ARTICLES

Evolvability and hierarchy in rewired bacterial gene networks

Mark Isalan[1], Caroline Lemerle[2], Konstantinos Michalodimitrakis[1], Carsten Horn[2], Pedro Beltrao[2], Emanuele Raineri[1], Mireia Garriga-Canut[1] & Luis Serrano[1]

Sequencing DNA from several organisms has revealed that duplication and drift of existing genes have primarily moulded the contents of a given genome. Though the effect of knocking out or overexpressing a particular gene has been studied in many organisms, no study has systematically explored the effect of adding new links in a biological network. To explore network evolvability, we constructed 598 recombinations of promoters (including regulatory regions) with different transcription or σ-factor genes in *Escherichia coli*, added over a wild-type genetic background. Here we show that ~95% of new networks are tolerated by the bacteria, that very few alter growth, and that expression level correlates with factor position in the wild-type network hierarchy. Most importantly, we find that certain networks consistently survive over the wild type under various selection pressures. Therefore new links in the network are rarely a barrier for evolution and can even confer a fitness advantage.

The *E. coli* genome codes for ~300 transcription factors (TFs)[1,2], organized hierarchically, with few master regulators[3–5] (Fig. 1). Only nine regulatory proteins (CRP, FNR, IHF, Fis, ArcA, NarL, H-NS, Fur and Lrp) control over half of all genes, through direct and indirect interactions[6,7]. Lower-tier nodes are more sparsely connected and the network structure has a scale-free power-law degree distribution[8,9]. It has been argued that such networks are particularly robust to random errors as only a few nodes are highly connected hubs, whose perturbation would affect the network drastically[10]. This conclusion is based on the effects of deleting or overexpressing individual nodes. However, the addition of new interactions is thought to be an equally important process for evolution, and the network responses to such changes remain to be systematically explored.

Genomes are moulded by gene duplication, transfer, mutation and loss. Duplication occurs rapidly in all species[11,12] and through mutation serves as material for innovation. This drives cellular network evolution[13,14], even though relatively few duplications become fixed in populations[11,12]. We therefore chose to reconstruct events where an open reading frame (ORF) or gene is duplicated and subsequently becomes linked to a new regulatory input. Thus, promoter region–ORF fusions were constructed on high copy number plasmids and a subset were stably integrated in the *E. coli* chromosome. Although evolution is unlikely to take such a direct approach, except in rare cases such as gene fusions in chromosomal rearrangements, our approach provides a systematic way to sample the viability of new connectivity. By adding new connections to the existing framework across different levels in the network hierarchy, including hub genes, we created a map of the network's robustness to change.

Rewired constructs and network robustness

Reconnected gene networks (598) were constructed using the genes for seven master TFs, seven σ-factors and eight downstream TFs[5] (Fig. 1). Each construct creates network paths which inherit the inputs to the regulatory region and connect these to the downstream outputs of the ORF. As new connections are added to the wild-type network, they can generate new network motifs[5], such as simple feedback loops. For example, if node A activates node B then a promoter-B:ORF-A fusion gives a direct positive feedback loop (for example, *fliA–flhD*; Fig. 1). Highly complex reconnections are also possible (for example, *csgD–crp*, where four *csgD* promoter inputs—CRP, RpoS, OmpR and CsgD—are connected to CRP output, creating more than four multi-layer feedback loops).

All 598 rewired high-copy plasmids were cloned, except for ~30 which gave either zero PCR positives in three cloning attempts (Fig. 2; black boxes) or gave positive colonies that died (Fig. 2; maroon boxes). Most clones had similar growth yields (37 °C in LB media, 16 h; 6 replicates): 94% had mean A_{600} (absorbance at 600 nm wavelength) within 2 standard deviations (s.d.) of the mean of 23 control plasmid (Co) colonies. As ~95% of the rewired networks could be maintained in *E. coli*, most added connections are well tolerated. Shuffling connections at the top of the network hierarchy could cause drastic changes, therefore the cells' tolerance is striking. For example, CRP is the most connected TF in *E. coli*, directly regulating ~400 genes[7], yet changing regulatory inputs is possible (Fig. 2; CRP columns). Similarly, σ-factors regulate transcription globally; σ^{70} and σ^{54} (RpoD and RpoN) control ~1,000 and ~100 genes, respectively[7] and also tolerate rewiring. Such hub genes[10] could have been less resilient than less-connected genes, but the bacteria can compensate. Therefore, at least when it comes to altering regulatory inputs, the hub genes do not appear to be the Achilles' heel of the network.

