The landscape of transcriptional and translational changes over 22 years of bacterial adaptation

John S. Favate[1*], Shun Liang[1], Alexander L. Cope[1,4], Srujana S. Yadavalli[1,3], Premal Shah[1,2,*]

***For correspondence:**
premal.shah@rutgers.edu (FMS);
john.favate@rutgers.edu (FS)

[1]Department of Genetics, Rutgers University, Piscataway NJ, USA; [2]Human Genetics Institute of New Jersey, Rutgers University, Piscataway NJ, USA; [3]Waksman Institute, Rutgers University, Piscataway NJ USA; [4]Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA

Abstract Organisms can adapt to an environment by taking multiple mutational paths. This redundancy at the genetic level, where many mutations have similar phenotypic and fitness effects, can make untangling the molecular mechanisms of complex adaptations difficult. Here we use the *E. coli* long-term evolution experiment (LTEE) as a model to address this challenge. To understand how different genomic changes could lead to parallel fitness gains, we characterize the landscape of transcriptional and translational changes across 12 replicate populations evolving in parallel for 50,000 generations. By quantifying absolute changes in mRNA abundances, we show that not only do all evolved lines have more mRNAs but that this increase in mRNA abundance scales with cell size. We also find that despite few shared mutations at the genetic level, clones from replicate populations in the LTEE are remarkably similar in their gene expression patterns at both the transcriptional and translational levels. Furthermore, we show that the majority of the expression changes are due to changes at the transcriptional level with very few translational changes. Finally, we show how mutations in transcriptional regulators lead to consistent and parallel changes in the expression levels of downstream genes. These results deepen our understanding of the molecular mechanisms underlying complex adaptations and provide insights into the repeatability of evolution.

Introduction

A key challenge in biology is understanding the relationships between genotype, phenotype, and evolutionary fitness. Comparative genomic approaches and large-scale mutation experiments have allowed us to map genetic changes to phenotypic changes underlying adaptation. For example, mutations that increase the affinity of hemoglobin for oxygen are adaptive in high-altitude dwelling deer mice (***Natarajan et al., 2013***), and mutations to the influenza haemagglutinin and neuraminidase proteins increase viral fitness (***Gong et al., 2013***; ***Lee et al., 2018***). Adaptive phenotypes can also result from changes in multiple genes, such as in yeast evolving under nutrient limitation (***Gresham et al., 2008***; ***Lauer et al., 2018***; ***Venkataram et al., 2016***), bacterial adaptation during infection (***Lieberman et al., 2011***) or to high temperature (***Tenaillon et al., 2012***), and in the evolution of smaller body sizes in Atlantic silversides under a size-selective fishing regime (***Therkildsen et al., 2019***). In many cases, similar adaptive phenotypes arise from different mutations to the same gene or regulatory region or from combinations of mutations to different genes and reg-

ulatory regions. This redundancy, where many genotypes produce similar phenotypes, makes it difficult to understand the molecular mechanisms behind adaptive phenotypes and is exacerbated by potential epistatic interactions among mutations. On the other hand, adaptive changes to expression have been shown to occur during the domestication of eggplants and tomatoes (***Koenig et al., 2013***; ***Page et al., 2019***), and in hybridization events between two weeds (***Kryvokhyzha et al., 2019***). Although not direct observations of adaptive changes to gene expression, recent comparative analyses of across-species gene expression suggest the expression levels of numerous genes are evolving under directional selection in vertebrates, fish, and butterflies (***Brawand et al., 2011***; ***Catalán et al., 2019***; ***Fukushima and Pollock, 2020***; ***Gillard et al., 2021***).

Here we use the long-term evolution experiment (LTEE) (***Lenski et al., 1991***) as a model to characterize the molecular changes underlying adaptation to a novel environment. In the LTEE, 12 replicate populations of *E. coli* have been adapting in parallel to a carbon-limited medium since 1988, growing over 75,000 generations thus far. As is common in lab-based evolution experiments, the replicate populations display similar phenotypic changes (***Blount et al., 2018***). Examples include increases in fitness (***Wiser et al., 2013***) and cell size (***Grant et al., 2021***; ***Philippe et al., 2009***). In contrast, a significant amount of diversity exists at the genomic level across the replicates (***Tenaillon et al., 2016***), with some lines having orders of magnitude more mutations than others due to the development of mutator phenotypes (***Good et al., 2017***). While few mutations are shared at the nucleotide level, some genes are commonly mutated across evolved lines (***Maddamsetti et al., 2017***; ***Woods et al., 2006***). Still, how most of the mutations affect fitness in the system is unknown.

Researchers have hypothesized that similar gene expression patterns might contribute to the parallel increases in fitness in the LTEE (***Fox and Lenski, 2015***). An earlier microarray-based study of transcriptional changes in LTEE showed parallel changes in mRNA abundances in clones from two evolved lines (Ara-1 and Ara+1) at 20,000 generations (***Cooper et al., 2003***). However, it remained unclear which mutations were responsible for these parallel changes and whether the remaining ten lines also had similar expression profiles.

Moreover, protein-coding mRNAs must be translated to perform their function. The majority of cellular biomass and energy expenditure is devoted to translation (***Bernier et al., 2018***), and the role of hierarchical regulation of gene expression in evolutionary processes has been a subject of debate in recent years (***Albert et al., 2014***; ***Artieri and Fraser, 2014***; ***McManus et al., 2014***). However, we know little of changes in gene expression at the translational levels in LTEE.

Here, we use both RNA-seq and ribo-seq (***Ingolia et al., 2009***) to profile the landscape of transcriptional and translational changes after 22 years (50,000 generations) of evolution in the LTEE to answer five fundamental questions: (i) do evolved lines show similar transcriptomic and translatomic changes after 50,000 generations despite acquiring mostly unique sets of mutations? (ii) how do changes in cell size affect changes in absolute expression levels? (iii) do changes in gene expression at the translational level buffer, augment, or match changes at the transcriptional level?, (iv) what classes of genes or pathways are altered in the evolved lines, and finally, (v) can we identify mutations responsible for parallel changes in gene expression across replicate populations?

Results

We generated RNA-seq and ribo-seq datasets for single clones grown in the exponential phase from each of the 12 evolved lines with sequenced genomes in ***Tenaillon et al.*** (***2016***) (see Methods section M1 for specific clone IDs) (Figure 1A). We aligned each clone's data to its unique genome and considered expression changes of 4131 transcripts from the ancestor. Due to concerns of contamination in our Ara+6 samples, we removed them from further analysis. We averaged between 151 and 1693 deduplicated reads per transcript across the 52 libraries (Figure 1-figure supplement 1A, Supplementary File 1), the distributions of read counts per transcript were similar across lines, replicates, and sequencing methods (Figure 1-figure supplement 1B), and correlations between biological replicates were high (Pearson correlation coefficient $R > 0.93$, Figure 1-figure supplement

1C). We also verified the presence of three-nucleotide periodicity in our ribo-seq datasets (Figure 1-figure supplement 1D). Previous studies have shown the existence of distinct ecotypes in the Ara-2 population (Plucain et al., 2014; Rozen et al., 2009). Based on an analysis of mutations, our Ara-2 clone comes from the L ecotype (see Appendix S1). Our Ara-3 clone can utilize citrate as a carbon source (Cit+). Finally, we note that both ancestral and evolved lines were grown in standard LTEE media supplemented with additional glucose to obtain enough starting material for paired RNA-seq and ribo-seq samples. We discuss the potential impacts of this difference in the supplement (Appendix A2).

Evolved lines show parallel transcriptomic changes

Gene expression levels are similar across evolved lines

Across the six evolved lines with non-mutator phenotypes in LTEE, we observe a modest degree of parallelism in genetic changes. We find that 22 genes share mutations in two or more evolved lines (***Tenaillon et al., 2016***). However, it remains unclear whether these parallel genetic changes are sufficient to explain the high degree of parallelism in fitness gains over 50,000 generations. We hypothesize that the evolved lines demonstrate a higher degree of parallel transcriptomic changes despite having unique genomes. To test this hypothesis, we compared the ancestors' and evolved lines' mRNA abundances (measured in transcripts per million, TPM). We find that the expression levels of most genes remain unchanged, leading to high correlations between ancestral and evolved strains (Spearman correlation coefficient r > 0.95 Figure 1B). Moreover, pairwise correlations between evolved strains were only marginally higher than the correlations between evolved strains and the ancestors. However, these increases were not statistically significant (KS-test, p-value = 0.28, Figure 1B). This suggests that transcriptomic changes are likely restricted to a small portion of the genome.

To more formally test the hypothesis that evolved lines show parallel changes in the transcriptome, we used DESeq2 (***Love et al., 2014***) to identify differentially expressed genes (DEGs) and quantify expression changes between each evolved line and the ancestor (for full results, see Supplementary File 2). A gene was considered differentially expressed between the evolved line and the ancestor if it reached a statistical threshold of q-value ≤ 0.01. We find that most fold-changes were small (Figure 1-figure supplement 2A) and consistent with our expectations; only a small proportion of the transcriptome was significantly altered (Figure 1-figure supplement 2B). On average, $\sim$ 270 genes (out of 4131) were differentially expressed in an evolved line across all 11 pairwise comparisons between each evolved line and the ancestor. In total, 2986 genes were differentially expressed, but this consisted of only 1273 unique genes, indicating that many differentially expressed genes are shared across evolved lines. The expression levels of these 1273 differentially expressed genes were more similar between evolved lines than between an evolved line and its ancestor (Figure 1B). Correlations based on fold-changes for DEGs were higher than those based on all genes (Figure 1C). Fold-changes for the set of 1273 DEGs were generally in the same direction regardless of their statistical significance (Figure 1D). Taken together, this is suggestive of parallelism in the evolution of gene expression across the evolved lines.

Quantifying the degree of parallelism of differentially expressed genes

To test if the number of observed parallel changes in gene expression across evolved lines differs from the number of parallel changes expected by random chance, we estimated the probability distribution representing the expected number of DEGs altered in the same direction given different proportions of up and down-regulated genes in each line. This null distribution is well-approximated by the distribution of the *Sum of Independent Non-Identical Binomial random variables* (SINIB), which we estimated using the R package sinib (***Liu and Quertermous, 2018***) by parameterizing the function with the number of up and downregulated DEGs from each line (Figure 1-figure supplement 2C). We find that the number of genes with expression changes in the same direction is significantly higher than expected by chance (KS-test, p-value $\sim$ 0.01, Figure 1E - bottom panel).

For example, if DEGs were randomly distributed across all lines, we would expect three genes to share expression changes in five or more lines. Instead, 117 genes are differentially expressed in the same direction in at least five lines.

Magnitude and direction of expression changes

Given the high correlations between expression levels of differentially expressed genes (DEGs) between evolved lines, it stands to reason that the correlation between fold-changes of DEGs genes will be higher than the correlation between fold-changes across all genes. Consistent with these expectations, we find that pairwise correlations between evolved lines of fold-changes in DEGs were higher than the fold-changes of all genes (Figure 1C). While the number of DEGs varies widely across lines (Figure 1-figure supplement 2B), 7 out of 11 evolved lines have more significantly downregulated DEGs than upregulated (Figure 1-figure supplement 2D, binomial test, p-value < 0.05). Furthermore, the magnitude of fold-changes of downregulated DEGs was significantly higher than fold-changes of upregulated DEGs in all 11 evolved lines (Figure 1-figure supplement 2D, KS-test, p-value < 0.0001).

Variation in expression changes across evolved lines

So far, we have considered the degree of parallelism in expression level changes across the evolved lines. However, the evolved lines differ not only in terms of their underlying mutations (***Tenaillon et al., 2016***) but also vary substantially at the phenotypic level. For instance, half of the evolved lines have developed a mutator phenotype, causing them to accumulate orders of magnitude more mutations than the non-mutator lines. Unlike the other 11 evolved lines, Ara-3 can utilize citrate as a carbon source (***Blount et al., 2012***), and Ara-2 has developed distinct, coexisting ecotypes ***Rozen et al.*** (***2009***). We wanted to characterize how phenotypic variation across evolved lines might correlate with variation in expression levels. Principal component analysis (PCA) based on all fold-changes mainly separates Ara-3 from the rest of the lines, whereas PC2 appears to separate at least some of the mutators from the non-mutators (Figure 1-figure supplement 2E). Variation in PC1 and PC2 seems primarily driven by deletions (Figure 1-figure supplement 2F), coded as downregulated genes (log2 fold-change = -10) in this analysis. The magnitude of encoded fold-changes of the deleted genes did not affect the groupings of the PCA between log2(fold-change) -1 and -10. Given the unique circumstances in Ara-3 and Ara-2, it is not surprising that these lines group separately from the others in the PCA.

Evolved lines are larger in cell size and carry more mRNAs

In the previous section, we discussed how changes in relative gene expression patterns across the evolved lines are similar. However, all evolved lines are significantly larger than their ancestors (***Grant et al., 2021***; ***Lenski and Mongold, 2000***; ***Mongold and Lenski, 1996***). Typically, bacterial cell volume depends on nutrient availability and growth rate (***Chien et al., 2012***; ***Schaechter et al., 1958***; ***Taheri-Araghi et al., 2015***) and the increase in cell volume in evolved lines appears to be under selection rather than solely due to increases in growth rate (***Mongold and Lenski, 1996***; ***Philippe et al., 2009***). As a result of these larger sizes, the cells in evolved lines have higher biomass and proportionally higher nucleic acid levels than the ancestors (***Turner et al., 2017***). Therefore, it is reasonable to expect that absolute abundances of mRNA molecules per cell should also increase with cell volume to maintain concentrations and reaction rates (***Padovan-Merhar et al., 2015***). To get a complete picture of transcriptional changes, we also quantified absolute changes in mRNA abundances.

