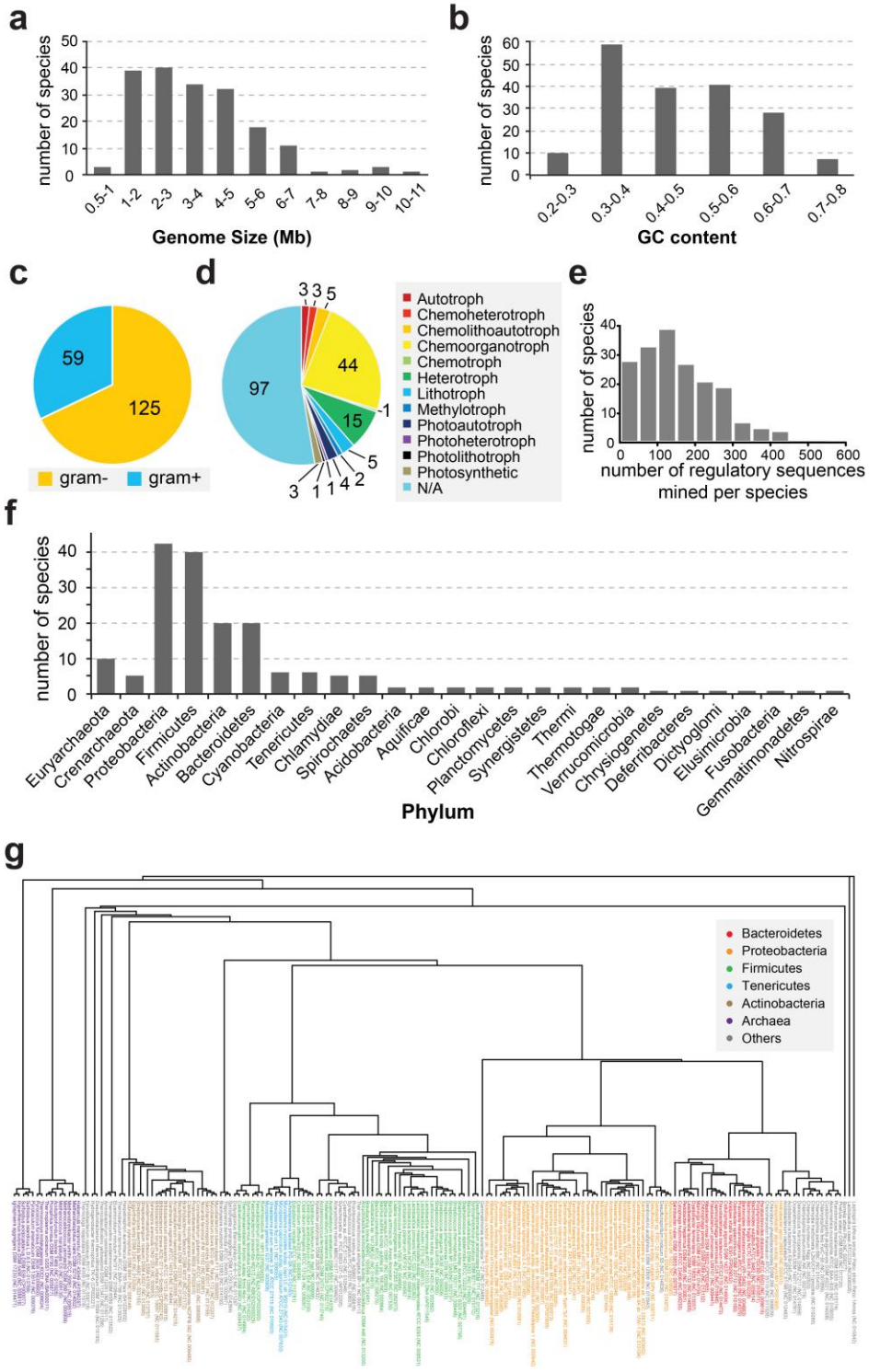
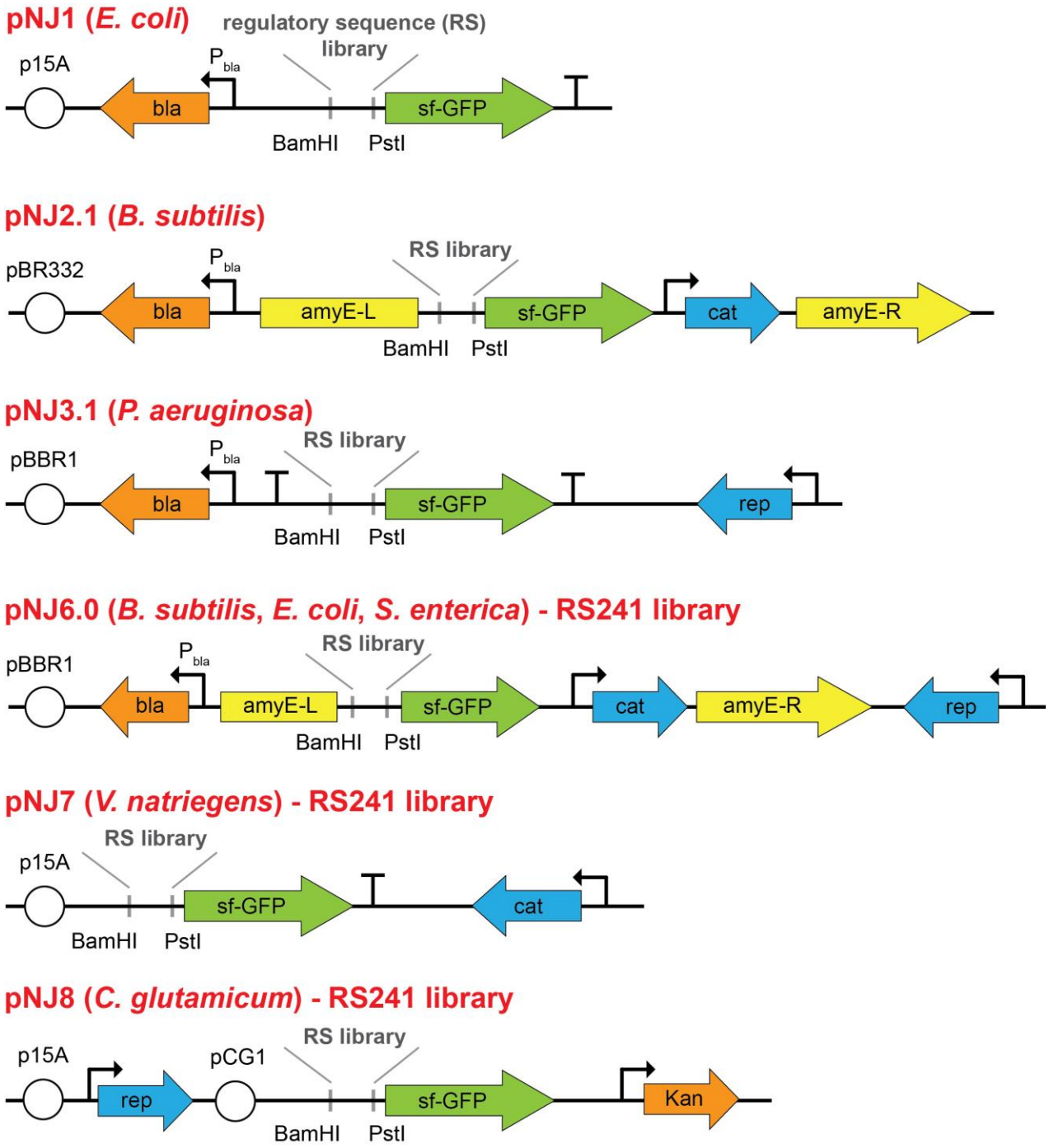


Supplementary Figure 1

Metadata of the 184 donor genomes used to derive the regulatory sequences used in this study.

