

Figures and captions

Figure S1

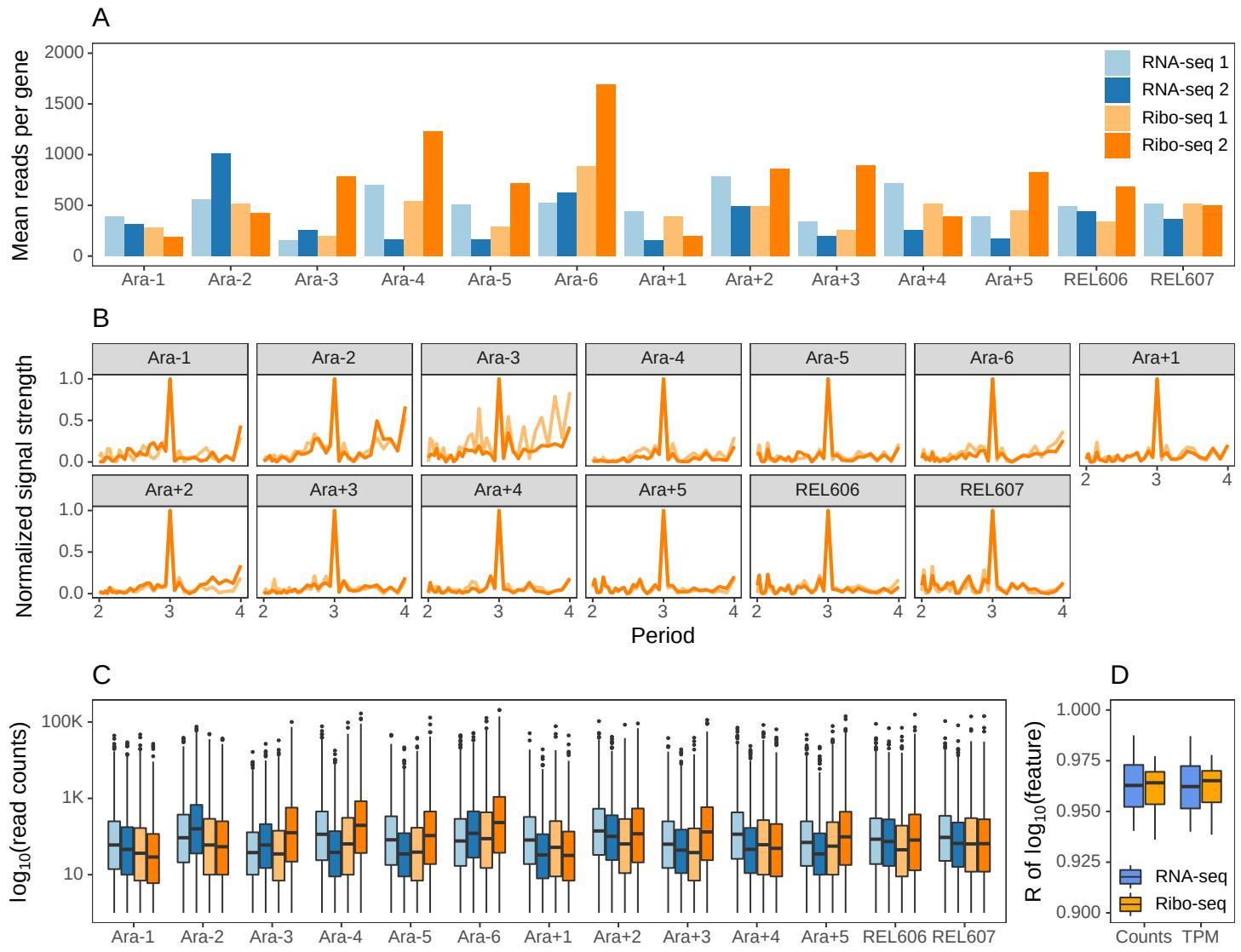


Figure S1: **(A)** The average number of reads aligned per gene by Kallisto for each sample. The color scheme remains the same for panels B and C. **(B)** The periodicity of the ribo-seq datasets determined using a fast Fourier transform (see methods). **(C)** Distributions of reads per gene in each sample. **(D)** Correlations between the replicates based on rounded counts or TPMs.

Figure 1

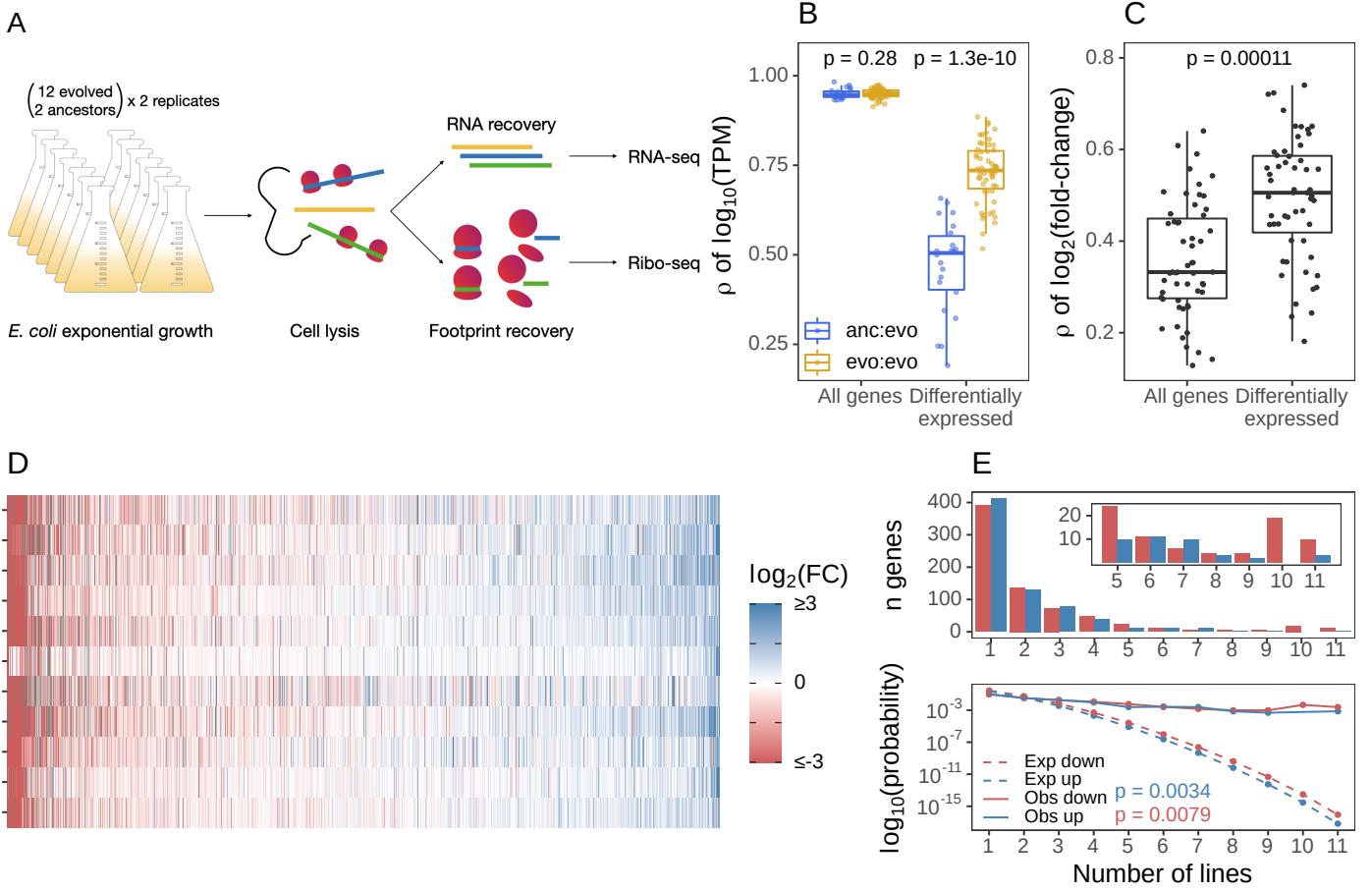


Figure 1: **(A)** Schematic of experimental setup. **(B)** Pairwise Pearson's correlations based on $\log_{10}(\text{TPM})$ (where TPM is the mean from replicates) separated by comparisons between evolved lines or from ancestors to evolved lines. P-values indicate the results of a Kolmogorov-Smirnov (KS) test. For differentially expressed genes (DESeq2 $q \leq 0.01$), 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 (All), and the union of the significant genes between two evolved lines. **(D)** Fold-changes for genes which were significant in at least one line are ordered left to right by increasing mean fold-change across the evolved lines. Genes containing deletions were 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.