We used phase-contrast microscopy to measure cell shape and estimate cell volume to confirm that our clones from evolved lines were larger than their ancestors (see Appendix A3). Consistent with earlier studies, we find that each evolved line is larger in volume compared to its ancestors (Figure 2A, Supplementary File 3). Our volume estimates are also consistent with measurements obtained using a Coulter counter from a recent study (***Grant et al., 2021***) (Figure Figure 2-figure sup-

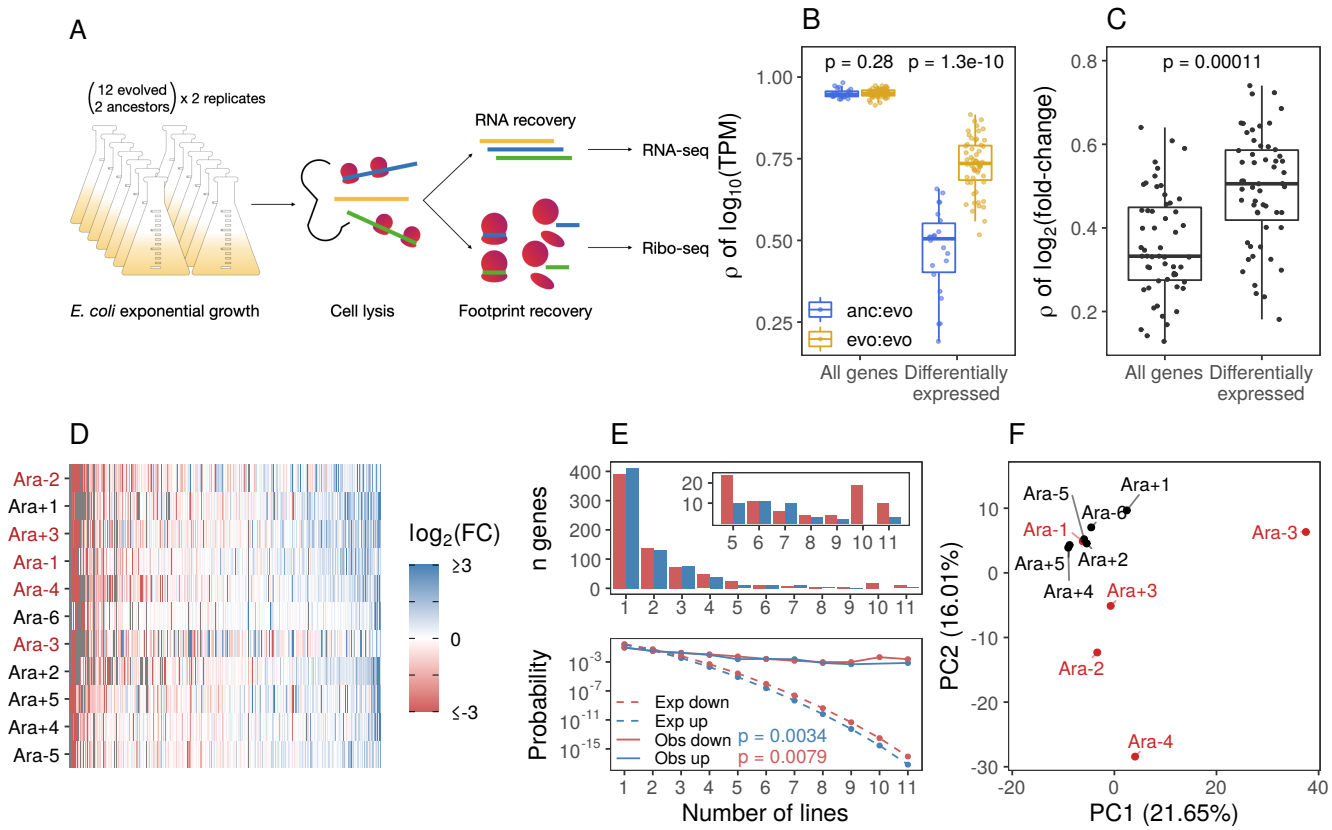


Figure 1. (A) Schematic diagram of the experimental setup. **(B)** Pairwise Pearsons's correlations based on $log_{10}(TPM)$ (where TPM is the mean from replicates) separated by comparisons between evolved lines or from ancestors to evolved lines. P-values indicate the results of a Kolmogorov-Smirnov (KS) test. For differentially expressed genes (DESeq2 q ≤ 0.01), evolved line were compared using the union of the significant genes from each line. When comparisons were between an evolved line and an ancestor, the significant genes from that evolved line were used. **(C)** Pairwise Spearman's correlations based on fold-changes from all genes, and the union of the significant genes between two evolved lines (Differentially expressed). **(D)** Fold-changes of differentially expressed genes that were significantly altered in at least one line. Genes are ordered left to right in increasing mean fold-change across all evolved lines. Genes containing deletions are not assigned a fold-change and are represented as grey spaces. Lines with a mutator phenotype are in red. **(E)** The upper panel shows the number of genes (y-axis) that were both statistically significant and had a fold-change in the same direction in a particular number of lines (x-axis). The bottom panel shows the expected (dashed) and observed (solid) probability of observing a particular result. P-values are the result of a KS test between the observed and expected distributions. **(F)** PCA based on all fold-changes. In this case, genes with some form of deletion (complete or indel) are assigned a fold-change of -10 to indicate severe downregulation because they are either completely absent from the genome or not expected to produce functional proteins.

Figure 1—figure supplement 1. Sequencing data statistics.

Figure 1—figure supplement 2. Magnitude and variation in mRNA fold-changes across evolved lines.

Figure 1—figure supplement 3. Comparison of expression changes between this study and ***Cooper et al.*** (***2003***)

plement 1A, Pearson correlation coefficient R = 0.87). Next, we estimated the absolute abundances of transcripts per CFU by comparison to known standards in our sequencing libraries. Specifically, we added the ERCC spike-in controls (***Baker et al., 2005***; ***External RNA Controls Consortium, 2005***) to our sequencing libraries and used a linear model to relate the number of molecules of a spike-in RNA added to its TPM in each sample. We find a linear relationship between molecules added and estimated TPM across all samples and replicates (Figure 2B, Figure 2-figure supplement 2A, Supplementary File 5). Finally, we measured the number of cells used in the generation of each sequencing library by counting colony-forming units (CFU) from each culture and accounting for sampling at each step of the library preparation (Supplementary File 4). Note that due to various factors, our estimates of CFU are likely underestimates (see Appendix A3, Figure 2-figure supplement 1C). Nonetheless, our gene-specific estimates of absolute abundances per CFU are highly similar across biological replicates (R > 0.93). Together, this allows us to measure absolute RNA abundance per CFU.

We find that most genes have increased mRNA abundance per CFU compared to the ancestor (Figure 2C, Figure 2-figure supplement 2B, Supplementary File 6) and that these differences were significantly larger than the differences between biological replicates (Figure 2D). Furthermore, the increases in total mRNA abundance scale with cellular volume, with larger evolved lines having more molecules per typical cell volume (Figure 2E). This suggests that the evolved lines have more mRNA per cell than the ancestors. Such an increase may be needed to maintain reaction rates in the face of increasing cell volumes. Another hypothesis is that stockpiling resources like mRNA and ribosomes might allow evolved lines to reduce the time spent in the lag phase after transfer to fresh medium. Indeed, reduced lag times occur in the LTEE (***Vasi et al., 1994***), and simulations suggest that bacteria can evolve to "anticipate" the regular transfer to fresh medium in a serial transfer regime (***van Dijk et al., 2019***).

Transcriptional changes drive translational changes

While mRNA abundances are an important molecular phenotype potentially linking genomic changes to adaptations, changes in mRNA abundances can themselves be buffered or augmented at other downstream regulatory processes such as translation (***Albert et al., 2014***; ***Artieri and Fraser, 2014***; ***McManus et al., 2014***). Translational regulation affects the rate at which an mRNA produces its protein product, and mRNAs vary widely in their translation efficiencies in both eukaryotes and prokaryotes (***Ingolia et al., 2009***; ***Li et al., 2014***; ***Picard et al., 2012***). However, the role of changes in translational regulation during adaptation and speciation remains poorly understood and, at least in yeast, is heavily debated (***Albert et al., 2014***; ***Artieri and Fraser, 2014***; ***McManus et al., 2014***). Moreover, because translation occupies the majority of cellular resources (***Bernier et al., 2018***), it may be a prime target for evolution in the LTEE. To study translational changes in LTEE, we performed ribo-seq in the evolved lines and their ancestors (Figure 1A).

We find that changes in ribosome densities are highly correlated with changes in mRNA abundances (Figure 3A, Figure 3-supplement 1A). This is somewhat surprising because changes in environmental conditions and small genetic perturbations usually result in large changes at the translational level (***Gerashchenko et al., 2012***; ***Rubio et al., 2021***; ***Woolstenhulme et al., 2015***). Despite the high correlation between mRNA and ribosome footprint fold-changes at the genomic level, individual genes might have altered ribosome densities. We used Riborex to quantify changes in ribosome densities (***Li et al., 2017***). Riborex quantifies changes in footprint densities while accounting for any changes in mRNA abundances. We considered a gene significantly altered if it reached a q-value ≤ 0.01. Only a handful of genes have altered ribosome densities, and none are shared between three or more lines (Figure 3B, Supplementary File 7). This suggests that over the course of the LTEE, most changes happen at the transcriptional level with insufficient evidence for significant changes at the translational level. We note that earlier studies have indicated that Riborex has limited power to detect small to moderate shifts in ribosome densities based on simulated data ***Li et al.*** (***2017***). Although comparing these simulations to our data is difficult, it is possible that we

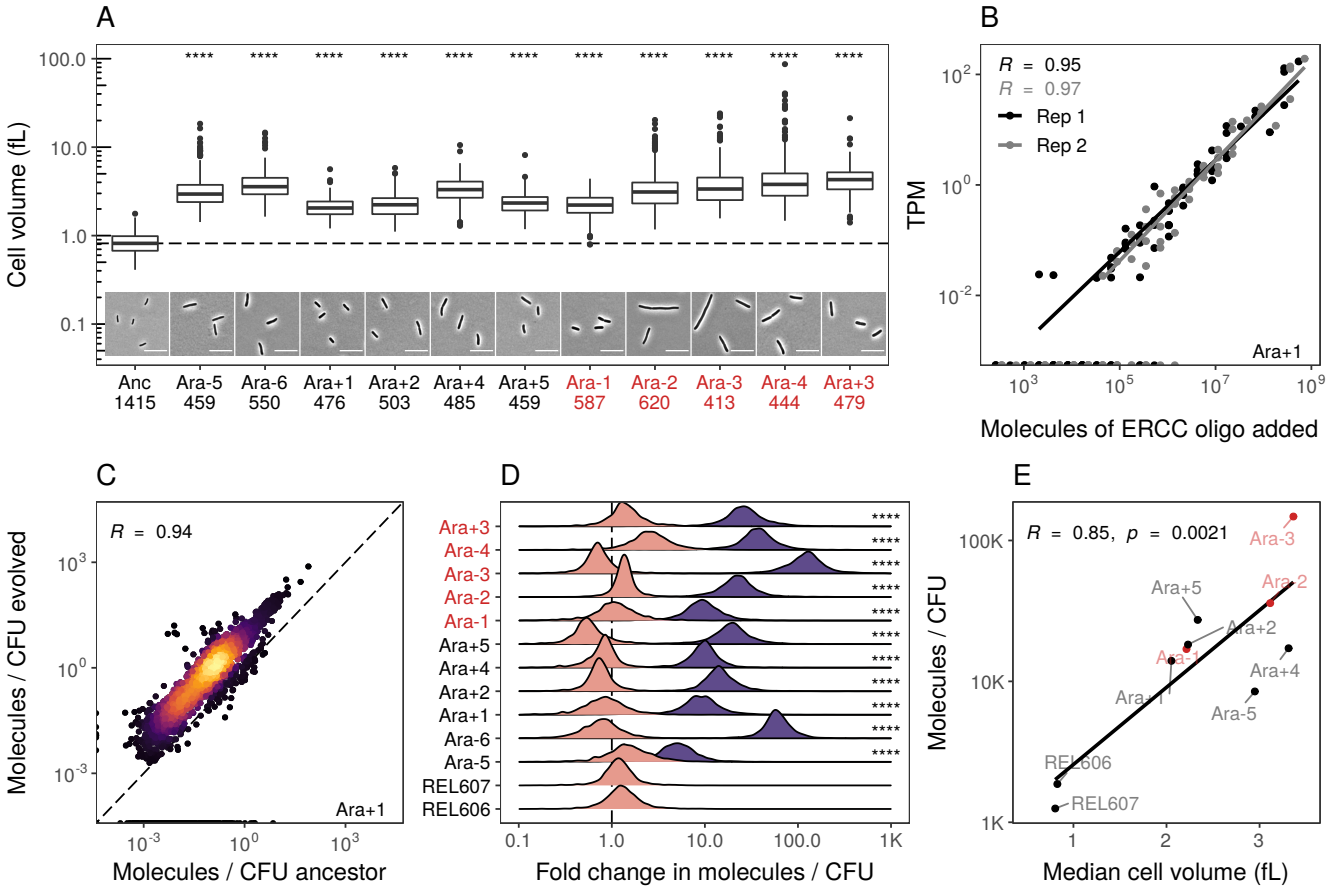


Figure 2. (A) All evolved lines are larger than the ancestral strain. Distributions of cellular volume as determined by phase-contrast microscopy and assuming sphero-cylindrical shape of *E. coli* along with representative images for each line. Numbers underneath a line's name indicates the total number of cells imaged (scale bar is 10um). The dashed line indicates the ancestral median, p-values indicate the results of a t-test when each line is compared to the ancestor, **** $p \leq 0.0001$. Lines listed in red have mutator phenotypes. **(B)** Abundances of Spike-in RNA control oligos are correlated with their estimates in sequencing data. Linear models relating the number of molecules of each ERCC control sequence added to their RNA-seq TPM (transcripts per million) in Ara+1 RNA-seq sample (see Figure 2 - figure supplement 2 for data for all lines). **(C)** Most genes have a higher absolute expression in evolved lines. Changes in the absolute number of mRNA molecules per CFU (colony forming unit) in the 50,000th generation of Ara+1 relative to the ancestor. The values plotted are the averages between 2 replicates of the evolved lines and both replicates from two ancestors (REL606 and REL607; see Figure 2 - figure supplement 2 for all lines). **(D)** Absolute changes in mRNA abundances of genes in evolved lines are significantly larger than the variation between biological replicates (KS-test, $p < .0001$ in all cases). Pink distributions indicate gene-specific fold-changes between biological replicates for each line (centered around 1). Purple distributions show the absolute fold-changes in molecules of RNA per CFU from the ancestor to each evolved line. Fold-changes are calculated in the same manner as in C. **(E)** Larger evolved lines have more mRNA per CFU. Relationship between the median cellular volume for each line and the total number of RNA molecules per CFU. Total molecules of RNA are calculated as the sum of the average number of molecules for each gene between replicates.

Figure 2—figure supplement 1. Relationship between cellular features and cell volume.

Figure 2—figure supplement 2. Absolute changes in mRNA abundances per CFU across all evolved lines.

are failing to detect some of these smaller shifts in gene-specific ribosome-densities. Regardless, our results indicate a greater role for changes in factors regulating mRNA abundances than factors regulating mRNA translation.

While Riborex can find gene-specific changes in ribosome densities, ribo-seq data can also provide codon level resolution, allowing us to perform a detailed analysis of the translation of specific codons or amino acids. We calculated genome-wide average codon-specific ribosome densities (see Methods M12, Supplementary File 8) in each of our ancestral and evolved lines and observed a high correlation between replicates (Pearson correlation coefficient R > 0.98). When comparing codon densities from each evolved line to the ancestor (Figure 3C), we find that densities at stop codons were lower in evolved lines than in the ancestor, indicating potentially faster translation termination. Importantly, ribosome densities estimated from the same evolved line are not truly independent, violating the assumption of independence for common statistical tests. We used a linear mixed model to account for possible evolved line-specific effects. The linear mixed model fit indicates an overall decrease in the ribosome density at stop codons relative to the sense codons, with a mean change in ribosome density (i.e., mean log2 fold-changes between evolved and ancestral lines) of -0.32 and 0.005, respectively. Note that these values represent the population-level fixed effect slope ($\beta_1 = -0.325$, p < 0.05) and population-level fixed effect intercept ($\beta_0 = 0.005$, p = 0.4423), respectively. The population-level fixed intercept ($\beta_0 = 0.005$, p = 0.4423) indicates the sense codons, on average, experienced no change in ribosome densities between the evolved and ancestral lines (i.e., the mean log fold changes of ribosome densities was 0). In contrast, the population-level fixed slope ($\beta_1 = -0.325$, p < 0.05) indicates that stop codons, on average, experienced a decrease in ribosome density between the evolved and ancestral lines (i.e., the mean log fold change of stop codon ribosome densities was -0.325 units lower than the mean log fold change of sense codon). Accounting for line-specific effects, the stop codon effect sizes for each evolved line range from -0.088 to -0.657 log fold change units (relative to sense codons), indicating that stop codons in all evolved lines have a decreased ribosome density compared to the ancestor. This suggests that the translation termination rate increased across all evolved lines (relative to the ancestral line), but this increase was greater in some evolved lines than others. For Ara-1, the TAG codon shows increased density, unlike other lines. This leads to a near-zero random effect size for this line.