GFP levels and the network structure

Each construct contains a downstream GFP ORF (Fig. 1a). Thus, GFP levels indirectly measure promoter transcription for all mutants, which can be related back to network properties (Fig. 2a). Spectrophotometer assays showed that 72% expressed GFP over 2 s.d. above mean Co (background). GFP (and A_{600}) results were also similar in minimal media with glucose, lactose or maltose as the sole carbon source, and in anaerobic conditions (Supplementary Data 1). In control RT–qPCR (reverse transcription real time quantitative PCR) assays on 84 selected clones, 70% expressed ORF transcripts >12-fold over Co (mean, 520-fold; range, 0.4 to 7,700-fold; Supplementary Fig. 2). Therefore most constructs are expressed and could potentially establish new network links. As expected,

[1]EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), UPF, 08003 Barcelona, Spain. [2]EMBL, Meyerhofstrasse 1, Heidelberg D-69117, Germany.

GFP levels vary with promoter region identity (rows, Fig. 2a). Surprisingly, there are also patterns between GFP levels and ORFs (columns, Fig. 2a). Therefore many TFs have associated expression levels that are partially promoter independent.

Analysis of variance testing confirmed that column GFP means are significantly different (one-way: F value (21 degrees of freedom, d.f.) = 8.8; P value $<2.2\times10^{-16}$) and that ORFs predict GFP levels better than promoters (two-way; ORFs: F value (21 d.f.) = 9.8, P value $<2.2\times10^{-16}$; promoters: F value (25 d.f.) = 3.5, P value $<5.7\times10^{-8}$). ORFs could set expression because each could have a particular RNA structure, affecting translation and degradation. Alternatively, the ORF TFs could be widely active, or autoregulating through ORF binding sites. The Ecocyc database[15] reports self-regulation for about two-thirds of our 22 TFs, although few ORF binding sites are currently known. Nonetheless, ORFs strongly affect expression in rewired networks.

The lowest ORFs in the wild-type hierarchy often had the lowest GFP expression (Figs 1, 2). Similarly, higher-tier factors have more interactions and significantly higher GFP (Spearman's rank correlation for GFP versus interactions: $r^2 = 0.410$; $P = 0.009$); as most network connections are positive, connecting a high-tier ORF to a low-tier promoter may increase the chance of downstream interactions indirectly activating the promoter, creating positive feedback. However, the mean GFP levels for predicted direct positive and negative feedback loops (+ and − in Fig. 2a) were not significantly different (one-sided t-test: $P = 0.393$). Thus, direct feedback loops can behave unexpectedly *in vivo*. This itself is informative, suggesting that other levels of network control can counteract direct feedback. Also, plasmid copies increase promoter concentration and thus even weak (non-physiological) TF–promoter interactions might create unpredicted loops. Overall, the results indicate a very complex rewired network response, suggesting that dissection into small network motifs may only lead to useful insights in some cases.

