Dynamics and genetic diversification of *Escherichia coli* during experimental adaptation to an anaerobic environment

Thomas J. Finn[1,2,3], Sonal Shewaramani[1,2,4], Sinead C. Leahy[1], Peter H. Janssen[1] and Christina D. Moon[1]

[1] Grasslands Research Centre, AgResearch Ltd, Palmerston North, New Zealand
[2] New Zealand Institute for Advanced Study, Massey University, Auckland, New Zealand
[3] Department of Biochemistry, University of Otago, Dunedin, New Zealand
[4] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, United States of America

ABSTRACT

Background. Many bacteria are facultative anaerobes, and can proliferate in both anoxic and oxic environments. Under anaerobic conditions, fermentation is the primary means of energy generation in contrast to respiration. Furthermore, the rates and spectra of spontaneous mutations that arise during anaerobic growth differ to those under aerobic growth. A long-term selection experiment was undertaken to investigate the genetic changes that underpin how the facultative anaerobe, *Escherichia coli*, adapts to anaerobic environments.

Methods. Twenty-one populations of *E. coli* REL4536, an aerobically evolved 10,000th generation descendent of the *E. coli* B strain, REL606, were established from a clonal ancestral culture. These were serially sub-cultured for 2,000 generations in a defined minimal glucose medium in strict aerobic and strict anaerobic environments, as well as in a treatment that fluctuated between the two environments. The competitive fitness of the evolving lineages was assessed at approximately 0, 1,000 and 2,000 generations, in both the environment of selection and the alternative environment. Whole genome re-sequencing was performed on random colonies from all lineages after 2,000-generations. Mutations were identified relative to the ancestral genome, and based on the extent of parallelism, traits that were likely to have contributed towards adaptation were inferred.

Results. There were increases in fitness relative to the ancestor among anaerobically evolved lineages when tested in the anaerobic environment, but no increases were found in the aerobic environment. For lineages that had evolved under the fluctuating regime, relative fitness increased significantly in the anaerobic environment, but did not increase in the aerobic environment. The aerobically-evolved lineages did not increase in fitness when tested in either the aerobic or anaerobic environments. The strictly anaerobic lineages adapted more rapidly to the anaerobic environment than did the fluctuating lineages. Two main strategies appeared to predominate during adaptation to the anaerobic environment: modification of energy generation pathways, and inactivation of non-essential functions. Fermentation pathways appeared to alter through selection for mutations in genes such as *nadR, adhE, dcuS/R*, and *pflB*. Mutations were frequently identified in genes for presumably dispensable functions such as toxin-antitoxin systems, prophages, virulence and amino acid transport. Adaptation of the

fluctuating lineages to the anaerobic environments involved mutations affecting traits similar to those observed in the anaerobically evolved lineages.

Discussion. There appeared to be strong selective pressure for activities that conferred cell yield advantages during anaerobic growth, which include restoring activities that had previously been inactivated under long-term continuous aerobic evolution of the ancestor.

INTRODUCTION

Advances in our understanding of adaptation to novel environments, and the population dynamics and genetic changes that underpin these, have been significantly enhanced by long-term experimental evolution (LTEE) studies (*Barrick et al., 2009*; *Blount, Borland & Lenski, 2008*; *Lenski et al., 1991*). Such studies have typically employed highly defined and reproducible experimental conditions with regard to growth media, temperature and pH. Only a few have examined the impact of oxygen availability on adaptation, where low oxygen contrasts have been used (*Manché, Notley-McRobb & Ferenci, 1999*; *Puentes-Téllez et al., 2013*). Facultatively anaerobic organisms, such as *E. coli*, grow in both aerobic and anaerobic environments. They are metabolically versatile, in particular, possessing distinct energy generating systems required to thrive in each environment (*Smith & Neidhardt, 1983*). In the presence of oxygen, energy is conserved in the form of ATP via aerobic respiration which involves glycolysis, the tricarboxylic acid (TCA) cycle, and the electron transport chain (*Gunsalus, 1992*; *Ingledew & Poole, 1984*) with oxygen as the terminal electron acceptor. However, the natural environment of *E. coli* is the lower gut, which is predominantly anaerobic, though microaerobic conditions are also apparent and where oxygen availability may fluctuate (*He et al., 1999*; *Jones et al., 2007*). Under anaerobic conditions, fermentation is the principle process for generating ATP in the absence of oxygen, or other terminal electron acceptors. Fermentation employs glycolysis in conjunction with substrate-level phosphorylation from 1,3-bisphosphoglycerate, phosphoenolpyruvate, and acetyl-phosphate (*Clark, 1989*). The energy yields during growth in aerobic and anaerobic environments differ, with 26 ATP molecules generated per glucose during aerobic respiration (*Kaleta et al., 2013*; *Haddock & Schairer, 1973*). In contrast, anaerobic fermentation of glucose by *E. coli* is less productive, yielding a maximum of three molecules of ATP per glucose, depending on which fermentation pathways are used (*Gunsalus & Park, 1994*). In some facultative anaerobes, organic compounds such as citrate can be metabolised under anaerobic conditions and thus increase energy yields when they are co-metabolised with other substrates e.g., glucose (*Bott, 1997*; *Lütgens & Gottschalk, 1980*).

While higher energy yields are obtained during aerobic respiration, natural by-products from this process include reactive oxygen species (ROS), which are mutagenic and can result

in distinct mutational spectra (*Maki, 2002*; *Sakai et al., 2006*; *Shewaramani et al., 2017*). Despite this, the spontaneous mutation rate of anaerobically grown *E. coli* is higher than that of aerobically grown *E. coli* (*Sakai et al., 2006*; *Shewaramani et al., 2017*), although the mutagenic agents in the anaerobic environment are not fully understood. It is hypothesised that the different mutation spectra and rates imposed by each environment are likely to impact the evolutionary trajectory of organisms during adaptation.

Independent *E. coli* populations that are allowed to evolve in a controlled experimental environment have been shown to take different adaptive paths to achieve increased fitness (*Elena & Lenski, 2003*; *Lenski & Travisano, 1994*). Functions accounting for these adaptations include DNA supercoiling (*Crozat et al., 2005*), modification of the stringent response (*Cooper, Rozen & Lenski, 2003*), increased substrate specificity (*Cooper et al., 2001*) and substrate switching (*Blount, Borland & Lenski, 2008*). The rate at which a population adapts to an environment, and the path(s) taken, are also influenced by the degree to which the population is exposed to an environment. Prolonged exposure to a constant environment is likely to promote the evolution of specialists that have high fitness in the environment of selection (*Cooper & Lenski, 2000*), but decreased fitness in alternative environments. This may arise due to selection for beneficial mutations in the environment that have detrimental effects in alternative environments (*Elena & Lenski, 2003*). However, varying exposure to a range of environments will likely result in the evolution of generalists (*Kassen, 2002*; *Kassen & Bell, 1998*), which may evolve unique adaptations that are beneficial in the range of environments to which it is exposed.

In the present study, we sought to better understand the adaptation of *E. coli* to environments that vary in oxygen availability, to provide insights into evolution in its natural habitat, which consists of passage through gastrointestinal tracts. Experimental growth of *E. coli* was modelled on the prominent LTEE study of Richard Lenski (*Lenski et al., 1991*), using an aerobically evolved derivative of his ancestral strain, and maintained on the same minimal glucose medium. However, parallel populations were also evolved under strictly anaerobic growth conditions, and under a treatment that fluctuated exposure to both aerobic and anaerobic environments. We report the genes and mutations that are important for adaptation to anaerobic environments, and provide new insights into the evolutionary pathways involved in adaptation during aerobic and anaerobic growth.

METHODS

Strains and growth media

The ancestral strain used in this study was *E. coli* strain REL4536 (*Barrick et al., 2009*), an aerobically evolved 10,000th generation descendent of the *E. coli* B strain, REL606 (*Lenski et al., 1991*). Cultures were propagated aerobically in a biosafety cabinet, and anaerobically in an anaerobic chamber (Coy Laboratories, Grass Lake, MI, USA) with a 95% CO_2:5% H_2 atmosphere. DM25 media (*Lenski et al., 1991*) based on Davis minimal media (DM; *Carlton & Brown, 1981*) were used to propagate populations in the experimental evolution experiment. For anaerobic preparation of DM25, the base DM broth (*Carlton & Brown, 1981*) was boiled and cooled to room temperature while bubbling CO_2 through it. Resazurin

(0.00002% wt/vol) was added as an oxygen indicator, and the medium was sterilised by autoclaving. Before use, L-cysteine HCl (0.025% wt/vol), was added as a reducing agent. Anaerobically-prepared magnesium sulfate (0.01% wt/vol, final concentration), thiamine (0.002% wt/vol final concentration), and glucose (0.0025% wt/vol final concentration) were filter sterilised, then added. DM1000 media used for the generation of a neutrally-marked strain for competition assays was prepared like DM25 but contained 0.1% (wt/vol) glucose.

DM agar plates were prepared as for DM media, with the addition of bacteriological grade 1.6% (wt/vol) agar and 0.4% (wt/vol) glucose, combined after sterilisation via autoclaving. DM plates were used for the isolation of randomly-selected clones for genome sequencing, where aerobic plates were incubated in air, and anaerobic plates were incubated in AnaeroJar™ gas canisters (Oxoid, Cheshire, UK), with oxygen indicator strips (Oxoid) to monitor oxygen conditions. The plates were placed in the jars in the anaerobic chamber, so the gas composition was the same as in the anaerobic chamber.

Lysogeny broth (LB) and LB agar plates (*Bertani, 1951*) were used for cultivating clones for genomic DNA extraction, and enumerating bacteria after competitive fitness experiments, respectively.

Long-term evolution experiment

A single randomly-selected colony of *E. coli* REL4536 was inoculated into 9.90 mL DM25, and incubated with shaking overnight at 37 °C to generate an ancestral REL4536 culture. From this, 21 independent *E. coli* populations were established for the LTEE, with seven populations propagated under each treatment regime: aerobic growth (AE), anaerobic growth (AN) and growth with temporally fluctuating exposure to aerobic and anaerobic environments (FL). Establishment of the long-term AE and FL populations took place aerobically, with inoculation of 10 μL of the ancestral culture into 990 μL of aerobic DM25 in 24-well plates (Becton, Dickinson and Co., Sparks, MD, USA). An uninoculated (medium-only) control well was included alongside the seven populations in each treatment to monitor for cross-contamination. The AN cultures were established in anaerobic DM25 medium, and cultures grown anaerobically in anaerobic gas boxes (AnaeroPack® 2.5 L, Mitsubishi Gas Company Inc., Tokyo Japan). All cultures were incubated at 37 °C with orbital shaking at 150 rpm, and propagated by daily serial subculture of 1/100th volume of the previous day's culture into fresh growth medium. The FL treatment was alternated daily between aerobic and anaerobic growth. The lineages were propagated for approximately 2,000 generations (ca. 300 days, at ∼6.67 generations per day). To monitor for external contaminants, growth within the medium-only wells was monitored, and routine testing with coliphage T5 and T6 was performed (*Lenski et al., 1991*). Aliquots of evolved lineages were stored in 70% (vol/vol) glycerol at −85 °C every two weeks.