Figure S2

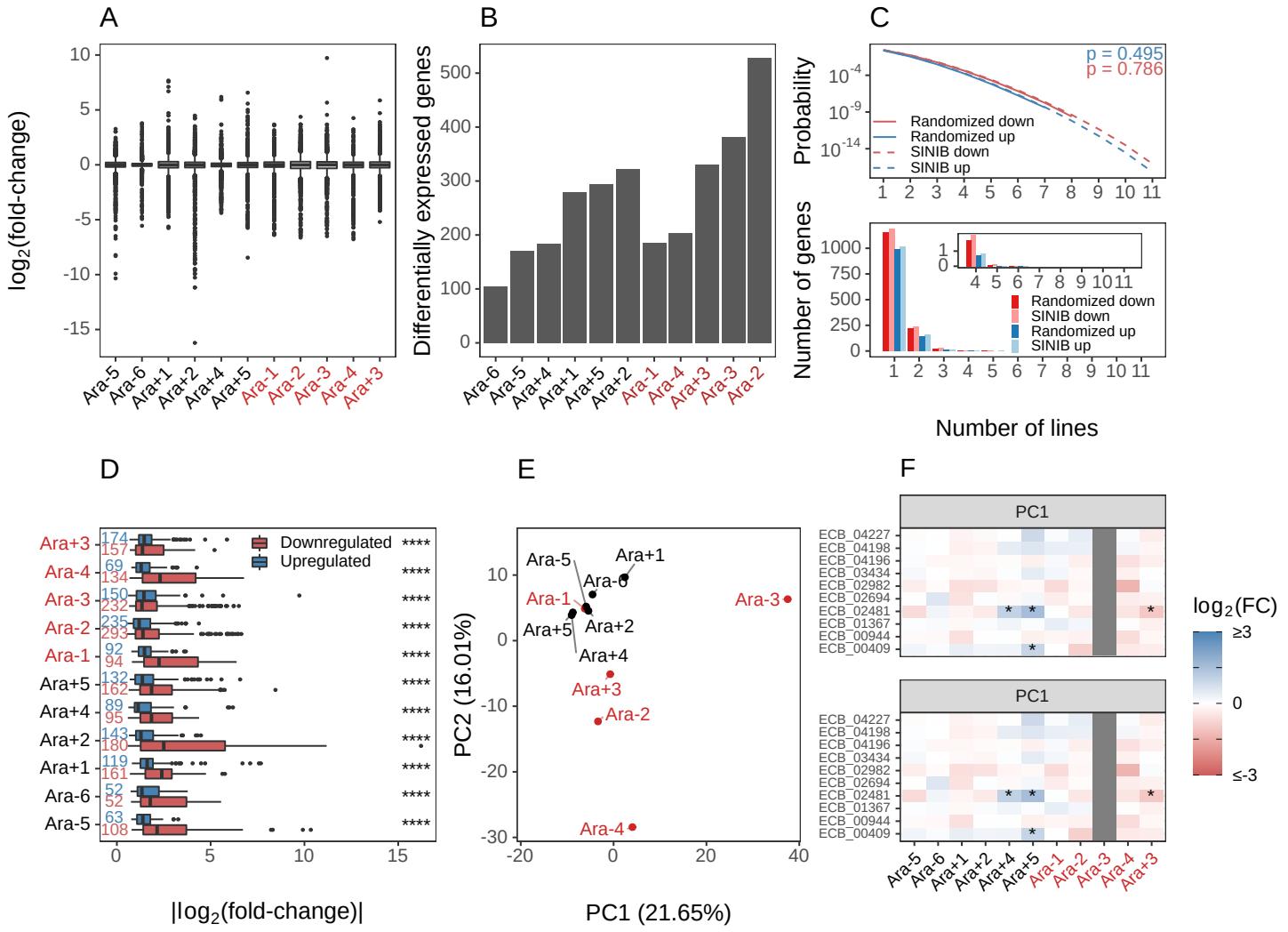


Figure S2: (A) Distributions of all DESeq2 fold-changes in each line. Lines with a mutator phenotype are in red. (B) The number of statistically significant (DESeq2 $q \leq 0.01$) genes in each line. (C) Upper panel shows the probabilities of observing a gene that was statistically significant and altered in the same direction in a given number of lines (x-axis). The solid lines represent mean probabilities derived from randomizing the gene names about their fold-changes and p-values one million times and the dashed lines represent the probabilities calculated using the SINIB method as shown in Fig 1E. P-values show the result of a KS test comparing the randomized to the SINIB distributions. The lower panel shows the number of genes expected to be statistically significant and altered in the same direction in a given number of lines (x-axis) based on the above probabilities. (D) Distributions of fold-changes for the statistically significant genes in each line. The number of transcripts in each box 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) PCA based on 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. The top 10 genes that contribute to PC1 and PC2 are listed to the right. (F) Heatmaps