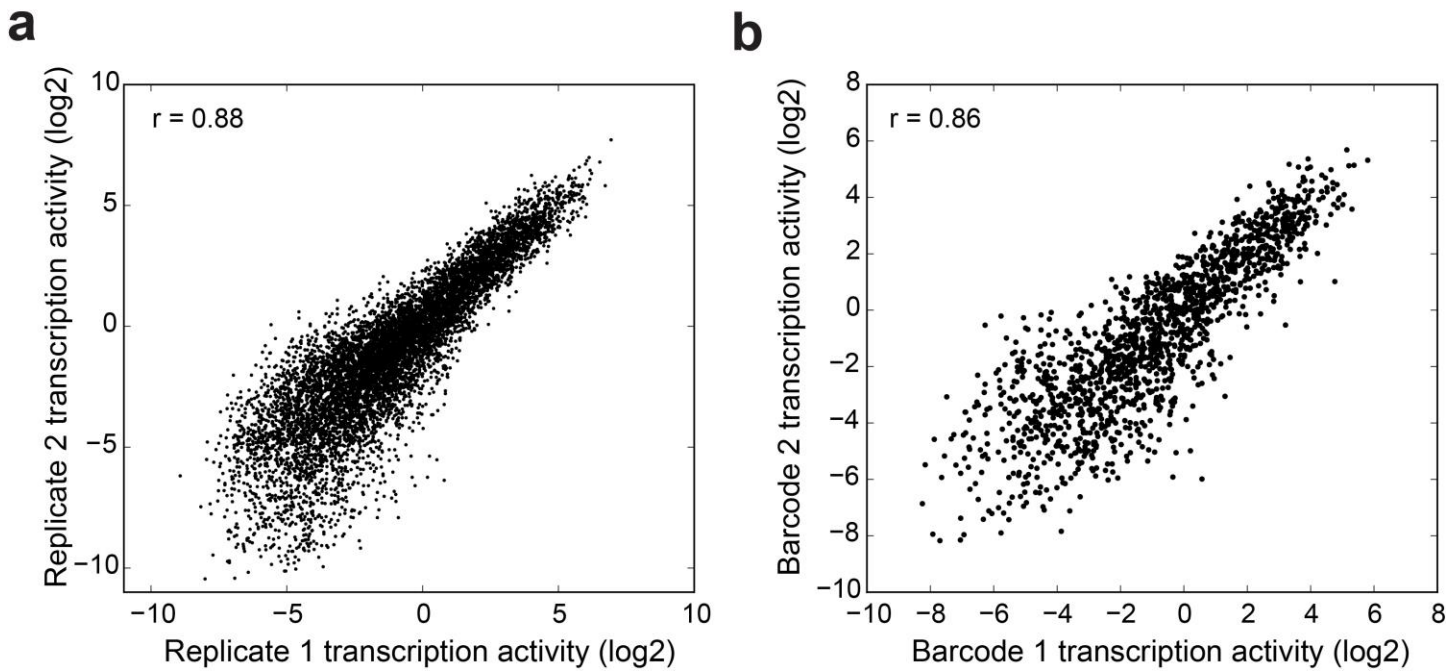
(a) genome size, **(b)** genomic GC content, **(c)** gram staining, **(d)** lifestyle, **(e)** number of regulatory sequences mined per genome, **(f)** the number of genomes per phylum, and **(g)** the 16S phylogenetic tree.



Supplementary Figure 2

Vector designs.

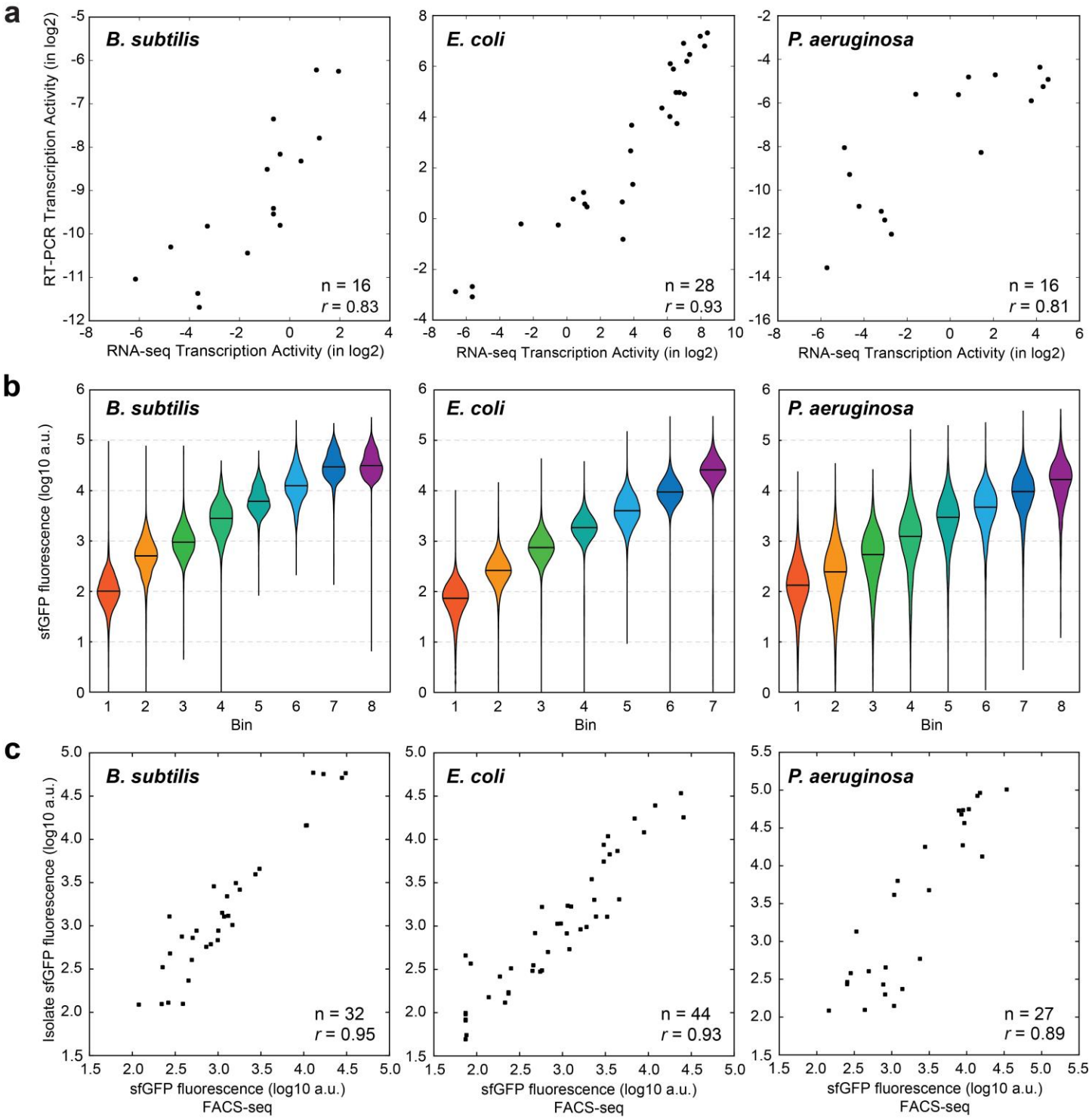
Vector maps for pNJ1, pNJ2.1, and pNJ3.1 used for expression measurements of metagenomic regulatory sequence library (RS) in *E. coli*, *B. subtilis*, and *P. aeruginosa* respectively and pNJ6.0, pNJ3.1, pNJ7 and pNJ8, which were used for RS241 library measurements in *E. coli*, *B. subtilis*, *P. aeruginosa, S. enterica*, *V. natriegens* and *C. glutamicum*.



Supplementary Figure 3

Replication experiments to validate method performance.

