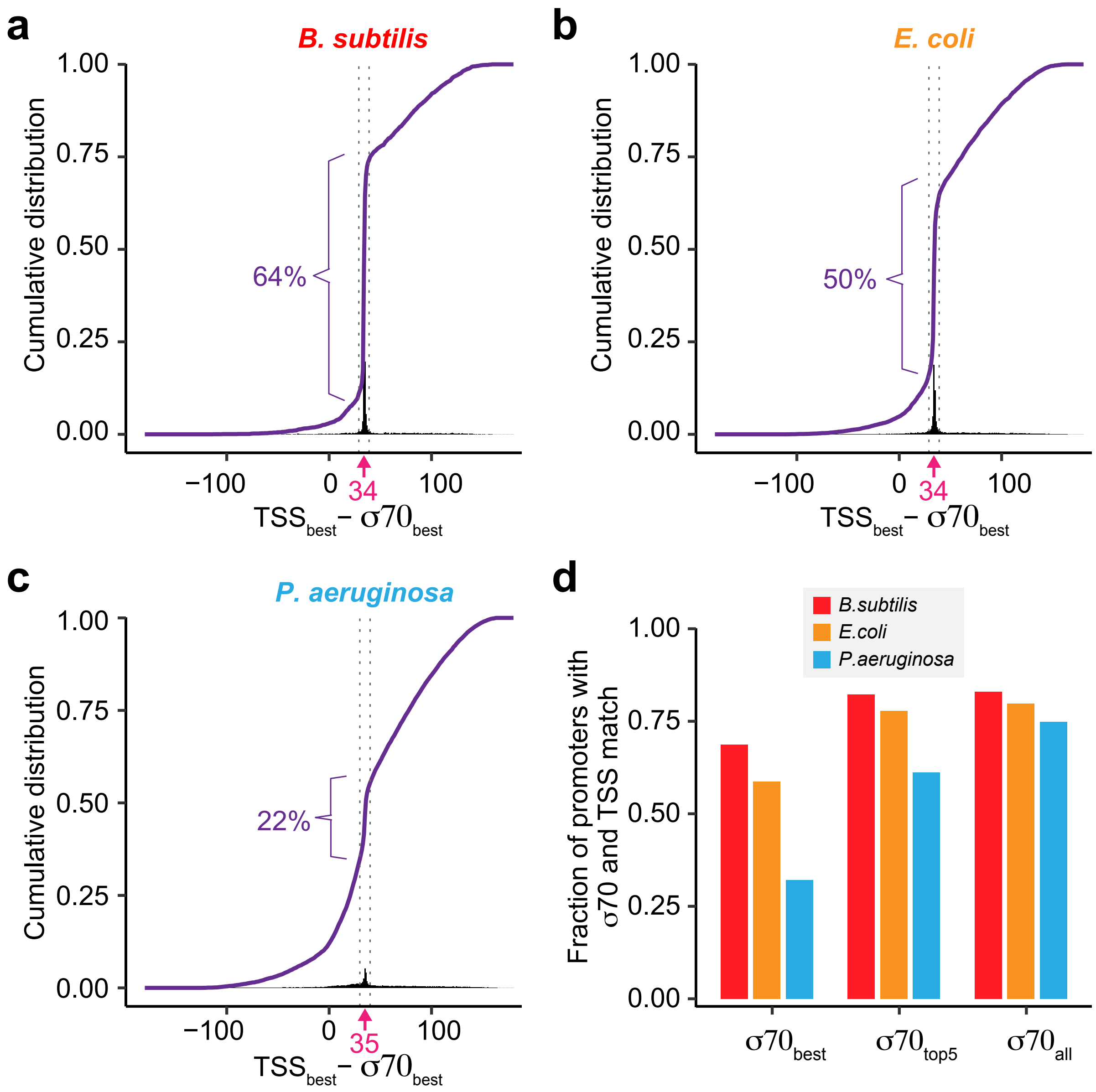
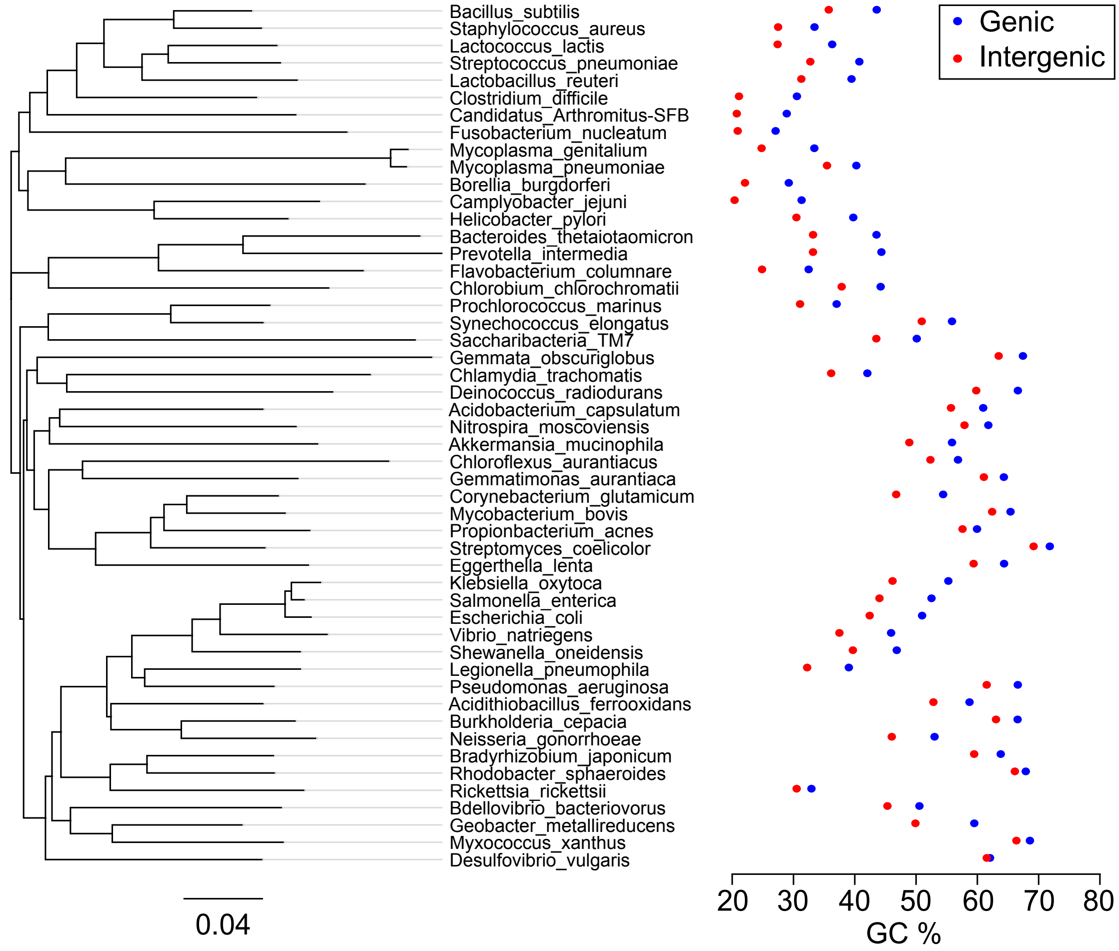
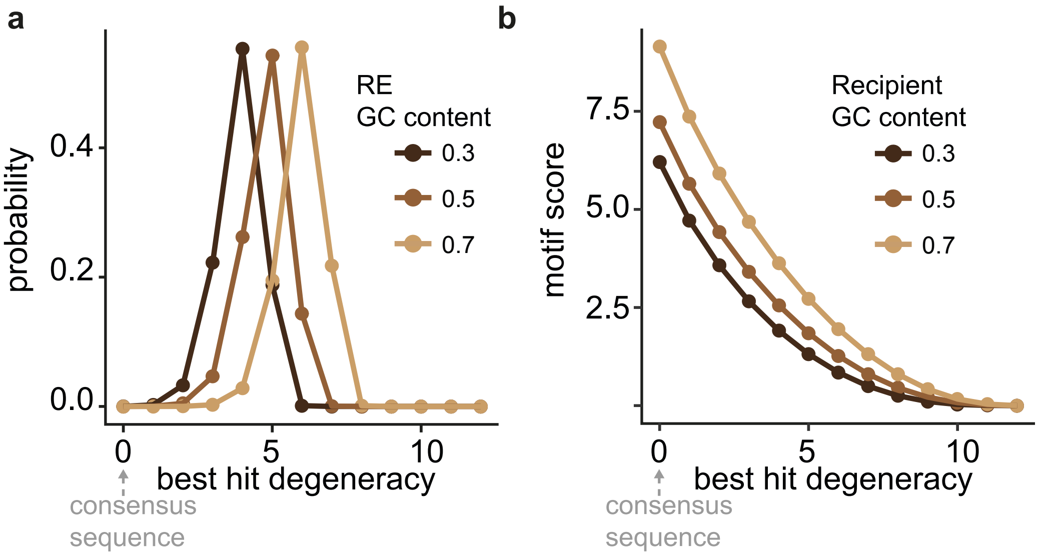


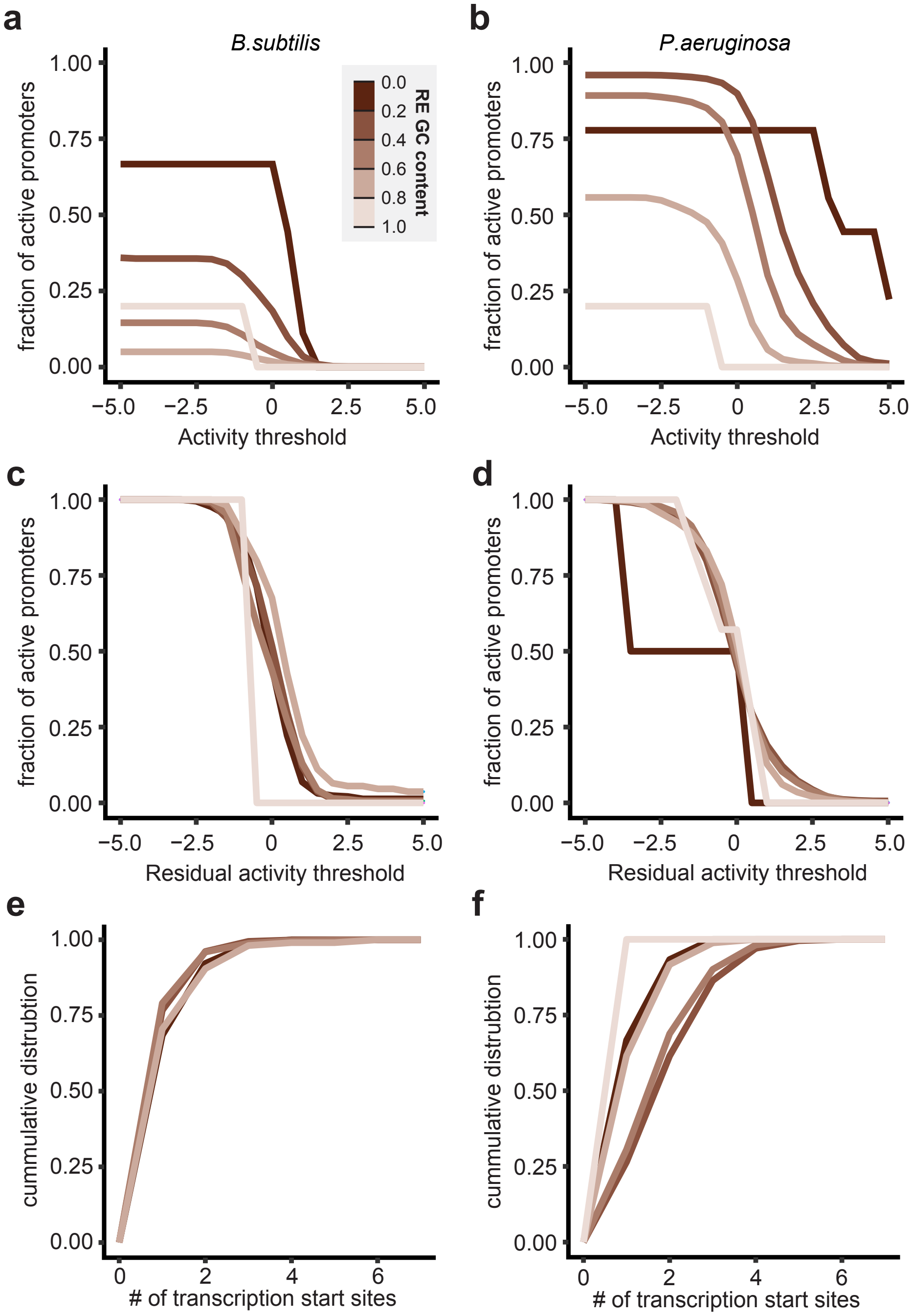
Supplementary Figure S1: Best σ70 hit can predict the primary TSS of a promoter. **(a-c)** The cumulative distribution of the distance between start location of best σ70 hit and strongest TSS is displayed for each recipient. The most likely distance is 34 bp for *B. subtilis* and *E. coli* and 35 bp for *P. aeruginosa*. Consistent with our main findings, *B. subtilis* shows the strongest relationship between σ70 motif and TSS location while *P. aeruginosa* shows the least. **(d)** The fraction of promoters that contains a match between σ70 motif and TSS is shown for each recipient. The x-axis labels indicate the number of σ70 hits used to predict TSS location. The labels $\sigma70_{best}$, $\sigma70_{top5}$, and $\sigma70_{all}$ correspond to the best, up top 5 and all σ70 hits respectively that were used to predict TSS matches across promoters.