Figure S3

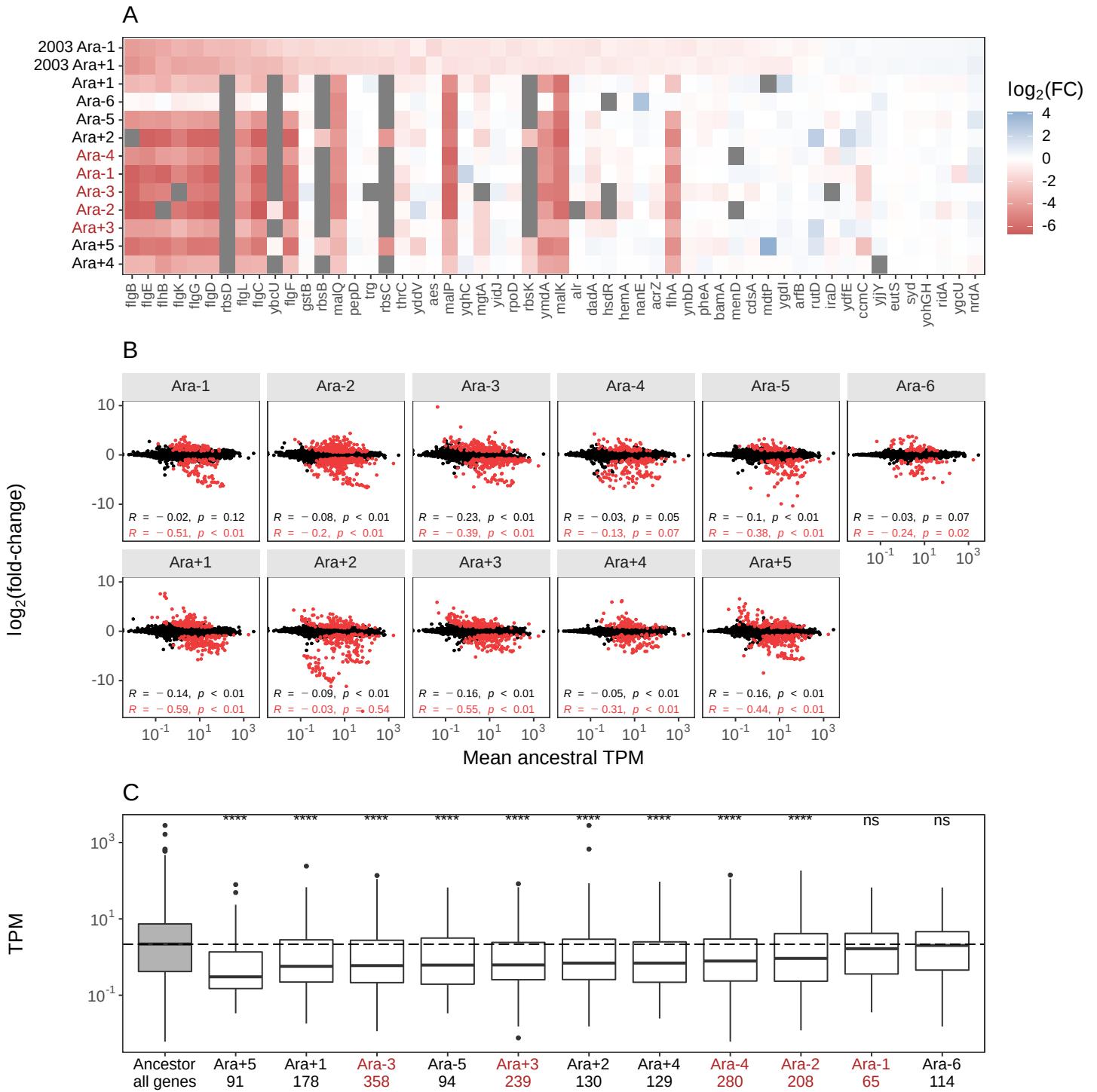


Figure S3: (A) The direction of expression changes from genes identified as differentially expressed from the Cooper 2003 study and the direction of changes for those genes in our dataset. While the two datasets share a color scale for fold-change, the data underlying the Cooper 2003 study was generated using a microarray whereas ours in RNAseq based. **(B)** Relationship between ancestral TPM (mean of all 4 ancestral samples) and fold-change in the indicated evolved line. The red points and correlations represent genes that were statistically significant ($\text{DESeq2 } q \leq 0.01$) in each line. The black points are the remaining genes, but the black correlation is based on all (both significant and not) genes. **(C)** Genes deleted in each evolved line had low TPM in the ancestor. The ancestor box represents the TPMs for all genes. Each evolved line box represents the ancestral TPM for genes which were deleted in that line. The number of genes deleted in each line is listed on the x-axis.

Figure 2

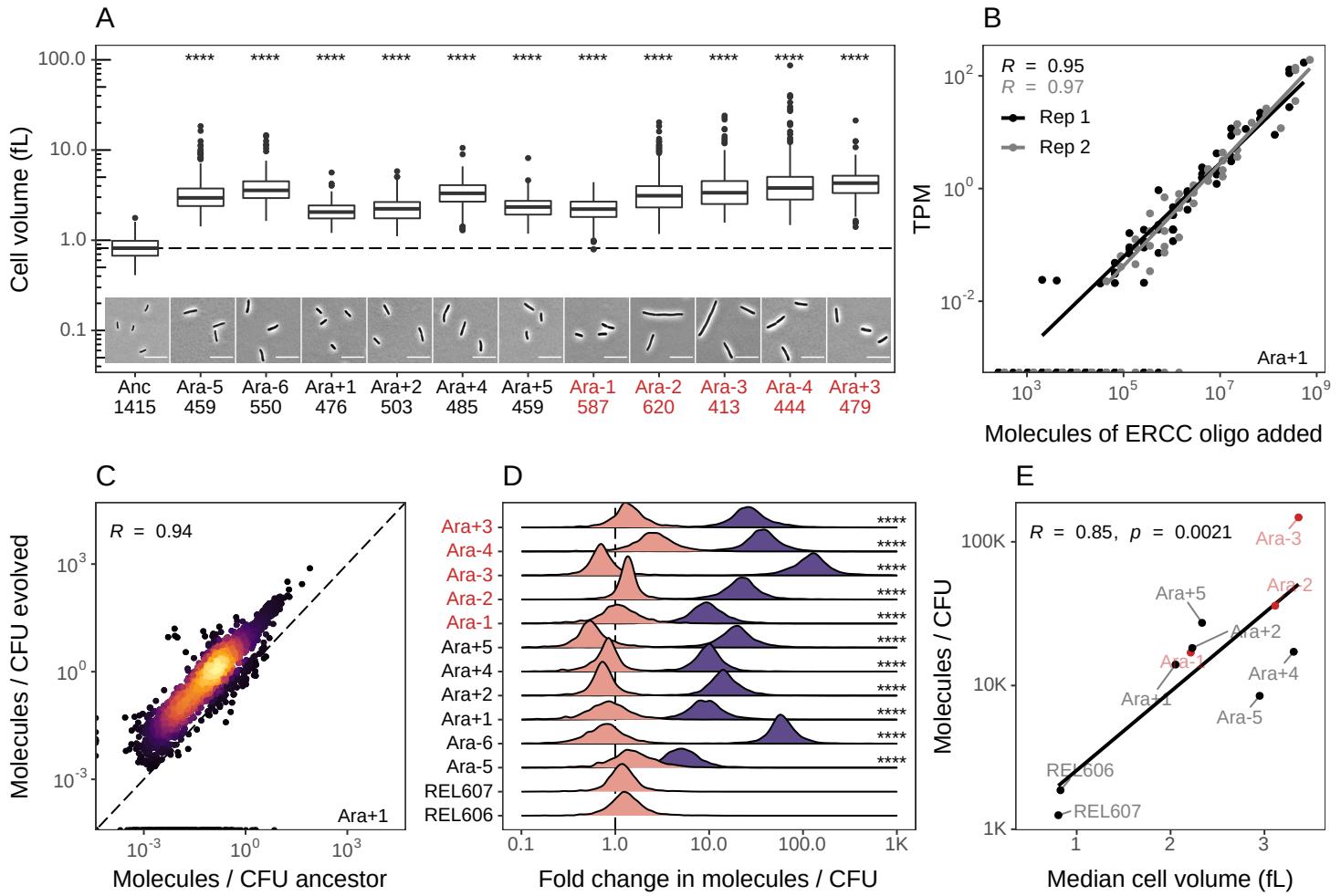


Figure 2: **(A)** Volume distributions (femtoliters, fL) determined by phase-contrast microscopy. Micrographs of each line are shown with a scale bar of 10um and the number of cells imaged is shown beneath each line name. The dotted line indicates the ancestral median and the asterisks indicate the results of a Kolmogorov–Smirnov test of each line to the ancestor. NS: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$. **(B)** Example of the model that estimates number of molecules from TPM using the ERCC spike-ins. **(C)** The mean number of molecules per CFU from all 4 ancestral samples compared to the mean from 2 replicates of Ara+1. **(D)** Fold-change distributions for all evolved lines. Purple curves represent the fold-change from ancestor to evolved lines, pink curves represent the fold-change between replicates. Fold-changes were calculated from 4 replicates of the ancestor and 2 replicates of each evolved line. Asterisks indicate the results of a Kolmogorov–Smirnov test between the fold-changes between replicates and fold-changes to the ancestors. **(E)** The relationship between median cellular volume and molecules of RNA per median cell volume.