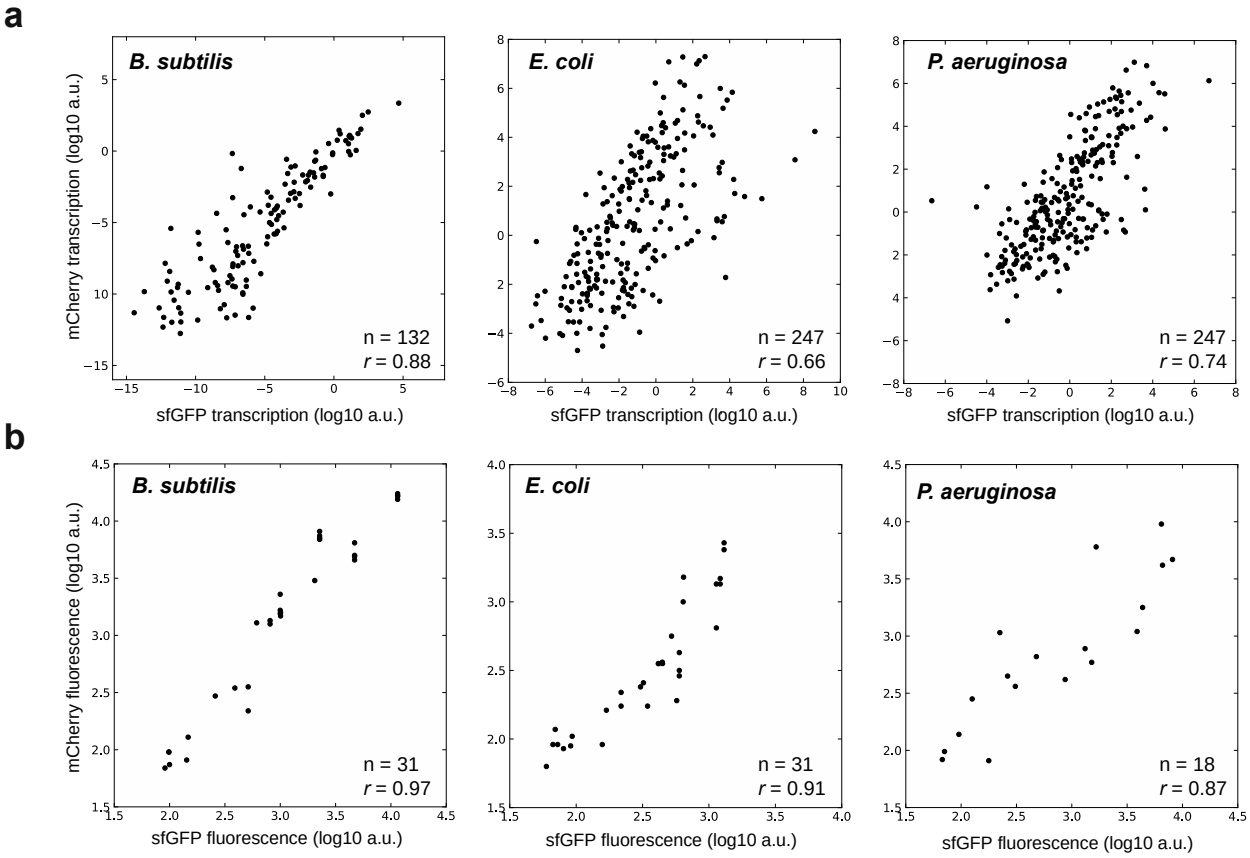
(a) Correlation of transcriptional measurements of RS library across two independent replicate cultures (>10 DNA counts across both replicates, n = 18,845) *in E. coli* performed on different days. **(b)** Correlation of transcriptional measurements of identical RSs with two different barcodes *in E. coli* (>10 DNA counts across both constructs, n = 2,273). Pearson correlation (*r*) is listed in each panel.



Supplementary Figure 4

Validations of gene expression measurements.

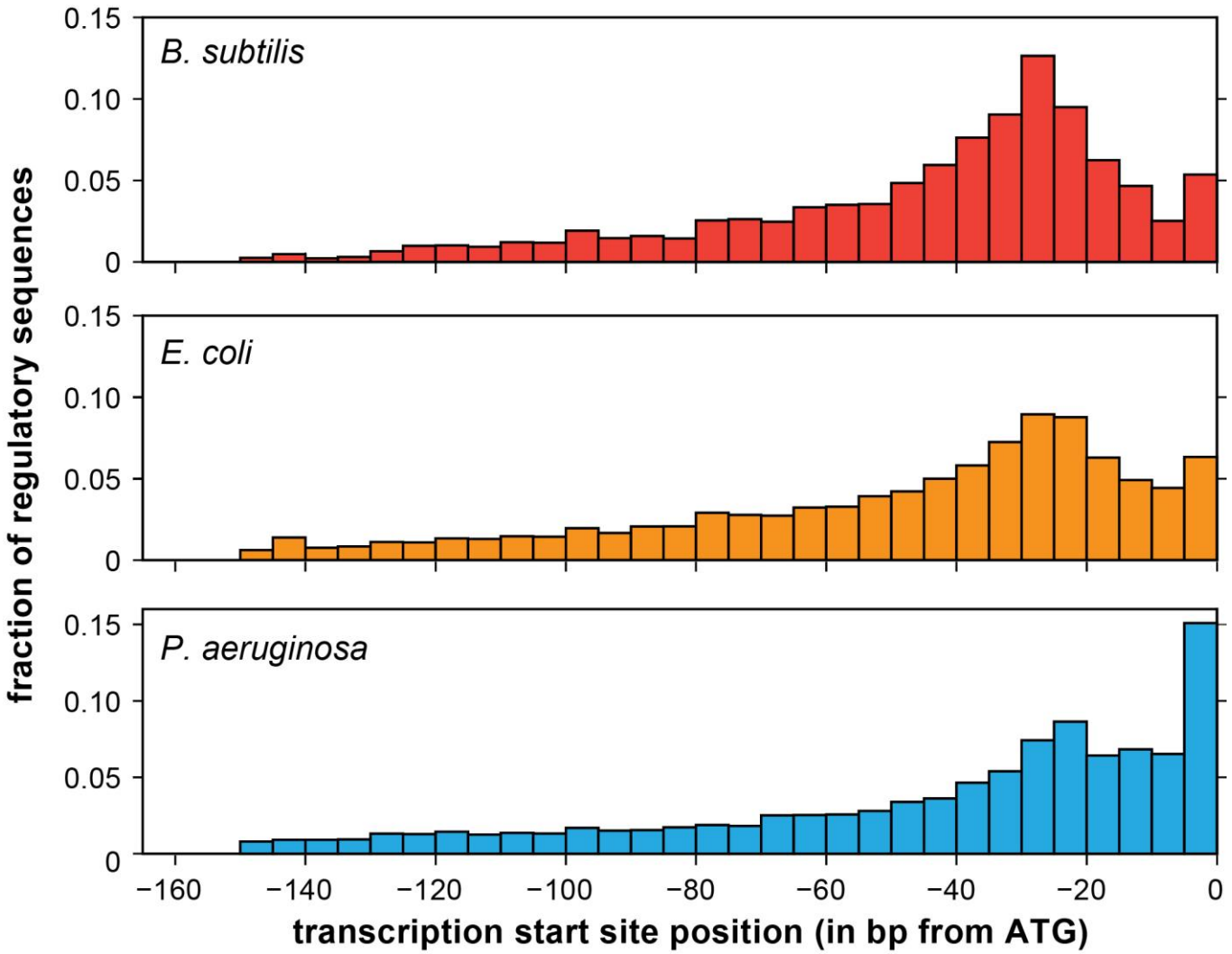
(a) Correlation of pooled RNA-seq measurements with individual RT-PCR data from isolate strains containing RS library members for three host species. **(b)** GFP fluorescence distributions of post-FACS RS library populations displayed as violin plots (n = 10,000 cells, mean value shown as horizontal bar). **(c)** Correlation of pooled FACS-seq measurements with individual flow cytometry measurements of isolate strains. Pearson correlation coefficients and sample sizes are listed for (*r* and n) listed in each subplot.