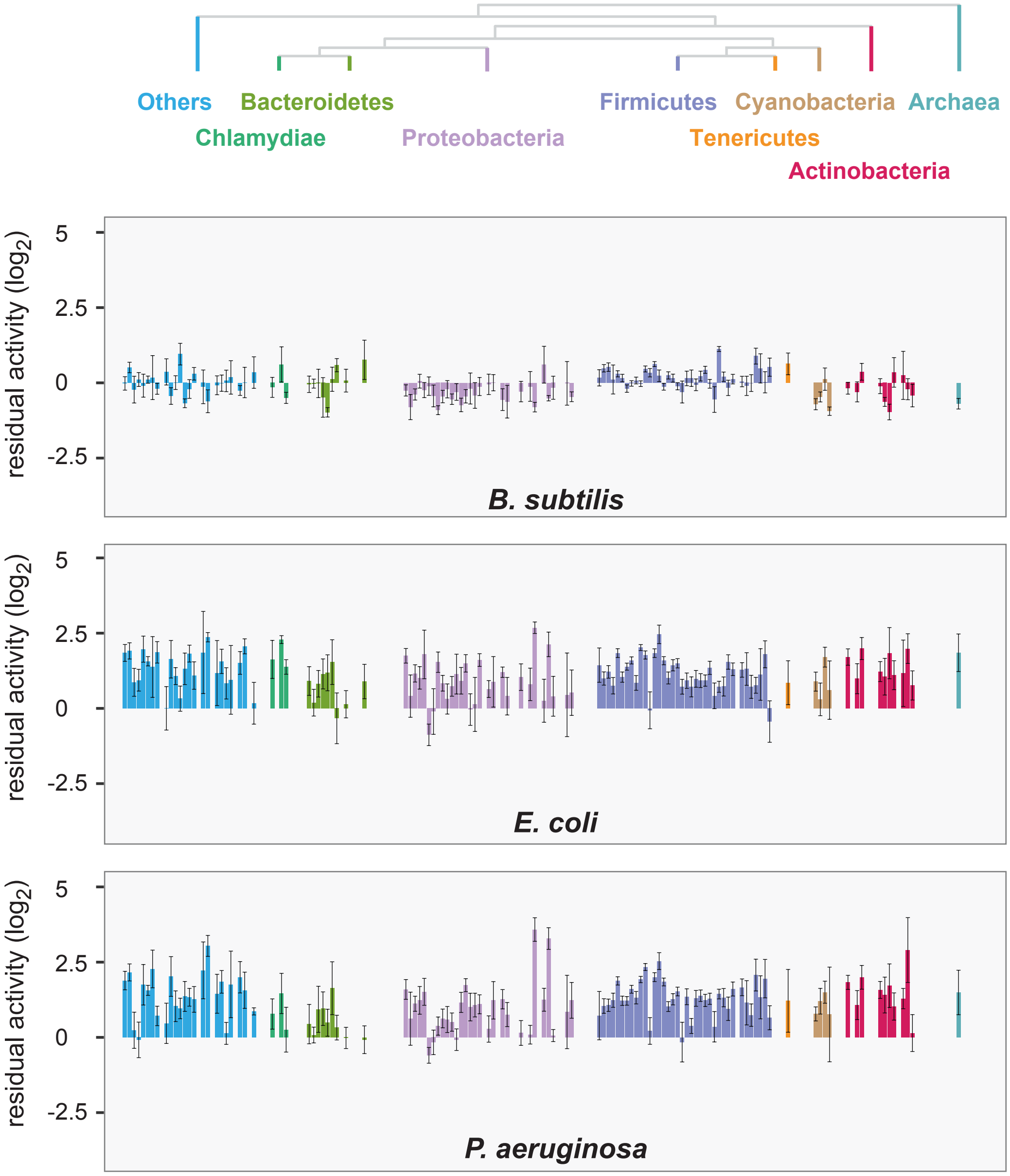
Supplementary Figure S2: Genic and intergenic GC content in 50 representative diverse bacterial genomes from major phyla. Species are arranged phylogenetically as a tree that was constructed using Geneious (11.0.5). Tree scale bar represents Jukes-Cantor distance of full-length 16S rRNA gene sequences.



