Stock culture heterogeneity rather than new mutational variation complicates short-term cell physiology studies of *Escherichia coli* K-12 MG1655 in continuous culture

By: Ranno Nahku, Karl Peebo, Kaspar Valgepea, Jeffrey E. Barrick, Kaarel Adamberg and Raivo Vilu

SUPPLEMENTARY METHODS

Genome sequence data analysis. Sequencing reads were compared with the *Escherichia coli* K-12 MG1655 reference genome (GenBank accession no. U00096.2) using the *breseq* analysis pipeline (version 1.00rc7). Source code for *breseq* is freely available online (http://barricklab.org/twiki/bin/view/Lab/ToolsBacterialGenomeResequencing). The online documentation describes the methods used to predict consensus point mutations, small indels, large deletions and new sequence junctions.

Predictions of base substitutions and single-base indels that were polymorphic in the population (present only in some individuals) were performed as detailed in the *breseq* documentation. Specifically, *breseq* was run with strict requirements for polymorphism prediction: only reads with alignments to the reference genome that spanned their entire sequence were considered, aligned bases with Phred quality scores of <30 were ignored, and only predictions supported by E values $\leq 10^{-4}$ were accepted. The empirical per-base error model employed by *breseq* was found to underestimate the rates of indel sequencing errors in homopolymer repeats. Therefore, we chose to exclude putative low-frequency subpopulations with indels in reference sequence homopolymer repeats of six or more bases after manual examination. IS5 insertions were predicted as polymorphic by counting how many reads matched the two new sequence junctions associated with insertion of a new copy of this sequence (one at each end) versus the sequence junction in the original genome.

Mutation validation with Sanger sequencing. Predicted mutations were validated by performing PCR followed by Sanger sequencing. Sanger sequencing was performed by the Estonian Biocentre using an Applied Biosystems 3730xl DNA analyser and the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems).

The PCR mixture contained 2.5 µl 10× PCR buffer [KCl, $(NH_4)_2SO_4$ and 20 mM $MgCl_2$], 2 µl of each deoxynucleotide triphosphate (2.5 mM), 0.1 µl DreamTaq DNA polymerase (5 U μl^{-1}) (all obtained from Fermentas), 2 µl of each primer (10 µM) (Microsynth) (Supplementary Table S2), template DNA (50 ng or single colony) and diethylpyrocarbonate (DEPC)-treated water up to a total volume of 25 µl. The PCR thermoprofile consisted of an initial denaturing temperature of 95 °C for 2 min, followed by 30 amplification cycles: denaturation at 95 °C for 30 s, primer annealing for 30 s at the temperature specified in Supplementary Table S2, extension at 72 °C for 1 min per kb of product, followed by a final

extension period of 10 min at 72 °C. Agarose gel electrophoresis [1% (w/v) agarose] of the PCR product was carried out, and bands were visualized using ethidium bromide (0.2 μg ml^{-1}). A GeneJET PCR Purification kit (Fermentas) was used for purification of PCR products prior to Sanger sequencing.

Subpopulations with base substitutions were confirmed by Sanger sequencing PCR products from mixed population samples (Supplementary Figs S1–S5). Mixed signal peaks in chromatograms at genome positions corresponding to *betA* (choline dehydrogenase), *cspH*/*cspG* (stress protein, member of the CspA family/cold-shock protein), *glyA* (serine hydroxymethyltransferase) and *dppD* (ATP-binding component of the dipeptide ABC transporter) mutations were found in population samples but not in negative control samples that did not contain the subpopulation. Low-scoring putative single-nucleotide polymorphisms (SNPs) in *allD* (ureidoglycolate dehydrogenase) and *recB* (exonuclease V, beta subunit) were rejected using this approach (Supplementary Fig. S6). For *yahE* (predicted protein), this validation procedure was inconclusive. Therefore, PCR products generated from 23 single colonies were combined in groups of three or four for subsequent sequence determination and chromatogram analysis as stated above. IS-related mutations: *flhD*/*uspC* (subunit of flagella regulator ($FlhD_2C_2$)/universal stress protein) and *yadL* (gene of predicted chaperone-usher fimbrial operon), were validated by screening colonies isolated from the stock culture for PCR product length differences.