Competitive fitness assays and calculating fitness

The fitness of the evolved populations relative to the ancestral strain was assessed both aerobically and anaerobically after approximately 1,000 and 2,000 generations. Spontaneous arabinose-utilising (Ara⁺) strains were generated from the ancestral strain to competitively assess fitness, following the approach used by *Lenski et al. (1991)*. However, these displayed

poor fitness in the anaerobic environment relative to the unmarked strain. Therefore, a rifampicin (Rif) resistant spontaneous mutant of REL4536 was obtained by plating cells on DM agar plates containing 100 μg/mL rifampicin (DM+Rif). The fitness of the Rif^r competitor relative to REL4536 was 0.97 ± 0.05 and 1.04 ± 0.05 under aerobic and anaerobic growth conditions, respectively, and deemed selectively neutral and used for all competitive fitness assays performed in this study.

For competition assays, competing strains were acclimated to the growth conditions by inoculating frozen stocks into wells of 24-well plates containing 1-ml volumes of either DM25 for competing strains, or DM25+Rif for the Rif^r competitor. Cultures were grown overnight at 37 °C with 150 rpm orbital shaking. At 0 h, 10 μL of each competitor culture was mixed at an estimated cell ratio of 1:1 (evolved strain:Rif^r ancestral strain) in fresh DM25 media, resulting in a 1 ml total competition volume. Competitors were enumerated at 0 and 24 h by plating serial dilutions in duplicate onto both LB and LB+Rif plates, followed by overnight incubation at 37 °C. The ancestor (Rif^r) cell densities were calculated directly from growth on LB+Rif plates, while the cell density of the competitor was derived from subtracting the Rif^r cell density from the total cell density (on LB plates). The Malthusian parameters (m) for each competitor was calculated as: $m = ln\left(\frac{\text{Cell density at 24 h}}{\text{Cell density at 0 h}}\right)$. Relative fitness ($\omega$) of the evolved strain relative to the ancestral was calculated as $\omega = \frac{m(\text{Evolved})}{m(\text{Ancestral})}$ (*Lenski et al., 1991*). Anaerobic competitive fitness assays were carried out for anaerobic and fluctuating lineages as above, except using anaerobic practices as described previously.

Whole genome sequencing and mutation detection

To identify mutations that arose during the LTEE, single, randomly selected clones were isolated from each of the 21 lineages at 2,000 generations. Genomic DNA was extracted from clones cultured in LB using a Qiagen Genomic-tip 100/G kit (Qiagen, Hilden, Germany). DNA was sequenced on an Illumina HiSeq 2000 instrument using paired-end sequencing with 90 bp reads of 500 bp paired-end libraries (BGI, Shenzhen, China). Approximately 1 Gb of clean sequence data (filtered to remove reads containing ≥10% unreadable bases, ≥20% low quality (≤Q20) bases, adapter contamination or duplicate read-pairs) were obtained for each sample, representing over 200 × fold coverage over each genome. The FastQC tool (*Andrews, 2010*), was used to assess FASTQ file quality.

The reference *E. coli* REL4536 sequence was obtained by manually editing the *E. coli* REL606 genome sequence (GenBank accession number NC_012967.1) to incorporate mutations that arose in REL4536 relative to REL606 (*Barrick et al., 2009*). Mutations in the genomes of the evolved strains were identified relative to REL4536 using breseq v.014 (*Barrick et al., 2009*). Biological processes affected by mutations in evolved lineages were inferred using the PANTHER protein (and gene) evolutionary and functional classification system with the "Functional classification viewed in gene list" analysis option (*Mi et al., 2016*). Genome sequence data generated in this study are available in NCBI BioProject PRJNA357967.
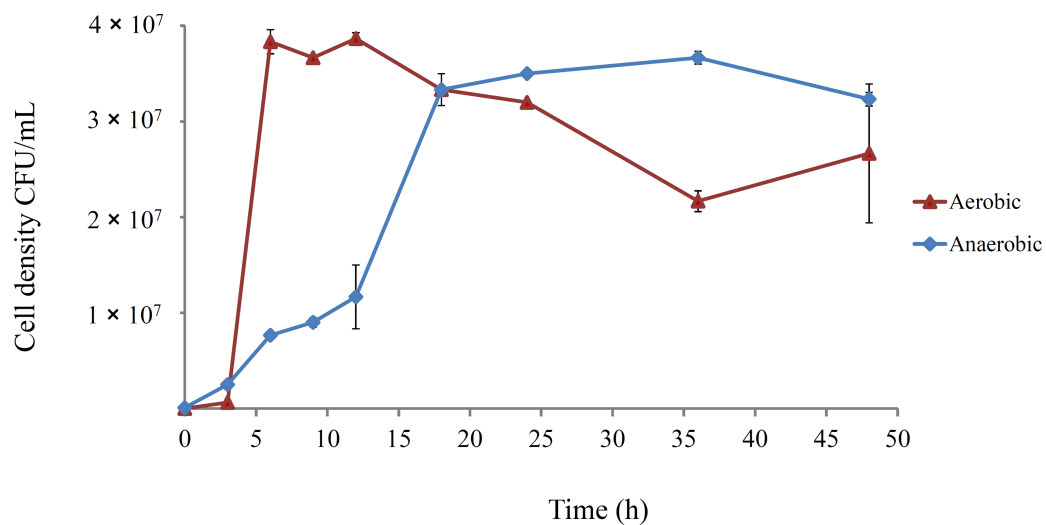
Figure 1 **Growth course of _E. coli_ REL4536 under aerobic and anaerobic conditions in DM25 media at 37 °C.** Data points represent mean values from three biological replicates. Error bars represent one standard error of the mean in each direction.

Statistical analyses

Fitness data for competition assays held in both aerobic and anaerobic environments were checked for normal distribution using a Shapiro–Wilk test (_Shapiro & Wilk, 1965_). To determine if lineages had adapted to the aerobic or anaerobic environments, repeated measurements analysis of variance (RM ANOVA) tests were performed, with relative fitness of evolved lineages in either environment as a response, and lineage and generation as factors. Fisher's Least Significant Difference (LSD) post hoc tests were performed to identify which lineages had increased in relative fitness. Differences in the rates at which adaptation occurred were examined using paired t-tests. To identify if the different lineage treatments had statistically significant mutation types, single factor ANOVAs were performed followed by LSD post hoc tests to identify which lineages had significant differences. Comparisons of the proportion of mutations was analysed with a Fisher's exact test, which is suitable for small sample sizes. Statistical analyses were performed in either GenStat (17th edition), Minitab (17th edition) or Microsoft Office Excel (version 14).

RESULTS AND DISCUSSION

Growth dynamics of REL4536 in the anaerobic environment

The growth dynamics of _E. coli_ B strains in DM25 under aerobic conditions are well-documented (_Lenski et al., 1998_; _Lenski et al., 1991_; _Vasi, Travisano & Lenski, 1994_). However, anaerobic growth has not been well characterised, and was investigated for REL4536 in DM25 (Fig. 1).

Aerobically-grown cultures reached a maximum density of approximately 4.0×10^7 CFU/mL within 6 h. In the anaerobic environment, the cell density reached approximately 3.5×10^7 CFU/mL over 20 h, and had a lower growth rate than the aerobically grown cells, due to metabolic differences under each condition (_Gray, Wimpenny & Mossman, 1966_;

Gunsalus & Park, 1994). However, the similar cell densities attained under the contrasting environments was unexpected, and likely due to the ability of *E. coli* to metabolise citrate when grown anaerobically (*Blount, Borland & Lenski, 2008*; *Lütgens & Gottschalk, 1980*; *Vaughn et al., 1950*) in DM25, which contains 1.7 mM citrate. REL4536 grew to a similar cell density in DM25 devoid of glucose (DM0), where citrate was the only carbon source available. Because both the aerobic and anaerobically grown cultures reached similar cell densities within 24 h (Fig. 1), serial sub-culturing was performed every 24 h in the LTEE of this study. Under this regime, the anaerobic lineages spent less time in stationary phase than the aerobic lineages as it was not practical to subculture more frequently (e.g., every 6–8 h). As such, interpretation of the data based on the different oxic treatments may be confounded by the differing lengths of time at stationary phase associated with each treatment. The aerobic lineages, being subject to longer periods of starvation, may have had more opportunity to generate growth advantage in stationary phase (GASP) mutations (*Zambrano & Kolter, 1996*). While such possibilities exist, GASP mutations are more commonly observed in cells that have been exposed to prolonged starvation over many days (*Finkel, 2006*). Furthermore, as significant fitness increases for the aerobic lineages were not observed (see below), the impact of the difference in stationary phase exposure between treatments was considered minimal.

Fitness dynamics of evolved populations in aerobic and anaerobic environments

The 21 independent *E. coli* REL4536 lineages were evolved for approximately 2,000 generations over a period of 300 days under the AE, AN and FL treatments of this study. To assess the fitness dynamics of the evolved lineages, the fitness of three randomly-selected lineages per treatment were determined relative to a Rif^r ancestral strain at approximately 0, 1,000 and 2,000 generations in both aerobic and anaerobic environments (Fig. 2).

In the aerobic environment, significant increases in mean relative fitness were not observed among the tested lineages in any treatments over the 2,000 generations of evolution (RM ANOVA, $F_{(4,12)} = 2.22$, $P = 0.163$) (Table 1). The lack of fitness increase among the AE and FL lineages is expected because REL4536 had previously been adapted to the same medium under aerobic growth conditions for over 10,000 generations (*Lenski et al., 1991*; *Barrick et al., 2009*). No increase in mean fitness was reported among lineages during a similar 2,000 generation timeframe, where mean fitness was ~1.77 at 10,000 generations and ~1.72 at 12,000 generations (*Barrick et al., 2009*). When the AN lineages were competed in the aerobic environment, no significant change in fitness was seen, suggesting the ancestral fitness level in the aerobic environment was largely retained among anaerobic lineages. This may have arisen due to the substitution of mutations with pleiotropic effects during anaerobic adaptation, suggesting that beneficial mutations that occurred in the anaerobic environment were presumably neutral, weakly beneficial or weakly deleterious in the aerobic environment, and therefore, did not adversely affect AN lineage performance in the aerobic environment.