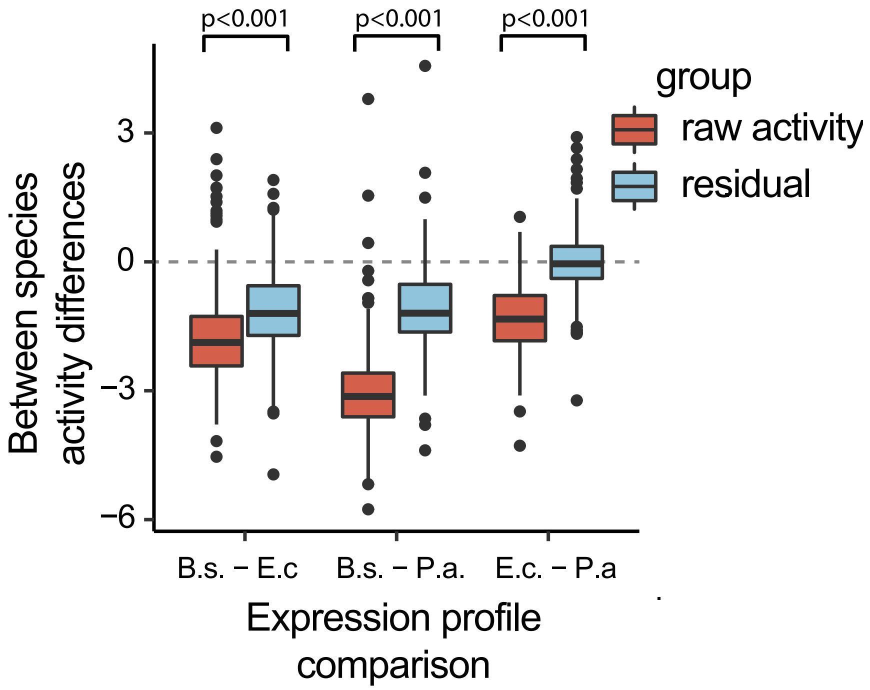
Supplementary Figure S3: Effect of promoter and recipient GC content in motif score. (a) Best kmer is less degenerated in low GC content promoter**.** The null distribution of kmer degeneracy is computed for different promoter GC content according to **Equation 11**. We used σ70 consensus sequence TTGACA(N_{17})TATAAT and assumed promoter region of length 165bp. (b) The motif score per kmer degeneracy (**Equation 9)** is stronger in high GC content recipients.