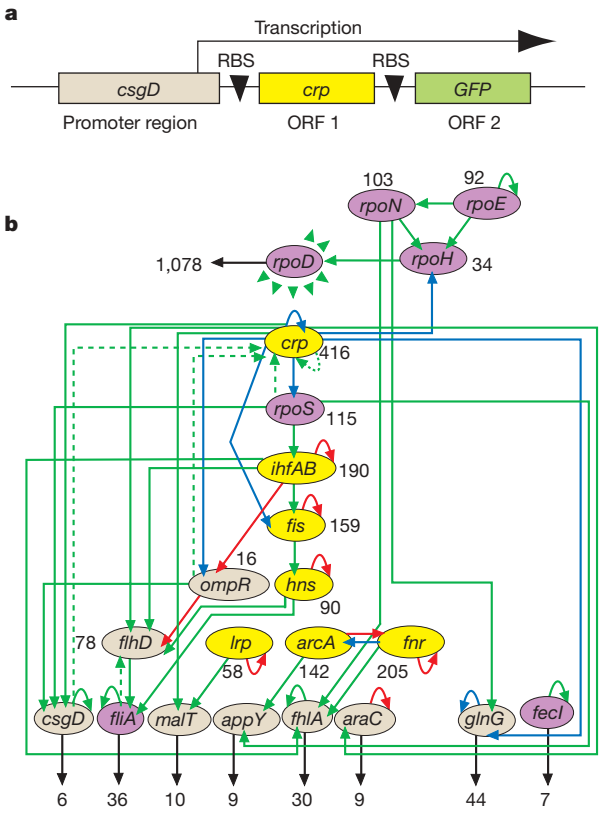

Figure 1 | Promoter–ORF network rewiring. a, Example of construct (*csgD–crp*), with two ribosome binding sites (RBS). **b**, Network diagram of the major transcription factor and σ-factor genes used. Green, red and blue arrows denote direct activating, repressing and dual interactions, respectively, from RegulonDB[6,7]. σ-factors, master regulators and lower-tier regulators are in purple, yellow and beige, respectively. Black numbers denote the total number of direct downstream ORF–gene interactions per node. The housekeeping σ-factor RpoD can activate all other nodes. Dotted arrows illustrate two rewired constructs (*fliA–flhD* and *csgD–crp*; for example, CRP, RpoS, OmpR and CsgD all regulate *csgD*, thus connecting four nodes to CRP in *csgD–crp*).

Growth signatures in rewired gene networks

To explore whether acquired network connections affect bacterial growth, A_{600} timecourses were measured. The A_{600} time derivative (estimated as linear regression slope for nine sequential A readings) gives a characteristic 'growth signature', reliably distinguishing between different *E. coli* strains (C.L., manuscript in preparation). Thus, growth signatures for all 598 constructs were calculated and the sums of least-squared distances ($\sum$l.s.d.), relative to mean control Co, indicate the scale of perturbations (Fig. 2b). Most constructs have little or no effect on growth: 84% are within the 95% confidence interval of 60 Co colonies ($(0–0.4)\times10^8$ A units^2). Therefore only 16% give distinct growth phenotypes (Fig. 2c, d). Interestingly, the corresponding genome-integrated constructs have similar but milder growth signature variations, perhaps because they are expressed 150-fold less on average (Supplementary Information).

Examining the outlier growth signatures, we noticed several patterns. For example, many constructs with *ihf A+B* ORFs have much-steeper late-growth signatures with reduced late-peaks (time, ~500 min; Fig. 2d and Supplementary Information). IHF gene products mediate the switch from exponential growth into stationary phase[16] and purified IHF binds to regulatory regions in stationary phase genes[16]. Thus the differently regulated expression of IHF in the rewired constructs may be affecting stationary phase entry. The *ihf A+B* clones were studied further using highly-detailed GFP timecourses, as developed by the Alon group; this has been achieved for 2,000 different promoters in *E. coli*, giving an unprecedented look at *E. coli* promoter activity[17]. GFP fluorescence dynamics show distinct expression profiles, with GFP expression peaking during stationary phase transition, and RT–qPCR analysis of different plasmid and integrated clones reveals dose-dependence of the phenotype (Supplementary Information).