In the anaerobic environment, there was a significant difference between the mean relative fitness of lineages over 2,000 generations (RM ANOVA, $F_{(4,12)} = 8.32$, $P = 0.002$)
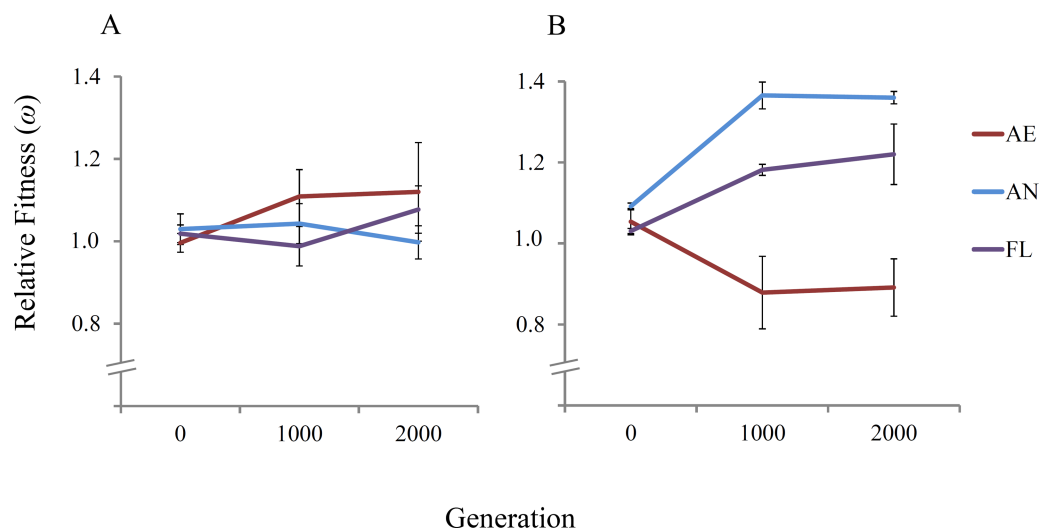
Figure 2 Mean relative fitness of lineages over 2,000 generations. Displayed are the mean relative fitness values of three randomly-selected lineages per treatment (AE2, AE3 and AE7; AN1, AN4 and AN6; and FL2, FL3 and FL7) relative to the Rif[r] ancestor in (A) the aerobic environment and (B) the anaerobic environment. Error bars represent one standard error of the mean in each direction.

Table 1 RM ANOVA of fitness of lineages over 2,000 generations in aerobic and anaerobic conditions.

Source	Degrees of freedom	Aerobic		Anaerobic	
		F	P	F	P[a]
Lineage	(2, 12)	0.31	0.747	28.58	<0.001*
Generation	(2, 12)	1.47	0.273	4.13	0.045*
Generation × Lineage	(4, 12)	2.22	0.163	8.32	0.002*

Notes.
[a] Asterisk indicates statistical significance at $P < 0.05$.

(Table 1). Fitness differences were detected at 1,000 generations, where AN and FL lineages had significantly higher mean fitness at 1,000 generations than at generation 0 (Table 2). The increased fitness of the AN and FL lineages under anaerobic conditions is consistent with rapid fitness increases observed by populations exposed to novel environments and suggests that many beneficial mutations were acquired in the first 1,000 generations of the experiment. Mean fitness increased from 1.091 to 1.408, ~30% over the first 1,000 generations for the AN lineages (Table 2), which is in line with fitness increases reported in LTEE studies that expose populations to a novel environment (*Puentes-Téllez et al., 2013*; *Lenski et al., 1991*; *Melnyk & Kassen, 2011*). Furthermore, the rapid adaptation under anaerobic conditions may be facilitated by the elevated spontaneous mutation rate during anaerobic growth where it is almost double that for aerobic growth for REL4536 (*Shewaramani et al., 2017*) where higher mutation rates can lead to more rapid rates of fitness increase (*Gerrish & Lenski, 1998*).

The FL lineages displayed increased fitness in both environments (8% in the aerobic environment and 22% in the anaerobic environment), but only the fitness increases in

Table 2 Fisher's LSD test of mean fitness of lineages over 1,000 generations relative to initial mean fitness levels in the anaerobic environment.

| Lineages	Mean fitness at generation 0 ($\omega \pm$ SEM)	Mean fitness at generation 1,000 ($\omega \pm$ SEM)	Difference[a]
AE	1.054 ± 0.032	0.879 ± 0.089	0.175*
AN	1.091 ± 0.008	1.408 ± 0.043	0.317*
FL	1.031 ± 0.006	1.182 ± 0.013	0.151*

Notes.

[a]Differences are significant if greater than 0.150, and are indicated with an asterisk.

the anaerobic environment were statistically significant ($P < 0.05$). Furthermore, fitness increases of the FL lineages in the anaerobic environment were not as great as for the AN specialists. This is consistent with evolutionary theory, and with other studies that have exposed experimental populations to novel environments in a variable manner (*Kassen, 2002*; *Elena & Lenski, 2003*; *Ketola et al., 2013*; *Leroi, Lenski & Bennet, 1994*). In the present study, FL lineage fitness increased by ~18% in the first 1,000 generations, as compared to 11% in a similar set up where lineages were exposed to oxygen-rich and oxygen-poor conditions in a rich medium (*Puentes-Téllez et al., 2013*), and 20% where lineages were exposed to alternative sugar sources in a non-constant manner (*Melnyk & Kassen, 2011*).

Mutation analysis of evolved lineages

To identify the full set of mutations that occurred during experimental evolution, whole genome sequencing of single randomly-selected 2,000 generation clones from each of the 21 lineages was performed. Relative to the REL4536 genome, 209 mutations were identified among the 21 genomes. Of these, 31, 94 and 84 were observed in the AE, AN and FL treatments, respectively (Tables S1, S2 and S3). Of the seven AN lineages, three of the randomly-selected sequenced clones (AN-4, AN-6 and AN-7) displayed a small colony morphology when incubated aerobically on solid LB medium. However, no mutations were found in common between these three genomes and so the genetic basis for this phenotype is unclear.

On average, AE, AN and FL lineages accumulated 4.4, 13.4 and 12 mutations per lineage respectively. When compared to other studies that exposed populations to novel environments, approximately six mutations per genome were identified within 2,000 generations during aerobic adaptation to minimal glucose medium (*Barrick et al., 2009*), and ~11 mutations per genome observed within 2,000 generations during adaptation to an environment with elevated temperature (*Tenaillon et al., 2012*). Within the present study, presuming none of these mutations directly impacted mutation rate, the much greater number of mutations in the AN and FL lineage genomes compared to the AE likely reflects selection for mutations that are adaptive in the novel environment, as compared to the aerobic environment, which the ancestor had been extensively pre-adapted (*Lenski et al., 1991*). Furthermore, the spontaneous mutation rate during anaerobic growth is almost double that during aerobic growth for REL4536 (*Shewaramani et al., 2017*), thus providing greater genetic variation in anaerobically grown cells upon which selection can act to facilitate adaptation.
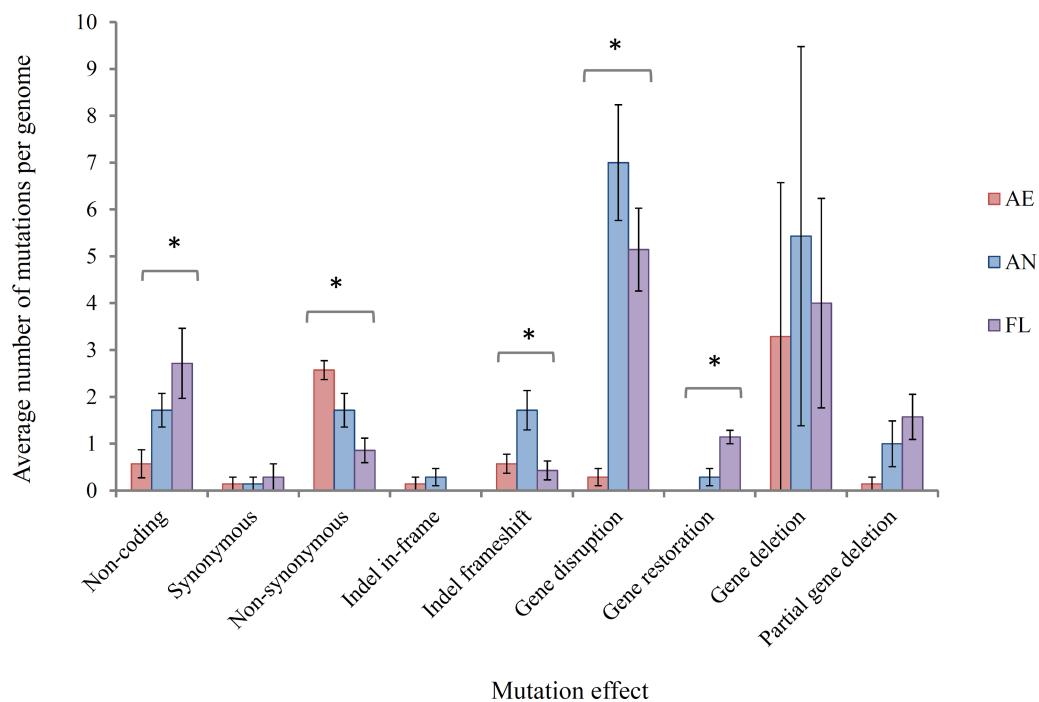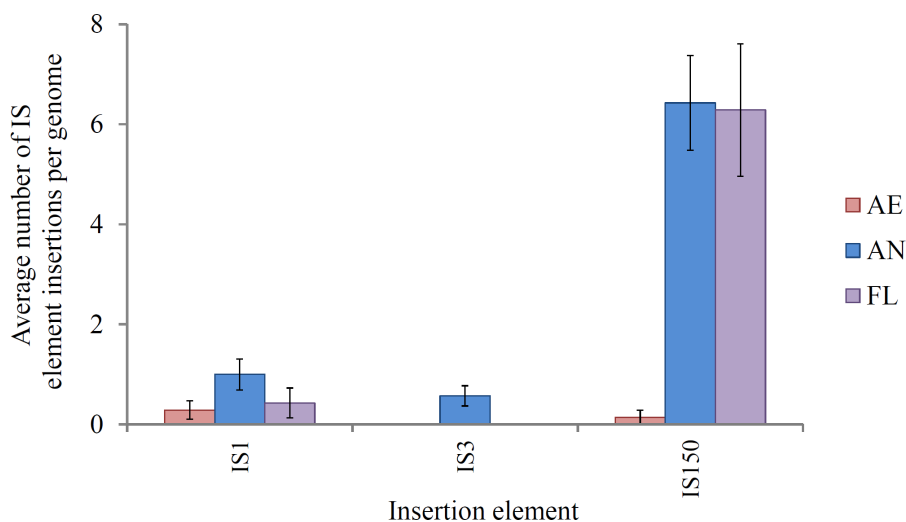
Figure 3 Average number of mutations per genome in lineages propagated under each treatment for 2,000 generations. Error bars represent one standard error of the mean of seven clones.* Significant at the $P = 0.05$ level, single factor ANOVA.

Overall, mutations were considered on three levels: (i) their effect on gene/protein function, (ii) their effect on biological processes and (iii) occurrence (parallelism) between lineages, implying an adaptive role.