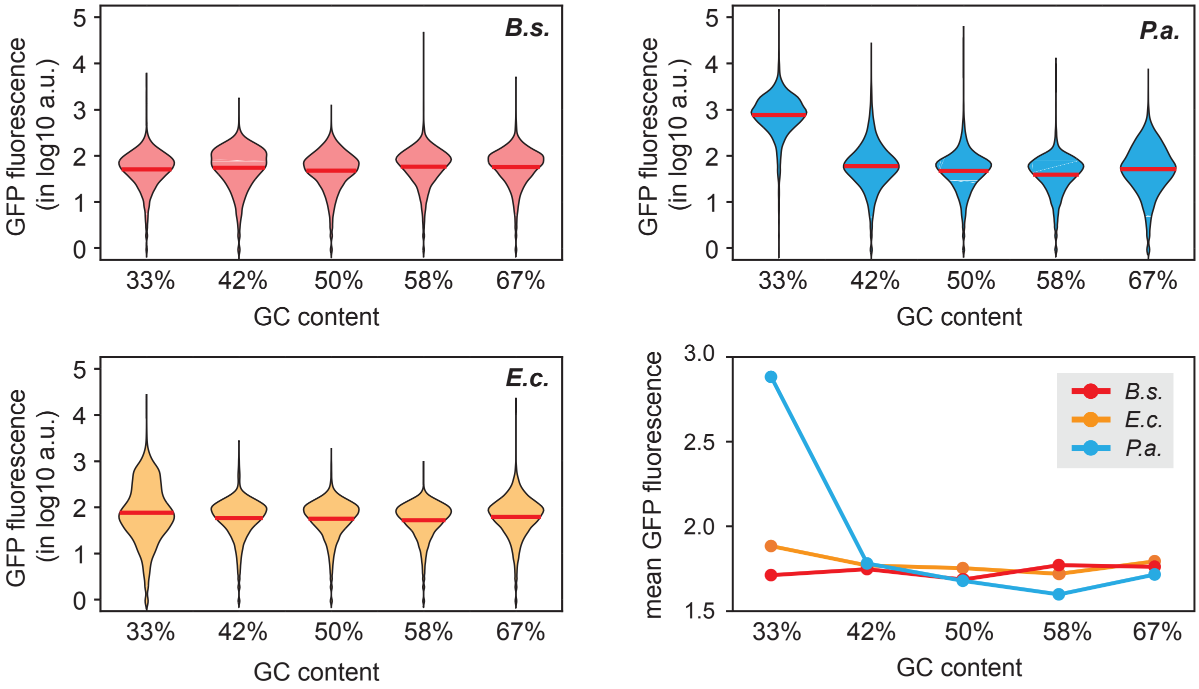
Supplementary Figure S4: Promoter promiscuity controlled by GC content for *B. subtilis* and *P. aeruginosa*. Fraction of active promoters as a function of activity threshold grouped per promoter GC content is plotted for *B. subtilis* **(a)** and *P. aeruginosa* **(b)**. Fraction of active promoters as a function of residual activity threshold grouped per promoter GC content is plotted for *B. subtilis* **(c)** and *P. aeruginosa* **(d)**. Residual activity is obtained from linear model that considers *RNA* stability and best σ70 motif score. Cumulative distribution of number of TSSs in *B. subtilis* **(e)** and *P. aeruginosa* **(f)**.



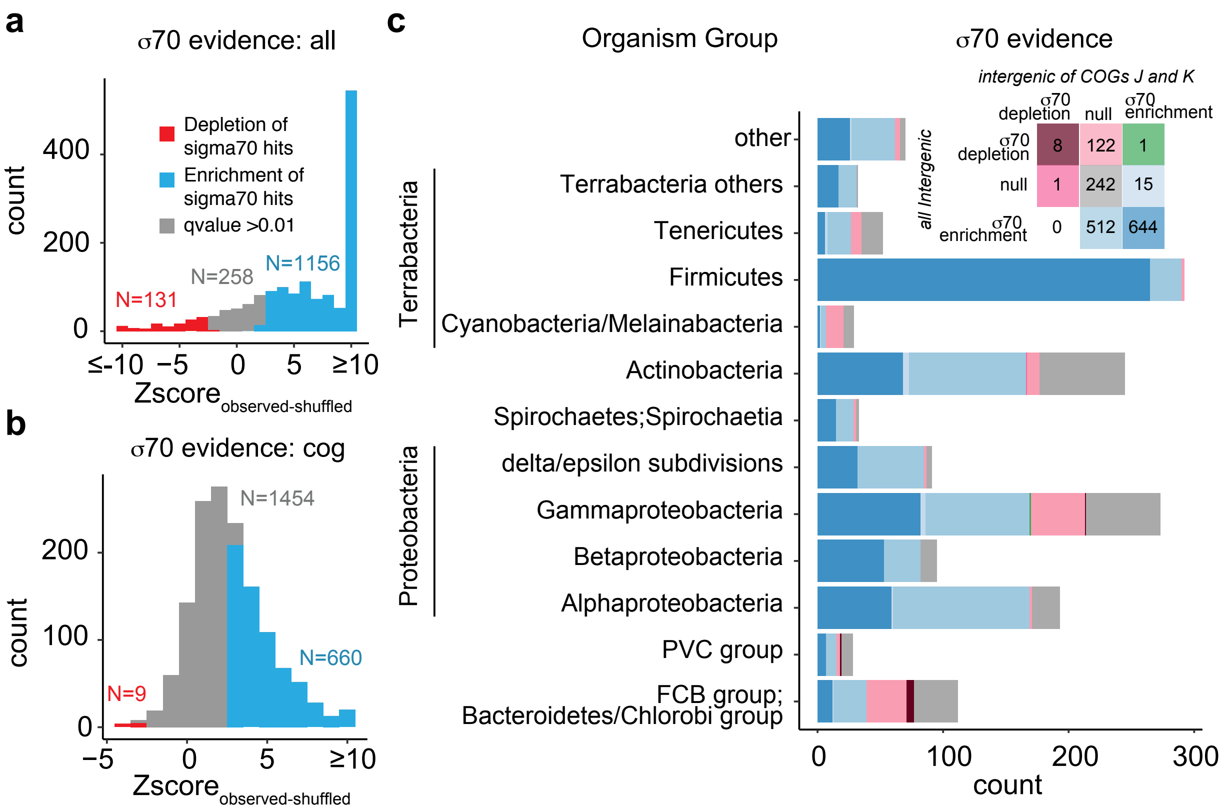
Supplementary Figure S5: Phylogenetic bias in expression per recipient is attenuated after correction for key factors that influences gene expression. The average residual activity of regulatory elements (Observed – Expected activity) is displayed as a bar plot arranged by donor organism phylogeny for each recipient (Compare with **Figure 1**). Error bars represent two standard error distance from the mean value. Expected activity is computed from a linear model that considers RE GC content, best $\sigma70$ motif score and mRNA 5’ stability (**Methods**).