To examine ORF overexpression versus rewiring effects, we cloned 21 ORFs into arabinose-inducible pBAD202 directional TOPO vector. *rpoE* did not clone in three attempts, which may reflect its apparent toxicity in certain rewired combinations. Different induced expression levels were quantitated using RT–qPCR (Supplementary Information). ORFs *ihfA+B*, *rpoD*, *fliA*, *appY* and *rpoE* show dose dependence, with higher expression being more deleterious to growth (Supplementary Data 2). Conversely, ORFs *fis*, *lrp*, *rpoS*, *rpoH*, *arcA*, *flhDC*, *malT* and *flhA* have cases where low or medium expression alters growth more in some promoter–ORF constructs, indicating a dominance of rewiring effects over high expression. ORFs *fecI*, *hns*, *fnr*, *araC*, *glnG*, *ompR* and *csgD* have very few different growth effects in all conditions. Overall, the growth phenotypes of only 7 of the 22 ORFs tested were explained primarily by overexpression effects. Growth phenotypes are ultimately a mixture of expression levels (dosage), timing and rewiring effects.

Evolvability in rewired gene networks

As most acquired network connections affect growth minimally, the first step in evolving a new network property is easily accessed. We therefore investigated whether rewired constructs themselves provide any potential for evolution. By pooling all cloned constructs (~570, plus a 23-fold molar excess of wild-type Co) and applying selective pressures, we searched for individuals with specific fitness advantages under three conditions: (1) serial passaging of bacteria in liquid culture; (2) longevity in extended periods at 37 °C; and (3) survival after 50 °C heat shock for 1 h. Serial passaging was done in seven replica flasks, transferring 1 μl of culture mixture into 120 ml

fresh medium, every 12–16 h. After 20 to 55 rounds, 12 network clones were repeatedly selected in independent flasks (Fig. 3a). The clones can be plotted as a 'selectability map' (marked 'S' in Fig. 2), and are associated with near-wild-type growth and low (but non-zero) GFP expression. Notably, certain *flhD* promoter–ORF combinations were enriched. *flhD* regulates flagellar genes and loss-of-function mutants increase cell division fivefold[18]. Flagellar genes are non-essential and cost the cell time, energy and materials. We speculate that unnaturally connected *flhD* promoters may repress flagellar biosynthesis, giving a selective advantage. Conversely, expressing FliA flagellar σ-factor can disrupt growth (for example, *malT–fliA*; Fig. 2d). This correlates with flagellar biosynthesis, as 27 of the 30 largest changes in *malT–fliA* are upregulated flagellar or taxis genes (Supplementary Data 3). Serial passaging can select for mutations and adaptations that optimize the bacteria to their environment[19,20], and our results show that reconnected gene networks themselves can provide a substrate for selection.

Two further selection pressures tested the rewired networks. Stationary phase library mixtures were incubated at 37 °C, for up to 8 days, in 10 replica flasks. Alternatively, stationary cultures were heat-shocked at 42 °C (15 min) and then at 50 °C (1 h)[21], using three rounds of heat selection, plating and harvesting. In both longevity and heat experiments, virtually all surviving clones were *rpoS–ompR* (Fig. 3b, c). As other *rpoS*–promoter and *ompR*–ORF clones were never selected, it appears that both are required together. Furthermore, *rpoS–ompR* integrated into the bacterial chromosome is selected over wild-type in heat-shock and longevity experiments, despite much lower expression in RT–qPCR: plasmid *rpoS–ompR* = 650-fold over Co; integrated = 2-fold. Integrated *rpoS–ompR* heat selection is weaker than for plasmid (Fig. 3d, e), while longevity selection is stronger, reaching 92% after 1 week at 37 °C (Fig. 3f). By contrast, 430-fold overexpressed *ompR* ORF (in pBAD-*ompR*) is not selected over a pBAD-empty control (Fig. 3g). Therefore selection requires the rewiring combination, functioning even with low expression. While individual pressures may select for overexpression or new mutations, we have not found evidence of this. As selections were reproducible in independent tubes, and with different copy numbers, extra mutations are probably not necessary. Therefore, even in a small library space of ~600 networks, acquired connections can themselves provide specific fitness advantages.