(i) Effects of mutations at gene/protein level At the level of the gene and protein function, mutations were classified, with predicted impacts on generating adaptive mutations, into the following main types: (a) non-coding DNA, i.e., mutations occurring in intergenic regions that may or may not impact the expression of adjacent genes, (b) synonymous SNPs, which are generally assumed not to alter gene function, (c) non-synonymous SNPs, which may impact gene function to varying degrees of severity from similar amino acid changes to gene truncation, (d) indel in-frame, which may result in altered gene function but does not result in frameshift, (e) indel frameshift, which likely disrupts gene function but, in cases, can restore the function of pseudogenes that have arisen through frameshift mutations, (f) insertion sequence (IS) element-mediated gene disruption, (g) gene restoration where the loss of an IS element restores a gene that had been previously disrupted, (h) complete gene deletion and (i) partial gene deletion. The average number of mutations per genome in each of these classes is shown in Fig. 3.

In this study, it was hypothesised that among the SNPs arising in AE evolved clones, an increase in G:C → T:A and A:T → C:G transversions would be expected due to ROS-mediated damage on DNA for aerobically grown clones relative to AN clones. However, these specific transversions were not found at a higher frequency (Fisher's exact test,

Figure 4 **Average number of IS1, IS3 and IS150 insertions per genome in lineages propagated under each treatment for 2,000 generations.** Error bars represent one standard error of the mean of seven clones.

$P > 0.05$ in all cases), and could likely be explained by highly efficient repair mechanisms for these mutations in the aerobic environment (*Shewaramani et al., 2017*; *Sakai et al., 2006*).

IS-mediated mutations (gene disruptions, restorations and deletions) comprised the largest class of mutations in lineages exposed to the anaerobic environment, consistent with their high spontaneous mutation rate (approximately 8×10^{-4} mutations per genome/generation; *Shewaramani et al., 2017*). This suggests that this class of mutation is likely to play an important role in adaptation. IS element activity is a key feature of the *E. coli* B genome (*Jeong et al., 2009*; *Papadopoulos et al., 1999*; *Studier et al., 2009*). In REL4536, nine different IS element families are present, each with characteristic size, frequency and modes of transposition (*Barrick et al., 2009*; *Mahillon & Chandler, 1998*; *Shewaramani et al., 2017*). Of these, three were active in this LTEE: IS1, IS3 and IS150 (Fig. 4).

IS150 insertions accounted for the majority of IS-mediated mutations in lineages exposed to the anaerobic environment, and this is likely a consequence of the ∼10-fold greater spontaneous mutation rate observed for IS150 during anaerobic growth as compared to aerobic growth (*Shewaramani et al., 2017*). In the study by *Shewaramani et al. (2017)* enhanced IS150 gene expression was observed during anaerobic growth, and it was speculated that further post-transcriptional regulation of transposase expression may be affected by anaerobic physiology, though the mechanism for this is unclear. Conceivably, the accumulation of fermentation end-products, altered pH and slower growth rate may impose stresses or conditions that alter IS150 activity, and these will require further investigation. IS-mediated mutation may offer genome plasticity during adaptation to heterogeneous environments due to their transient nature and ability to insert and remove themselves from various genes within the host genome (*Casacuberta & González, 2013*; *Gaffé et al., 2011*). Movement of IS elements can affect genome structure, and also gene expression by inserting in regulatory elements (*Kinnersley, Holben & Rosenzweig, 2009*).

Genomic hotspots for IS element insertion within REL4536 were apparent as many identical IS150 mutations were identified among different lineages, between different treatments (particularly AN and FL), and even in completely independent studies where identical IS150 mutations were observed during anaerobic mutation accumulation (*Shewaramani et al., 2017*). For example, IS-element mediated deletions of up to 29 genes (approximately 0.74% of genes in REL 4536) were common in the AN and FL lineages. Such mutations may have deleted non-essential DNA and resulted in a smaller genome that may allow the cell to increase their replication rate during anaerobic growth (*Giovannoni et al., 2005*; *Lee & Marx, 2012*; *Mira, Ochman & Moran, 2001*). The prevalence of IS150-mediated mutation among the adaptive lineages suggests such mutations are a strong driving force behind adaptation. Our observations are in contrast to a traditional view where IS elements have been considered genetic parasites that relocate and proliferate within host genomes (*Charlesworth, Sniegowski & Stephan, 1994*) causing potentially deleterious effects on the host cell. However, growing evidence suggests that IS element transposition has profound effects on the adaptation of organisms to their environments (*Casacuberta & González, 2013*; *Cooper et al., 2001*; *Kazazian, 2004*; *Schneider et al., 2000*; *Schneider & Lenski, 2004*). In this LTEE, the AN and FL lineages had the highest numbers of IS mutations, and were also those with the greatest fitness responses in the anaerobic environment.

(ii) Effects of mutations at the biological process level The genome ontology (GO) classification of biological processes that were impacted by mutation during evolution were similar across the three treatments (Fig. 5A).

For all three treatments, metabolic and transport processes accounted for the highest proportion of processes whose genes were impacted by mutation overall. Within the AE lineages, metabolic processes such as anabolism and catabolism, cellular processes such as the cell cycle and transport of biomolecules, and biological regulatory processes such as catalytic activities accounted for 53.8%, 7.7%, 23.1% and 15.4% of affected processes, respectively (Fig. 5). However, the relative proportions and diversity of categories at the molecular function level varied much more among AN and FL lineages than for AE (Fig. 5B).

A wider range of metabolic processes were impacted by mutation in lineages exposed to the anaerobic environment (AN and FL) than in AE lineages, where only primary metabolic processes were affected (Fig. 5B). For lineages exposed to the anaerobic environment, the diversity of transport systems impacted differed from AE lineages, where only amino acid and carbohydrate transport were affected (Fig. 5B). Furthermore, FL lineages had the highest percentage of mutations involved in transport processes (27%), where a notably large proportion was due to mutations in genes for ion transport, such as *gltP*, *btuC* and *zntB* (Table 3). Overall, the FL lineages had a greater diversity of protein functional subcategories impacted, which is consistent with the idea that the fluctuating treatment selects for mutations that underpin broad generalisation to maintain fitness across varied environments.

In contrast to the AE treatment, the AN and FL lineages had mutations in genes in response to stimulus class (Fig. 5A). In AN lineages, mutations were in genes involved in stress response, such as *appY* (*Atlung, Knudsen & Heerfordt, 1997*), and *envY* (*Lundrigan &*
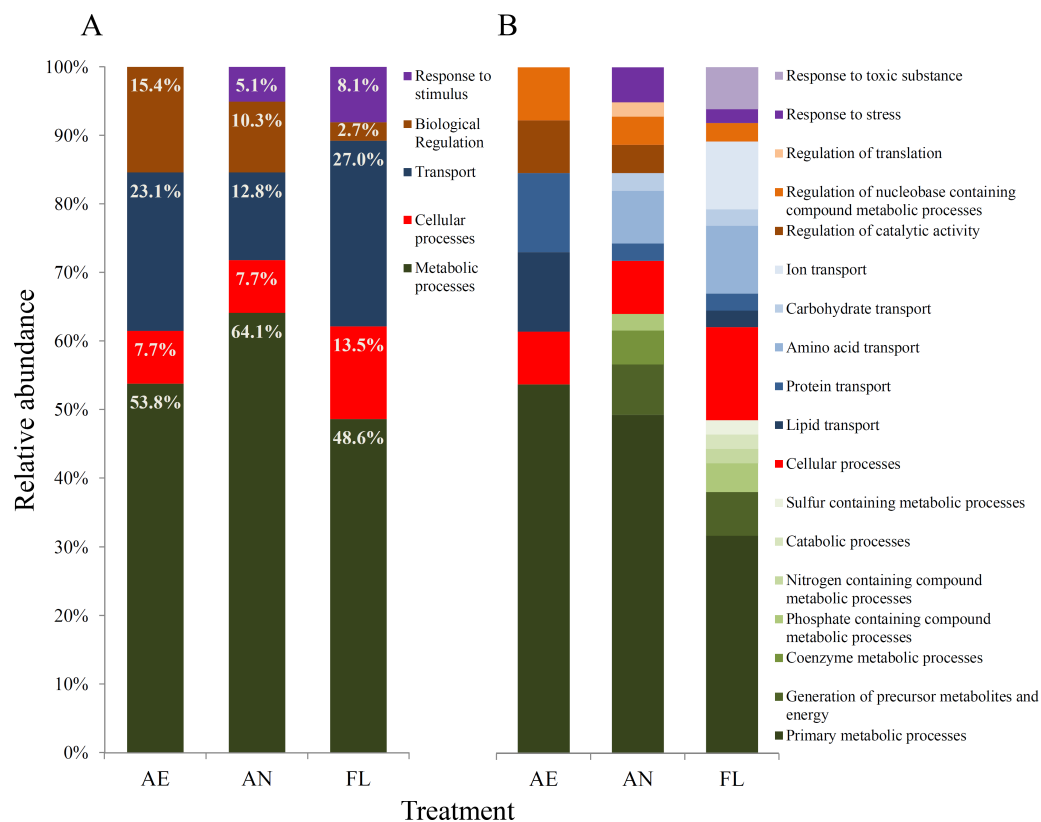
Figure 5 Gene ontology (GO) biological processes (A) and GO molecular functions (B) affected by mutations within genomes of evolved lineages, as defined using the PANTHER protein classification system (*Mi et al., 2016*).

Earhart, 1984), while in FL lineages, additional mutations were found in mismatch repair gene, *mutL,* and genes classified in response to toxic substances, e.g., *btuE* which encodes a putative glutathione peroxidase (Fig. 5B).

(iii) Parallelism between lineages The predicted roles that mutations had on the fitness of evolving lineages was explored. Adaptive mutations in evolving lineages were inferred based on the extent of parallelism among mutations, and in consideration of the treatments under which they were identified (Table 3). Two general mechanisms of adaptation are hypothesised from these analyses: altered metabolism under the environment of selection, and the deletion or disruption of redundant functions.

Altered metabolism

 (a) **Modification of fermentation** Mutations in genes involved in fermentation were found only among lineages that were exposed to the anaerobic environment. In particular, in both the AN and FL lineages, mutations that are predicted to (a) disrupt *nadR* and *adhE,* and (b) restore *dcuS,* and *pflB,* were identified multiple times in independent lineages (Table 3). The roles of these genes in fermentation are shown in Fig. 6. It is further noted that mutations in similar genes and functions (e.g., *adhE* and NAD$^+$ metabolism) also

Table 3 Adaptive mutations reported in AE, AN and FL lineages.