Translation termination is one of the rate-limiting steps in translation and is typically much slower than codon elongation rates. Therefore, faster termination might increase the ribosome recycling rates and eventually allow faster translation initiation and protein production (***Andersson and Kurland, 1990***; ***Plotkin and Kudla, 2011***; ***Shah et al., 2013***). We wondered if faster termination was due to changes in the expression of translation termination factors. While some termination factors show increased expression in some lines, no single gene shows a consistent pattern across all lines (Figure 3D). Notably, while faster translation termination may increase ribosome recycling and enable faster growth, it may come at the expense of altering a key regulatory mechanism in translational control. As a result, it remains unclear if these regulatory changes can evolve in more complex environments.

Functional characterization of differentially expressed genes

Thus far, we have only considered the magnitude and source of parallelism in expression changes. In this section, we attempt to functionally characterize the altered genes, identify mutations that might be driving some of these expression changes, and determine how much higher-order entities such as metabolic pathways are altered across the evolved lines. To identify altered functional categories and pathways, we use function and pathway analysis tools such as GO (***Ashburner et al., 2000***), KEGG (***Kanehisa and Goto, 2000***), and the BioCyc database pathway perturbation score (PPS, higher numbers indicate stronger alterations to a pathway) (***Karp et al., 2017***) to assess these features (see Methods M13). Because our data suggest that changes in mRNA abundances are the driving force of change in the system, we present results from our RNA-seq data but note that sim-

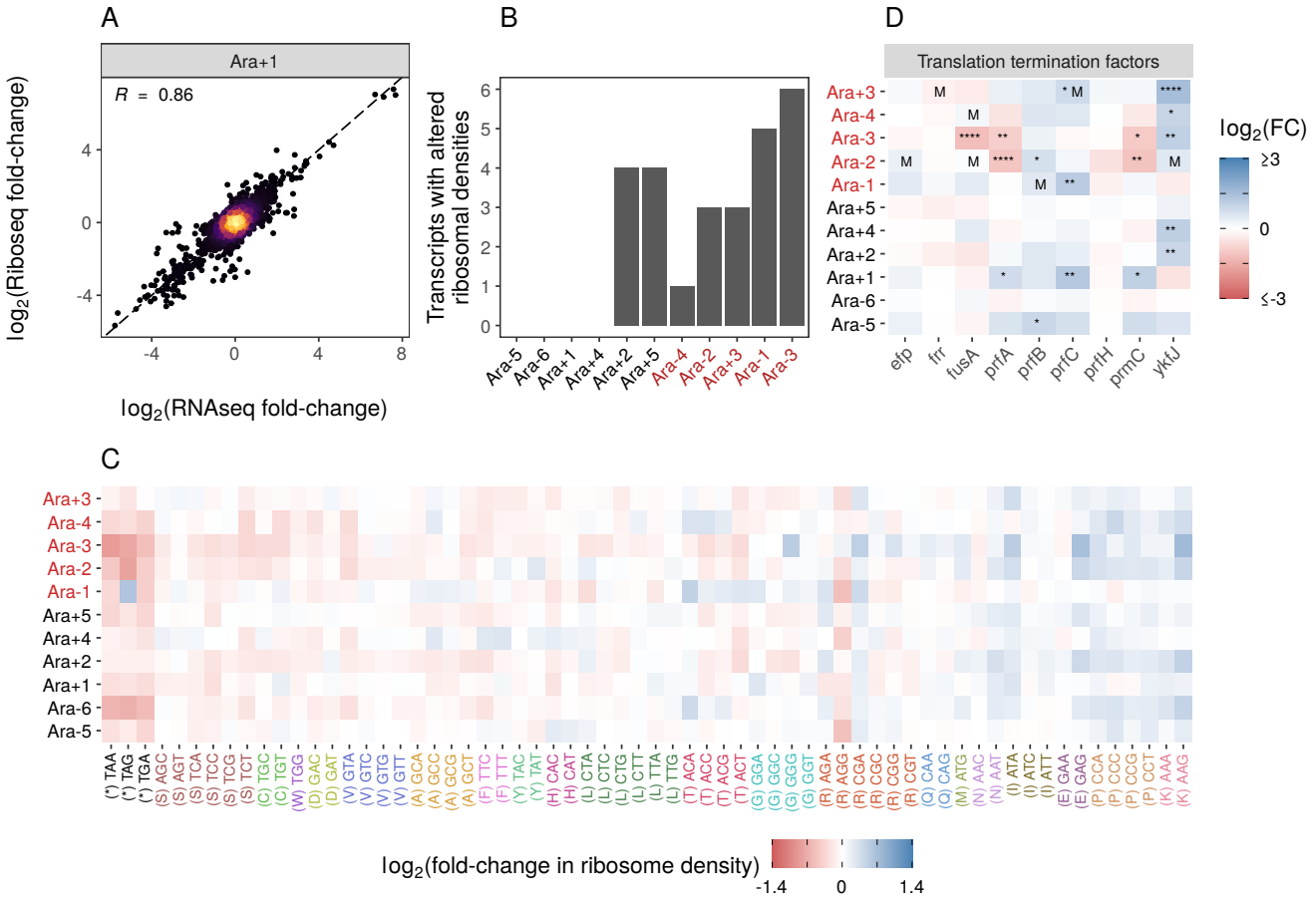


Figure 3. (A) Translational changes are correlated with transcriptional changes. The relationship between RNA-seq and ribo-seq fold-changes in Ara+1 (see Figure 3-supplement 1A for all evolved lines). **(B)** The distribution of genes with significantly altered ribosome densities ($q \leq 0.01$) estimated using Riborex ($q \leq 0.01$). **(C)** Evolved lines have faster translation termination. Stop codons had lowered ribosome density compared to all sense codons. Changes in codon-specific ribosome densities in each of the evolved lines relative to the ancestor. Codons are colored according to the amino acid they code for. Amino acids are ordered left to right in order of mean fold-change across the lines. **(D)** Fold-changes in mRNA abundances of translation termination factors and related genes *ykfJ*, *prfH*, *prfA*, *prmC*, *prfB*, *fusA*, *efp*, *prfC*. RNA-seq fold-changes for termination factors, asterisks indicate DESeq2 q-values blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$ and an "M" indicates a SNP in that gene.

Figure 3—figure supplement 1. Relationship between RNAseq and Ribo-seq fold-changes in evolved lines.

ilar results are obtained when using the ribo-seq data as well (Figure 4-figure supplement 1 and 2). For this section, we treat genes that experienced some form of deletion (complete or containing indels) as downregulated (log2 fold-change = -10) because they no longer produce functional proteins.

Many functional categories were altered across the lines in the KEGG analysis (Supplementary File 9). Consistent with earlier microarray experiments (***Cooper et al., 2003***), we find that the flagellar assembly genes are significantly downregulated in 10 out of 11 evolved lines (Figure 4A). Consistent with increased growth rates, we also find many categories related to biosynthetic and metabolic processes involving sugars or amino acids are upregulated. The biosynthesis of nucleotide sugars appears downregulated mainly due to the deletion of many of the genes involved in creating sugars which eventually lead to O-antigen biosynthesis. Many of these sugars are involved in constructing the cell membrane or walls; this could be related to known changes in cell shape and size (***Grant et al., 2021***). Overall, we find that changes in functional categories were mostly similar across all evolved lines (Figure 4B).

While KEGG pathway analysis encompasses molecular interactions and reaction networks, we wondered which specific metabolic reactions were altered across all lines and which ones remained unchanged over 50,000 generations. Because *E. coli* REL606 is annotated in the Biocyc collection of databases, we used their metabolic mapping tool to score pathway alterations with a pathway perturbation score (PPS) in each of the evolved lines (see Methods M13 for a detailed explanation of the scoring). Similar to the KEGG pathway analysis, we find a high degree of parallelism, even at the level of specific metabolic reactions (Figure 4C, 4D, Figure 4—figure supplement 2D). Interestingly, 4 out of 5 most altered pathways are involved in lipopolysaccharides (LPS) biosynthesis, a major component of Gram-negative bacteria's outer membrane. This suggests that the composition of the evolved lines' outer membrane has significantly changed in addition to changes in cell size and shape. Nonetheless, there is a core set of unaltered pathways, even in clones with a mutator phenotype. Pathways with low PPS scores, indicating low levels of alteration included D-serine degradation (mean RNAseq PPS = 0.13, σ = 0.07), pseudouridine degradation (mean RNAseq PPS = 0.12, σ = 0.06), and others (see Supplementary File 11 for complete PPS scores). These may represent pathways with activity levels that cannot be altered or whose alteration provides little to no fitness benefit.

Mutations to transcriptional regulators explain many parallel expression changes

Given the high degree of parallelism in evolved lines at the gene expression level, we wondered whether some of these patterns could be explained by a parallel set of mutations at the genetic level. Because KEGG, PPS, and GO analyses all identified metabolism and catabolism of various sugars to be significantly altered, we looked at mutations to genes involved in these functional categories. Previous work has shown that depending on the generation sampled, evolved clones grow poorly (20,000th generation) or not at all (50,000th generation) on maltose (***Leiby and Marx, 2014***). Because maltose is absent from the growth media in the LTEE, maintenance of these transporters is likely unnecessary (***Pelosi et al., 2006***). Additionally, at 20,000 generations, the transcriptional activator of the operon responsible for maltose metabolism, *malT*, was the frequent target of mutations that reduced its ability to act as a transcriptional factor, and the introduction of *malT* mutations in the ancestor had a fitness benefit (***Pelosi et al., 2006***). In *E. coli*, *malT* regulates the transcription of several operons - *malEFG* (maltose ABC transporter), *malK-lamB-malM* (*malK*, part of maltose ABC transporter; *lamB*, maltose transporter; *malM*, conserved gene of unknown function, *malPQ* (two enzymes involved in maltose metabolism), and the genes *malZ* (maltodextrin glucosidase) and *malS* (an α-amylase). We find that each of these operons was consistently and significantly downregulated across all lines (Figure 5A). Changes to the *lamB* gene have also been shown to affect susceptibility to phage infection in the LTEE (***Meyer et al., 2010***).

Many categories related to the molecule nicotinamide adenine dinucleotide (NAD) appeared in our PPS (Figure 4C) and GO results (Figure 4—figure supplement 2A). In the LTEE, *nadR*, a transcrip-

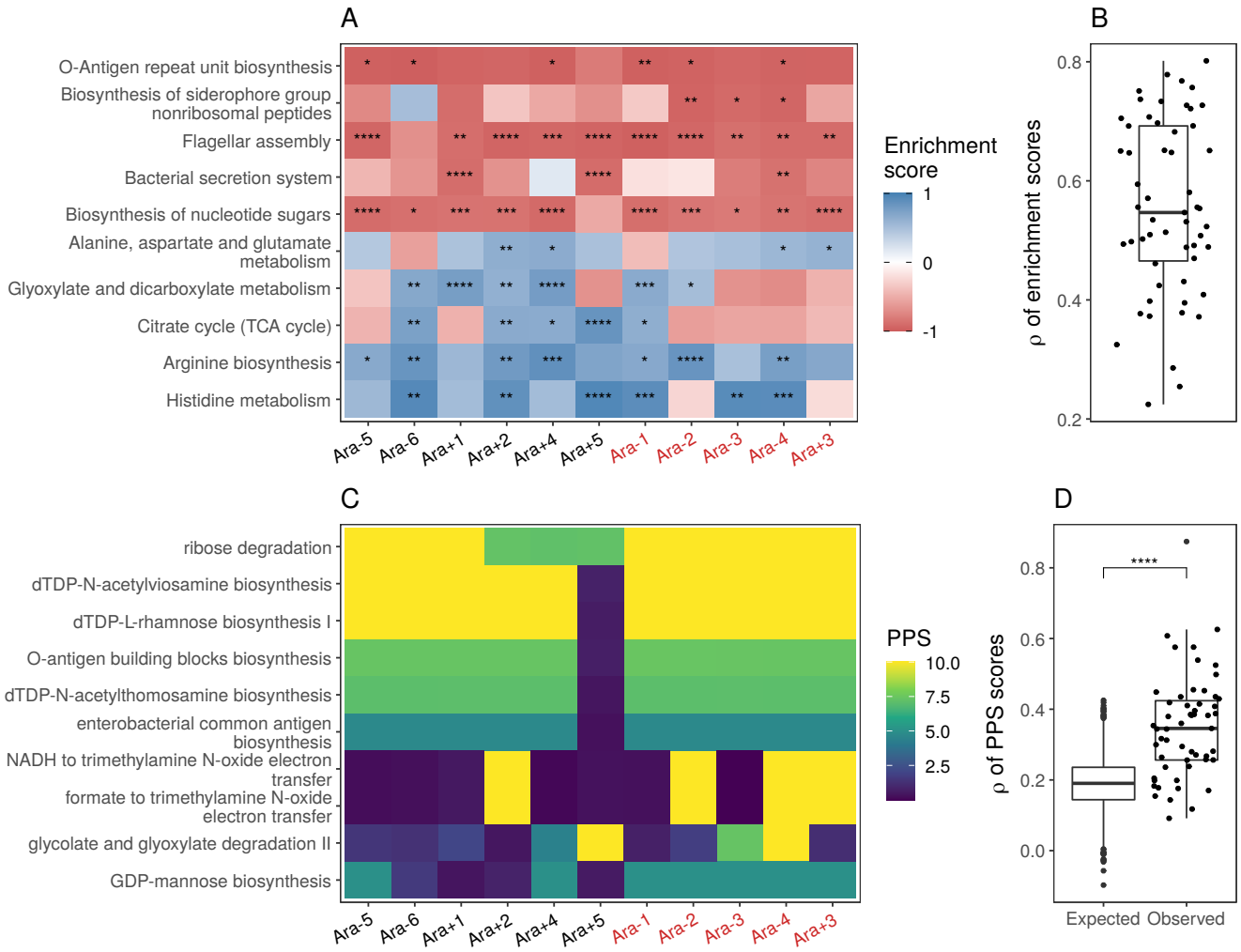


Figure 4. (A) Parallel changes in biological processes and pathways. The top 10 KEGG pathways that were significantly altered ($FDR \leq 0.05$) based on RNA-seq data. Enrichment score represents the degree to which a pathway was up (positive) or downregulated (negative). Functional categories are ordered by increasing mean enrichment score across the lines. Enrichment score represents the degree to which a pathway was up (positive) or downregulated (negative). **(B)** Distribution of pairwise Spearman's correlations of enrichment scores of all significantly altered functional categories ($FDR \leq 0.05$). **(C)** The top 10 pathways with the highest mean Pathway perturbation scores (PPS) calculated from RNA-seq fold changes. Higher PPS indicates larger degrees of alteration but does not indicate directionality. **(D)** Distribution of pairwise Spearman's correlations based on all PPS scores (observed) compared to 1000 sets of correlations generated from PPS scores calculated after randomization of fold-changes (expected). The p-value is the result of a Kolmogorov–Smirnov test. blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$.

