Methods

Data collection Data of *Escherichia coli* genes were extracted from the Ecogene database[1], which contains annotations for 4308 genes. These genes were united into transcription units (TUs) based on data from the regulonDB database[2] and from the experiments of Tjaden et al.[3], giving a total of 3405 TUs.

Data regarding transcription regulation interactions in *E. coli* were extracted from the regulonDB database[2], the EcoTFs database[4], and from Shen-Orr et al[5]. Interactions between a TF and its target gene were converted to interactions between the TU encoding the TF and the TU encoding the target gene. In total, our dataset contained 549 non-redundant interactions between 106 regulating TUs (encoding 111 TFs) and 337 regulated TUs (encoding 737 genes). Data regarding the location of TF binding sites were extracted from regulonDB. *E. coli* gene classifications were downloaded from the GenProtEC database [6]. Some less informative classifications, such as the localization of gene products, were ignored. We consider the list of classifications of a TU to include all the classifications belonging to the genes encoded by that TU.

Data regarding *Saccharomyces cerevisiae* genes were extracted from the SGD database[7]. Only 4140 genes which were annotated as "verified" were used. Data regarding regulation interactions were extracted from Yeger-Lotem *et a*l.[8], and included 1272 interactions between 126 TFs and 558 regulated genes. The clustering of *S. cerevisiae* genes to 'transcription modules' was based on the study of Ihmels *et al.* [9].

Data regarding *Bacillus subtilis* was extracted from the DBTBS database [10]. *B. subtilis* genes were assigned to TUs based on the operon predictions of De Hoon *et al.* [11]. Genes were considered to be in the same operon if the assigned probability that they

belong to the same operon was 0.5 or higher. Data regarding experimentally verified regulation interactions in *B. subtilis* were extracted from the DBTBS database [10].

Detecting *k*-node network motifs. This analysis was performed per organism. The network representation of the TUs and their relationships regarding chromosomal adjacency and transcriptional regulation is described in Fig. 1 of the paper. Motif analysis was performed as in Yeger-Lotem et al.[8]: All connected sub-networks containing k nodes in the organism network were collated into isomorphic patterns and the number of times each pattern occurred was counted. If the number of occurrences was at least 5 and was statistically significantly higher than in randomized networks ($p\text{-value} \leq 0.05$), the pattern was considered as a *network motif*. The statistical significance test was performed by generating 1000 randomized networks[8], and computing the fraction of randomized networks in which the pattern appeared at least as often as in the organism network. Auto-regulation edges were ignored, except for the case of k=2 in which they were retained.

Duplication analysis. In order to determine whether pairs of co-regulated neighbors arose through duplication, the protein sequences of the two pair-mates were compared. In the case of *E. coli*, each gene encoded by one of the co-regulated TUs was compared to each of the genes encoded by its neighboring co-regulated TUs. Sequence comparisons were performed using a locally installed version of the BLAST 2 sequences algorithm [12]. We set the theoretical database size for the calculation of the E-value to the size of the *E. coli* genome for the *E. coli* comparisons, and to the size of the *S. cerevisiae* genome for comparisons in *S. cereivisiae*. A pair of sequences was considered to share sequence similarity if the E-value of their alignment was 0.01 or lower.

Analysis of *E. coli* gene expression profiles. Gene expression data were extracted from the study of Allen et al[13] (http://asap.ahabs.wisc.edu/annotation/php/logon.php). Data were used from 32 experiments representing 16 conditions (see Supplementary Material). Only genes whose expression level was obtained in all 32 experiments were included in the analysis. These amounted to 3623 of the 4308 genes in Ecogene[1].

The expression level of a TU in a certain experiment was calculated as the average expression of the genes it encodes and for which data were available. In total, 2880 out of the 3405 TUs in *E. coli* were used in the expression correlation analysis.

Spearman correlations between expression profiles of TUs in 100,000 random TU pairs were calculated. We set the threshold for statistically significant positive (negative) correlation to be the correlation value above (below) which only 2.5% of the random pairs reside. The thresholds for significant positive and negative correlations were $r_s>=0.71$ and $r_s<=-0.38$, respectively. In order to determine whether a group of TU pairs has more significantly correlated pairs than expected at random we used a binomial test, taking the expected probability to be 0.05.