Supplementary Figure 5

Alternative reporter gene experiments**.**

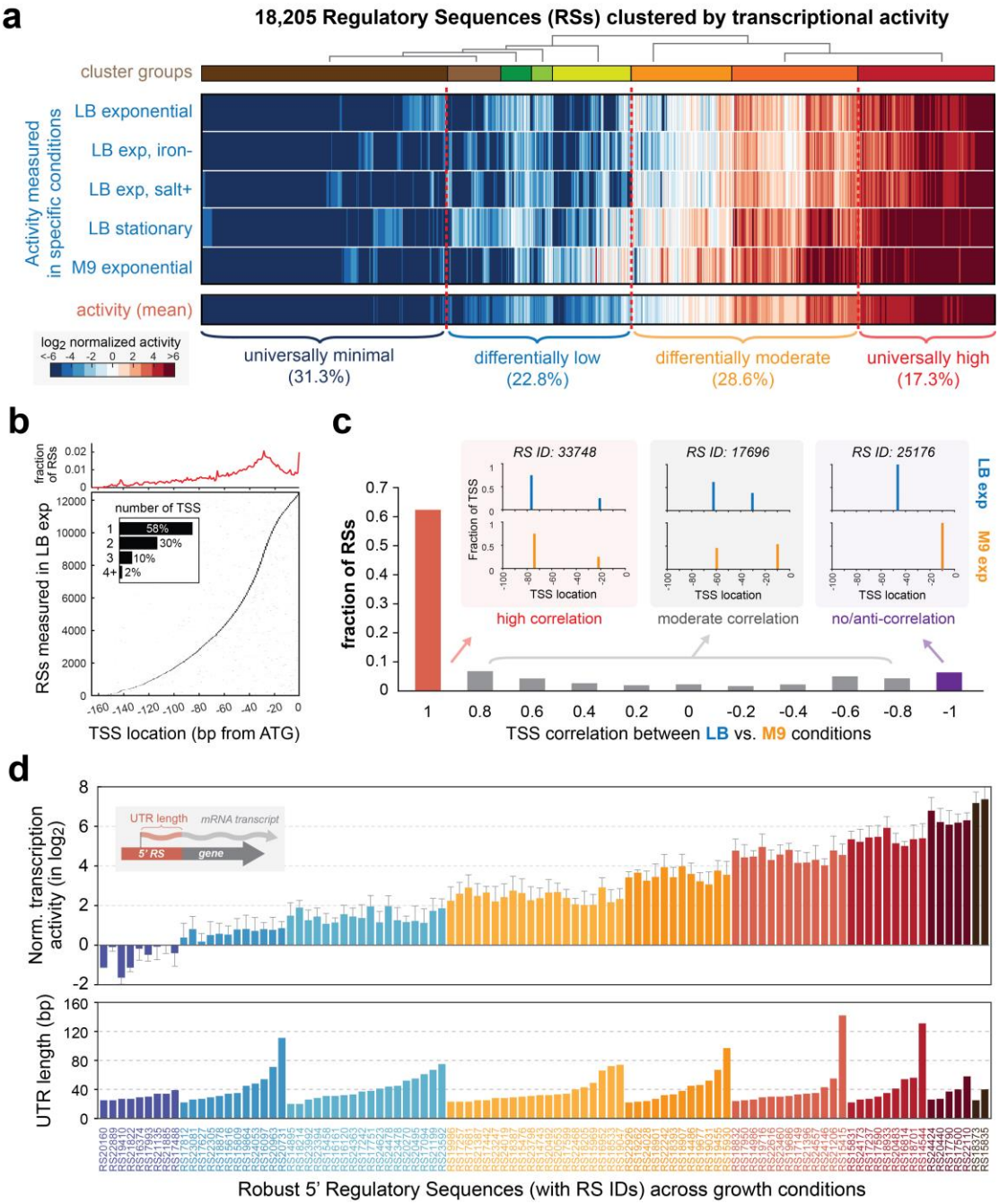
Correlation between transcription **(a)** and translation **(b)** data measured using sfGFP and an alternate reporter mCherry. Sample sizes (n) and Pearson correlation coefficients (*r*) are listed in the lower right of each plot.



Supplementary Figure 6

Transcription start sites in three species.

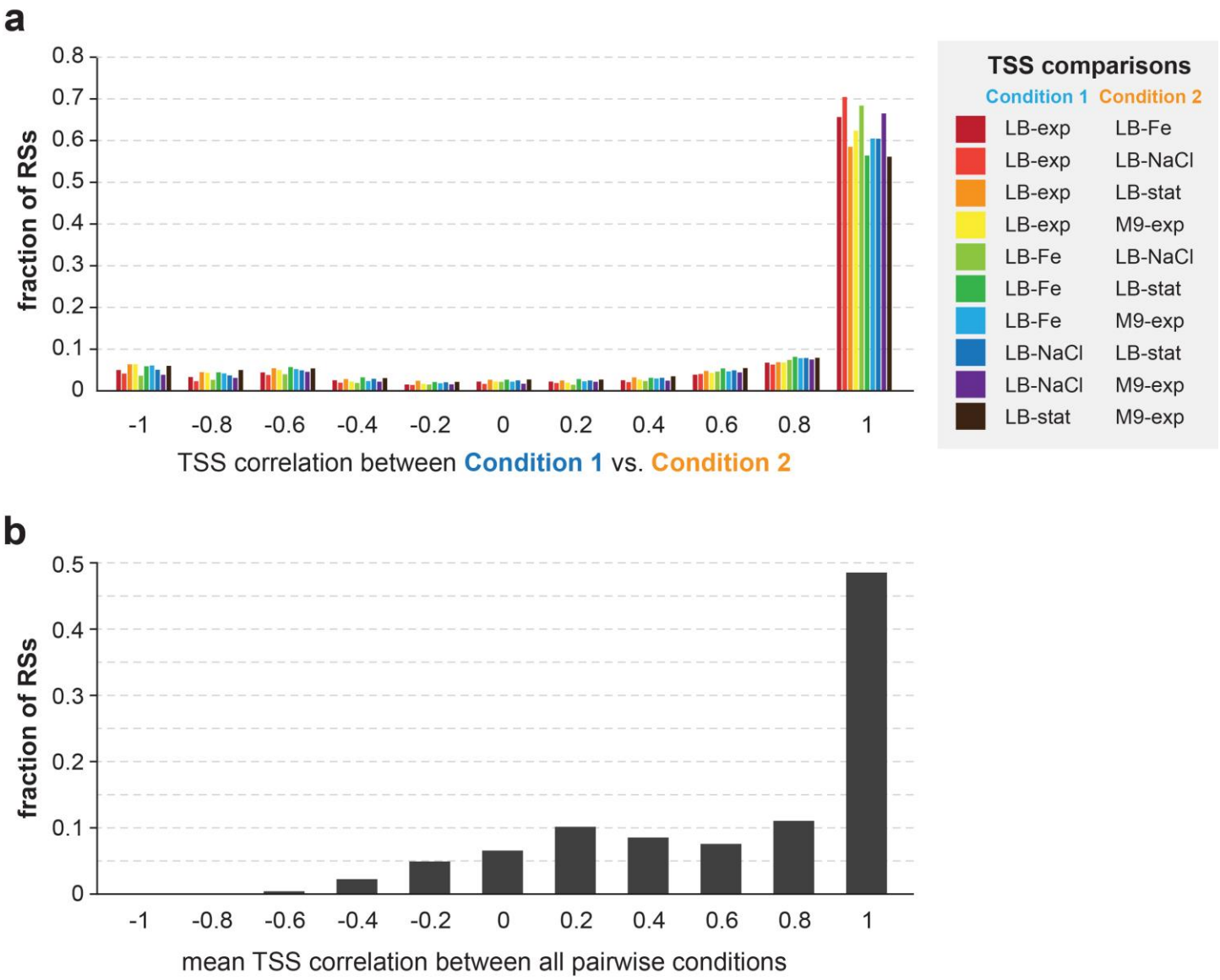
Distribution of transcription start sites (TSSs) for active regulatory sequences containing one primary TSS with >70% of reads starting within +/- 5 bp. Most TSSs occur between 20-50 bp upstream of the start codon for *B. subtilis, E. coli*, and *P. aeruginosa*.



Supplementary Figure 7

Alternative-growth-condition transcription data.