Figure 4—figure supplement 1. Parallel changes in biological processes and pathways based on Ribo-seq data

Figure 4—figure supplement 2. GO and other functional analyses of differentially expressed genes

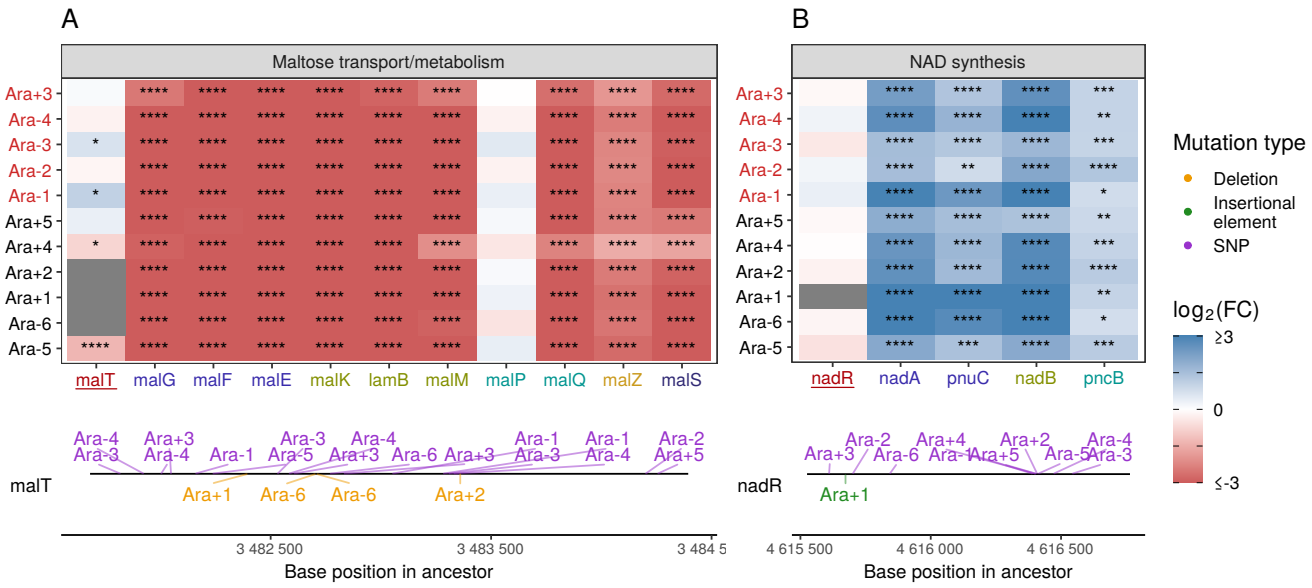


Figure 5. Mutations in transcriptional regulators lead to parallel changes in gene expression. RNA-seq fold changes for genes belonging to **A**. maltose-transport/metabolism and **B**. NAD biosynthesis. Gene names in each category are colored based on their operon membership. Mutations in transcriptional activator malT decrease expression of its downstream genes/operons. Mutations in transcriptional repressor nadR increase expression of its downstream genes/operons. Asterisks indicate statistical significance of fold-changes, (blank: $q > 0.05$, *: $q \leq 0.05$, **: $q \leq 0.01$, ***: $q \leq 0.001$ ****: $q \leq 0.0001$). Grey panels in the heatmap indicate gene deletion. Lower panels show the type and location of mutations in each transcription factor.

Figure 5—figure supplement 1. Analysis of additional pathways

tional repressor of genes involved in NAD biosynthesis, is frequently mutated, with many mutations occurring in its DNA binding domain (***Ostrowski et al., 2008***; ***Woods et al., 2006***). All evolved clones used in this study are known to have some mutation in *nadR* (***Tenaillon et al., 2016***). Given the high frequency of parallel inactivating mutations in *nadR*, these mutations are likely adaptive as they might increase intracellular NAD concentrations leading to faster growth (***Ostrowski et al., 2008***; ***Woods et al., 2006***). We find that genes directly under the regulation of *nadR* - the *nadAP* operon consisting of *nadA* (quinolinate synthase) and *pnuC* (nicotinamide riboside transporter), and genes - *nadB* (L-aspartate oxidase) and *pncB* (nicotinate phosphoribosyltransferase, were significantly upregulated in all lines (Figure 5B). Interestingly, four genes *nadCDEK*, which play various NAD biosynthesis roles in other pathways and are not regulated by *nadR*, were largely unaltered (Figure 4—figure supplement 2C). Concordantly, their transcriptional regulator, *nac*, is rarely mutated, suggesting that there is some specificity to how NAD levels may be increased in the cell.

In addition to linking the effects of specific mutations on gene expression changes in maltose and NAD regulation, we have also identified mutations that likely change the expression of genes involved in arginine biosynthesis, glyoxylate bypass system, and copper homeostasis (Figure 5-figure supplement 1, see Appendix A4). However, several functionally-related sets of genes exist, such as flagellar assembly, sulfur homeostasis, and the glycine cleavage system – that have parallel changes in expression levels without any obvious sets of parallel mutations linking these changes (Figure 5-figure supplement 1). The data generated in this study will likely prove to be a rich resource for understanding the metabolic changes that occur over long periods of evolution in a simple environment such as in the LTEE, thereby adding a new dimension to the well-studied mutational changes and gene-expression changes described here.

Discussion

Adaptation to novel environments often takes unique mutational paths even when the tempo and mode of adaptation are similar across populations (***Cheng, 1998***; ***Levy et al., 2015***; ***Meyer et al., 2012***; ***Tenaillon et al., 2012***, ***2016***; ***Therkildsen et al., 2019***). This is due, in part, to the fact that most genetic networks are highly redundant and that many mutations have pleiotropic effects. To begin to bridge the gap between parallel fitness gains in a system with mostly unique genetic changes, we wanted to study gene expression – a key link between genotype and fitness. Two key findings from our work are that (i) most of the transcriptome remains unaltered in its relative expression levels and (ii) genes with altered expression levels have remarkably similar changes (magnitude and direction of changes, pathways targeted, etc.) across all evolved lines after 50,000 generations. While parallel changes in expression profiles are perhaps not surprising given the strong selection in a well-specified environment, our work suggests that expression profiles serve as a link between the disparate mutations and similar fitness gains observed in the LTEE. Although our results do not directly implicate these parallel changes in gene expression to improved fitness, the high degree of parallelism across independently evolved populations warrants further investigation into the fitness consequences of these changes. More importantly, this suggests an optimal expression profile in any particular media that supports maximum growth. Expression profile optimization may be a mode of adaptation with each fixed mutation bringing the expression profile closer to this optimum. Nonetheless, the specific mechanisms by which the evolved lines in the LTEE have achieved similar changes in expression remain unclear. Below, we review three key proposed mechanisms that each might contribute partly to the overall story of parallelism in gene expression changes in LTEE - (i) key-regulator hypothesis, (ii) chromosomal architecture and DNA supercoiling, and (iii) growth-rate dependent changes.

Mechanisms driving parallel expression changes

According to the "key regulator" hypothesis, changes to one or a few genes can regulate the activity of many other genes responsible for most of the expression changes. In an earlier study of expression changes in the LTEE (***Cooper et al., 2003***), it was suggested that mutations to *spoT* observed in 8 out of 12 lines were responsible for many of the observed expression changes. *spoT* is involved in the stringent response pathways (***Traxler et al., 2008***) and regulates the activity of many genes. However, of the two lines whose expression was surveyed, Ara+1 and Ara-1, only Ara-1 contained a *spoT* mutation. When transferred to the ancestor, the Ara-1 *spoT* mutation did increase fitness by reducing the duration of the lag phase and increasing growth rates and caused similar expression changes in 11 of the 59 genes found to be altered in both Ara+1 and Ara-1. This means that other mutations in both lines were necessary to achieve changes in the remaining genes. Like *spoT*, ribosomal proteins and *rpoD* (the beta subunit of RNA polymerase) have also evolved faster than other genes in the LTEE (***Maddamsetti et al., 2017***). Mutations in these genes can have large pleiotropic effects and might contribute substantially to parallelism in observed expression changes.

DNA supercoiling is known to play a strong role in regulating transcription (***El Houdaigui et al., 2019***). All the evolved lines have mutations in genes related to chromosomal architecture, such as *fis*, topoisomerase A and B, or other genes which contribute to parallel changes in DNA superhelicity (***Crozat et al., 2010***). *Fis* was also part of the set of fast-evolving genes (***Maddamsetti et al., 2017***), suggesting that changes to chromosomal architecture are a target of selection in the system. Parallel mutations in genes affecting chromosomal architecture might also explain why we observe parallel expression changes in several pathways, such as sulfur homeostasis, despite the lack of parallel mutations in transcription factors that directly regulate them (Figure 5-supplement 1D).

While the above two mechanisms might be driving many parallel changes in expression levels, changes in the expression of some genes might simply be a consequence of faster growth. Expression levels of many genes in bacteria scale with growth rate (***Klumpp et al., 2009***; ***Macklin et al., 2020***) to maintain stoichiometric concentrations. As a result, simply increasing the growth rate of

replicate cultures of bacteria might produce similar expression profiles. Disentangling the effects of growth rate and genetic changes on gene expression is difficult, and therefore, we need to be cautious in over-interpreting the role of mutations in driving parallel expression changes.

On the lack of observed translational changes

Given the universality and importance of translation to life (***Bernier et al., 2018***), it is surprising that we detect few translational changes over 50,000 generations of adaptation. Bacteria possess polycistronic genes, where many proteins are translated from a single mRNA, typically belong to the same pathway or protein complex, and are translationally regulated (***Li et al., 2014***). Therefore, it is likely that any additional translational changes to genes in an operon might disrupt the stoichiometric balance of proteins in a metabolic pathway or protein complex. It is also likely that the dynamic range of translational changes is smaller than transcriptional changes in bacteria (***Cambray et al., 2018***; ***Li et al., 2014***; ***Goodman et al., 2013***) or that it might take much longer than the time scales of LTEE to observe such changes.

Conclusions

The LTEE remains a rich source for studies of evolution. Our work suggests that alterations to the global transcriptional profile is a mode of adaptation in the LTEE and that specific categories of genes have undergone similar expression changes across the lines. However, as described above, relating gene expression changes to specific mutations in LTEE is far from perfect. This is further compounded by the fact that half of the evolved lines in LTEE have a hypermutable phenotype. These genotypes have 100-fold higher mutational load than their non-mutator counterparts. It is remarkable that despite a higher mutational burden, expression patterns between mutator and non-mutator lines are highly correlated, suggesting that the bulk of the additional mutations are indeed passenger mutations (***Good et al., 2017***). While our current study has focused on expression patterns in the exponential phase, populations in the LTEE spend a significant amount of time in the stationary phase before serial transfer. However, it remains unclear if we would observe a similar level of parallelism in the stationary growth phase or how similar the expression profiles might be across distinct growth phases. Finally, the analyses undertaken here have focused on single clones from each of the evolved lines. However, each evolved population has many distinct genotypes and segregating mutations. Taking a single-cell sequencing approach, while still challenging in bacteria (***Imdahl and Saliba, 2020***), should provide a better understanding of gene expression evolution in LTEE. Lab evolution experiments combined with high-throughput multi-level sequencing approaches offer a rich resource for studying the molecular mechanisms underlying complex adaptations and provide insights into the repeatability of evolution.

Acknowledgements

We thank Richard Lenski for generously providing clones from ancestral and 50,000 generations of the LTEE. We thank Olivier Tenaillon and Richard Lenski for helpful discussions. P.S. is supported by NIH/NIGMS grant R35 GM124976, NSF DBI 1936046, subcontracts from NIH/NIDDK R01 DK056645, R01 DK109714, and R01 DK124369, as well as start-up funds from the Human Genetics Institute of New Jersey at Rutgers University. A.L.C is supported by the INSPIRE (IRACDA New Jersey/New York for Science Partnerships in Research and Education) Postdoctoral Program (NIH PAR-19-366). S.S.Y is supported by start-up funds from the Waksman Institute and Rutgers University.

Competing interests statement

P.S. is a scientific advisory board member of Trestle Biosciences and consults for Ribo-Therapeutics. P.S. is also a director at an RNA-therapeutics startup.

Methods and Materials

M1. Bacterial cell culture, recovery, and lysis

We used the following clones for generating RNA-seq and ribo-seq datasets: Ara-1 - 11330, Ara+1 - 11392, Ara-2 - 11333, Ara+2 - 11342, Ara-3 - 11364, Ara+3 - 11345, Ara-4 - 11336, Ara+4 - 11348, Ara-5 - 11339, Ara+5 - 11367, Ara-6 - 11389, Ara+6 - 11370. Bacteria were cultured in medium as per the recipe on the LTEE website (http://myxo.css.msu.edu/ecoli/dm25liquid.html) supplemented with 4 g/L glucose instead of the typical 25mg/L. Each culture was grown in 50 mL in a shaking incubator at 37 C at 125 rpm until an OD600 of 0.4-0.5 was reached. This took between 1.5-4 hr, depending on the line. Cells were recovered via vacuum filtration and immediately frozen in liquid nitrogen (LN2). Frozen pellets were stored at -80 C until lysis. A mortar and pestle were chilled to cryogenic temperatures with LN2 for lysis. The pellet was ground to a powder while submerged in LN2. Once pulverized, 650 uL of lysis buffer was added to each sample and ground further. Lysis buffer contained the following: 20 mM Tris pH 8, 10 mM MgCl2, 100 mM NH4Cl, 5 mM CaCl2, 1 mM chloramphenicol, 0.1% v/v sodium deoxycholate, 0.4% v/v Triton X-100, 100 U/mL DNase I, 1 uL/mL SUPERase-In (Thermo Fisher Scientific AM2694). The frozen lysate was allowed to thaw until liquid, then incubated for 10 min on ice to allow complete lysis. Afterward, the lysate was centrifuged at 20,000g for 10 minutes at 4 C, and the supernatant recovered and transferred to a new tube. Each sample was split into two for RNA-seq and Ribo-seq libraries.