Predicted mechanism[a]	Category[b]	Gene/Region	Gene function[c]	Treatment[d]	Mutation	Position[e]	Lineage[f]
Altered metabolism	Fermentation pathways	*nadR*	Upregulates fermentation network *via* NAD	Anaerobic	IS150 insertion	4,581,545	AN-2K-3, 4, 5 and 6
Altered metabolism	Fermentation pathways	*adhE*	Disrupts ethanol production	Anaerobic	SNP	1,438,030	AN-2K-1, 3, 4 and 5
Altered metabolism	Fermentation pathways	*adhE*	Disrupts ethanol production	Anaerobic	SNP	1,439,673	AN-2K-2 and 7
Altered metabolism	Fermentation pathways	*dcuS* [g]	Induces succinate production	Anaerobic	Deletion	4,295,377	AN-2K-1, 2, 3, 4, 5 and 6
Altered metabolism	Fermentation pathways	*pflB*	Formate production	Anaerobic and fluctuating	IS150 deletion	1,764,888	AN-2K-2, 3 and FL-2K-1, 2, 3, 5, 6 and 7
Inactivation of redundant functions	Toxin-antitoxin system	*hokC/nhaA*	*hok/sok* system	Anaerobic and fluctuating	IS150 insertion	16,972	AN-2K-1, 4 and FL-2K-4
Inactivation of redundant functions	Toxin-antitoxin system	*hokC/nhaA*	*hok/sok* system	Anaerobic and fluctuating	IS150 insertion	16,989	FL-2K-7
Inactivation of redundant functions	Toxin-antitoxin system	*insA-7/hokE*	*hok/sok* system	Fluctuating	IS150 insertion	582,237	FL-2K-7
Inactivation of redundant functions	Toxin-antitoxin system	*trg/mokB*	*hok/sok* system	Anaerobic and fluctuating	IS150 insertion	1,272,468	AN-2K-1, 4 and FL-2K- 2, 6 and 7
Inactivation of redundant functions	Toxin-antitoxin system	*chaA/ldrC*	*ldr* system	Fluctuating	IS150 insertion	1,464,061	FL-2K-7
Inactivation of redundant functions	Toxin-antitoxin system	*ldrC/ldrB*	*ldr* system	Fluctuating	IS150 insertion	1,464,678	FL-2K-2
Inactivation of redundant functions	Toxin-antitoxin system	*ldrC/ldrB*	*ldr* system	Aerobic	IS150 insertion	1,464,679	AE-2K-1
Inactivation of redundant functions	Prophage	P22	–	Anaerobic and fluctuating	7 gene deletion	*insB-6-ompY*	FL-2K-7
Inactivation of redundant functions	Prophage	P22	–	Anaerobic and fluctuating	30 gene deletion	*insB-6-ybdK*	AN-2K-7
Inactivation of redundant functions	Prophage	Qin	–	Aerobic and fluctuating	5 gene deletion	*ydfX-ECB_01533*	FL-2K-2

Table 3 (*continued*)

Predicted mechanism[a]	Category[b]	Gene/Region	Gene function[c]	Treatment[d]	Mutation	Position[e]	Lineage[f]
Inactivation of redundant functions	Prophage	Qin	–	Aerobic and fluctuating	25 gene deletion	*ECB_01536-insE-3*	AE-2K-6
Inactivation of redundant functions	Amino acid transporters	*brnQ*	Branched chain amino acid transporter	Fluctuating	IS150 insertion	388,275	FL-2K-4 and 6
Inactivation of redundant functions	Amino acid transporters	*brnQ*	Branched chain amino acid transporter	Fluctuating	IS150 insertion	388,543	FL-2K-2
Inactivation of redundant functions	Amino acid transporters	*brnQ*	Branched chain amino acid transporter	Fluctuating	Insertion	388,020	FL-2K-3
Inactivation of redundant functions	Amino acid transporters	*brnQ*	Branched chain amino acid transporter	Fluctuating	6 gene deletion	*[araJ]-[brnQ]*	FL-2K-7
Inactivation of redundant functions	Amino acid transporters	*gltK*	Glutamate and aspartate transporter	Fluctuating	SNP	2,069,532	FL-2K-7
Inactivation of redundant functions	Amino acid transporters	*gltP*	Glutamate and aspartate transporter	Fluctuating	IS150 insertion	4,239,784	FL-2K-7
Inactivation of redundant functions	Amino acid transporters	*cycA*	Short chain amino acid permease	Anaerobic and fluctuating	IS150 insertion	4,381,583	AN-2K-1, 3, 4, 5, 6 and 7 and FL-2K-1, 2, 3, 4, 5, 6 and 7
Inactivation of redundant functions	Virulence genes	*agn43*	Outer membrane auto-transporter	Aerobic and anaerobic	SNP	2,972,858	AN-2K-7
Inactivation of redundant functions	Virulence genes	*agn43*	Outer membrane auto-transporter	Aerobic and anaerobic	SNP	2,973,574	AN-2K-3
Inactivation of redundant functions	Virulence genes	*agn43*	Outer membrane auto-transporter	Aerobic and anaerobic	IS1 insertion	2,972,936	AE-2K-5 and 7
Inactivation of redundant functions	Virulence genes	*kpsT*	Inner membrane polysaccharide transport	Aerobic	Deletion	2,999,898	AE-2K-6
Inactivation of redundant functions	Virulence genes	*kpsT*	Inner membrane polysaccharide transport	Aerobic	SNP	3,000,095	AE-2K-3 and 4

(*continued on next page*)

Table 3 (*continued*)

Predicted mechanism[a]	Category[b]	Gene/Region	Gene function[c]	Treatment[d]	Mutation	Position[e]	Lineage[f]
Inactivation of redundant functions	Virulence genes	*kpsT*	Inner membrane polysaccharide transport	Aerobic	SNP	3,000,161	AE-2K-1
Inactivation of redundant functions	Virulence genes	*kpsT*	Inner membrane polysaccharide transport	Aerobic	SNP	3,000,346	AE-2K-5 and 7
Inactivation of redundant functions	Virulence genes	*kpsE*	Polymer translocation	Aerobic	SNP	2,987,334	AE-2K-2
Inactivation of redundant functions	Virulence genes	*kpsD*	Polymer translocation	Aerobic	SNP	2,988,653	AE-2K-3 and 4
Inactivation of redundant functions	Virulence genes	*kpsM*	Inner membrane polysaccharide transport	Fluctuating	IS3 insertion	3,000,519	FL-2K-3
Inactivation of redundant functions	Virulence genes	*kpsS*	Post translational modification	Anaerobic	IS1 insertion	2,992,382	AN-2K-1, 3, 4, 5 and 6

Notes.
[a] The predicted mechanism of adaptation.
[b] The category of trait mutated.
[c] The function of the wild type gene.
[d] The treatment under which the mutations arose.
[e] Mutation location using REL4536 reference genome, except for prophage excision mutations which are listed as the genes in which the deletions begin and end.
[f] Lineages are identified by the treatment (AE, FL, or AN), followed by the generation (in all cases, approximately 2,000, designated 2K), and the individual lineage number, 1 to 7.
[g] Gene *dcuS* is restored by the 5 bp deletion slippage mutation.
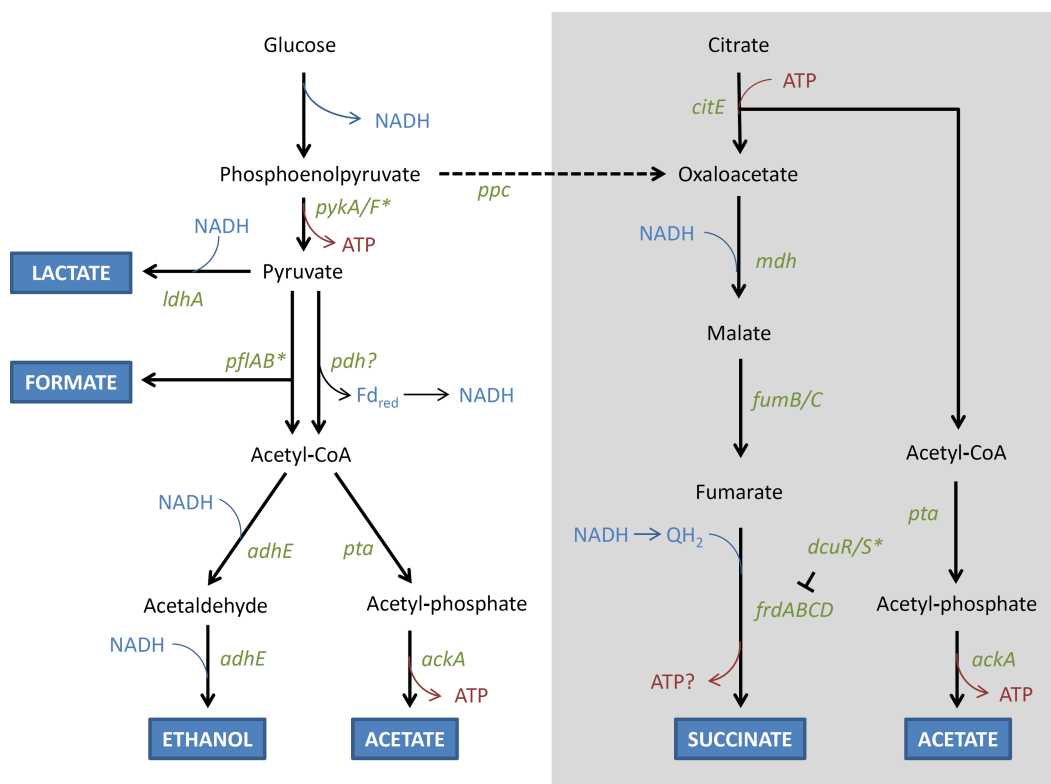
Figure 6 Schematic (unbalanced) diagram of the anaerobic fermentation pathways and genes involved in glucose fermentation by *E. coli.* Thick branching arrows represent cleavage into two smaller products. Where there are two thick arrows from one compound, these represent alternative fates. The grey box contains the pathway of citrate metabolism, and the dashed arrow represents the flow from glucose fermentation to the succinate pathway in mixed acid fermentation. Genes that are disrupted in REL4536 are represented by asterisks. The T symbol represents regulation, Fd_{red} represents ferredoxin, and QH_2 represents reduced quinone.

arose when *E. coli* MC1000 was adapted to oxygen rich and oxygen poor conditions in a rich medium (*Puentes-Téllez et al., 2013*).

NadR is a negative transcriptional regulator of NAD biosynthetic genes, whose products catalyse the initial steps of the NAD biosynthetic pathway (*San et al., 2002*). Disruption of NadR is predicted to increase the intracellular concentration of NAD via the constitutive expression of *nadA* and *nadB* (*Woods et al., 2006*). NAD is an essential coenzyme that plays a central role in metabolism by acting as electron acceptors and donors in its oxidised (NAD^+) and reduced (NADH) states in many biochemical redox reactions (*Gerasimova & Gelfand, 2005*). Fermentation pathways are highly responsive to the intracellular $NADH/NAD^+$ ratio (*Leonardo, Cunningham & Clark, 1993*; *Wimpenny & Firth, 1972*). An increase in NAD concentration is likely to globally stimulate fermentation pathways under anaerobic growth conditions by increasing electron flow (*Liu et al., 2006*; *San et al., 2002*).