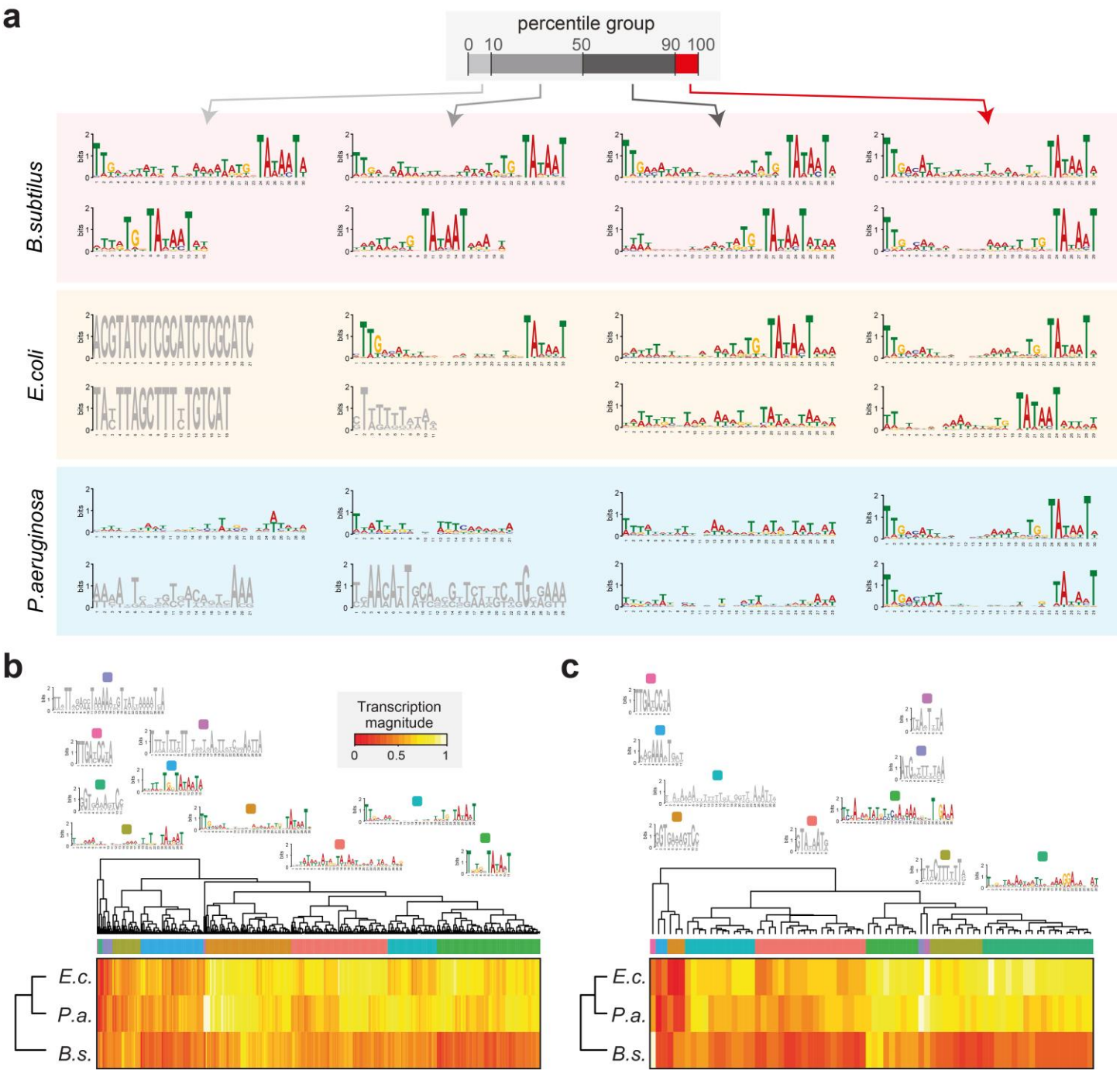
(a) Transcription activity for 18,205 members of the RS library across multiple growth conditions in *E. coli* is clustered and shown as a heatmap. Transcription levels are $\log_2$ (RNA/DNA) ratios normalized by the mean activity of control sequences (see Methods). **(b)** Ranked TSS locations of each RS measured in *E. coli* during LB exponential phase are shown, along with the TSS distribution (top panel) and the frequency of multiple TSSs (inset) of the RS library. **(c)** Frequency of matching TSS positions for RSs in LB and M9 growth media. Pearson correlation of 1 signifies perfectly matched TSS between conditions and -1 denoting no or anti-correlation. Intermediate values denote partial TSS matching. Example RSs with high, moderate, and no correlation in TSS positions in LB and M9 are shown in the inset (n = 18,205). **(d)** A subset of 100 robust RSs with condition-invariant transcription levels of different strengths (top panel) generated from a single TSS of different untranslated region (UTR) lengths (bottom panel) is provided as a useful community resource.



Supplementary Figure 8

Comparison of TSS data for regulatory sequences (RSs) across growth conditions in *E. coli*.

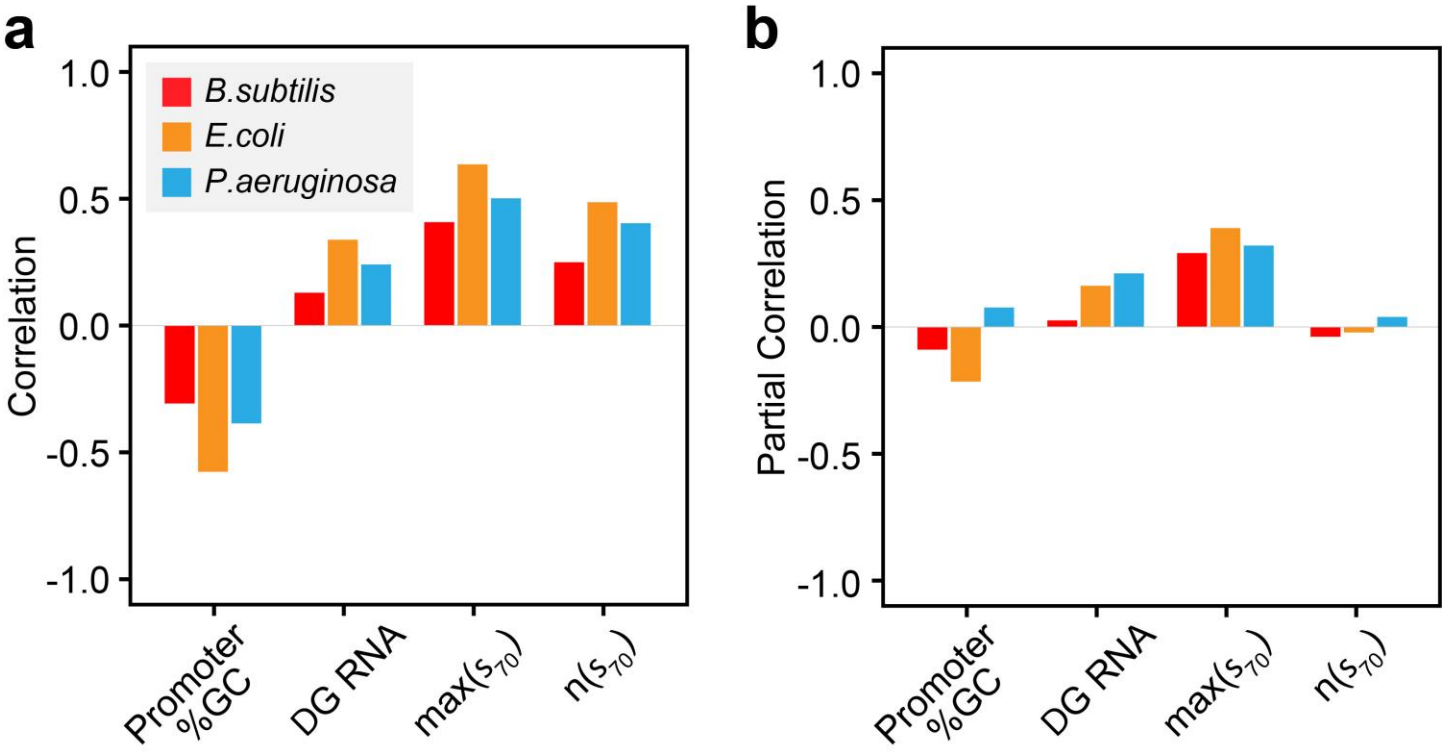
(a) A histogram of the distribution of all 10 pairwise comparisons of TSS position of regulatory sequences measured in 5 growth conditions (LB exponential growth phase, LB-exp; LB exponential with iron depletion, LB-Fe; LB exponential with high salt, LB-NaCl; LB stationary phase, LB-stat; M9 minimal media exponential phase, M9-exp) is shown (n = 18,205). Perfectly matched TSSs in two conditions have a Pearson correlation of 1, while an un-matched pair of TSSs has a correlation of -1. **(b)** A histogram of the mean TSS correlations (Pearson *r*) of all RSs across all pairwise conditions show almost half of RSs have the same TSS across all 5 conditions (n = 18,205).