M2. RNA-seq library preparation

Lysate destined for RNA-seq libraries was subjected to total RNA extraction using the Trizol method (Thermo Fisher Scientific 15596026) as per the manufacturer's instructions. RNA was quantified using UV spectrophotometry. We used the ERCC RNA Spike-In Mix (Thermo Fisher Scientific 4456740) in library preparation. For RNA-seq libraries, 3 uL of a 1:100 dilution of the set 1 oligos was added to the first replicate and 4 uL to the second replicate. The spike-ins were added directly to the lysate destined for RNA-seq before Trizol based RNA extraction. 2 ug of RNA with ERCC controls were subjected to fragmentation in a buffer containing final concentrations of 1 mM EDTA, 6 mM Na2CO3, and 44 mM NaHCO3 in a 10 uL reaction volume for 15 minutes at 95 C. 5 uL of loading buffer (final concentrations of 32% v/v formamide, 3.3 mM EDTA, 100 ug/mL bromophenol blue) was added to each sample, and the resulting 15 uL mixture was separated by gel electrophoresis with a 15% polyacrylamide TBE-urea gel (Invitrogen EC68852BOX) at 200 V for 30 minutes. Gels were stained for 3 minutes with SYBR Gold (Thermo Fisher Scientific S11494), and the region corresponding to the 18-50 nucleotide fragments was excised. We excised this region so that we would have similarly sized fragments for both RNA-seq and Ribo-seq libraries. RNA was recovered from the extracted fragments by adding 400 uL a buffer containing 300 mM sodium acetate, 1 mM EDTA, and .25% w/v SDS, and freezing the samples on dry ice for 30 minutes. Then, samples were incubated overnight on a shaker at 22 C. 1.5 uL of GlycoBlue (Thermo Fisher Scientific AM9515) was added as a co-precipitant, followed by 500 uL of 100% isopropanol. The samples were chilled on ice for 1 hour and then centrifuged for 30 minutes at 20,000g at 4 C. The supernatant was removed, and the pellet was allowed to air dry for 10 minutes. The pellet was resuspended in 5 uL of water, and 1 uL was used to check RNA concentration via UV spectrophotometry.

M3. Ribo-seq library preparation

Lysate destined for Ribo-seq was incubated with 1500 units of micrococcal nuclease purchased from Roche (cataog number 10107921001) and 6 uL of SUPERase-In at 25 C for 1 hour and shaken at 1400 rpm. 2 uL of .5 M EGTA pH 8 was added to quench the reaction, which was then placed on ice. The reaction was centrifuged over a 900uL sucrose cushion (final concentrations of 20 mM Tris pH 8, 10 mM MgCl2, 100 mM NH4Cl, 1 mM chloramphenicol, 2 mM DTT, .9 M sucrose, 20 U/mL SUPERase-In) using a Beckman Coulter TLA100 rotor at 70,000 rpm at 4 C for 2 hours in a 13 mm x 51 mm polycarbonate ultracentrifuge tube (Beckman Coulter 349622). The sucrose solution was

removed from the tube, and the pellet was resuspended in 300 uL of Trizol, mixed by vortexing, and RNA was extracted according to the manufacturer's protocol. Samples were then separated by gel electrophoresis and purified in the same manner as for RNA-seq.

M4. Unified library preparation

Once fragments were obtained from RNA-seq and Ribo-seq samples, they could be subject to a unified library preparation protocol as in (***Chatterji et al., 2018***; ***Gupta et al., 2019***). In total, 8 pooled libraries were prepared, with a final library structure of 5' adapter - 4 random bases - insert - 5 random bases - sample barcode - 3' adapter. The randomized bases function as UMIs for deduplication.

M5. ERCC spike-in controls and modeling

The ERCC RNA Spike-In Mix (Thermo Fisher Scientific 4456740) was used in library preparation. For RNA-seq libraries, 3 uL of a 1:100 dilution of the set 1 oligos was added to the first replicate and 4 uL to the second replicate. The spike-ins were added directly to the lysate destined for RNA-seq before Trizol based RNA extraction. The file "absolute_counts.Rmd" contains the code for the linear modeling using the ERCC data.

M6. CFU determination

Before recovery, 1mL of culture was extracted for CFU determination. LB agar plates were used for colony growth. We performed a dilution series of that 1mL culture from 1:10 to 1:1e6 in increments of 10. 100uL of each dilution was spread on a plate and incubated overnight at 37C. We determined CFU counts manually from the most appropriate dilution for each culture, usually between 1:1e3 and 1:1e6 dilutions.

M7. Optical microscopy

Liquid cultures were grown at 37C with aeration, unless otherwise indicated, in DM25 medium (Davis minimal broth supplemented with glucose at a concentration of 25 mg/L (***Lenski et al., 1991***)). Before each experiment, clones were grown in liquid cultures in DM25 medium overnight at 37C with aeration. OD600 of the cultures were 0.1–0.3. Microscope slides were prepared with 1% agarose pads, and cells were imaged by microscopy. Phase-contrast microscopy was performed using an Olympus IX81 microscope with a 100W mercury lamp and 100x NA 1.35 objective lens. 16-bit images were acquired with a SensiCam QE cooled charge-coupled device camera (Cooke Corp.) and IPLab version 3.7 software (Scanalytics) with 2 × 2 binning. Analysis of the images was performed with ImageJ (***Abràmoff et al., 2004***) and the MicrobeJ plugin (***Ducret et al., 2016***).

M8. Sequencing data processing

Raw sequencing data is deposited in the GEO database under the ascension GSE164308. Code for all data processing and subsequent analysis can be found in a series of R markdown documents uploaded to github (https://github.com/shahlab/LTEE_gene_expression_2). The file titled "data_processing.Rmd" contains the code for processing the raw sequencing data. Briefly, the following tools were used to remove adapters (cutadapt, (***Martin, 2011***)), deduplicate (BBtools dedupe.sh script), and demultiplex (FASTX-toolkit barcode splitter script) the data. Only reads of at least 24 nucleotides in length after trimming were retained for alignment. Transcript quantification for both sequencing types datasets was performed with kallisto (***Bray et al., 2016***). hisat2 (***Kim et al., 2019***) was used to align ribo-seq data for analyzing changes at specific codons. For this analysis, alignment was performed against a custom transcriptome that padded each coding region with 25nt on the 3' and 5' ends to allow for better mapping of ribosomes at the start and stop codons.

M9. Differential expression analysis of gene expression

Code for this section can be found in the file "DEseq2.Rmd". We used DEseq2 (***Love et al., 2014***) with the "apeglm" normalization (***Zhu et al., 2019***) for differential expression. In estimating fold-changes, we compared the 4 replicates of the ancestors (2 each from ancestors of Ara+ and Ara-) to 2 replicates of each of the evolved lines. Because some genes in some lines contained indels or were deleted entirely, some transcripts were missing from the transcriptome fastas used to create indices for alignment. We added these genes back to Kallisto's counts with estimated counts of 0 and assigned them fold-changes of NA. Count matrices containing identical complements of transcripts were used in the differential expression analysis for each line, such that all evolved lines had the same complement of genes as the ancestors.

M9. Change in ribosomal density analysis

We used Riborex (***Li et al., 2017***) to analyze changes in ribosomal density. The same count matrices used for DEseq2 were used here, and comparisons were made in the same manner of 4 ancestral samples (2 lines, 2 replicates each) to 2 evolved clones (1 line, 2 replicates). The code for this section can be found in the file "riborex.Rmd".

M11. Linear mixed modeling for changes in ribosome density

Code for this section can be found in "fig_3.Rmd" under the "Modeling" heading. Briefly, we fit linear mixed models using the "lme" function from the R package "nlme" to test if stop codons showed a larger decrease in ribosome densities (relative to the ancestor) as compared to the sense codons. Briefly, linear mixed models perform linear regression allowing for fixed effects (i.e. a population-level effect) and potential random effects (i.e. effects restricted to pre-specified subpopulations of the data). In this case, the random effects correspond to evolved line-specific effects on log2 ribosome density fold changes. We fit various linear mixed models allowing for different constraints on the random effect slopes and intercepts, as well as an ordinary linear regression (i.e., no random effects across evolved lines) as the null model. Models were compared using the Akaike Information Criterion (AIC): the model with the lowest AIC score is generally considered the best model. Although we identified 3 linear mixed model fits that had similar performance based on the AIC score (i.e., the difference in AIC scores was less than 2), we chose to use the simplest model, which allowed for uncorrelated random effect intercepts and slopes. This model also happened to be the model with the lowest AIC score. For comparison, this model was approximately 27 AIC units better than the ordinary linear regression.

M12. Codon-specific positioning of Ribo-seq data

Code for this section can be found in the file "codon_specific_densities.Rmd". It has been shown that mapping bacterial Ribo-seq reads by their 3' ends is more accurate than 5' mapping (***Mohammad et al., 2019***), so we mapped the A-site position of a read by using a fixed offset of 37nt (12nt offset + 25nt addition to transcript ends). To calculate ribosome densities on a codon for a gene, the number of reads mapping to a codon was normalized to the total number of reads mapping to that gene in a replicate and line-specific manner. Genome-wide codon density is calculated by taking genes with at least 100 reads mapping to them and taking the average number of normalized reads mapping to each codon across that set of genes as the genome-wide codon density. Three nucleotide periodicity is determined in the file "3nt_periodicity.Rmd".

M13. Functional analysis

We used three different functional analysis methods – GO (using the R package topGO), KEGG (using the R package clusterprofiler (***Yu et al., 2012***), and PPS (***Karp et al., 2017***). The code for each of these analyses can be found in the Rmd files named "go.Rmd", "kegg_analysis.Rmd", and "manual_PPS.Rmd," respectively. PPS scores are calculated as follows: each pathway is composed of at least one reaction, and each reaction is completed by at least one enzyme. First, a reaction

perturbation score is calculated for each reaction in a pathway, defined as the absolute value of the largest fold-change of an enzyme associated with that reaction. To calculate PPS, for a pathway having N reactions, PPS = sqrt((ΣRPS^2) / N). Additionally, a document titled "kegg_sensitivity.Rmd" tests the effects of adding deletions to our analysis.

Description of supplementary tables

Supplementary File 1: The file "table_s1_read_counts.csv" contains the results of the kallisto alignment for all samples. Counts in this file were first rounded, and new TPMs were calculated based on rounded counts. This file was generated using "data_cleaning.Rmd".

Supplementary File 2: The file "table_s2_fold_changes.csv" contains the results from DESeq2 for all samples and was generated from "DESeq2.Rmd".

Supplementary File 3: The file "table_s3_cell_size.csv" contains the quantifications from our optical microscopy. This table is supplied and is not generated from the code.

Supplementary File 4: The file "table_s4_colony_counts.csv" contains our CFU numbers. This table is supplied and is not generated from the code.

Supplementary File 5: The file "table_s5_ercc_molecules_per_sample.csv" details the amounts of ERCC spike-ins added to each sample and their abundance in the sequencing libraries. This table is supplied and is not generated from the code.

Supplementary File 6: The file "table_s6_mRNAs_per_cfu.csv" contains the measures of mRNA abundance per CFU and is generated from "absolute_counts.Rmd".

Supplementary File 7: The file "table_s7_riborex_results.csv" contains the results from riborex and is generated from "riborex.Rmd"

Supplementary File 8: The file "table_s8_genome_wide_codon_densities.csv" contains the calculated genome-wide codon densities and is generated from "codon_specific_densities.Rmd"

Supplementary File 9: The file "table_s9_kegg_results.csv" contains the KEGG search results and is generated from "kegg_analysis.Rmd"

Supplementary File 10: The file "table_s10_go_results.csv" contains the GO search results and is generated from "go.Rmd"

Supplementary File 11: The file "table_s11_pps_scores.csv" contains the PPS calculations and is generated from "manual_pps.Rmd"

Supplementary File 12: The file "table_s12_mutations.csv" contains the mutation data for our clones as downloaded from https://barricklab.org/shiny/LTEE-Ecoli/. This file is supplied and not generated from the code or can be downloaded from the website.

References

Abràmoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ. Biophotonics international. 2004; 11(7):36–42.

Albert FW, Muzzey D, Weissman JS, Kruglyak L. Genetic influences on translation in yeast. PLoS Genet. 2014 Oct; 10(10):e1004692.

Andersson SG, Kurland CG. Codon preferences in free-living microorganisms. Microbiol Rev. 1990 Jun; 54(2):198–210.

Artieri CG, Fraser HB. Evolution at two levels of gene expression in yeast. Genome Res. 2014 Mar; 24(3):411–421.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May; 25(1):25–29.

Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R, Fero M, Foy C, Fuscoe J, Gao X, Gerhold DL, Gilles P, Goodsaid F, Guo X, Hackett J, Hockett RD, Ikonomi P, et al. The External RNA Controls Consortium: a progress report. Nat Methods. 2005 Oct; 2(10):731–734.

Bernier CR, Petrov AS, Kovacs NA, Penev PI, Williams LD. Translation: The Universal Structural Core of Life. Mol Biol Evol. 2018 Aug; 35(8):2065–2076.

Blount ZD, Barrick JE, Davidson CJ, Lenski RE. Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature. 2012 Sep; 489(7417):513–518.

Blount ZD, Lenski RE, Losos JB. Contingency and determinism in evolution: Replaying life's tape. Science. 2018 Nov; 362(6415):eaam5979.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H. The evolution of gene expression levels in mammalian organs. Nature. 2011 Oct; 478(7369):343–348.

Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016 May; 34(5):525–527.

Caldara M, Charlier D, Cunin R. The arginine regulon of Escherichia coli: whole-system transcriptome analysis discovers new genes and provides an integrated view of arginine regulation. Microbiology. 2006 Nov; 152(Pt 11):3343–3354.

Cambray G, Guimaraes JC, Arkin AP. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. Nat Biotechnol. 2018 Nov; 36(10):1005–1015.

Catalán A, Briscoe AD, Höhna S. Drift and Directional Selection Are the Evolutionary Forces Driving Gene Expression Divergence in Eye and Brain Tissue of Heliconius Butterflies. Genetics. 2019 Oct; 213(2):581–594.

Chatterji P, Hamilton KE, Liang S, Andres SF, Wijeratne HRS, Mizuno R, Simon LA, Hicks PD, Foley SW, Pitarresi JR, Klein-Szanto AJ, Mah AT, Van Landeghem L, Gregory BD, Lengner CJ, Madison BB, Shah P, Rustgi AK. The LIN28B-IMP1 post-transcriptional regulon has opposing effects on oncogenic signaling in the intestine. Genes Dev. 2018 Aug; 32(15-16):1020–1034.

Cheng CH. Evolution of the diverse antifreeze proteins. Curr Opin Genet Dev. 1998 Dec; 8(6):715–720.

Chien AC, Hill NS, Levin PA. Cell size control in bacteria. Curr Biol. 2012 May; 22(9):R340–9.

Chonoles Imlay KR, Korshunov S, Imlay JA. Physiological Roles and Adverse Effects of the Two Cystine Importers of Escherichia coli. J Bacteriol. 2015 Dec; 197(23):3629–3644.

Cooper TF, Rozen DE, Lenski RE. Parallel changes in gene expression after 20,000 generations of evolution in Escherichia coli. Proc Natl Acad Sci U S A. 2003 Feb; 100(3):1072–1077.

Crozat E, Winkworth C, Gaffé J, Hallin PF, Riley MA, Lenski RE, Schneider D. Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of Escherichia coli. Mol Biol Evol. 2010 Sep; 27(9):2113–2128.

van Dijk B, Meijer J, Cuypers TD, Hogeweg P. Trusting the hand that feeds: microbes evolve to anticipate a serial transfer protocol as individuals or collectives. BMC Evol Biol. 2019 Nov; 19(1):201.

Ducret A, Quardokus EM, Brun YV. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat Microbiol. 2016 Jun; 1(7):16077.

Edwards RJ, Sockett RE, Brookfield JFY. A simple method for genome-wide screening for advantageous insertions of mobile DNAs in Escherichia coli. Curr Biol. 2002 May; 12(10):863–867.

Eichhorn E, van der Ploeg JR, Leisinger T. Deletion analysis of the Escherichia coli taurine and alkanesulfonate transport systems. J Bacteriol. 2000 May; 182(10):2687–2695.