Figure S4

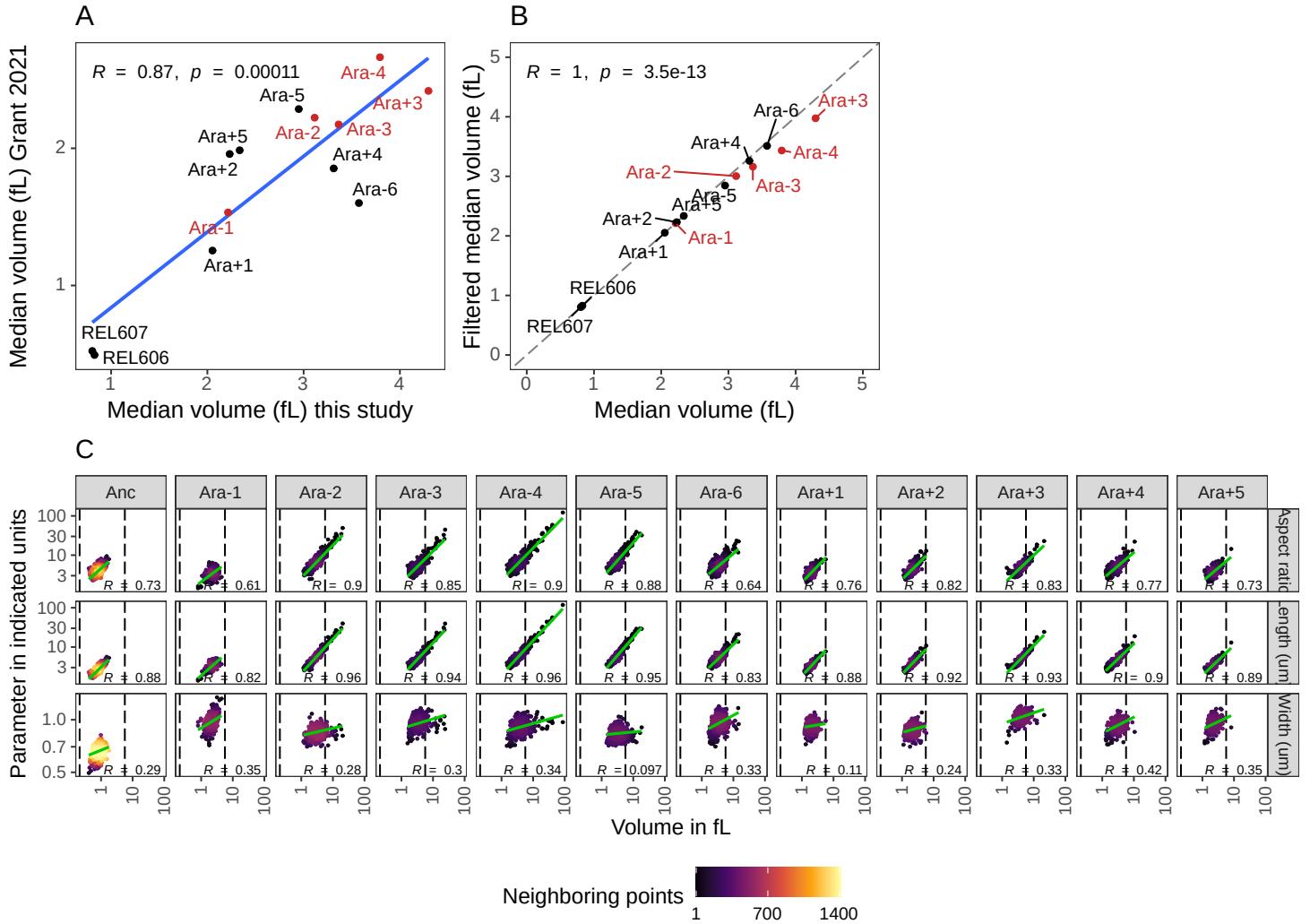


Figure S4: (A) Comparison of median volumes from this manuscript to those from Grant et. al. 2021. (B) Changes in our median cell volume after only cells between 0.21 fL and 5.66 fL (the same filtering range as in Grant 2021) were considered when calculating medians. (C) The relationship between volume and the parameter indicated on the right of the graph. The dotted line indicates volumes of 0.21 fL and 5.66 fL.