Supplementary Figure 9

De novo motif search.

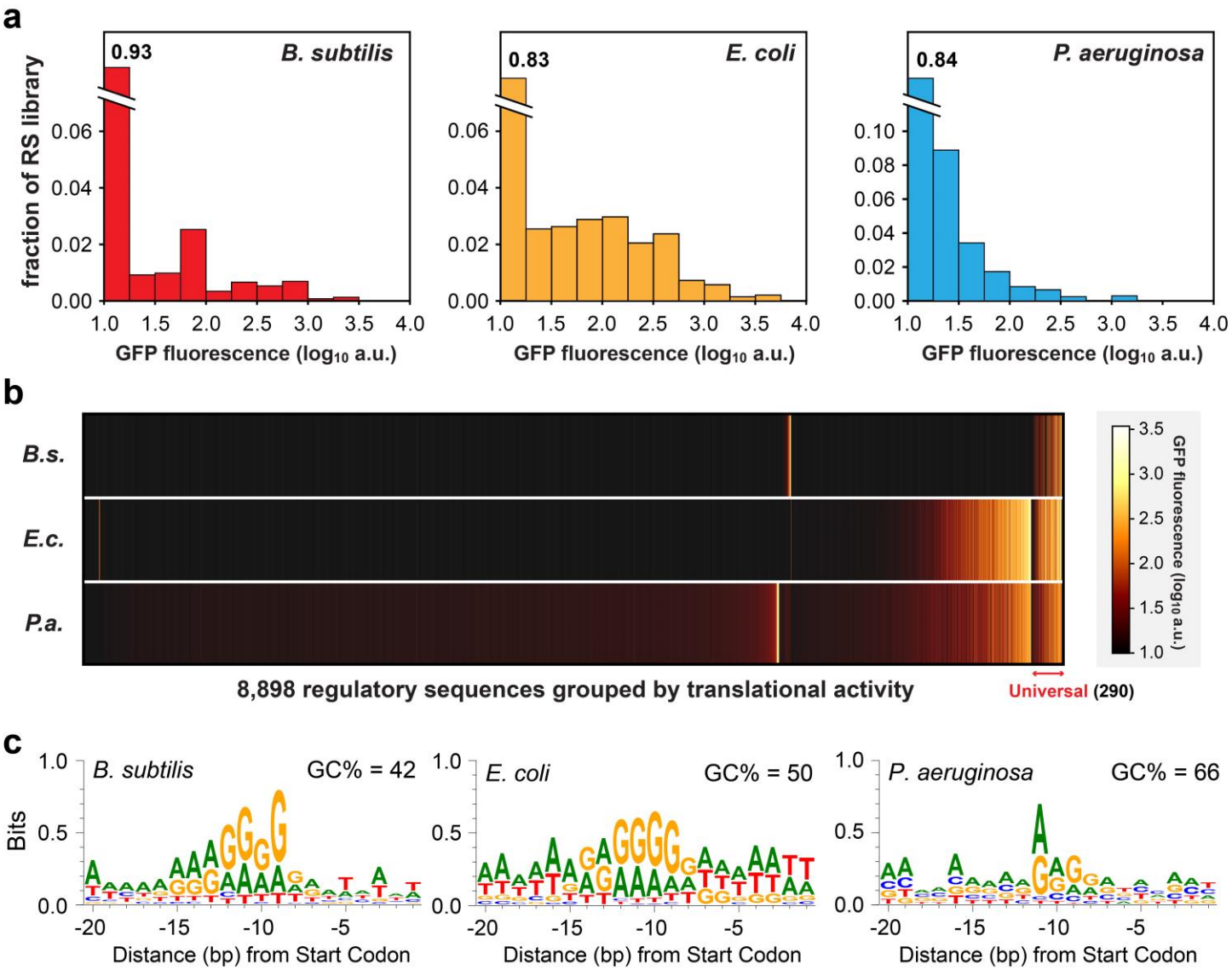
(a) Motif analysis of promoters binned by activity levels. The top two motifs identified by MEME for each recipient at the four activity bins (low, medium low, medium high, high) are shown. All motifs resembled the *σ70* motif or its degenerate versions. Statistically non-significant motifs are displayed in gray color. Additional MEME motif outputs are not shown since none were significantly different from *σ70*-like motifs. **(b)** Transcriptional activity heatmap grouped by hierarchical clustering (n=395). Motif finding was performed to identify motifs across ten clusters. The corresponding motif for each cluster is indicated by colored circle. **(c)** Removal of regulatory sequences containing the *σ70* motif from the dataset and repeating the analysis performed in **a** did not reveal additional non-*σ70*-like motifs (n=76). Statistically non-significant motifs (MEME E-value > 1e-2) are displayed in gray color in **b** and **c**.



Supplementary Figure 10

The σ^{70} motif is the dominant factor governing transcriptional activity of horizontally acquired regulatory sequences.

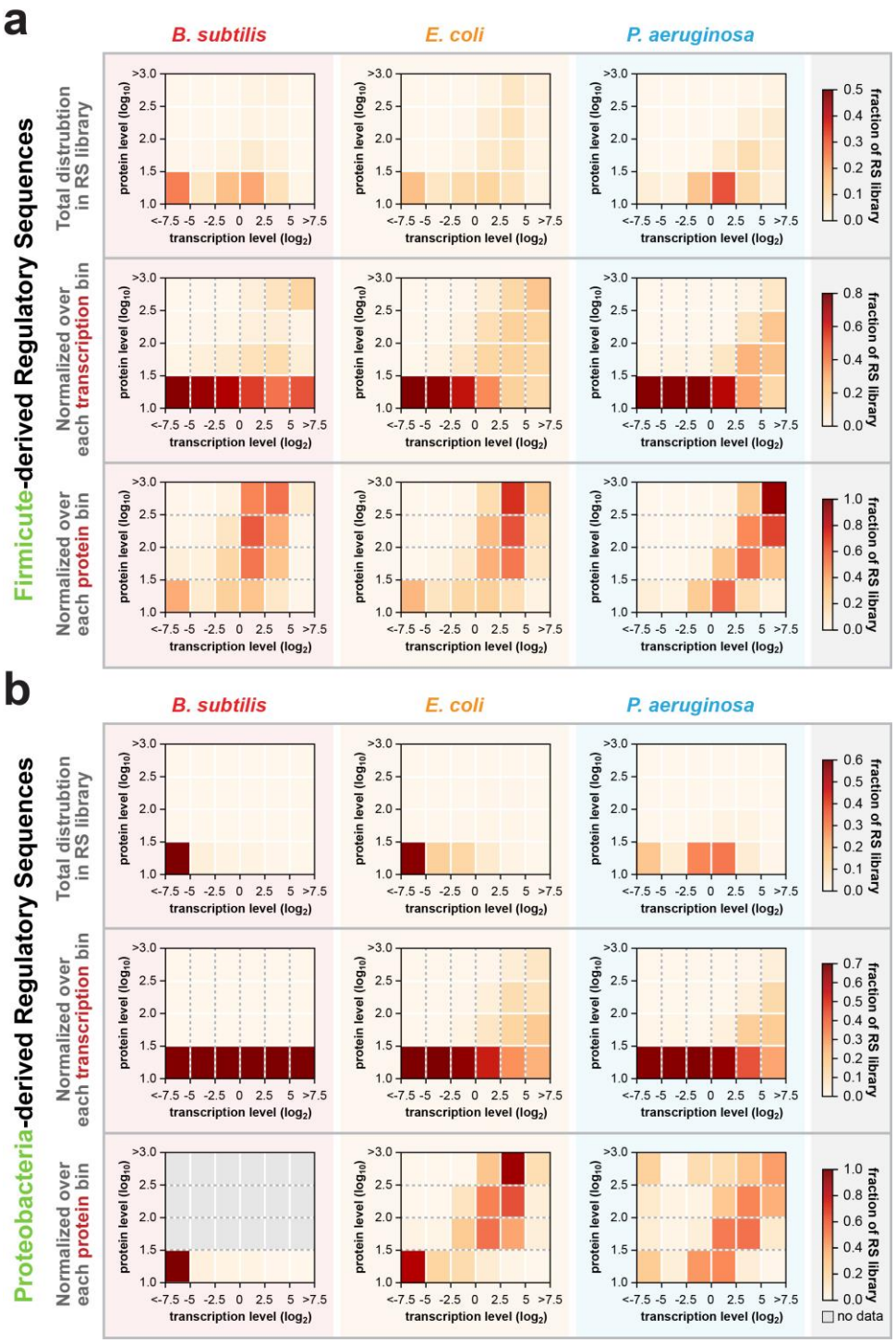
(a) Pearson correlation of transcriptional activity versus promoter GC content (%GC), RNA structural stability (ΔG RNA), best *σ70* match score (*max(σ70)*) and number of *σ70* matches (*n(σ70)*) are displayed per recipient species. **(b)** Partial correlation displays activity versus variable by controlling to the other variables. Sample sizes (n) are 4314, 14809, and 17787 regulatory sequences for *B. subtilis*, *E. coli*, and *P. aeruginosa* respectively.



