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 TABLES

Supplementary Table S1. Predicted and validated subpopulations in the A-stat experiment

Top *breseq* predictions of base substitution and single-base indel polymorphisms in the sequenced populations are listed. Positions where a polymorphism with a significant score that passed bias tests was detected in any of the three samples are shown with statistics for all three samples. The methods behind these predictions are described in the *breseq* (1.00rc7) documentation (http://barricklab.org/twiki/bin/view/Lab/ToolsBacterialGenomeResequencing). Predictions are shaded based on whether they passed two specific test criteria: (1) $-\log_{10}$ E value score ≥4.0 for the likelihood ratio test of a mixed-base model (that there was a true subpopulation with the variant base) versus a single-base model (that discrepancies from the reference were sequencing errors), and (2) *P* value <0.05 for Fisher's exact test for bias in the strands of reads supporting the variant subpopulation versus the reference sequence. Green passed both tests, blue failed just the polymorphism test, red failed just the strand bias test, and no variant bases were observed for grey text. Overall scores were calculated as sums of the scores for predictions across all samples. Columns: **?**, results of experimentally testing specific predictions, validated as detailed in the main text and supplement (√), not detected by Sanger sequencing of the mixed-population sample and therefore likely spurious (X) or not experimentally tested because of low Mut % and Overall score (NT); Position**,** in the reference genome sequence; Mut, change in sequence from reference genome; Overall score, as explained above; Mut %, maximum-likelihood predicted frequency of subpopulation with mutation; Score, $-\log_{10}$ E value for polymorphism test; Q30 Mut, number of bases with Phred quality scores ≥30 supporting the new subpopulation in reads on the top/bottom genome strand; Q30 Ref, number of bases with Phred quality scores ≥30 supporting the reference base in reads on the top/bottom genome strand; Annotation, amino acid and codon changes for base substitution mutations in genes, location of mutations in reading frame for indels, or distance to the two neighbouring genes for intergenic predictions with –/+ for orientation upstream or downstream of those genes, respectively; Gene, gene containing mutation or two neighbouring genes separated by a solidus (/); Description, description of gene containing mutation or two neighbouring genes separated by a solidus.

				Stock (0 gen)				Dilution rate (*D*)=0.11 h^{-1} (18 gen)				Dilution rate (*D*)=0.48 h^{-1} (34 gen)						
?	**Position**	**Mut**	**Overall score**	**Mut %**	**Score**	**Q30 Mut**	**Q30 Ref**	**Mut %**	**Score**	**Q30 Mut**	**Q30 Ref**	**Mut %**	**Score**	**Q30 Mut**	**Q30 Ref**	**Annotation**	**Gene**	**Description**
√	335361	T→C	79.6	8.1%	12.5	3/3	44/24	9.5%	45.3	9/7	98/55	7.0%	21.8	6/3	91/28	Coding F71F (TT**T**→TT**C**)	*yahE*	Predicted protein
√	1050465	T→A	65.8	2.9%	3.9	2/1	61/39	4.4%	24.8	9/0	129/68	10.7%	37.1	6/6	64/36	Intergenic (–67/–219)	*cspH*/*cspG*	Stress protein/DNA-binding transcriptional regulator
√	3701283	G→A	40.8	17.0%	19.4	6/2	27/12	10.3%	15.4	4/3	44/17	7.8%	6.0	1/3	29/18	Coding L197L (**C**TG→**T**TG)	*dppD*	Dipeptide transporter
√	2683035	G→A	30.2	–	–	0/0	19/12	9.2%	15.2	4/3	34/35	14.9%	15.0	4/3	19/21	Coding H165H (CA**C**→CA**T**)	*glyA*	Serine hydroxymethyltransferase
√	326446	C→T	28.8	2.3%	–	0/1	13/29	5.7%	16.5	1/7	33/108	9.7%	12.3	4/2	14/42	Coding G9D (G**G**T→G**A**T)	*betA*	Choline dehydrogenase
NT	3410579	T→A	26.5	4.4%	6.6	2/2	62/25	1.5%	3.1	2/1	110/86	4.2%	16.8	2/5	86/74	Intergenic (+20/–64)	*yhdJ*/*yhdU*	Predicted methyltransferase/ predicted membrane protein
NT	3237537	T→G	23.1	2.9%	1.0	1/1	21/45	4.1%	21.0	6/1	79/86	1.7%	1.1	2/0	52/65	Coding N312K (AA**T**→AA**G**)	*alx*	Predicted inner membrane protein
X	2952301	T→G	5.9	7.9%	5.9	2/1	9/26	–	–	0/0	39/52	1.5%	–	0/1	24/33	Coding E575D (GA**A**→GA**C**)	*recB*	Exonuclease V, beta subunit
NT	3590376	C→T	5.6	1.3%	–	0/2	76/71	2.3%	5.6	2/3	84/127	1.1%	–	1/1	88/99	Intergenic (–28/+371)	*ugpB*/*livF*	Glycerol-3-phosphate transporter subunit/branched-chain amino acid transporter subunit
NT	432228	Δ1 bp	5.5	–	–	0/0	33/38	0.9%	–	0/1	63/48	4.1%	5.5	1/2	25/45	Coding (3/450 nt)	*nrdR*	Conserved protein
NT	3701921	G→A	4.8	6.1%	4.8	1/3	43/18	–	–	0/0	65/30	–	–	0/0	38/22	Coding *288C (TGA→TG**T**)	*dppC*	Dipeptide transporter
X	545057	A→G	4.4	6.5%	2.5	1/2	22/21	7.1%	1.9	0/3	16/23	3.6%	–	0/1	10/17	Coding T177T (AC**T**→AC**C**)	*allD*	Ureidoglycolate dehydrogenase
NT	1962621	C→A	4.2	–		0/0	16/26	4.6%	4.2	1/3	38/45	–	–	0/0	19/29	Coding G152W (**G**GG→**T**GG)	*flhA*	Predicted flagellar export pore protein