Figure S5

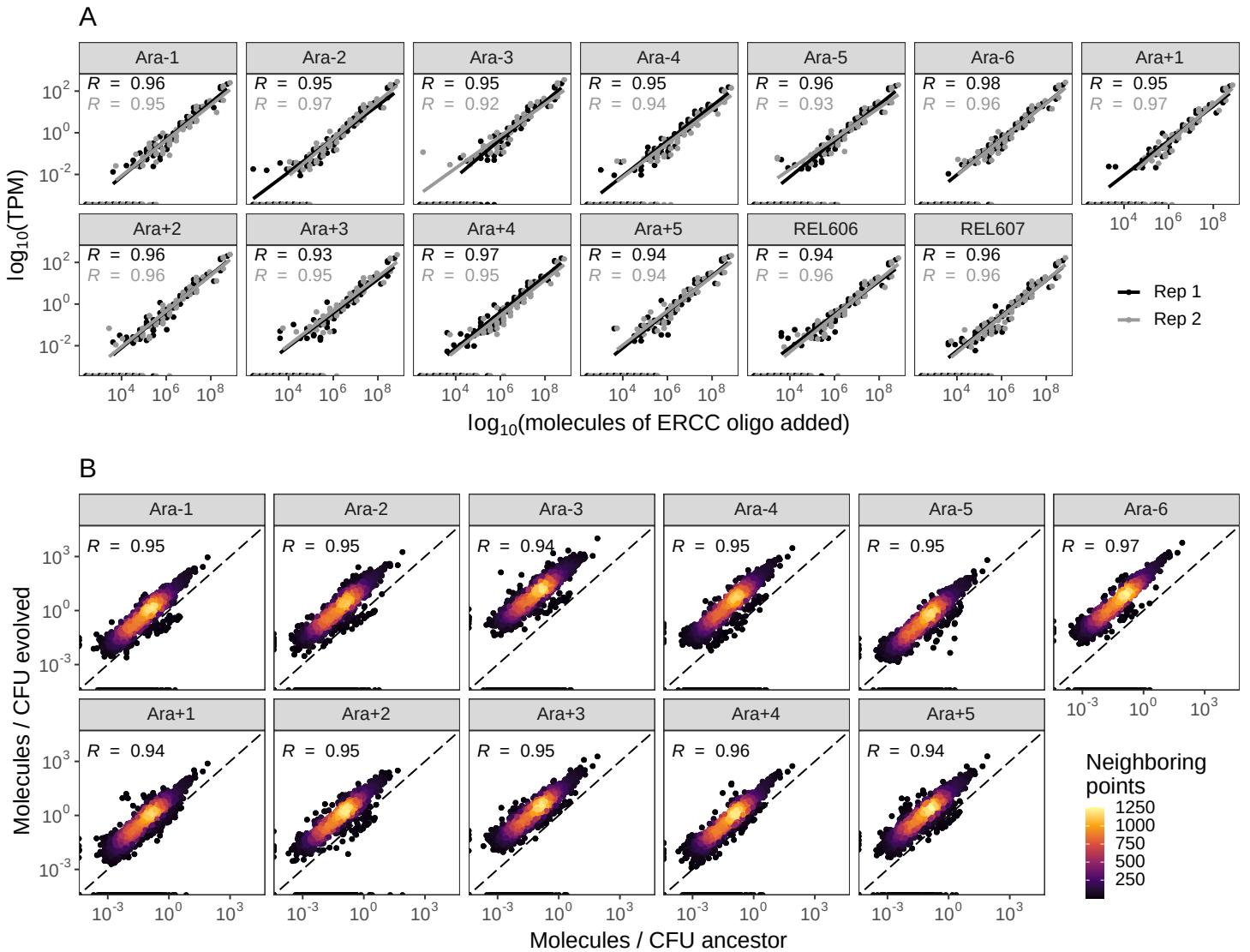


Figure S5: **(A)** The relationship between the number of molecules of ERCC oligos added and their TPM in each sample. **(B)** Slopes and intercepts for all models are similar. **(C)** Correlations of replicates based on $\log_{10}(\text{number of molecules})$ estimated from each replicates respective model. **(D)** The mean number of molecules per cell volume from all 4 ancestral samples compared to the mean from 2 replicates of an evolved line.

Figure 3

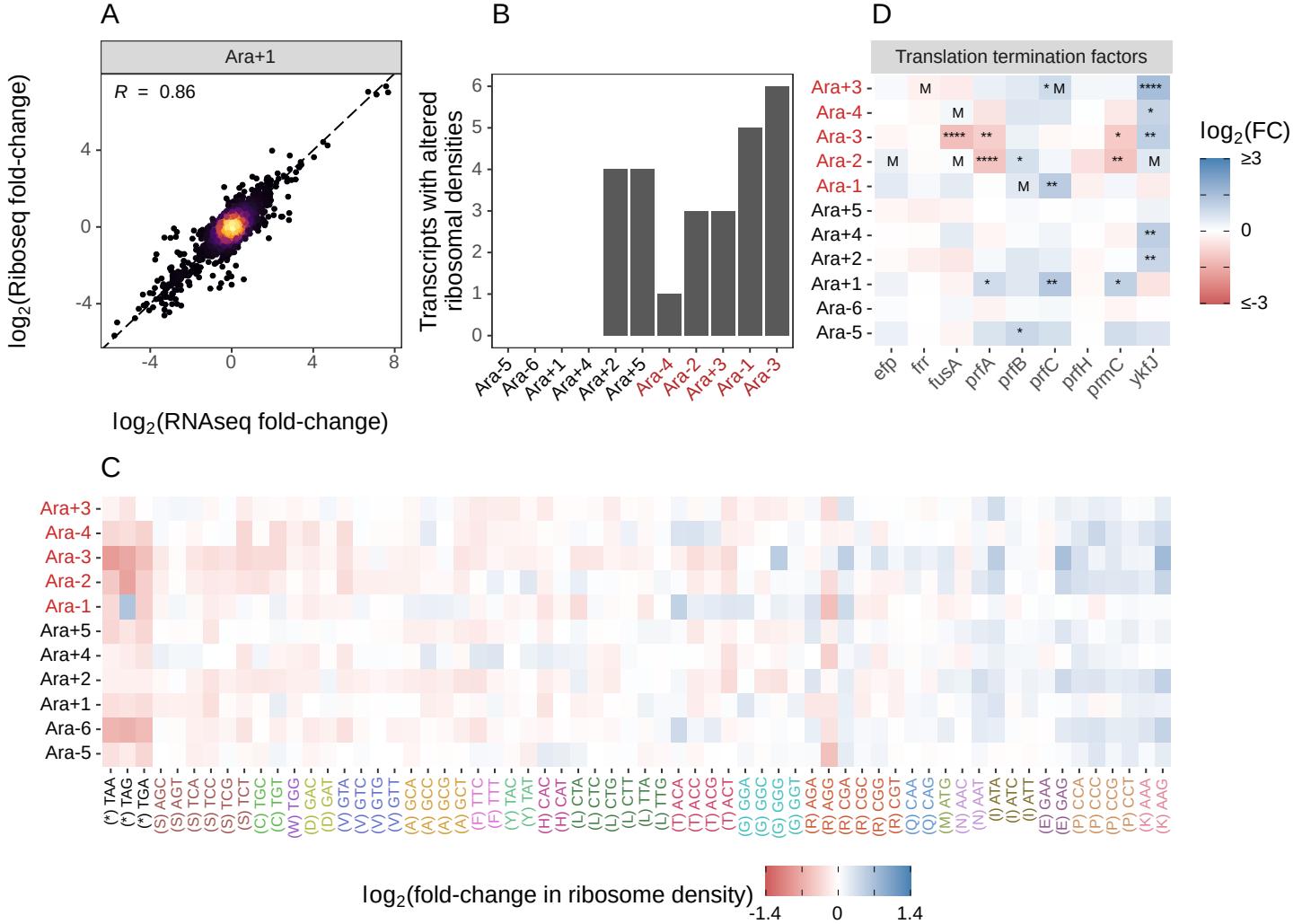


Figure 3: **(A)** Example of correlation between RNA-seq and ribo-seq fold-changes from Ara+1, see Fig S6 for other lines. **(B)** The number of genes which reached statistical significance in the Riborex analysis ($q \leq 0.01$). **(C)** Ratio of codon specific ribosome density relative to the ancestor for each line. Red indicates decreased density in an evolved line relative to the ancestor and blue indicates increased density. 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)** RNA-seq fold-changes for termination factors, asterisks indicate DESeq2 q-values blank: > 0.05, *: ≤ 0.05 , **: ≤ 0.01 , ***: ≤ 0.001 ****: ≤ 0.0001 and an “M” indicates a SNP in that gene.