Supplementary Figure 11

Regulatory sequence translation levels determined by FACS-seq in *B. subtilis*, *E. coli*, and *P. aeruginosa.*

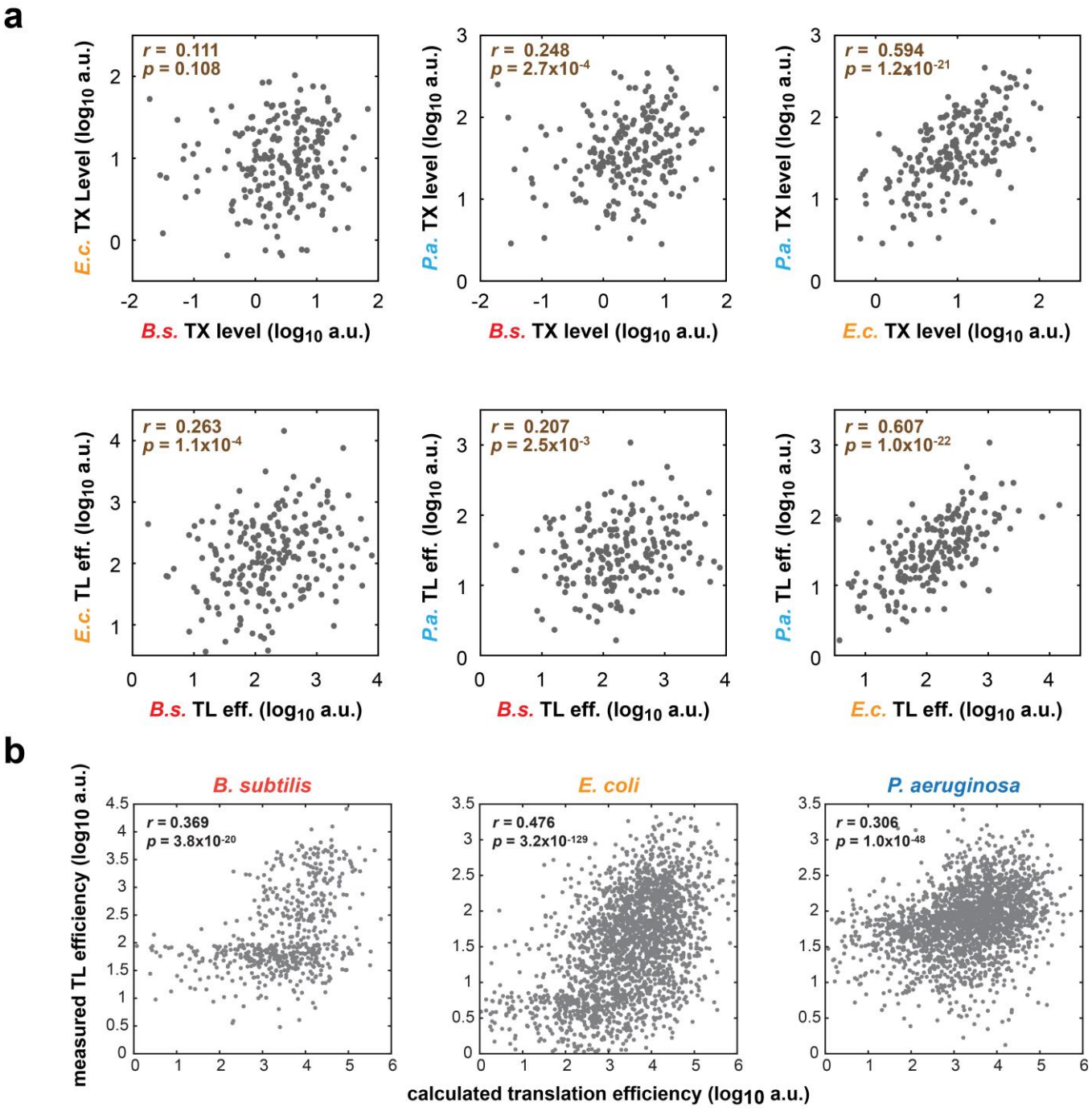
(a) The distribution of GFP fluorescence values of the regulatory sequence library in each recipient. **(b)** Translational activity of 8,898 regulatory sequences with measurable GFP fluorescence data across all three recipients. **(c)** Analysis of ribosome binding site sequence motifs in highly translated constructs. Motif logos were constructed using WebLogo v3.5.0. The genomic GC content of each species was used for background nucleotide frequency models and are listed in each subplot.



Supplementary Figure 12

Protein expression from Firmicute and Proteobacterial regulatory sequences.