The REL4536 ancestral strain contains an IS150 disruption of *pflB* (*Barrick et al., 2009*), which encodes pyruvate formate lyase and is required for the fermentation of pyruvate to formate (Fig. 6). *E. coli* contains multiple genes coding for putative pyruvate formate lyases

(*Sawers & Watson, 1998*), but only *pflB* seems to be involved in formate production as a catabolic product from glucose (*Falke et al., 2016*). Deletion of *pflB* (and not its homologue *pflD*) results in increased lactate formation (*Zhu & Shimizu, 2005*). Deletion of *pflB* also abolishes the ability to produce formate, acetate, and ethanol. Therefore, it is predicted that the ancestor may undertake fermentation mainly via the lactate pathway, yielding only two ATP per glucose, because it lacks the ability to produce formate as a catabolic product. In the AN-2K-2, AN-2K-3 and all FL-2K genomes, deletion of this IS150 copy is predicted to restore *pflB*, and thus allow a mixed acid fermentation that would increase ATP yield to three per glucose. Selection for this mutation under the AN and FL treatments further suggests its beneficial nature for growth in anaerobic environments.

Normally, pyruvate dehydrogenase (Pdh) is not active under anaerobic conditions as, in the absence of an electron sink, NADH and reduced ferredoxin (Fd_{red}) would accumulate and metabolism would cease. However, with citrate acting as a potential electron sink, re-oxidation of NADH can take place and Pdh activity would not be limited by the build-up of reduced co-factors. Pdh is also allosterically inhibited by NADH, but this inhibition would be expected to be lower if there was an electron sink, like citrate, to ensure re-oxidation of NADH. The inhibition is also overcome by competition by NAD^+ (*Hansen & Henning, 1966*), which may also explain why an active *nadR* was apparently selected for in some of the AN lineages.

Alcohol dehydrogenase (encoded by *adhE*) is required for the reduction of acetyl-CoA to ethanol in *E. coli* (*Leonardo, Cunningham & Clark, 1993*), and disruption of this gene appeared to be strongly selected for during anaerobic growth, with mutations detected in the majority of the AN genomes. Mutations in *adhE* may alleviate the build-up of ethanol (*Zhu & Shimizu, 2005*), which is toxic at high concentrations. However, they may also allow for electron flow to be diverted to other, more energetically favorable, pathways. As citrate is supplied in the medium, it may be used as an electron sink, and the citrate-succinate pathway is predicted to result in greater ATP yield (Fig. 6). However, in the ancestral strain, *dcuS* was disrupted by a frameshift caused by a 5-bp slippage event. The *dcuS* gene encodes part of the DcuR/S two-component regulatory system that regulates fumarate reductase (*Zientz, Bongaerts & Unden, 1998*) required for succinate formation. Restoration of *dcuS* was strongly selected for, with deletion of the 5-bp repeated motif being observed in six of the AN genomes (Table 3). The resulting reactivation of the DcuR/S system is predicted to enable use of citrate as an electron sink, by increasing fumarate reductase expression (*Zientz, Bongaerts & Unden, 1998*). Furthermore, the diversion of fermentation from ethanol to succinate is predicted to result in greater ATP yields that would be highly advantageous under anaerobic growth (*Kwon et al., 2011*).

It was considered that the production of acetate during fermentation may provide an alternative carbon source for lineages exposed to the anaerobic environment of this LTEE as *E. coli* is capable of metabolising acetate as an alternative to glucose (*Le Gac et al., 2008*; *Rosenzweig et al., 1994*). Acetate metabolism among evolving *E. coli* experimental lineages has been found in other LTEE studies performed in aerobic environments and is hypothesised to mediate a cross-feeding dynamic among co-evolving populations (*Treves,*

Manning & Adams, 1998). However, this is unlikely to be advantageous under strictly anaerobic growth conditions due to lack of an electron acceptor.

It is noted that the inactivation of both *dcuS* and *pflB* in the REL4536 ancestor was likely to have had little impact during the long-term aerobic selection from which it had derived (*Barrick et al., 2009*). However, in the present LTEE, the functions of both genes were restored in lineages exposed to the anaerobic environment, suggesting that they are important for adaptation in this genetic background.

The current study has highlighted that many mutations that arose in the Lenski LTEE impacted genes that are beneficial for anaerobic growth, such as *pflB, pflC, pykF, arcB* and *nadR* (*Barrick et al., 2009*). In Lenski's study, the low glucose concentration in the growth medium gives rise to low cell densities, thus oxygen availability is not likely limiting for growth. This is in contrast to typical shaken batch broth cultures, where anaerobicity is typically reached and a switch to fermentation is required for further growth (*Vasala et al., 2006*). Furthermore, end products are unlikely to accumulate to inhibitory levels in such low density cultures. As such, the disruption of genes involved in anaerobic growth was unlikely to negatively impact fitness during selection in Lenski's LTEE (*Barrick et al., 2009*). In the current study, the use of a 10,000th generation evolved clone from Lenski's LTEE was primarily to overcome adaptation to the growth medium. However, anaerobic adaptation has selected for mutations that reverse the effects of mutations that arose under prolonged aerobic selection in Lenski's LTEE. Overall, the mutations in genes affecting fermentation pathways identified in this study are likely to alter the metabolic activity of the organism by changing its ability to dispose of electrons, and in doing so, improve the efficiency of energy production during anaerobic growth. Testing the impacts of these mutations and pathways on fermentation and energy production efficiency will be the subject of ongoing research.

Disruption of redundant functions

The second general mechanism of adaptation observed in this LTEE involved mutations that are predicted to eliminate physiological functions that are likely not to be required in the environment of selection. Indeed, this seems to have happened in the evolution of the REL4536 ancestor strain used here, as described above. The majority of the new mutations in this study did not exclusively occur in one treatment or environment, and so it is likely that these contribute to adaption via conditions that were common between the treatments, such as growth media. Furthermore, these mutations likely accounted for a comparatively lower fitness response than the mutations that altered metabolic function, as they were present in AE lineages, which did not increase in fitness to the same extent as the AN or FL lineages. The functions affected by such mutations included inactivation of toxin-antitoxin (TA) systems, prophage excisions, inactivation of certain amino acid transporters, and inactivation of virulence associated genes.

(a) Inactivation of TA systems Multiple independent mutations were found in the intergenic regions adjacent to *hok/sok* and *ldr* TA system genes, predominantly in the FL genomes, but also in some AE and AN clones (Table 3). TA systems are widely distributed among prokaryotes where, typically, they are located on plasmids and involved

in post segregational killing of plasmid-free daughter cells (*Gerdes & Wagner, 2007*). A large proportion of the non-coding DNA mutations identified (Fig. 3) were in the cis-regulatory regions of TA system genes, such as the upstream complementary box (*ubc*), or translational activator element (*tae*) (*Pedersen, Christensen & Gerdes, 2002*). Plasmid maintenance may be advantageous when plasmids encode beneficial functions such as antibiotic resistance. However, the ancestral strain in this study did not contain plasmids, thus post segregational killing is presumed to be redundant in this context. Alternatively, it has been proposed that post segregational killing may no longer be the primary role of these TA systems (*Mruk & Kobayashi, 2014*) and that chromosomal TA systems may function in cell cycle arrest or programmed cell death, triggered under stressful conditions, such as oxygen limitation (*Hayes, 2003*; *Mruk & Kobayashi, 2014*). Disruption of TA gene function may be a strategy to ensure programmed cell death is forgone in favour of adaptation during exposure to the anaerobic environment.

(b) Prophage excisions Prophages are often maintained in bacterial genomes because they may confer beneficial phenotypes such as resistance to antibiotics and increased oxidative stress tolerance (*Osman et al., 2010*; *Wang et al., 2010*; *Wang, Kim & Wood, 2009*). However, prophages can encode activities that can cause profound metabolic changes in the cell (*Edlin, Lin & Bitner, 1977*). Prophage excisions accounted for 36% of the large deletions reported in this LTEE, and were responsible for 75% of genes deleted. Specifically, deletions impacted cryptic prophages, or the remnants of prophages Qin and P22. Qin and P22 sequences account for 47 kb, or ∼1%, of the REL4536 genome, thus their deletions may be selected in order to lower the metabolic burden on the cell during replication.

(c) Inactivation of amino acid transporters Mutations were found in genes *brnQ*, *gltK*, *gltP* and *cycA*, which encode various amino acid transporters, in lineages that evolved in the anaerobic environment (Table 3). In particular, a high level of parallelism was observed in *brnQ* among FL lineages (Table 3), which encodes a symporter for the uptake of LIV family branched chain amino acids (Table 3) (*Saier Jr, 2000*; *Stucky et al., 1995*). The *glt* genes encode one of four different transporters for the uptake of glutamate and aspartate in *E. coli* (*Schellenberg & Furlong, 1977*). Strong selection for the IS150 disruption of *cycA* under anaerobic conditions occurred in 13 AN and FL lineages (Table 3). This gene encodes a transporter involved in the uptake of glycine, D-serine and D-alanine (*Baisa, Stabo & Welch, 2013*). Among the three treatments, FL lineages had the highest proportion of deleterious mutations in amino acid uptake systems in an environment where free amino acids are limited, and may be a mechanism of economising the cell's resources, allowing it to better cope with starvation and/or fluctuating conditions (*D'Souza et al., 2015*).

