplementary Table 3. Comparison of amino acid sequences of EfeU protein of the '*efeU* repaired' strain with reference

Strain	Sequence identity	Amino acid 30 to 36								Changes with respect to reference
Reference		Q	R	G	R	W	I	G		
ALE_36.86	99.6%	Q	R	G	R	W	I	Δ		One deletion
ALE_37.25	98.2%	Q	R	A	R	P	M	D	C	4 substitution & 1 insertion
ALE_39.91	97.5%	Y	P	A	R	P	M	D	C	6 substitution & 1 insertion
ALE_38.28	97.5%	Y	P	A	R	P	M	D	C	6 substitution & 1 insertion

Supplementary Table 4. *efeU* nucleotide sequence of *E. coli* 562.11502 (PATRIC ID)

PATRIC Genome ID	*efeU* sequence
562.11502	atgtttgttccgtttctcattatgttgcgcgaaggacttgaagccgcgctgattgtcagtttgattgccagctatcttaagcgtacc cagcgaggccgatggattggtgtgatgtggattggcgtgttgcttgccgctgcgttgtgcctgggcttggggatcttcattaac gaaaccaccggcgaatttccgcaaaaagagcaggaactgtttgaaggtatcgtggcggtgatcgccgtggtgatccttacct ggatggttttctggatgcgcaaagtgtcgcgc_acgtcaaagtgcaactggaacaggcagtcgatagcgcattg**c**cagc gtggaaatcatcatggctgggcgctggtgatgatggtctttttgccgttgcaagggaagggctggagtcggtctttttcctgct ggcggcatttcaacaagatgtcgggatctggccgccgctgggtgcaatgctcggtcttgctactgccgtggtgctcggcttcc tgctctactggggcggtattcgcctcaatcttggtgcatttttaaatggaccagcctgtttattctcttcgtcgccgcagggctgg cggctggtgccattcgcgcatttcatgaagccggattgtggaaccactttcaggaaatcgccttcgatatgagcgcggtcctct caactcactcgctgtttggcacgctgatggaagggatttttggctatcaggaagcaccgagcgtcagcgaagtcgccgtctg gtttatttatctcatcccggcgctggtggcatttgccctgccaccacgcgcaggggcgacagcgtctcgctccgcgtag

Supplementary Table 5. Cases showing a case of the potential natural frame restoration of *efeU* ORF in the *E. coli* (PATRIC ID 562.11502)

Sequence details	Translated peptide (sequence length)
EfeU from genome ID: 562.11502	Complete (276 AA)
Case-1: without A286 deletion and C328 insertion	Complete (276 AA)
Case-2: A286 deletion but no C328 insertion	Truncated (119 AA)
Case-3: C328 insertion but no A286 deletion	Truncated (119 AA)

Supplementary Table 6. List of primers used in this study

Kanamycin specific primers	
k1	CAGTCATAGCCGAATAGCCT
k2	CGGTGCCCTGAATGAACTGC

Gene specific primers	
entC (U)	GGCGCAGGACATCACATTGC
entC (D)	CTACACGCGAGGTTATCCGC
menF (U)	ACTATCGGGCGAAGCAGGCA
menF (D)	TTAACGGTGTAGAACGCGAG
ubiC (U)	CTGGCATCCTGGACGGTGAT
ubiC (D)	CCGGCAGCGCGCATCAGCCA
*efeU*_1 (U)	ACACCCGCTTATCAGTTTTA
*efeU*_1 (D)	GACTCCAGCCCTTCCCTTGC
efeU GTAC base insertion specific primer	
Forward primer (used with *efeU*_1 (D)	GCTATCTTAAGCGTACGTAC
Reverse primer (used with *efeU*_1 (U)	CGGCCTCGCTGGGTACGTAC

(U: Upstream primer; D: Downstream primer)

Supplementary reference

1. R. A. LaCroix *et al.*, *Appl Environ Microbiol* **81**, 17 (Jan, 2015).