El Houdaigui B, Forquet R, Hindré T, Schneider D, Nasser W, Reverchon S, Meyer S. Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. Nucleic Acids Res. 2019 Jun; 47(11):5648–5657.

External RNA Controls Consortium. Proposed methods for testing and selecting the ERCC external RNA controls. BMC Genomics. 2005 Nov; 6:150.

Fox JW, Lenski RE. From Here to Eternity–The Theory and Practice of a Really Long Experiment. PLoS Biol. 2015 Jun; 13(6):e1002185.

Fukushima K, Pollock DD. Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution. Nat Commun. 2020 Sep; 11(1):4459.

Gerashchenko MV, Lobanov AV, Gladyshev VN. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc Natl Acad Sci U S A. 2012 Oct; 109(43):17394–17399.

Gillard GB, Grønvold L, Røsæg LL, Holen MM, Monsen Ø, Koop BF, Rondeau EB, Gundappa MK, Mendoza J, Macqueen DJ, Rohlfs RV, Sandve SR, Hvidsten TR. Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. Genome Biol. 2021 Apr; 22(1):103.

Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. Elife. 2013 May; 2:e00631.

Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. The dynamics of molecular evolution over 60,000 generations. Nature. 2017 Nov; 551(7678):45–50.

Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. Science. 2013; 342(6157):475–479.

Grant NA, Abdel Magid A, Franklin J, Dufour Y, Lenski RE. Changes in Cell Size and Shape during 50,000 Generations of Experimental Evolution with Escherichia coli. J Bacteriol. 2021 Apr; 203(10).

Grass G, Rensing C. Genes involved in copper homeostasis in Escherichia coli. J Bacteriol. 2001 Mar; 183(6):2145–2147.

Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 2008 Dec; 4(12):e1000303.

Gupta R, Walvekar AS, Liang S, Rashida Z, Shah P, Laxman S. A tRNA modification balances carbon and nitrogen metabolism by regulating phosphate homeostasis. Elife. 2019 Jul; 8.

Imdahl F, Saliba AE. Advances and challenges in single-cell RNA-seq of microbial communities. Curr Opin Microbiol. 2020 Oct; 57:102–110.

Ingle AP, Duran N, Rai M. Bioactivity, mechanism of action, and cytotoxicity of copper-based nanoparticles: a review. Appl Microbiol Biotechnol. 2014 Feb; 98(3):1001–1009.

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009 Apr; 324(5924):218–223.

Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Ségurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF. Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). J Mol Biol. 2009 Dec; 394(4):644–652.

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan; 28(1):27–30.

Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, Ong WK, Paley SM, Subhraveti P. The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform. 2017 Aug; .

Keseler IM, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, Paulsen IT, Peralta-Gil M, Karp PD. EcoCyc: a comprehensive database resource for Escherichia coli. Nucleic Acids Res. 2005 Jan; 33(Database issue):D334–7.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug; 37(8):907–915.

Klumpp S, Zhang Z, Hwa T. Growth rate-dependent global effects on gene expression in bacteria. Cell. 2009 Dec; 139(7):1366–1375.

Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, Tohge T, Bolger A, Schneeberger K, Ossowski S, Lanz C, Xiong G, Taylor-Teeples M, Brady SM, Pauly M, Weigel D, et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci U S A. 2013 Jul; 110(28):E2655–62.

Kryvokhyzha D, Milesi P, Duan T, Orsucci M, Wright SI, Glémin S, Lascoux M. Towards the new normal: Transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (Capsella bursa-pastoris). PLoS Genet. 2019 May; 15(5):e1008131.

Lauer S, Avecilla G, Spealman P, Sethia G, Brandt N, Levy SF, Gresham D. Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. PLoS Biol. 2018 Dec; 16(12):e3000069.

Le Gac M, Plucain J, Hindré T, Lenski RE, Schneider D. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with Escherichia coli. Proc Natl Acad Sci U S A. 2012 Jun; 109(24):9487–9492.

Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. Proc Natl Acad Sci U S A. 2018 Aug; 115(35):E8276–E8285.

Leiby N, Marx CJ. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in Escherichia coli. PLoS Biol. 2014 Feb; 12(2):e1001789.

Lenski RE, Mongold JA. Cell size, shape, and fitness in evolving populations of bacteria. In: *Scaling in biology* USA: Oxford University Press, Inc.; 2000.p. 221–235.

Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. Am Nat. 1991 Dec; 138(6):1315–1341.

Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature. 2015 Mar; 519(7542):181–186.

Li GW, Burkhardt D, Gross C, Weissman JS. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell. 2014 Apr; 157(3):624–635.

Li W, Wang W, Uren PJ, Penalva LOF, Smith AD. Riborex: fast and flexible identification of differential translation from Ribo-seq data. Bioinformatics. 2017 Jun; 33(11):1735–1737.

Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis MR Jr, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, McAdam AJ, Priebe GP, Kishony R. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat Genet. 2011 Nov; 43(12):1275–1280.

Liu B, Quertermous T. Approximating the sum of independent non-identical binomial random variables. R J. 2018; 10(1):472.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15(12):550.

Macklin DN, Ahn-Horst TA, Choi H, Ruggero NA, Carrera J, Mason JC, Sun G, Agmon E, DeFelice MM, Maayan I, Lane K, Spangler RK, Gillies TE, Paull ML, Akhter S, Bray SR, Weaver DS, Keseler IM, Karp PD, Morrison JH, et al. Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. Science. 2020 Jul; 369(6502).

Maddamsetti R, Hatcher PJ, Green AG, Williams BL, Marks DS, Lenski RE. Core Genes Evolve Rapidly in the Long-Term Evolution Experiment with Escherichia coli. Genome Biol Evol. 2017 Apr; 9(4):1072–1083.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011 May; 17(1):10–12.

McManus CJ, May GE, Spealman P, Shteyman A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 2014 Mar; 24(3):422–430.

Meydan S, Klepacki D, Karthikeyan S, Margus T, Thomas P, Jones JE, Khan Y, Briggs J, Dinman JD, Vázquez-Laslop N, Mankin AS. Programmed Ribosomal Frameshifting Generates a Copper Transporter and a Copper Chaperone from the Same Gene. Mol Cell. 2017 Jan; 65(2):207–219.

Meyer JR, Agrawal AA, Quick RT, Dobias DT, Schneider D, Lenski RE. Parallel changes in host resistance to viral infection during 45,000 generations of relaxed selection. Evolution. 2010 Oct; 64(10):3024–3034.

Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. Repeatability and contingency in the evolution of a key innovation in phage lambda. Science. 2012 Jan; 335(6067):428–432.

Mohammad F, Green R, Buskirk AR. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. Elife. 2019 Feb; 8.

Mongold JA, Lenski RE. Experimental rejection of a nonadaptive explanation for increased cell size in Escherichia coli. J Bacteriol. 1996 Sep; 178(17):5333–5334.

Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan LK, Meng J, Durham BP, Shen C, Varaljay VA, Smith CB, Yager PL, Hopkinson BM. Sizing up metatranscriptomics. ISME J. 2013 Feb; 7(2):237–243.

Natarajan C, Inoguchi N, Weber RE, Fago A, Moriyama H, Storz JF. Epistasis among adaptive mutations in deer mouse hemoglobin. Science. 2013 Jun; 340(6138):1324–1327.

Okamura-Ikeda K, Ohmura Y, Fujiwara K, Motokawa Y. Cloning and nucleotide sequence of the gcv operon encoding the Escherichia coli glycine-cleavage system. Eur J Biochem. 1993 Sep; 216(2):539–548.

Ostrowski EA, Woods RJ, Lenski RE. The genetic basis of parallel and divergent phenotypic responses in evolving populations of Escherichia coli. Proc Biol Sci. 2008 Feb; 275(1632):277–284.

Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, Foley SW, Wu AR, Churchman LS, Singh A, Raj A. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol Cell. 2015 Apr; 58(2):339–352.

Page A, Gibson J, Meyer RS, Chapman MA. Eggplant Domestication: Pervasive Gene Flow, Feralization, and Transcriptomic Divergence. Mol Biol Evol. 2019 Jul; 36(7):1359–1372.

Pelosi L, Kühn L, Guetta D, Garin J, Geiselmann J, Lenski RE, Schneider D. Parallel changes in global protein profiles during long-term experimental evolution in Escherichia coli. Genetics. 2006 Aug; 173(4):1851–1869.

Philippe N, Pelosi L, Lenski RE, Schneider D. Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with Escherichia coli. J Bacteriol. 2009 Feb; 191(3):909–921.

Picard F, Milhem H, Loubière P, Laurent B, Cocaign-Bousquet M, Girbal L. Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. BMC Genomics. 2012 Oct; 13:528.

Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 2011 Jan; 12(1):32–42.

Plucain J, Hindré T, Le Gac M, Tenaillon O, Cruveiller S, Médigue C, Leiby N, Harcombe WR, Marx CJ, Lenski RE, Schneider D. Epistasis and allele specificity in the emergence of a stable polymorphism in Escherichia coli. Science. 2014 Mar; 343(6177):1366–1369.

Quandt EM, Gollihar J, Blount ZD, Ellington AD, Georgiou G, Barrick JE. Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. Elife. 2015 Oct; 4.

Rozen DE, Philippe N, Arjan de Visser J, Lenski RE, Schneider D. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. Ecol Lett. 2009 Jan; 12(1):34–44.

Rubio A, Ghosh S, Mülleder M, Ralser M, Mata J. Ribosome profiling reveals ribosome stalling on tryptophan codons and ribosome queuing upon oxidative stress in fission yeast. Nucleic Acids Res. 2021 Jan; 49(1):383–399.

Schaechter M, MaalOe O, Kjeldgaard NO. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. J Gen Microbiol. 1958 Dec; 19(3):592–606.

Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-limiting steps in yeast protein translation. Cell. 2013 Jun; 153(7):1589–1601.

Sirko A, Zatyka M, Sadowy E, Hulanicka D. Sulfate and thiosulfate transport in Escherichia coli K-12: evidence for a functional overlapping of sulfate- and thiosulfate-binding proteins. J Bacteriol. 1995 Jul; 177(14):4134–4136.

Soutourina OA, Bertin PN. Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol Rev. 2003 Oct; 27(4):505–523.

Suzuki H, Koyanagi T, Izuka S, Onishi A, Kumagai H. The yliA, -B, -C, and -D genes of Escherichia coli K-12 encode a novel glutathione importer with an ATP-binding cassette. J Bacteriol. 2005 Sep; 187(17):5861–5867.

Taheri-Araghi S, Bradde S, Sauls JT, Hill NS, Levin PA, Paulsson J, Vergassola M, Jun S. Cell-size control and homeostasis in bacteria. Curr Biol. 2015 Feb; 25(3):385–391.

Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, Wu GC, Wielgoss S, Cruveiller S, Médigue C, Schneider D, Lenski RE. Tempo and mode of genome evolution in a 50,000-generation experiment. Nature. 2016 Aug; 536(7615):165–170.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. The molecular diversity of adaptive convergence. Science. 2012 Jan; 335(6067):457–461.

Therkildsen NO, Wilder AP, Conover DO, Munch SB, Baumann H, Palumbi SR. Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. Science. 2019 Aug; 365(6452):487–490.

Traxler MF, Summers SM, Nguyen HT, Zacharia VM, Hightower GA, Smith JT, Conway T. The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli. Mol Microbiol. 2008 Jun; 68(5):1128–1148.

Turner CB, Wade BD, Meyer JR, Sommerfeld BA, Lenski RE. Evolution of organismal stoichiometry in a long-term experiment with Escherichia coli. R Soc Open Sci. 2017 Jul; 4(7):170497.

Vasi F, Travisano M, Lenski RE. Long-Term Experimental Evolution in Escherichia coli. II. Changes in Life-History Traits During Adaptation to a Seasonal Environment. Am Nat. 1994 Sep; 144(3):432–456.

Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Hérissant L, Blundell JR, Levy SF, Fisher DS, Sherlock G, Petrov DA. Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. Cell. 2016 Sep; 166(6):1585–1596.e22.

Wilson RL, Steiert PS, Stauffer GV. Positive regulation of the Escherichia coli glycine cleavage enzyme system. J Bacteriol. 1993 Feb; 175(3):902–904.

Wiser MJ, Ribeck N, Lenski RE. Long-Term Dynamics of Adaptation in Asexual Populations. Science. 2013 Dec; 342(6164):1364–1367.

Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. Proc Natl Acad Sci U S A. 2006 Jun; 103(24):9107–9112.

Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep. 2015 Apr; 11(1):13–21.

Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 May; 16(5):284–287.

Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019 Jun; 35(12):2084–2092.

A1. Determination of Ara-2 ecotype

Analysis for determination of Ara-2 ecotype can be found in the file "araM2_ecotype.Rmd". Briefly, we compared mutations in our clones to the mutations determined in ***Plucain et al.*** (***2014***). Our clone of Ara-2 does not possess mutations in the arcA or gntR genes. We also compared mutations in our clone against the list of mutations unique to the S or L ecotype and found that our clone possesses many mutations unique to the L type but not the S type. Finally, ***Le Gac et al.*** (***2012***) found two large 35 and 41 kilobase deletions in the S lineage at 40,000 generations, neither of which are present in our clone at 50,000 generations.

A2. The potential effects of increased sugar in the culture medium

The LTEE media recipe uses 25 mg/L glucose. However, this low glucose environment leads to low cell-densities and constrains our ability to generate matched RNA-seq and ribo-seq samples with sufficient depth to perform genome-wide analyses from the same culture. To overcome limitations of cell-densities, we used 4 g/L, the amount of sugar specified in the agar recipe used for solid growth assays on the LTEE website (http://myxo.css.msu.edu/ecoli/dmagar.html). The increased glucose level in our medium is expected to affect the final cell density rather than the growth rate

during the exponential phase. Additionally, though our experiment takes place 30,000 generations after the **Cooper et al.** (**2003**) study, we observe similar patterns in expression changes (Figure 1—figure supplement 3A). This suggests that some patterns may have reached fixation long ago and that bacteria may behave similarly across the two experiments. Finally, even in the case where the increased glucose has altered the physiology of cells in our cultures, the fact that we see parallel patterns of differential expression relative to the ancestor in each evolved line indicates that we are observing heritable differences from the ancestor.

A3. Absolute abundances and CFU counts

We used colony forming units (CFUs) of our cultures as a measure of cell densities to generate each library. However, filamentation of cells in our cultures can bias our estimates of cell-densities since it remains unclear whether a colony was initiated from a single cell or a filament. In our data, volume increases are best correlated with length or aspect ratio as opposed to width (Figure 2—figure supplement 1C). This suggests that while some volume increases are truly individual cells getting larger, exceptionally large cells are likely chains. In the absence of absolute changes, simply undercounting the number of cells would also produce the observed results. Removal of large, presumably filamentous cells using the same filtering metric as in **Grant et al.** (**2021**) (0.21 fL $\leq$ volume $\leq$ 5.66 fL, Figure 2—figure supplement 1B) has little effect on our median cell volumes and hence does not affect results that use the median volume, such as those in figure 2E. That said, the amount of transcripts estimated from our data is well over what is believed to be present inside a bacterium (**Moran et al., 2013**), so CFUs likely underrepresent the number of cells used to make each library. Moreover, a CFU assay only considers living cells, whereas dead cells, depending on their time of death relative to collection time, could also contribute to RNA abundance but not CFUs.