Figure S6

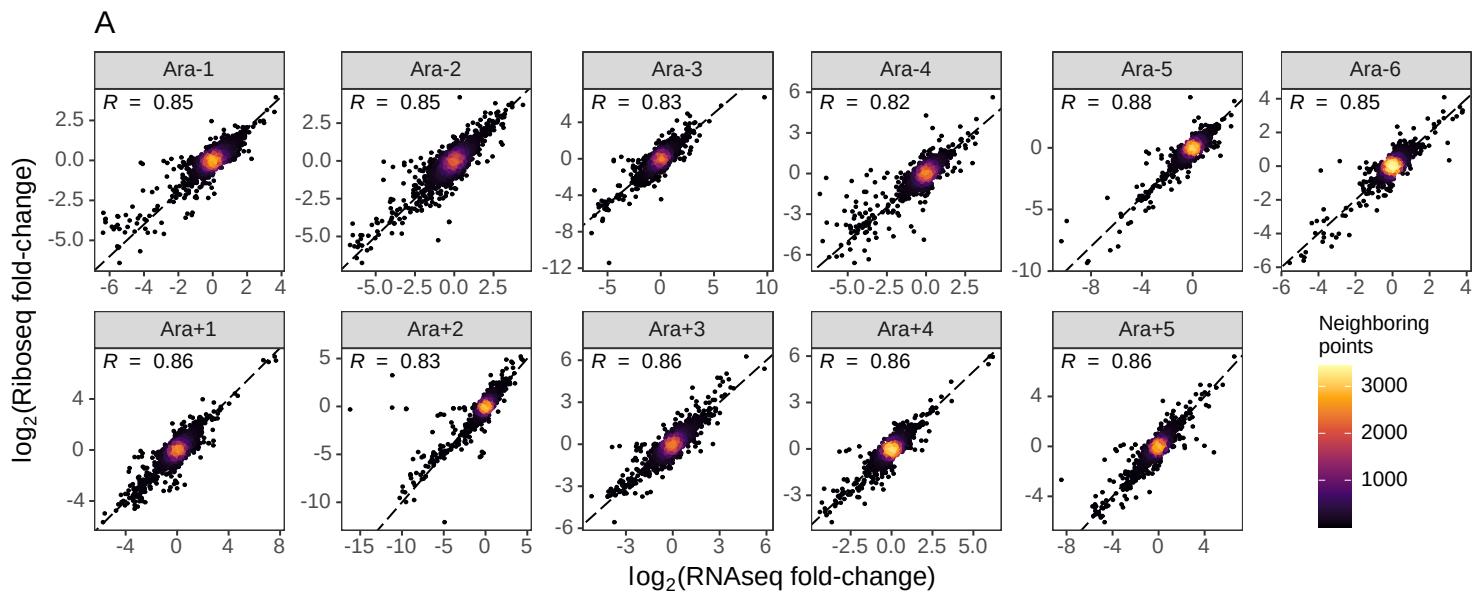


Figure S6: (A) The relationship between RNA-seq fold-change and ribo-seq fold-change.

Figure 4

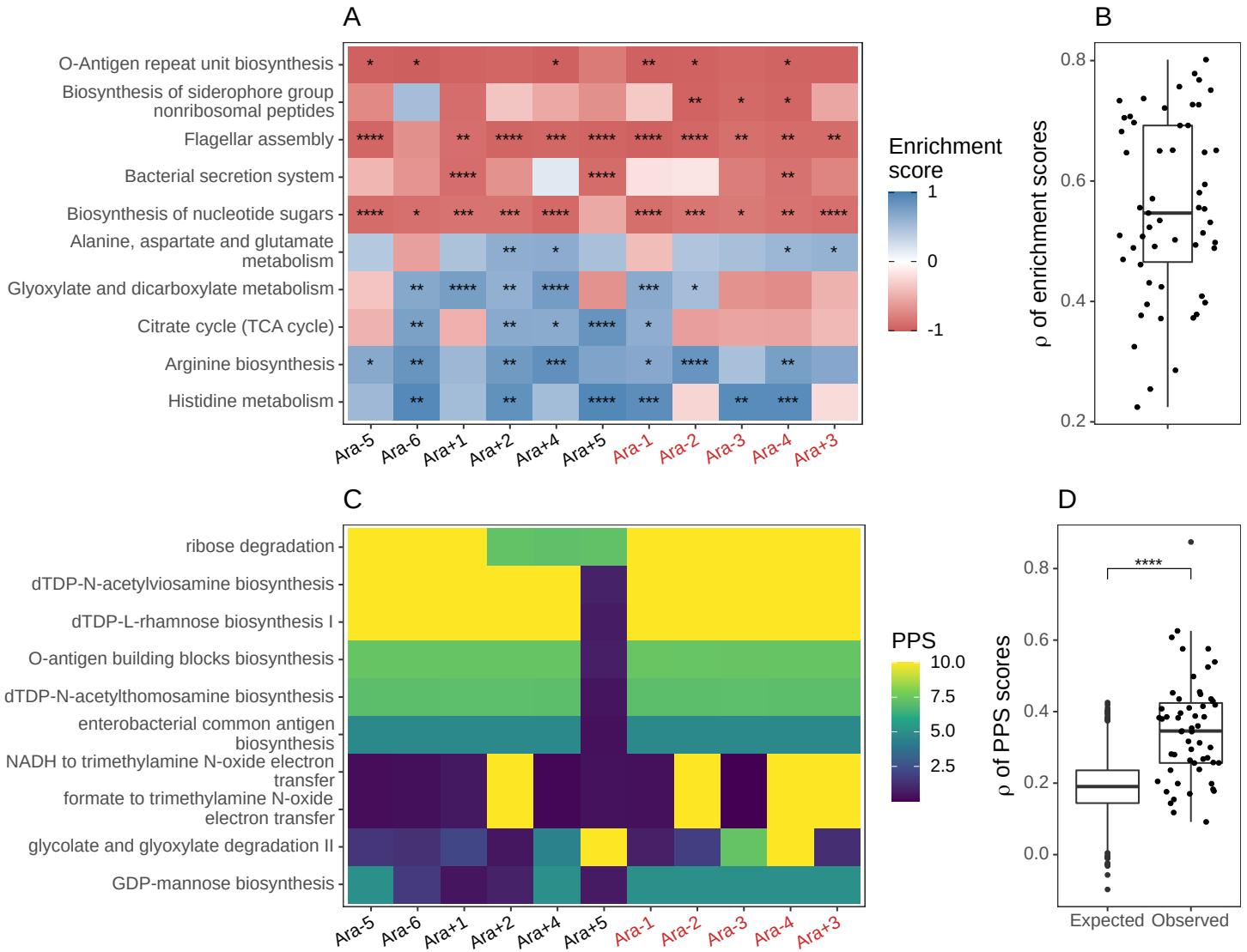


Figure 4: **(A)** The top 10 KEGG pathways that were statistically significant ($FDR \leq 0.05$) in the most number of lines, positive enrichment scores indicate upregulation of genes in that category. **(B)** Distribution of line to line Spearman's correlations based on enrichment scores from the set of categories that reached statistical significance ($FDR \leq 0.05$) in at least one line. **(C)** The top 10 pathways with the highest mean PPS across the lines are arranged from top to bottom with higher PPS indicate larger degrees of alteration. **(D)** Line to line pairwise Spearman's correlations based on all PPS scores (observed) are 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: > 0.05 , *: ≤ 0.05 , **: ≤ 0.01 , ***: ≤ 0.001 , ****: ≤ 0.0001 .