Supplementary Figure S6: Regression model increases expression similarities among recipients. The expression dissimilarity, defined as the difference in transcriptional activity, is computed for each recipient pair. The dissimilarity is computed according to donor organism for raw (**Figure 1**) and residual (**Figure S5**) activity levels. Each point represents REs from a donor organism. For all recipient pairs, residual activity increases similarities among recipients (p<0.001).

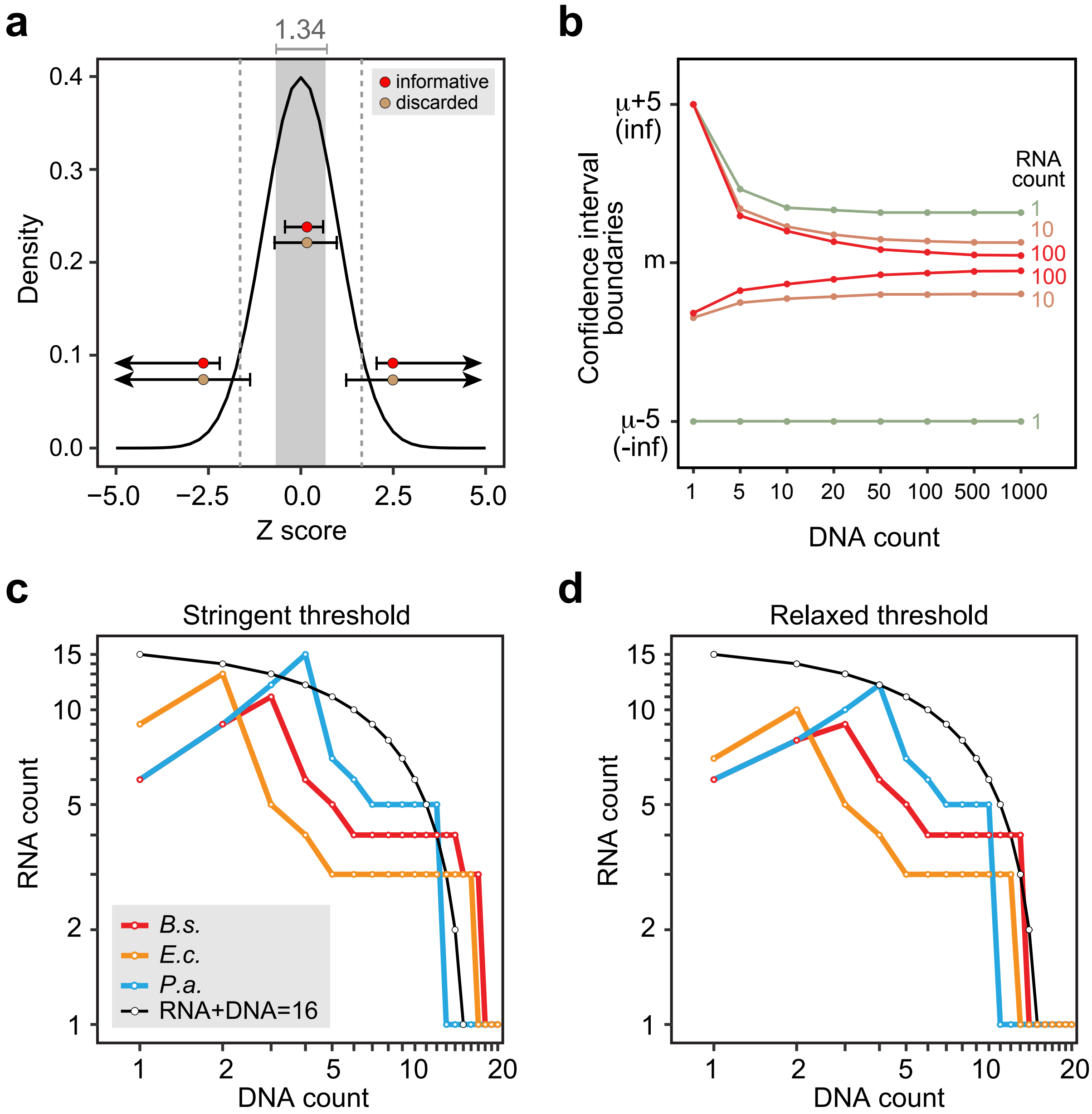


Supplementary Figure S7: Expression libraries of libraries of degenerate low GC regulatory sequences show high transcription in a GC-rich recipient. GFP reporter gene expression from a set of five RE libraries originating from oligos containing 140 ambiguous bases with varying expected GC compositions (33%, 42%, 50%, 58%, 67%) measured in *B. subtilis*, *E. coli* and *P. aeruginosa* using flow cytometry. The lowest random regulatory library (33% GC) showed elevated activity in the GC-rich recipient, *P. aeruginosa* (67% genomic GC).



all intergenic \ intergenic of COGs J and K	σ70 depletion	null	σ70 enrichment
σ70 depletion	8	122	1
null	1	242	15
σ70 enrichment	0	512	644

Supplementary Figure S8: Evidence for σ70 motif signature in intergenic regions is observed in the majority of representative bacteria. We estimated the Z-score of conservation of σ70 motif in natural vs shuffled intergenic regions and observed evidence of σ70 motif in 84.3% (1303/1545) of this set at an FDR value of 0.01. We looked for evidence using all non-convergent intergenic regions (**a**) and the subset that is associated with a gene that belongs to the housekeeping COG categories *J* (translation) and *K* (transcription) (**b**). Evidence for σ70 motif is displayed in terms of type of evidence, either enrichment or depletion of σ70 hits, for different groups of bacteria.



Supplementary Figure S9: Definition of statistically informative promoters. (a) A promoter is classified as informative when the 95% confidence interval of its activity measurement is narrow enough (less than 1.34 of standard deviation of reference distribution) or when its maximum or minimum boundary lies at the low or high end of the reference distribution. The panel displays instances of accepted (red dots) and discarded points (orange dots) at each location along the distribution. **(b)** The upper and lower limit of the 95% confidence interval have a monotonically decreasing distance to its mean value when plotted as a function of either DNA or RNA counts.

This property is important to efficiently compute an approximated confidence interval for any pair of DNA and RNA counts (see Methods). Values with distance to mean greater than 5 are clustered at μ-5 or μ+5. **(c-d)** The threshold of RNA and DNA counts to obtain an informative promoter is displayed for each recipient at stringent **(c)** and inclusive **(d)** thresholds. The curve with total count equal to 16 displays the threshold used for analysis in this paper. The stringent threshold assumes $L_{i,max}<q_{0.30}$ or $L_{i,min} > q_{0.70}$ (Equation 8). The relaxed threshold assumes $L_{i,max}<q_{0.35}$ or $L_{i,min} > q_{0.65}$.