Supplementary Table S2. PCR primers used in HT DNA sequencing data validation

Genome position of mutation*	Mutation	Related gene(s)†	Left primer	Right primer	Product length (bp)	Product length in case of insertion/deletion (bp)	Annealing temperature (°C)
326446	C→T	*betA*	CATCTGGGTGCGGAAGTC	GCTACAAACACTCCGGCATT	264		54
335361	T→C	*yahE*	CAGAAAGACGGTGTGTGTGG	CTGAATAACTGCCCGTCGTT	470		60
547694	A→G	*ylbE_1*	GCAATAAGGGGTTCCCATGT	GCATCACTTTGCCGATACCT	421		55
547835	+G	*ylbE_1*					
1050465	T→A	*cspH/cspG*	CGGAGGGGATAATGAATCCT	GGGGATAATGCCTGATAATGA	272		54
2683035	G→A	*glyA*	AACAACGTGAGCATGAGGAA	ACCAGGTGATACCGTTCTGG	337		54
3422257	ATC→CAT	*rrlD*	ATGACAACCCGAACACCAGT	TCGCTCAACGGATAAAAGGT	283		55
3701283	G→A	*dppD*	GGAACGACACCTGGCAAC	AGCGATTAAGGTGCATCAGG	483		60
3957957	C→T	*ppiC/yifN*	ATATGCAGTGCTGCTGCTGT	ACATACAGAGGGCGGTCATC	280		55
1976527–1977302	IS*1* deletion	*flhD/uspC*	TCGACGCAACTGTACTCGTC	GCAACATCCCATTTCGATTA	988	213	50
1977510–1977513	IS*5* insertion	*flhD/uspC*	CGCTAGCAAGCAAAAATGAA	CCCGGTGATCATATAATTTCAA	286	1481	54
4294305– 4294415	Δ111 bp	*gltP/yjcO*	TGGCTTATGAGCGTGAAGTG	CACTGGTTGAACCTGAGCTG	769	658	54
4294291	T→C	*gltP/yjcO*					
152081–152084	IS*5* insertion	*yadL*	GGCGACAGCACTTGCTTTAT	GTTGCGCTACGTGAAACAGA	482	1677	55
545057	A→G	*allD*	CCGAGTAAGACGCCTGAGAG	GCAATTTTACATGCCGACAA	494		54
2952301	T→G	*recB*	CCAGATTTGCCGATAACCAT	TGAAAATGTGGCTGATGGAA	461		54

*Positions of IS*5* insertions give the target site nucleotides that were duplicated upon insertion of the new IS copy. Both new IS*5* copies inserted in the forward (+) orientation in the genome.

†*Gene1*/*Gene2* indicates that the mutation was in the intergenic region.