Figure 5

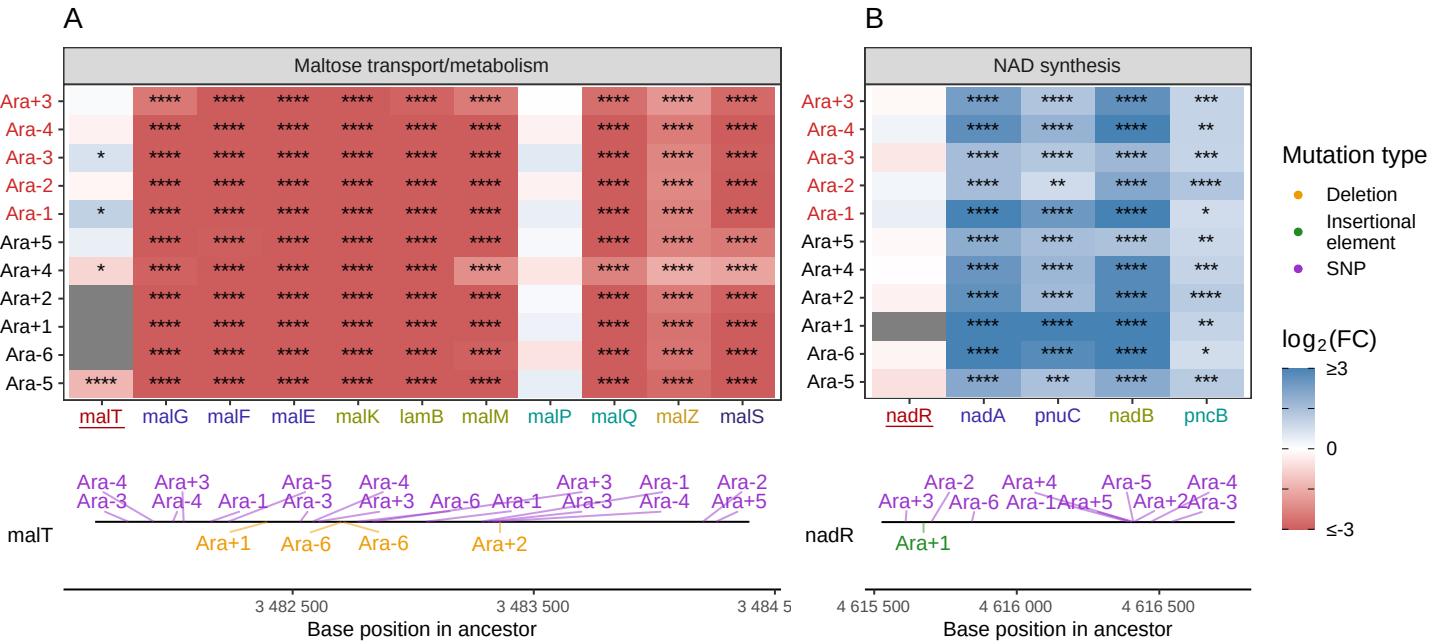


Figure 5: **(A-B)** RNA-seq fold-changes and DESeq2 q-values (blank: $q > 0.05$, *: $q \leq 0.05$, **: $q \leq 0.01$, ***: $q \leq 0.001$ ****: $q \leq 0.0001$) for genes involved in various pathways are shown. In each plot, genes along the x-axis are colored based on operon membership and coloring from one plot is unrelated to the others. Transcription factors that regulate the other genes in the plot are underlined and the location and type of mutation to them are shown beneath each heatmap.

Figure S7

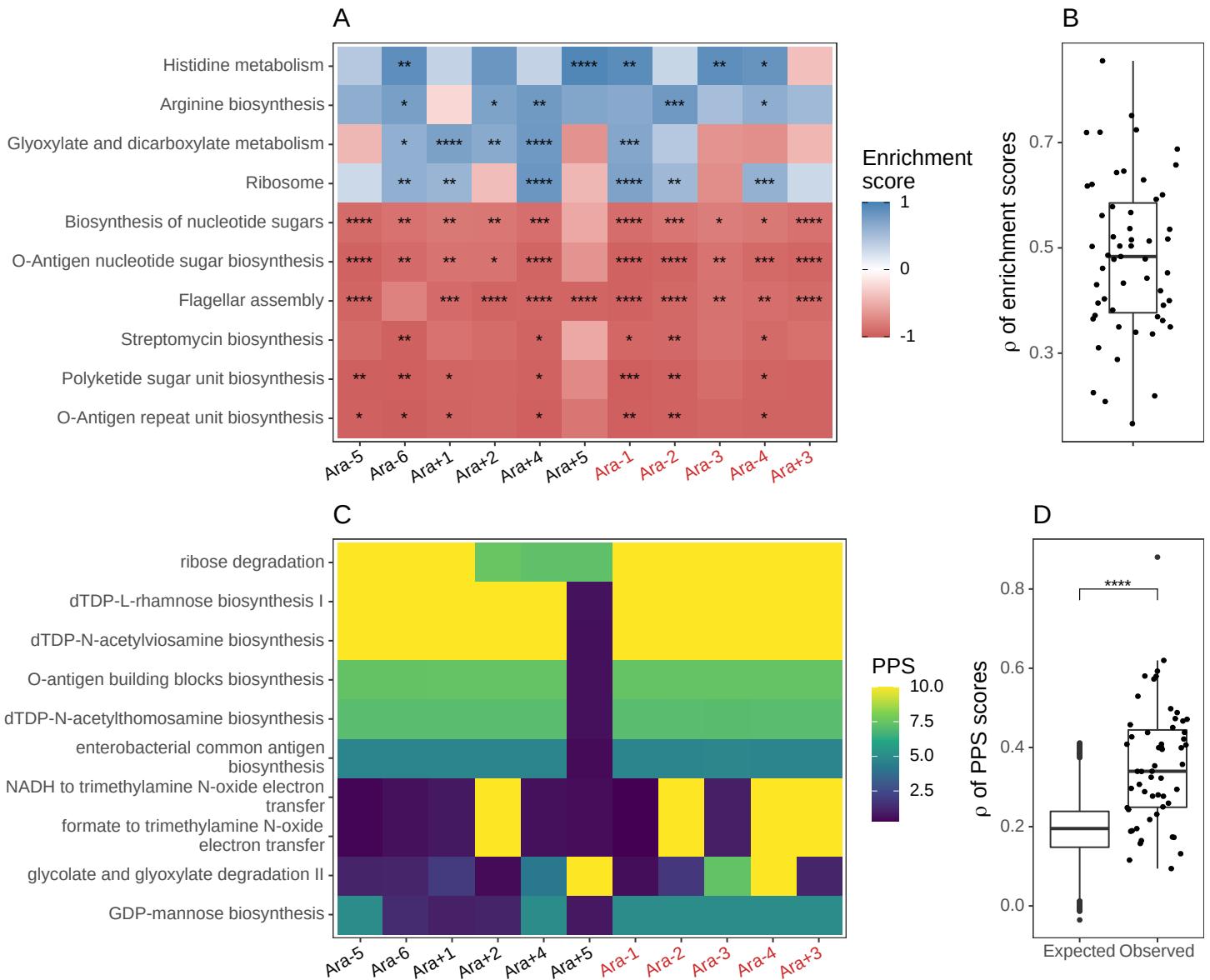


Figure S7: Ribo-seq KEGG and PPS data. **(A)** The top 10 KEGG pathways that were statistically significant ($FDR \leq 0.05$) in the most number of lines, positive enrichment scores indicate upregulation of genes in that category. **(B)** Distribution of line to line Spearman's correlations based on enrichment scores from the set of categories that reached statistical significance ($FDR \leq 0.05$) in at least one line. **(C)** The top 10 pathways with the highest mean PPS across the lines are arranged from top to bottom with higher PPS indicate larger degrees of alteration. **(D)** Line to line pairwise Spearman's correlations based on all PPS scores (observed) are 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: > 0.05 , *: ≤ 0.05 , **: ≤ 0.01 , ***: ≤ 0.001 , ****: ≤ 0.0001 .