(d) Inactivation of virulence genes Numerous mutations were identified in genes associated with virulence activity in *E. coli*, specifically in *agn43* (previously known as *flu*) and the *kps* operon (Table 3). Antigen 43 (Ag43) is an outer membrane auto-transporter, and is a virulence factor associated with infection in pathogenic strains of *E. coli* (*Van der Woude & Henderson, 2008*). Mutations were predicted to disrupt the α domain of Ag43 which confers the auto-aggregation phenotype of the virulent protein (*Van der Woude & Henderson, 2008*). The *kps* operon is involved in the assembly of an extracellular capsid
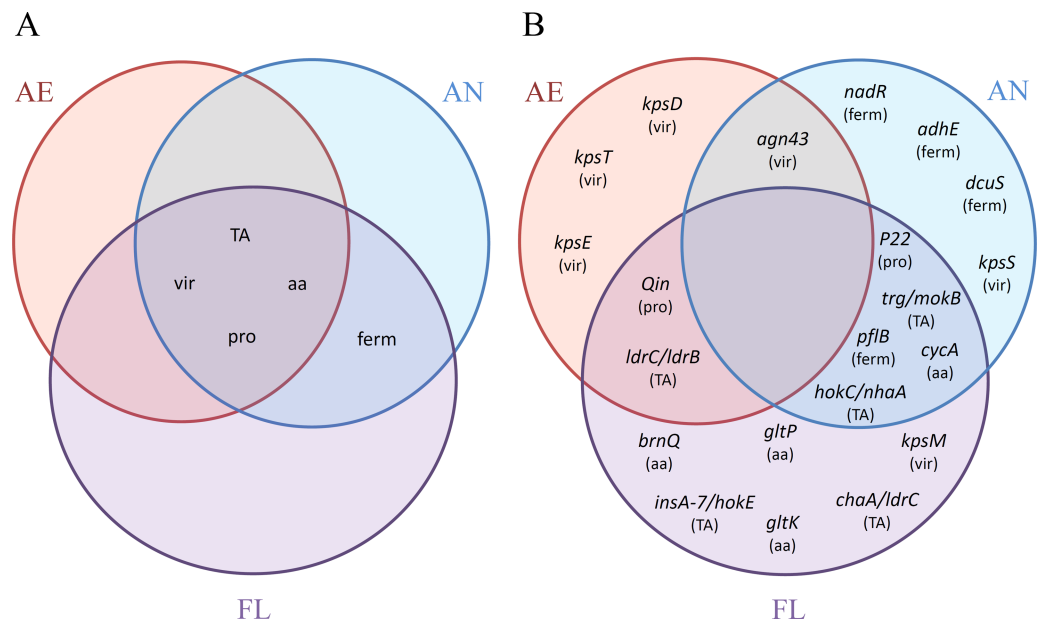
Figure 7 Venn diagrams of the distributions of (A) categories (TA, TA systems; vir, virulence genes; aa, amino acid transporters; pro, prophage excision; ferm, fermentation network) and (B) adaptive mutations, with corresponding categories affected in parenthesis, as reported among the AE, AN and FL genomes after 2,000 generations.

found in pathogenic strains of *E. coli* (*Scholl, Adhya & Merril, 2005*; *Schwan et al., 2005*). This feature is likely to be advantageous in the wild, but not under the conditions of this LTEE.

Adaptation to aerobic and anaerobic environments

A summary of the adaptive traits and impacted genes affected by mutation in the LTEE for each treatment group is shown in Fig. 7. None of the general traits impacted were found solely within a single treatment group (Fig. 7A), but were either shared among all three treatments (inactivation of virulence genes, TA systems and prophage excisions), or shared between treatments that were exposed to the anaerobic environment (i.e., the AN and FL lineages only). However, upon examination of specific genes within each category, mutations in particular genes appeared to have greater treatment specificity (Fig. 7B). For example, mutations to *dcuS* were only found in the AN evolved lineages, and *hok/sok* TA system-based disruptions were only found in lineages exposed to anaerobic conditions, whereas the *ldr* TA system mutations were only found in lineages exposed to the aerobic environment. Whether these are environment-specific adaptations requires further investigation.

CONCLUSIONS

We examined the experimental evolution of *E. coli* REL4536 in aerobic and anaerobic environments with constant and non-constant exposure. Selection in the strictly anaerobic environment resulted in fitness increases of up to 40% over 2,000 generations of selection.

Mutations that had arisen during the 2,000 generations of evolution were identified, where IS150 element-mediated mutations in non-coding regions and gene disruptions played a large role in the anaerobically-grown cells. Biological processes impacted by mutation were relatively similar among treatments, though metabolic and transport processes were affected more in the AN and FL lineages, respectively. Adaptation to the anaerobic environment was mediated by mutations that were predicted to improve energy yield during fermentation and eliminate functions redundant to the anaerobic environment. Our results demonstrate the dynamic nature of genomes in response to natural selection in the anaerobic environment and reinforce the importance of studying diverse environments to enhance our understanding of evolution in the wild.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Grant Disclosures

Competing Interests

Sinead C. Leahy, Peter H. Janssen and Christina D. Moon are employees of AgResearch Ltd, Palmerston North, New Zealand. The authors declare there are no competing interests.

Author Contributions

- Thomas J. Finn conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Sonal Shewaramani performed the experiments, reviewed drafts of the paper.
- Sinead C. Leahy analyzed the data, reviewed drafts of the paper.
- Peter H. Janssen analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Christina D. Moon conceived and designed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

Data Availability

The following information was supplied regarding data availability:

The raw sequence data generated in this study are accessible via the GenBank BioProject ID PRJNA357967.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.3244#supplemental-information.

REFERENCES

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. *Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

Atlung T, Knudsen K, Heerfordt L. 1997. Effects of σ^s and the transcriptional activator AppY on induction of the *Escherichia coli hya* and *cbdAB-appA* operons in response to carbon and phosphate starvation. *Journal of Bacteriology* **179(7)**:2141–2146 DOI 10.1128/jb.179.7.2141-2146.1997.

Baisa G, Stabo NJ, Welch RA. 2013. Characterisation of *Escherichia coli* D-cycloserine transport and resistant mutants. *Journal of Bacteriology* **195**:1389–1399 DOI 10.1128/JB.01598-12.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**:1243–1247 DOI 10.1038/nature08480.

Bertani G. 1951. Studies on lysogenesis I: the mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology* **62**:293–300.

Blount ZD, Borland CZ, Lenski RE. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **105**:7899–7906 DOI 10.1073/pnas.0803151105.

Bott M. 1997. Anaerobic citrate metabolism and its regulation in enterobacteria. *Archives of Microbiology* **167**:78–88 DOI 10.1007/s002030050419.

Carlton B, Brown B. 1981. Gene mutation. In: Gerhardt P, Murray RGE, Costilow RN, Nester EW, Wood WA, Krieg NR, Phillips GB, eds. *Manual of methods for general bacteriology*. Washington, D.C.: American Society for Microbiology, 222–242.

Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. *Molecular Ecology* **22**:1503–1517 DOI 10.1111/mec.12170.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**:215–220 DOI 10.1038/371215a0.

Clark DP. 1989. The fermentation pathways of *Escherichia coli*. *FEMS Microbiology Letters* **63**:223–234.

Cooper VS, Lenski RE. 2000. The population genetics of ecological specialisation in evolving *Escherichia coli* populations. *Earth Planet Science Letters* **12**:411–418 DOI 10.1038/35037572.

Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **100**:1072–1077 DOI 10.1073/pnas.0334340100.

Cooper VS, Schneider D, Blot M, Lenski RE. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *Journal of Bacteriology* **183**:2834–2841 DOI 10.1128/JB.183.9.2834-2841.2001.

Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* **169**:523–532 DOI 10.1534/genetics.104.035717.

D'Souza G, Waschina S, Kaleta C, Kost C. 2015. Plasticity and epistasis strongly affect bacterial fitness after losing multiple metabolic genes. *Evolution* **69**:1244–1254 DOI 10.1111/evo.12640.

Edlin G, Lin L, Bitner R. 1977. Reproductive fitness of P1, P2, and Mu lysogens of *Escherichia coli*. *Journal of Virology* **21**:560–564.

Elena SF, Lenski RE. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* **4**:457–469 DOI 10.1038/nrg1088.

Falke D, Doberenz C, Hunger D, Sawers RG. 2016. The glycyl-radical enzyme 2-ketobutyrate formate-lyase, TdcE, interacts specifically with the formate-translocating FNT-channel protein FocA. *Biochemistry and Biophysics Reports* **6**:185–189 DOI 10.1016/j.bbrep.2016.04.005.

Finkel S. 2006. Long-term survival during stationary phase: evolution and the GASP phenotype. *Nature Reviews Microbiology* **4**:113–120 DOI 10.1038/nrmicro1340.

Gaffé J, McKenzie C, Maharjan RP, Coursange E, Ferenci T, Schneider D. 2011. Insertion sequence-driven evolution of *Escherichia coli* in chemostats. *Journal of Molecular Evolution* **72**:398–412 DOI 10.1007/s00239-011-9439-2.

Gerasimova AV, Gelfand MS. 2005. Evolution of the NadR regulon in Enterobacteriaceae. *Journal of Bioinformatics and Computational Biology* **3**:1007–1019 DOI 10.1142/S0219720005001387.

Gerdes K, Wagner EGH. 2007. RNA antitoxins. *Current Opinion in Microbiology* **10**:117–124 DOI 10.1016/j.mib.2007.03.003.

Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* **102/103**:127–144. DOI 10.1023/A:1017067816551.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**:1242–1245 DOI 10.1126/science.1114057.

Gray C, Wimpenny J, Mossman MR. 1966. Regulation of metabolism in facultative bacteria II. Effects of aerobiosis, anaerobiosis and nutrition on the formation of Krebs cycle enzymes in *Escherichia coli*. *Biochimica et Biophysica Acta* **117**:33–41 DOI 10.1016/0304-4165(66)90149-8.

Gunsalus RP. 1992. Control of electron flow in *Escherichia coli*: coordinated transcription of respiratory pathway genes. *Journal of Bacteriology* **174**:7069–7074 DOI 10.1128/jb.174.22.7069-7074.1992.

Gunsalus RP, Park S-J. 1994. Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the ArcAB and Fnr regulons. *Research in Microbiology* **145**:437–450 DOI 10.1016/0923-2508(94)90092-2.

Haddock BA, Schairer HU. 1973. Electron-transport chains of *Escherichia coli*. *European Journal of Biochemistry* **35**:34–45.

Hansen RG, Henning U. 1966. Regulation of pyruvate dehydrogenase activity in *Escherichia coli* K12. *Biochimica Biophysica et Acta* **122**:355–358 DOI 10.1016/0926-6593(66)90076-2.

Hayes F. 2003. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* **301**:1496–1499 DOI 10.1126/science.1088157.

He G, Shankar RA, Chzhan M, Samouilov A, Kuppusamy P, Zweier JL. 1999. Non-invasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. *Proceedings of the National Academy of Sciences of the United States of America* **96**:4586–4591 DOI 10.1073/pnas.96.8.4586.

Ingledew W, Poole R. 1984. The respiratory chains of *Escherichia coli*. *Microbiology and Molecular Biology Reviews* **48**:222–271.

Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi S-H, Couloux A, Lee S-W, Yoon SH, Cattolico L. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3). *Journal of Molecular Biology* **394**:644–652 DOI 10.1016/j.jmb.2009.09.052.

Jones SA, Chowdhury FZ, Fabich AJ, Anderson A, Schreiner DM, House AL, Autieri SM, Leatham MP, Lins JJ, Jorgensen M, Cohen PS, Conway T. Respiration of *Escherichia coli* in the mouse intestine. *Infection and Immunity* **75**:4891–4899.

Kaleta U, Schauble C, Rinas S, Schuster S. 2013. Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnology Journal* **8**:1105–1114 DOI 10.1002/biot.201200267.

Kassen R. 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. *Journal of Evolutionary Biology* **15**:173–190 DOI 10.1046/j.1420-9101.2002.00377.x.

Kassen R, Bell G. 1998. Experimental evolution in *Chlamydomonas*. IV. Selection in environments that vary through time at different scales. *Heredity* **80**:732–741 DOI 10.1046/j.1365-2540.1998.00329.x.

Kazazian HH. 2004. Mobile elements: drivers of genome evolution. *Science* **303**:1626–1632 DOI 10.1126/science.1089670.