Expression array experiment conditions

1. Standard growth conditions - Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log phase (OD600=0.2)
2. Cold shock -Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log phase (OD600=0.2) then moved to a 15 degrees bath. RNA was extracted 10 minutes later.

3. Acid shock 1 min – Base growth medium: Mops minimal. carbon source: glucose. Grown at 37 degrees to log phase (OD600=0.2) then moved to pH=3.8 and RNA was extracted 1 minute later.
4. Acid shock 4 min – Base growth medium Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then moved to pH=3.8 and RNA was extracted 4 minutes later.
5. Acid shock 8 min – Base growth medium Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then moved to pH=3.8 and RNA was extracted 8 minutes later.
6. Acid shock 14 min – Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then moved to pH=3.8 and RNA was extracted 14 minutes later.
7. Acid shock 20 min – Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then moved to pH=3.8 and RNA was extracted 20 minutes later.
8. Ciprofloxacin 20 ng/ml 5 minutes – Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then ciprofloxacin 20 ng/ml was added and RNA was extracted 5 minutes later.
9. Ciprofloxacin 20 ng/ml, 10 minutes – Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then ciprofloxacin 20 ng/ml was added and RNA was extracted 10 minutes later.

10. Ciprofloxacin 20 ug/ml, 4 minutes -Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then ciprofloxacin 20 ug/ml was added and RNA was extracted 4 minutes later.
11. Ciprofloxacin 20 ug/ml, 10 minutes - Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then ciprofloxacin 20 ug/ml was added and RNA was extracted 10 minutes later.
12. Ciprofloxacin 20 ug/ml, 14 minutes – Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to log (OD600=0.2) then Ciprofloxacin 20 ug/ml was added and RNA was extracted 14 minutes later.
13. Late log phase, 90 minutes – Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to late log phase (OD600=0.5) RNA was extracted 90 minutes after culture reached OD600=0.2
14. Transition to stationary phase, 105 minutes – Base growth medium: Mops minimal, Carbon source: glucose. Grown at 37 degrees to late log phase (OD600=0.67) RNA was extracted 105 minutes after culture reached OD600=0.2
15. Stationary phase, 135 minutes 1-3– Base growth medium: Mops minimal. Carbon source: glucose. Grown at 37 degrees to stationary phase (OD600=0.75) RNA was extracted 135 minutes after culture reached OD600=0.2
16. LB, log phase growth - Cultured in LB with no additive in 37 degrees to log phase OD600=0.8

References

1. Rudd, K.E. (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res* 28, 60-64
2. Salgado, H. *et al.* (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res* 32, D303-306
3. Tjaden, B. *et al.* (2002) Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res* 30, 3732-3738
4. Wall, M.E. *et al.* (2004) Design of gene circuits: lessons from bacteria. *Nat Rev Genet* 5, 34-42
5. Shen-Orr, S.S. *et al.* (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 31, 64-68
6. Serres, M.H. *et al.* (2004) GenProtEC: an updated and improved analysis of functions of *Escherichia coli* K-12 proteins. *Nucleic Acids Res* 32 Database issue, D300-302
7. Christie, K.R. *et al.* (2004) *Saccharomyces* Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res* 32 Database issue, D311-314
8. Yeger-Lotem, E. *et al.* (2004) Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc Natl Acad Sci U S A* 101, 5934-5939
9. Ihmels, J. *et al.* (2002) Revealing modular organization in the yeast transcriptional network. *Nat Genet* 31, 370-377
10. Makita, Y. *et al.* (2004) DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. *Nucleic Acids Res* 32 Database issue, D75-77
11. De Hoon, M.J. *et al.* (2004) Predicting the operon structure of *Bacillus subtilis* using operon length, intergene distance, and gene expression information. *Pac Symp Biocomput*, 276-287
12. Tatusova, T.A. and Madden, T.L. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174, 247-250
13. Allen, T.E. *et al.* (2003) Genome-scale analysis of the uses of the *Escherichia coli* genome: model-driven analysis of heterogeneous data sets. *J Bacteriol* 185, 6392-6399