Figure S8

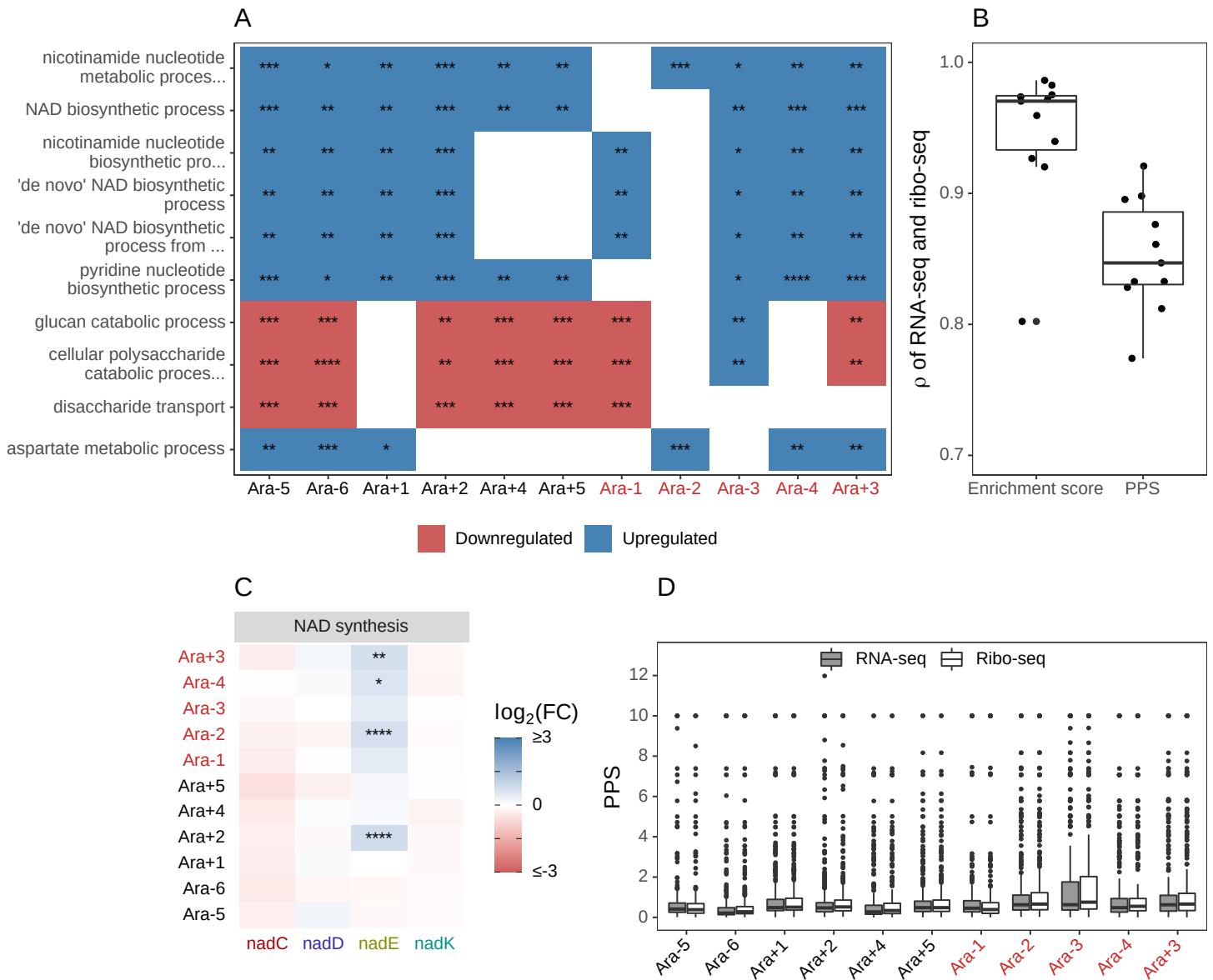


Figure S8: (A) The top 10 GO biological process categories that reached statistical significance (Fisher's exact test ≤ 0.05) in the most number of evolved lines. The color indicates whether a category was found using down or upregulated genes, white spaces indicate that the category was not significant in that line. (B) Spearman's correlations between the RNA-seq and ribo-seq 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. (D) RNA-seq and ribo-seq PPS distributions for each line. Blank: > 0.05 , *: ≤ 0.05 , **: ≤ 0.01 , ***: ≤ 0.001 , ****: ≤ 0.0001 .

Figure S9

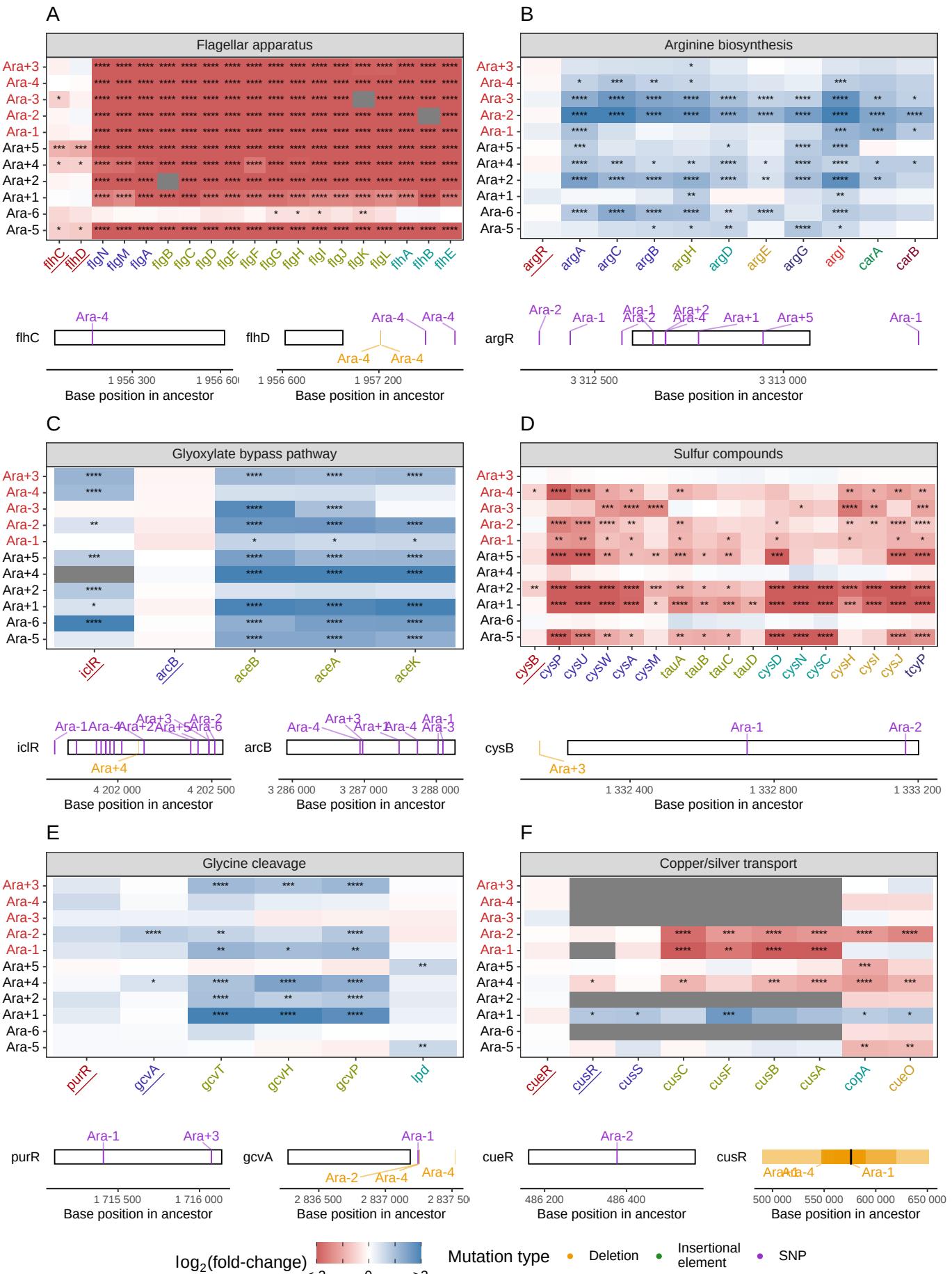


Figure S9: **(A-F)** RNA-seq fold-changes and DESeq2 q-values (blank: $q > 0.05$, *: $q \leq 0.05$, **: $q \leq 0.01$, ***: $q \leq 0.001$ ****: $q \leq 0.0001$) for genes involved in various pathways are shown. In each plot, genes along the x-axis are colored based on operon membership and coloring from one plot is unrelated to the others. Transcription factors that regulate the other genes in the plot are underlined and the location and type of mutation to them are shown beneath each heatmap.