A4. Analysis of altered pathways

Flagellar assembly was the top category in the KEGG results, and categories relating to motility or flagella were frequent in the PPS and GO analyses. Flagella are used for motility and allow bacteria to move to new environments when necessary. Downregulation of flagellar genes is a common adaptation in laboratory-based evolution experiments (**Edwards et al., 2002**) and was a principle finding in **Cooper et al.** (**2003**). We also observed downregulation of the *flgBCDEFGHIJK*, *flgAMN*, and *flhABE* operons in all but one evolved line (Figure Figure 5—figure supplement 1A, upper panel). These operons contribute various proteins to the flagellar apparatus and are regulated in part by the transcription factors *flhC* and *flhD*, which themselves have complicated regulation dictated by various environmental factors (**Soutourina and Bertin, 2003**). *flhC* and *flhD* are downregulated in 3 of the evolved lines but mostly unaltered in the others. These genes are rarely mutated in the clones used in this study (Figure 5—figure supplement 1A, lower panel). Because *E. coli* B is thought to be non-motile (**Jeong et al., 2009**), it's likely that the downregulation of these genes is due to the removal of an unnecessary function and was fixed early on in the experiment. The lack of parallel changes in transcriptional regulators *flhCD* suggests that other mechanisms may play a part in causing the downregulation of these genes.

Terms relating to arginine and other amino acids were common in our results. We found that genes related to arginine synthesis were statistically significant and upregulated in many lines (Figure 5—figure supplement 1B). Upregulation of genes in amino acid synthesis pathways could increase intracellular amino acid amounts, allowing faster translation and leading to faster growth. Alternatively, the arginine synthesis pathways have many intermediate molecules which can be fed into other metabolic pathways, one of which could also allow faster growth. *argR*, which represses transcription of these genes when L-arginine is abundant (**Caldara et al., 2006**), frequently contains mutations in or around its coding sequence and is unaltered in its expression. As such, some of these mutations may have disabled the repressive ability of *argR*, leading to the increased expression we observe here.

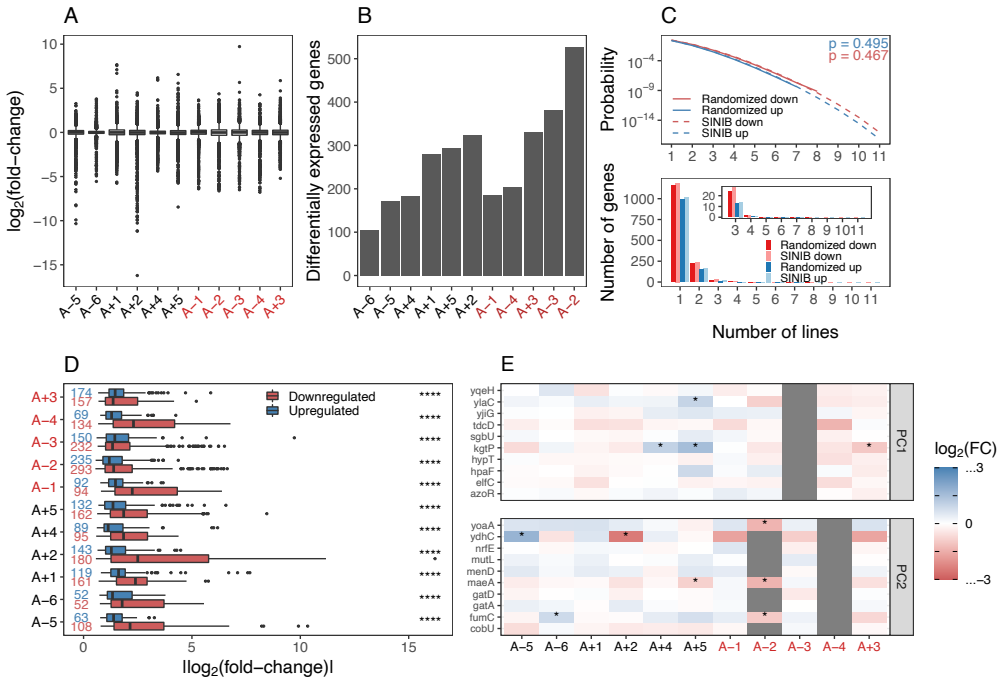
The glyoxylate bypass system allows *E. coli* to utilize acetate as a carbon source. It is composed of the *aceBAK* operon and is regulated by *iclR* and *arcAB* (**Okamura-Ikeda et al., 1993**). Acetate is a metabolic by-product but can be returned to central carbon metabolism for biosynthetic reactions by this system. Previous studies have shown that mutations in *iclR* and *arcB* cause derepression of their target genes are beneficial in the LTEE (**Quandt et al., 2015**). Consistent with these results, we found that the *aceBAK* operon was upregulated in 9 of 11 evolved lines (Figure 5—figure supplement 1C). This confirms the hypothesis from **Quandt et al.** (**2015**) that mutations to *iclR* and *arcB* derepress enzymes involved in acetate metabolism.

Sulfur is a critical component of many biological molecules, like amino acids, and participates in creating other structures like iron-sulfur cluster proteins. Organic sulfur is transported across the cell membrane by proteins from the *cysPUWAM* operon, which encodes for a sulfate/thiosulfate importer (**Sirko et al., 1995**), the *gsiABCD* operon which encodes for a glutathione importer (**Suzuki et al., 2005**), the *tauABCD* operon which codes for a taurine importer (**Eichhorn et al., 2000**), and *tcyP*, the major L-cysteine importer (**Chonoles Imlay et al., 2015**). We found that many of these genes were downregulated in many lines (Figure 5—figure supplement 1E). The *cysB* gene positively regulates these genes and was downregulated in most lines and contained few mutations. The sources of organic sulfur in the medium used in the LTEE are ammonium and magnesium sulfate, for which the *cysPUWAM* operon functions as the importer. The mechanism and reasons for alterations to these operons remain unclear. One hypothesis is that the amount of organic sulfur in the medium is sufficient to allow the downregulation of sulfur transport systems without impacting downstream pathways that require sulfur and negatively impacting growth, thus saving energy by not transcribing or translating them.

Glycine plays a role in protein construction and can be a building block for other metabolic pathways such as one-carbon metabolism or serine synthesis (**Okamura-Ikeda et al., 1993**; **Wilson et al., 1993**). We found that the *gcvTHP* operon, which encodes for proteins in the glycine cleavage system, was upregulated in many of the evolved lines. Increases in the levels of compounds involved in this set of reactions may directly increase growth rates. Though some mutations exist in and around transcriptional regulators of these genes, their effects are unclear. Whether changes to these genes are due to changes in their transcription factors or other changes, the upregulation of these genes in many lines suggests that it may be beneficial.

Copper and silver have antibacterial properties (**Ingle et al., 2014**), and bacteria have evolved systems to mitigate toxicity from these elements. The *cusCFBA* operon, regulated by the *cusRS* sensor kinase, codes for proteins that transport copper and silver ions out of the cell (Nies, 2003). Additionally, the cytoplasmic copper chaperone *copA*, regulated by *cueR* (**Meydan et al., 2017**), and *cueO* (multicopper oxidase (**Grass and Rensing, 2001**)) regulate copper homeostasis in the cell. These genes contained deletions in 5 of our clones and were downregulated in three of the six lines where they remained (Figure 5—figure supplement 1F). Overall, eight of the eleven lines surveyed here had defects in these systems. This suggests that these genes may be selected for removal or downregulation. In contrast to natural environments, the laboratory environment is likely free of copper and silver, rendering these systems dispensable. That said, because many of these genes are casualties of large deletions, it's not obvious which genes, if any, provide a fitness benefit in the system.

Figure 1—figure supplement 1. (A) The average number of reads aligned per gene using Kallisto for each library. The color scheme remains the same in panels B and D. **(B)** Distributions of mapped and deduplicated read counts per gene in each sample. **(C)** Correlations between the replicates based on rounded counts or TPMs. **(D)** The periodicity of the ribo-seq datasets determined using a fast Fourier transform (see Methods M12).



F

PC1 genes	Product
kgtP	alpha−ketoglutarate:H+ symporter
eltC	putative fimbrial usher protein EltC
tdcD	propionate kinase
sgbU	predicted L−xylulose 5−phosphate 3−epimerase
yqeH	orf
yjiG	conserved inner membrane protein
ylaC	predicted inner membrane protein
azoR	FMN dependent NADH:quinone oxidoreductase
hypT	DNA−binding transcriptional dual regulator HypT
hpaF	5−carboxymethyl−2−hydroxymuconate delta−isomerase

PC2 genes	Product
mutL	MutHLS complex, methyl−directed mismatch repair
yoaA	putative 5' to 3' DNA helicase implicated in DNA repair
fumC	fumarase C
nrfE	putative cytochrome c−type biogenesis protein NrfE
cobU	cobinamide−P guanylyltransferase / cobinamide kinase
maeA	NAD+−dependent malate dehydrogenase
menD	2−succinyl−5−enolpyruvyl−6−hydroxy−3−cyclohexene−1−carboxylate synthase
gatD	galactitol−1−phosphate 5−dehydrogenase
gatA	NA
ydhC	predicted transport protein YdhC

Figure 1—figure supplement 2. **(A)** Distributions of all mRNA fold-changes (using DESeq2) in each line. Lines with a mutator phenotype are in red. **(B)** The number of differentially expressed genes (DESeq2 $q \leq 0.01$) in each line. **(C)** Upper panel shows the probabilities of observing a gene that was differentially expressed and altered in the same direction in a given number of lines (x-axis).The solid lines represent mean probabilities derived from randomizing the fold-changes of genes in each line one million times and the dashed lines represent the probabilities calculated using the SINIB method as shown in Figure 1E. P-values show the result of a KS test comparing the randomized to the SINIB distributions. The lower panel shows the expected number of differentially expressed genes that are shared and altered in the same direction in a given number of lines (x-axis) based on the above probabilities. **(D)** Distributions of absolute fold-changes of differentially expressed genes in each line. The number of DEGs in each evolved line is indicated. Asterisks indicate the results of a Kolmogorov-Smirnov test comparing distributions of the magnitudes of positive and negative fold-changes in each line NS: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$. **(E)** The list of top 10 genes contributing to variation in each principle component, grey spaces represent deletions which were encoded as having a $log_2(fold - change) = -10$. **(F)** The genes and descriptions of genes contributing to first two principal components retrieved from EcoCyc (***Keseler et al., 2005***).

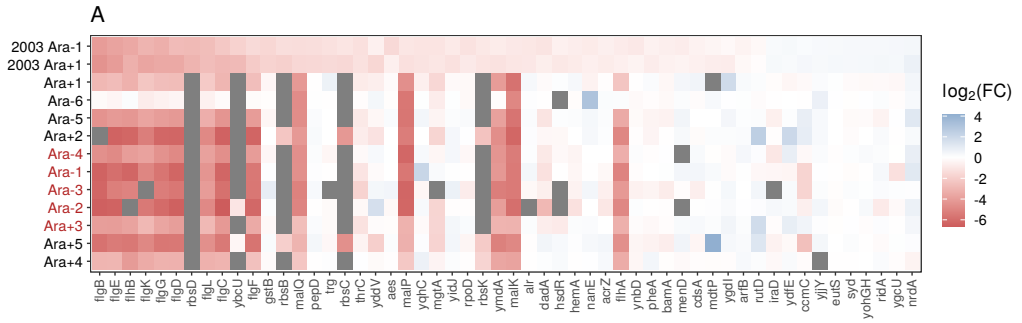


Figure 1—figure supplement 3. (A) The direction and magnitude of expression changes in genes identified as differentially expressed in ***Cooper et al.*** (***2003***) study and the direction of changes for those genes in our dataset. While the two datasets share a color scale for fold-change, the data underlying the ***Cooper et al.*** (***2003***) study was generated using a microarray compared to RNAseq data in the current study.

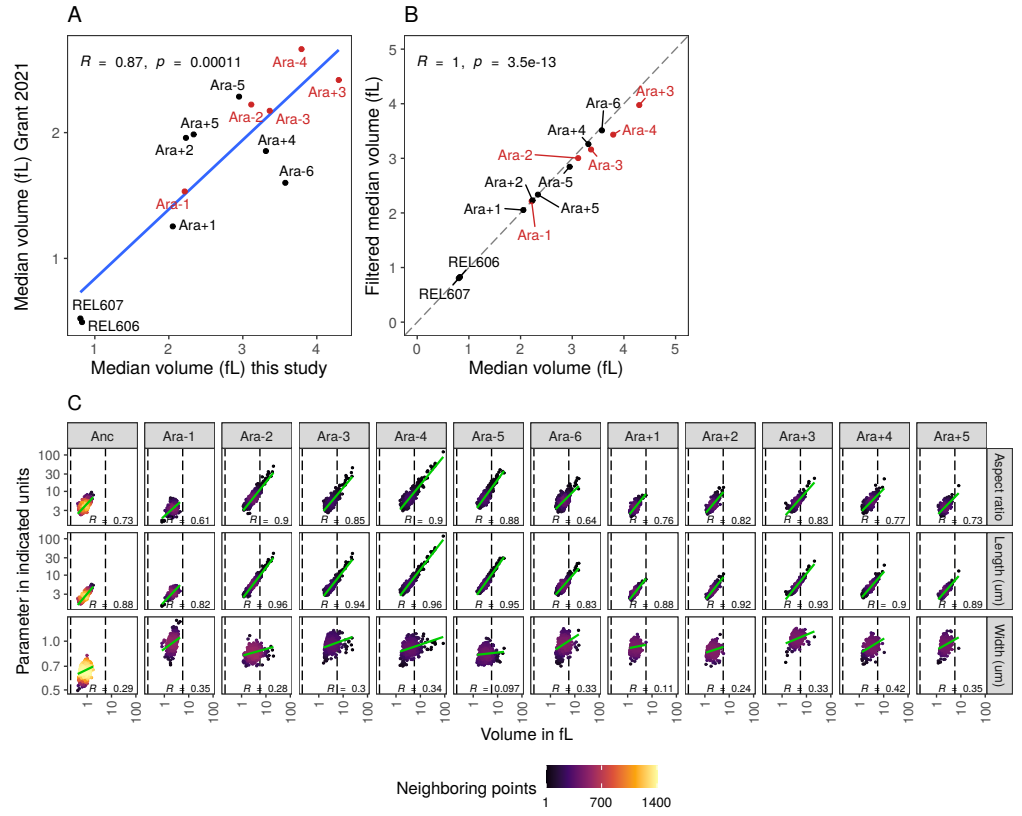


Figure 2—figure supplement 1. (A) Comparison of median volumes of each evolved line from this manuscript to estimates of cellular volumes from ***Grant et al.*** (***2021***)). **(B)** Relationship between median cell volumes of all cells comapred to median cell volume of filtered cells between 0.21 fL and 5.66 fL used in ***Grant et al.*** (***2021***)). **(C)** Increase in cell volume is more strongly correlated with cell length compared to cell width. The dotted lines indicate volumes of 0.21 fL and 5.66 fL.