Ketola T, Mikonranta L, Zhang J, Saarinen K, Örmälä A, Friman V, Mappes J, Laakso J. 2013. Fluctuating temperature leads to evolution of thermal generalism and preadaptation to novel environments. *Evolution* **67**:2936–2944 DOI 10.1111/evo.12148.

Kinnersley MA, Holben WE, Rosenzweig F. 2009. *E Unibus Plurum*: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli*. *PLOS Genetics* **5**:e1000713 DOI 10.1371/journal.pgen.1000713.

Kwon S, Kim S, Lee S, Kim P. 2011. Long-term continuous adaptation of *Escherichia coli* to high succinate stress and transcriptome analysis of the tolerant strain. *Journal of Bioscience and Bioengineering* **111**:26–30 DOI 10.1016/j.jbiosc.2010.08.007.

Lee M-C, Marx CJ. 2012. Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLOS Genetics* **8**:e1002651 DOI 10.1371/journal.pgen.1002651.

Le Gac M, Brazas MD, Bertrand M, Tyerman JG, Spencer CC, Hancock RE, Doebeli M. 2008. Metabolic changes associated with adaptive diversification in *Escherichia coli*. *Genetics* **178**:1049–1060 DOI 10.1534/genetics.107.082040.

Lenski RE, Mongold JA, Sniegowski PD, Travisano M, Vasi F, Gerrish PJ, Schmidt TM. 1998. Evolution of competitive fitness in experimental populations of *Escherichia coli*: what makes one genotype a better competitor than another? *Antonie van Leeuwenhoek* **73**:35–47 DOI 10.1023/A:1000675521611.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *American Naturalist* **138**:1315–1341 DOI 10.1086/285289.

Lenski RE, Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences of the United States of America* **91**:6808–6814 DOI 10.1073/pnas.91.15.6808.

Leonardo MR, Cunningham P, Clark D. 1993. Anaerobic regulation of the *adhE* gene, encoding the fermentative alcohol dehydrogenase of *Escherichia coli*. *Journal of Bacteriology* **175**:870–878 DOI 10.1128/jb.175.3.870-878.1993.

Leroi AM, Lenski RE, Bennet AR. 1994. Evolutionary adaptation to temperature. III. Adaptation of *Escherichia coli* to a temporally varying environment. *Evolution* **48**:1222–1229 DOI 10.2307/2410380.

Liu L, Li Y, Shi Z, Du G, Chen J. 2006. Enhancement of pyruvate productivity in *Torulopsis glabrata*: increase of NAD^+ availability. *Journal of Biotechnology* **126**:173–185 DOI 10.1016/j.jbiotec.2006.04.014.

Lundrigan M, Earhart C. 1984. Gene *envY* of *Escherichia coli* K-12 affects thermoregulation of major porin expression. *Journal of Bacteriology* **157(1)**:262–268.

Lütgens M, Gottschalk G. 1980. Why a co-substrate is required for anaerobic growth of *Escherichia coli* on citrate. *Journal of General Microbiology* **119**:63–70 DOI 10.1099/00221287-119-1-63.

Mahillon J, Chandler M. 1998. Insertion sequences. *Microbiology and Molecular Biology Reviews* **62**:725–774.

Maki H. 2002. Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annual Review of Genetics* **36**:279–303 DOI 10.1146/annurev.genet.36.042602.094806.

Manché K, Notley-McRobb L, Ferenci T. 1999. Mutational adaptation of *Escherichia coli* to glucose limitation involves distinct evolutionary pathways in aerobic and oxygen-limited environments. *Genetics* **153**:5–12.

Melnyk AH, Kassen R. 2011. Adaptive landscapes in evolving populations of *Pseudomonas fluorescens*. *Evolution* **65**:3048–3059 DOI 10.1111/j.1558-5646.2011.01333.x.

Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. 2016. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Research* **44**:D336–D342 DOI 10.1093/nar/gkv1194.

Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. *Trends in Genetics* **17**:589–596 DOI 10.1016/S0168-9525(01)02447-7.

Mruk I, Kobayashi I. 2014. To be or not to be: regulation of restriction–modification systems and other toxin–antitoxin systems. *Nucleic Acids Research* **42**:70–86 DOI 10.1093/nar/gkt711.

Osman D, Waldron KJ, Denton H, Taylor CM, Grant AJ, Mastroeni P, Robinson NJ, Cavet JS. 2010. Copper homeostasis in *Salmonella* is atypical and copper-CueP is a major periplasmic metal complex. *Journal of Biological Chemistry* **285**:25259–25268 DOI 10.1074/jbc.M110.145953.

Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M. 1999. Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **96**:3807–3812 DOI 10.1073/pnas.96.7.3807.

Pedersen K, Christensen SK, Gerdes K. 2002. Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Molecular Microbiology* **45**:501–510 DOI 10.1046/j.1365-2958.2002.03027.x.

Puentes-Téllez PE, Hansen MA, Sørensen SJ, Van Elsas JD. 2013. Adaptation and heterogeneity of *Escherichia coli* MC1000 growing in complex environments. *Applied and Environmental Microbiology* **79**:1008–1017 DOI 10.1128/AEM.02920-12.

Rosenzweig RF, Sharp R, Treves DS, Adams J. 1994. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* **137**:903–917.

Saier Jr MH. 2000. Families of transmembrane transporters selective for amino acids and their derivatives. *Microbiology* **146**:1775–1795 DOI 10.1099/00221287-146-8-1775.

Sakai A, Nakanishi M, Yoshiyama K, Maki H. 2006. Impact of reactive oxygen species on spontaneous mutagenesis in *Escherichia coli*. *Genes to Cells* **11**:767–778 DOI 10.1111/j.1365-2443.2006.00982.x.

San K-Y, Bennett GN, Berríos-Rivera SJ, Vadali RV, Yang Y-T, Horton E, Rudolph FB, Sariyar B, Blackwood K. 2002. Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. *Metabolic Engineering* **4**:182–192 DOI 10.1006/mben.2001.0220.

Sawers G, Watson G. 1998. A glycyl radical solution: oxygen-dependent interconversion of pyruvate formate-lyase. *Molecular Microbiology* **29**:945–954 DOI 10.1046/j.1365-2958.1998.00941.x.

Schellenberg G, Furlong C. 1977. Resolution of the multiplicity of the glutamate and aspartate transport systems of *Escherichia coli*. *Journal of Biological Chemistry* **252**:9055–9064.

Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M. 2000. Long-term experimental evolution in *Escherichia coli*. IX. Characterisation of insertion sequence-mediated mutations and rearrangements. *Genetics* **156**:477–488.

Schneider D, Lenski RE. 2004. Dynamics of insertion sequence elements during experimental evolution of bacteria. *Research in Microbiology* **155**:319–327 DOI 10.1016/j.resmic.2003.12.008.

Scholl D, Adhya S, Merril C. 2005. *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Applied and Environmental Microbiology* **71**:4872–4874 DOI 10.1128/AEM.71.8.4872-4874.2005.

Schwan WR, Beck MT, Hultgren SJ, Pinkner J, Woolever NL, Larson T. 2005. Down-regulation of the *kps* region 1 capsular assembly operon following attachment of *Escherichia coli* type 1 fimbriae to D-mannose receptors. *Infection and Immunity* **73**:1226–1231 DOI 10.1128/IAI.73.2.1226-1231.2005.

Shapiro SS, Wilk MB. 1965. An analysis of variance test for normality (complete samples). *Biometrika* **52**:591–611 DOI 10.1093/biomet/52.3-4.591.

Shewaramani S, Finn TJ, Leahy SC, Kassen R, Rainey PB, Moon CD. 2017. Anaerobically grown *Escherichia coli* have an enhanced mutation rate and distinct mutational spectra. *PLOS Genetics* **13**:e1006570 DOI 10.1371/journal.pgen.1006570.

Smith MW, Neidhardt FC. 1983. Proteins induced by anaerobiosis in *Escherichia coli*. *Journal of Bacteriology* **154**:336–343.

Stucky K, Hagting A, Klein JR, Matern H, Henrich B, Konings WN, Plapp R. 1995. Cloning and characterisation of *brnQ*, a gene encoding a low-affinity, branched-chain amino acid carrier in *Lactobacillus delbrückii* subsp. *lactis* DSM7290. *Molecular and General Genetics* **249**:682–690 DOI 10.1007/BF00418038.

Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF. 2009. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3) and comparison of the *Escherichia coli* B and K-12 genomes. *Journal of Molecular Biology* **394**:653–680 DOI 10.1016/j.jmb.2009.09.021.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* **335**:457–461 DOI 10.1126/science.1212986.

Treves DS, Manning S, Adams J. 1998. Repeated evolution of an acetate crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Molecular Biology and Evolution* **15**:789–797 DOI 10.1093/oxfordjournals.molbev.a025984.

Van der Woude MW, Henderson IR. 2008. Regulation and function of Ag43 (*flu*). *Annual Reviews in Microbiology* **62**:153–169 DOI 10.1146/annurev.micro.62.081307.162938.

Vasala A, Panula J, Bollók M, Illmann L, Hälsig C, Neubauer P. 2006. A new wireless system for decentralised measurement of physiological parameters from shake flasks. *Microbial Cell Factories* **5**:8–14 DOI 10.1186/1475-2859-5-8.

Vasi F, Travisano M, Lenski RE. 1994. Long-term experimental evolution in *Escherichia coli.* II. Changes in life-history traits during adaptation to a seasonal environment. *American Naturalist* **144**:432–456 DOI 10.1086/285685.

Vaughn RH, Osborne JT, Wedding GT, Tabachnick J, Beisel CG, Braxton T. 1950. The utilisation of citrate by *Escherichia coli. Journal of Bacteriology* **60**:119–127.

Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic prophages help bacteria cope with adverse environments. *Nature Communications* **1**:147–156 DOI 10.1038/ncomms1146.

Wang X, Kim Y, Wood TK. 2009. Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *ISME Journal* **3**:1164–1179 DOI 10.1038/ismej.2009.59.

Wimpenny JW, Firth A. 1972. Levels of nicotinamide adenine dinucleotide and reduced nicotinamide adenine dinucleotide in facultative bacteria and the effect of oxygen. *Journal of Bacteriology* **111**:24–32.

Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America* **103**:9107–9112 DOI 10.1073/pnas.0602917103.

Zambrano MM, Kolter R. 1996. GASPing for life in stationary phase. *Cell* **86**:181–184 DOI 10.1016/S0092-8674(00)80089-6.

Zhu J, Shimizu K. 2005. Effect of a single-gene knockout on the metabolic regulation in *Escherichia coli* for D-lactate production under microaerobic condition. *Metabolic Engineering* **7**:104–115 DOI 10.1016/j.ymben.2004.10.004.

Zientz E, Bongaerts J, Unden G. 1998. Fumarate regulation of gene expression in *Escherichia coli* by the DcuSR (*dcuSR* genes) two-component regulatory system. *Journal of Bacteriology* **180**:5421–5425.