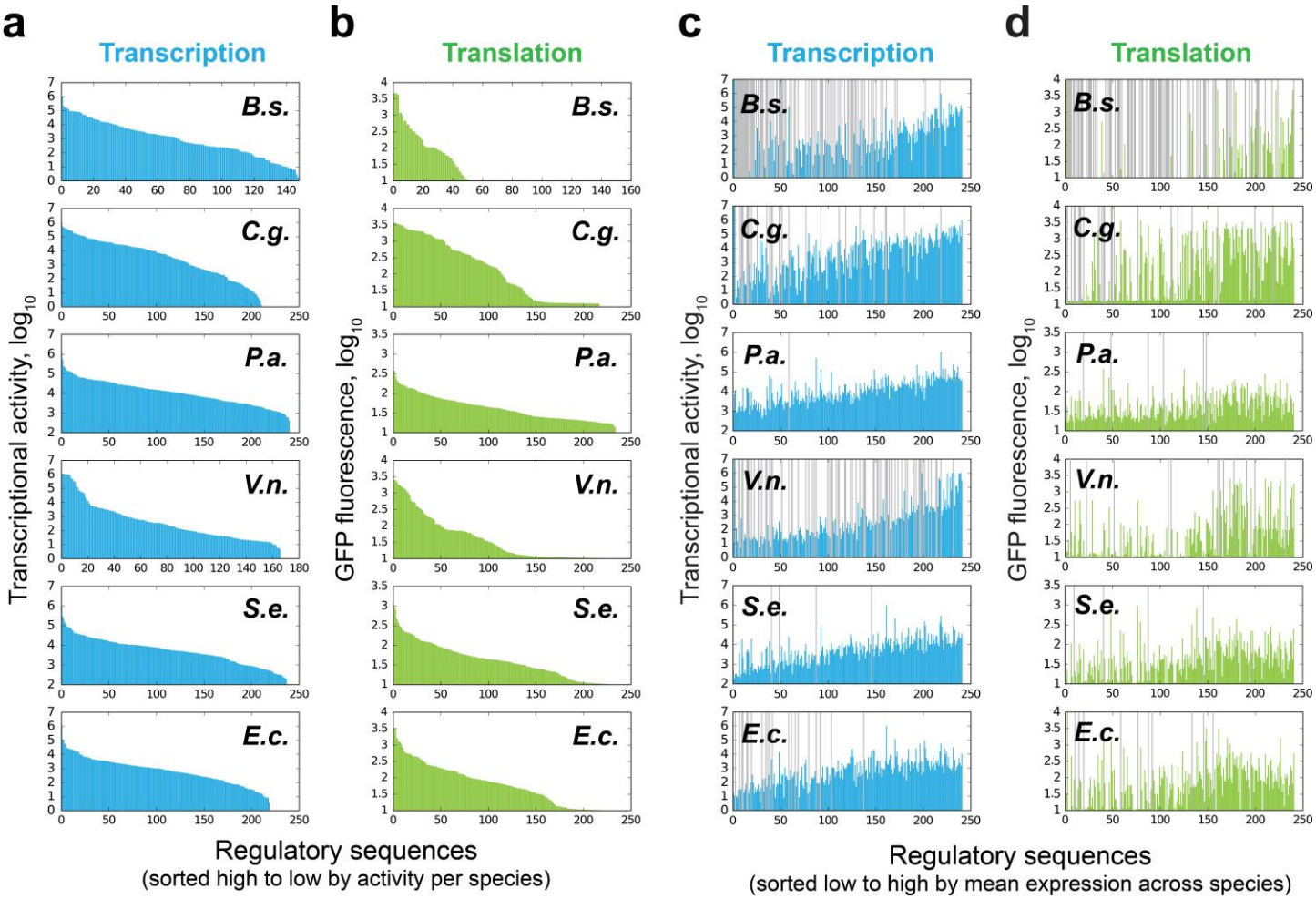
Heatmap panels show the fraction of RS library distributed across bins of transcription and translation levels in three recipients (colored columns). Donor RSs from Firmicutes genomes are shown in **(a)** and from Proteobacteria genomes in **(b)**. The top row of each heatmap subpanels use values normalized by the total number of regulatory sequences. The middle row use values normalized by each column bin corresponding to transcription windows. The bottom row use values normalized by each row bin corresponding to translation windows. Grey colored rows indicate data points with fewer than 10 RSs in total and insufficient for analysis.



Supplementary Figure 13

Cross-species and *in silico* comparisons of gene expression levels.

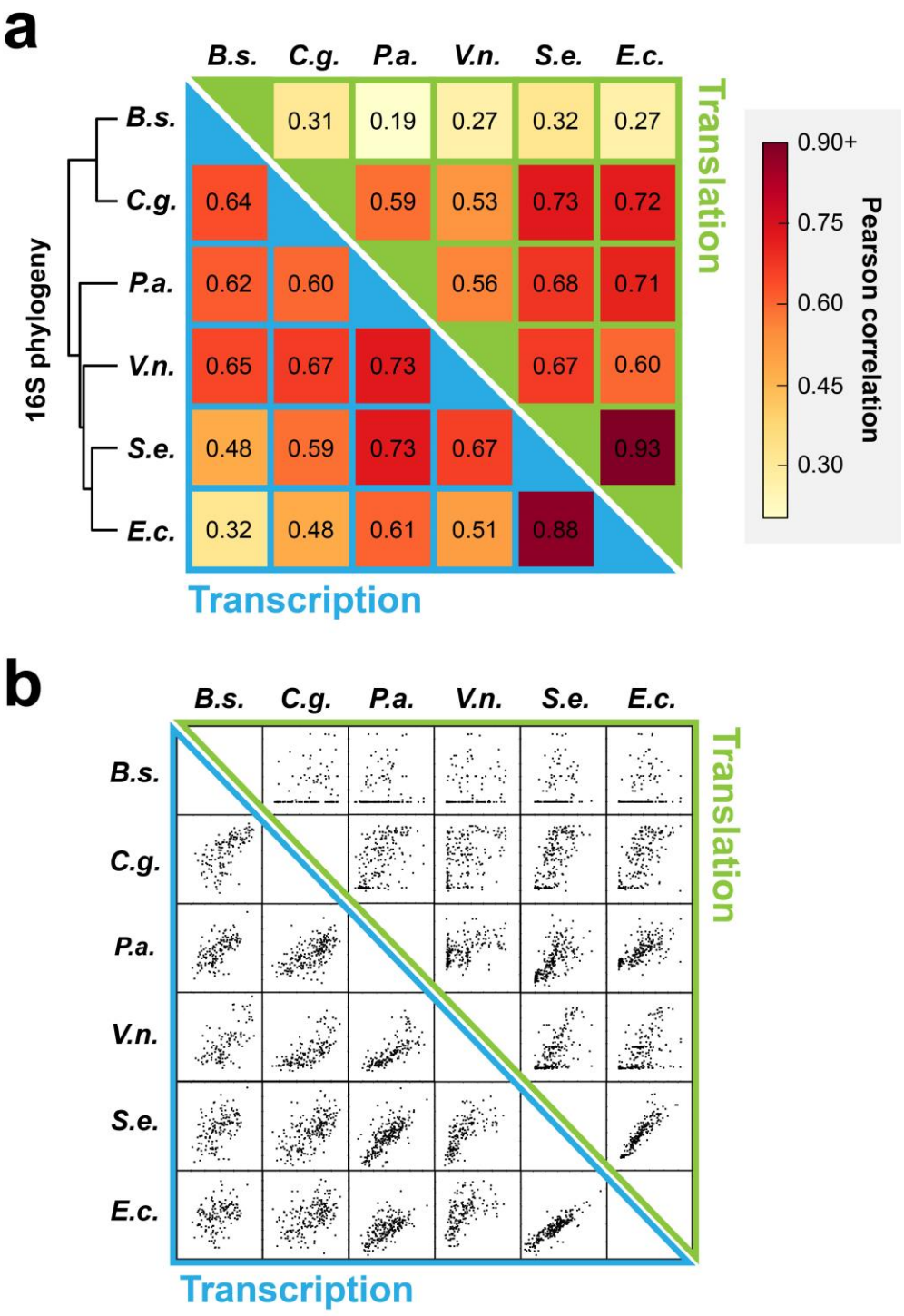
(**a**) Correlation of regulatory sequence activity in terms of transcription level and translation efficiency (calculated as the ratio of GFP protein levels and transcription levels) between recipient species. Each point corresponds to a single regulatory sequence that has measurable transcription and translation data. Pearson correlation coefficient (*r*) and statistical significance values (*p*) are shown for each subplot (n=212 for all six panels). **(b)** Correlation between calculated translation (TL) efficiency based on the RBS calculator and our measured translation efficiency across highly transcribed regulatory sequences (top 15%) in each recipient species (n = 581, 2276, and 2198 for *B. subtilis*, *E. coli*, and *P. aeruginosa* respectively).



Supplementary Figure 14

Regulatory activity of RS241 library in six bacterial species.

Regulatory sequences are sorted by activity (from high to low) per species by **(a)** transcription or **(b)** translation levels. Regulatory sequences are re-sorted by mean transcription levels (from low to high) across all species and plotted for **(c)** transcription and **(d)** translation levels. Transcriptional values were normalized with the highest expression construct having a value of 10^6. Gray lines correspond to sequences where no data was available. Species names are abbreviated as: *B. subtilis*, *B.s.*; *C. glutanicum*, *C.g.*; *P. aeruginosa*, *P.a.*; *V. natriegens*, *V.n.*; *S. enterica*, *S.e.*; *E. coli*, *E.c*.



Supplementary Figure 15

Cross-species transcription and translation level correlations.

(a) Pairwise Pearson correlation of transcription (blue triangle) and translation (green triangle) activity profiles of the RS241 library across six host species. Species are arranged based their 16S phylogenetic similarity. Numbers in each box correspond to the Pearson correlation coefficients (n = 241). **(b)** Scatter plot showing each pairwise correlation described in **(a)**.