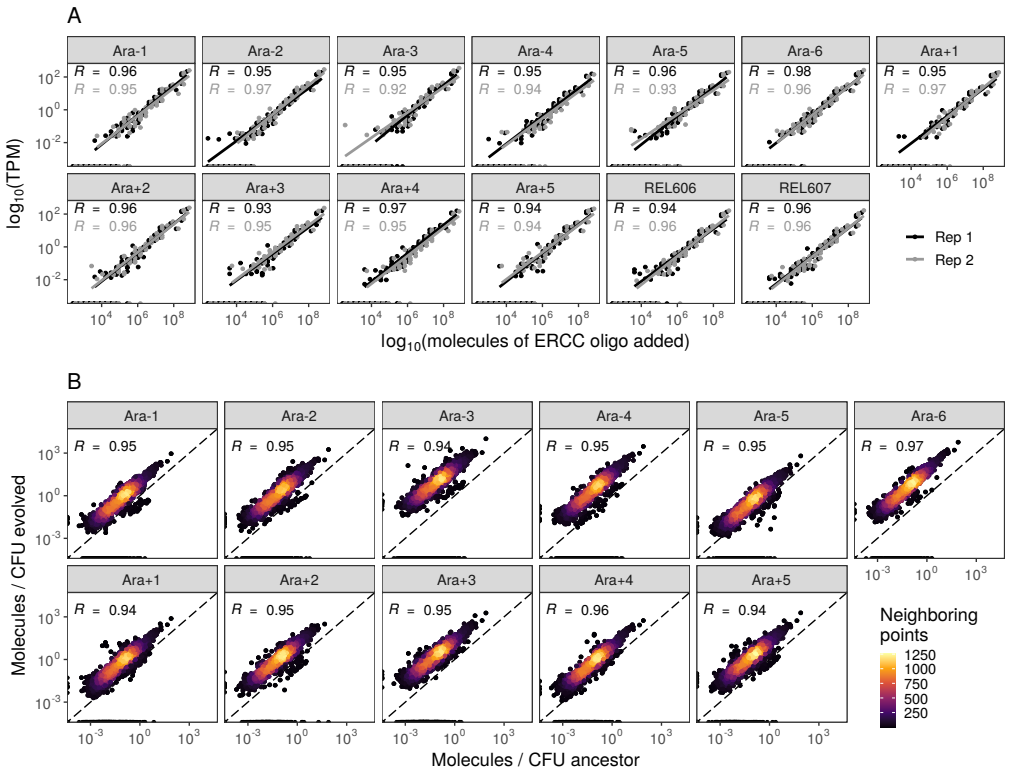


Figure 2—figure supplement 2. (A) Abundances of Spike-in RNA control oligos are correlated with their estimates in sequencing data. Linear models relating the number of molecules of each ERCC control sequence added to their RNA-seq TPM (transcripts per million).**(B)** Most genes have a higher absolute expression in evolved lines. Changes in the absolute number of mRNA molecules per CFU (colony forming unit) in the 50,000th generation of each line relative to the ancestor. The values plotted are the average between 2 replicates of the evolved lines and both replicates from both ancestors. REL606 and REL607 are ancestral strains.

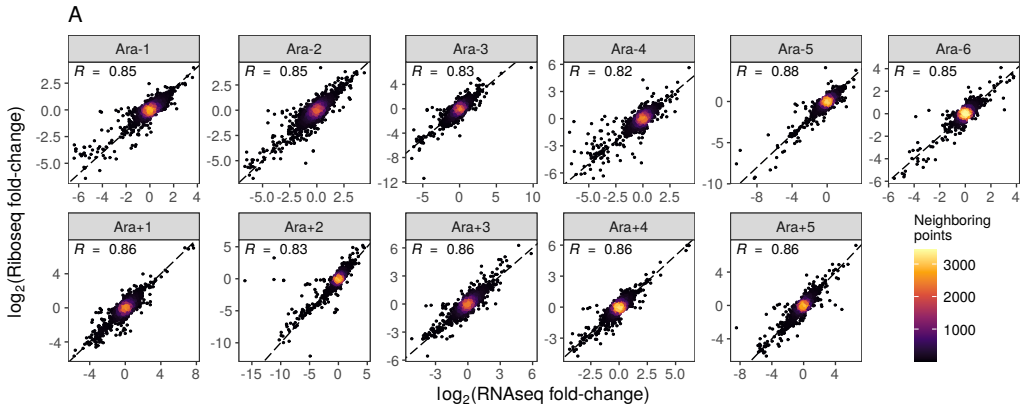


Figure 3—figure supplement 1. (A) The relationship between RNA-seq fold-changes and ribo-seq fold-changes in evolved lines.

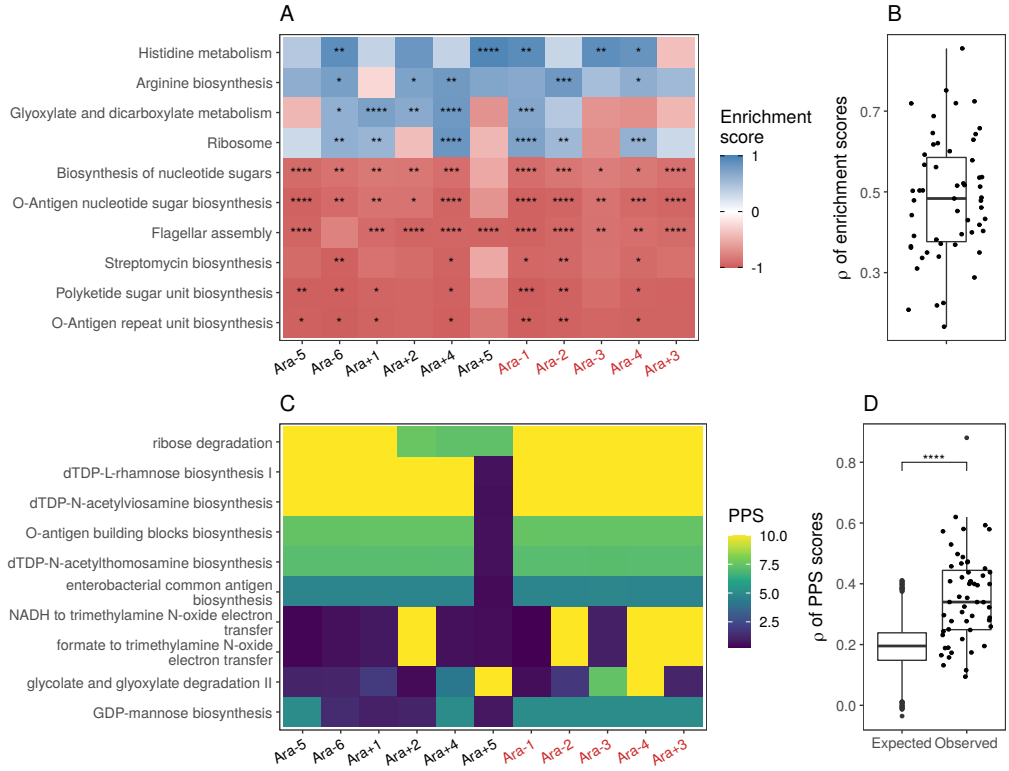


Figure 4—figure supplement 1. (A) Parallel changes in biological processes and pathways. The top 10 KEGG pathways that were significantly altered ($FDR \leq 0.05$) based on Ribo-seq data. Enrichment score represents the degree to which a pathway was up (positive) or downregulated (negative). Functional categories are ordered by increasing mean enrichment score across the lines. Enrichment score represents the degree to which a pathway was up (positive) or downregulated (negative). **(B)** Distribution of pairwise Spearman's correlations of enrichment scores of all significantly altered functional categories ($FDR \leq 0.05$). **(C)** The top 10 pathways with the highest mean Pathway perturbation scores (PPS) calculated from Ribo-seq fold changes. Higher PPS indicates larger degrees of alteration but does not indicate directionality. **(D)** Distribution of pairwise Spearman's correlations based on all PPS scores (observed) compared to 1000 sets of correlations generated from PPS scores calculated after randomization of fold-changes (expected). The p-value is the result of a Kolmogorov–Smirnov test. blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$.

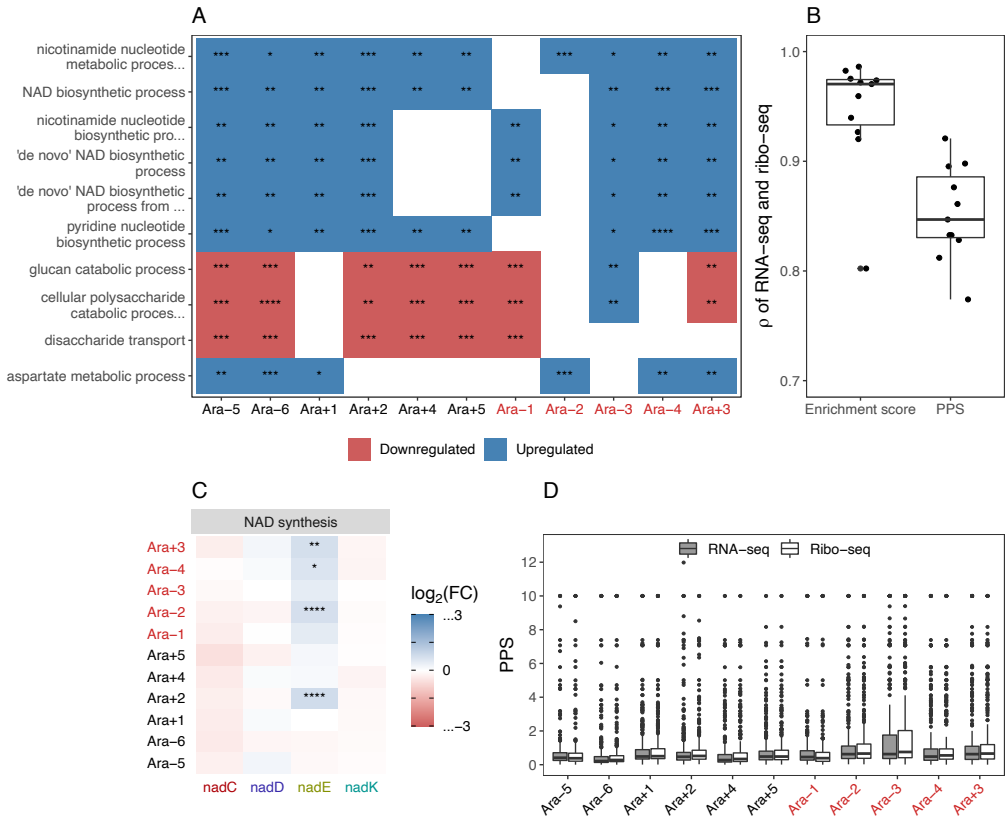
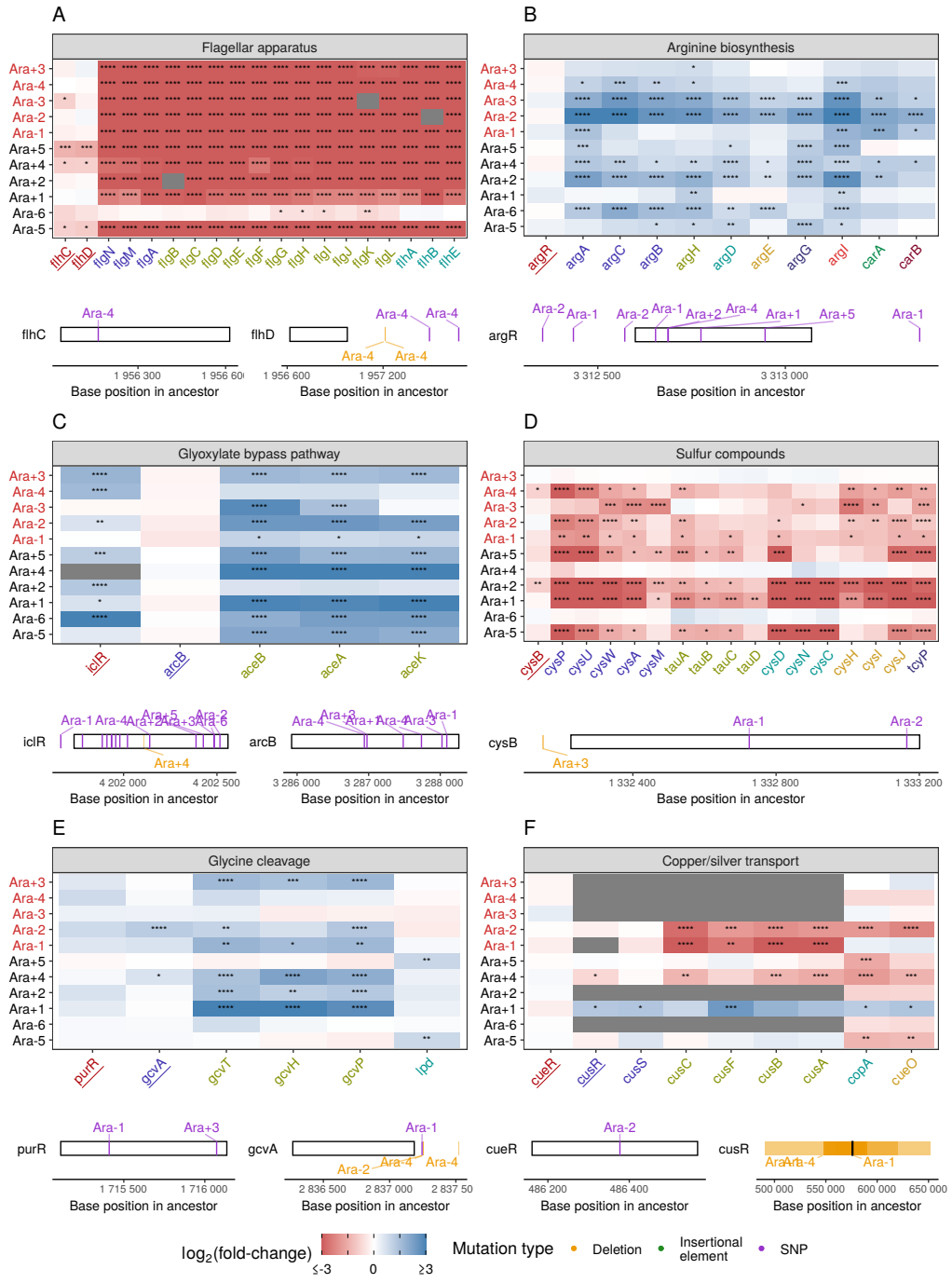


Figure 4—figure supplement 2. (A) The top 10 GO biological process categories that were significantly altered (Fisher's exact test ≤ 0.05). White spaces indicate that the category was not significantly altered in that line. **(B)** Spearman's correlations between the RNA-seq and Ribo-seq enrichment scores within each line. **(C)** RNA-seq fold-changes and DESeq2 q-values for the remaining genes in the NAD synthesis pathway shown in figure 5. Gene names along the x-axis are colored based on operon membership. blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$. **(D)** Distribution of RNA-seq and Ribo-seq pathway perturbation scores (PPS) distributions for each line.



971

Figure 5—figure supplement 1. **(A-F)** Mutations in transcriptional regulators lead to parallel changes in gene expression (RNA-seq). Gene names in each category are colored based on their operon membership. Transcription factors for each class of genes are underlined. Asterisks indicate statistical significance of fold-changes, (blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$). Grey panels in the heatmap indicate gene deletion. Lower panels show the type and location of mutations in each transcription factor.