DNA chip analysis of rewired gene networks

Affymetrix *E. coli* genome 2.0 arrays were used to get a transcriptome-wide view of rewired networks (three replicas per sample). The genes

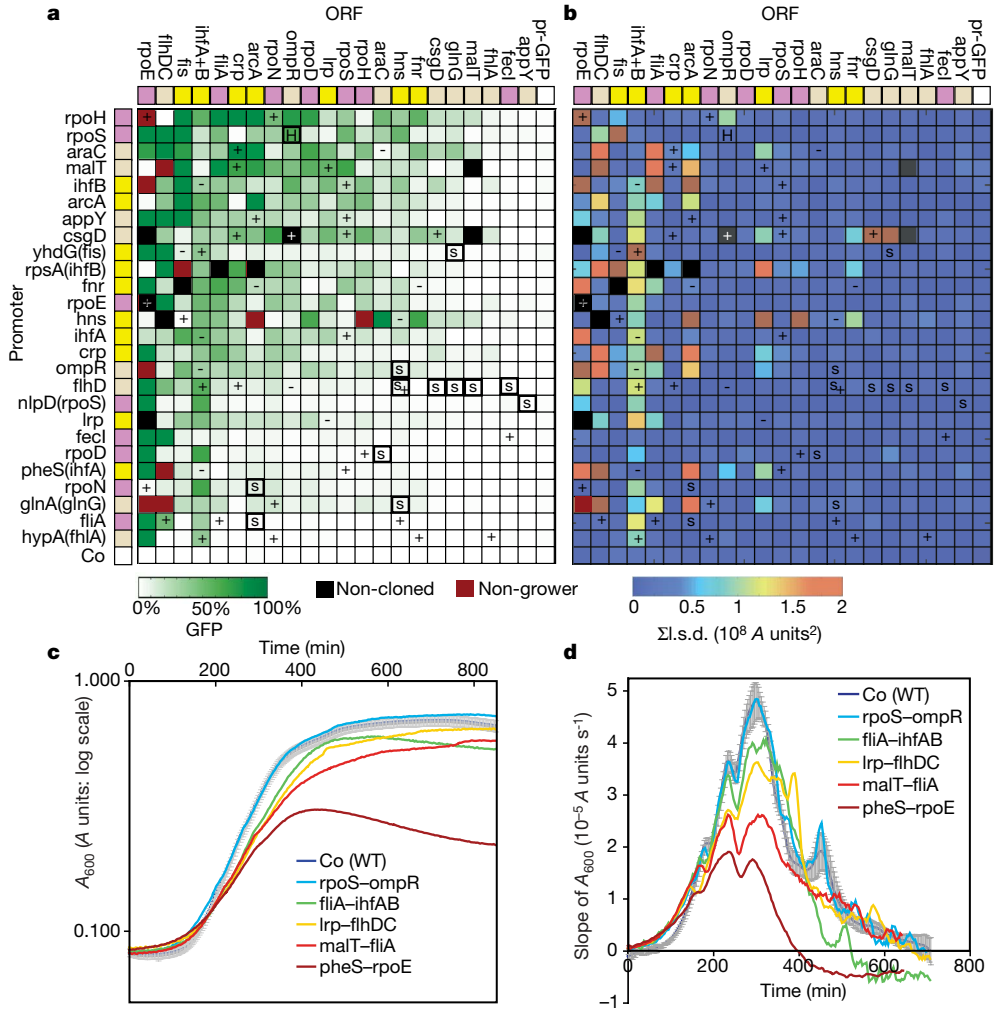

Figure 2 | GFP expression and growth signatures of promoter–ORF recombinants. σ-factors, master regulators and lower-tier regulators are in purple, yellow and beige, respectively. **a**, Green squares show mean GFP expression, ranked by columns, left-to-right, and rows, top-to-bottom (six repeats, normalized by A_{600}); non-cloned constructs are black; non-growers (maroon) vary slightly between independent colonies (panels **a** and **b**). Controls include promoter–GFP fusions (pr–GFP) and promoter-less GFP (Co). Direct positive and negative feedback loops are marked '+' or '−' and selection results by 'S' (serial passaging) or 'H' (50 °C heat survival). **b**, A_{600} slopes (time-derivatives) give characteristic profiles (growth signatures), displayed by plotting the sum of least squared difference ($\sum$l.s.d.), relative to mean wild-type. **c**, **d**, selected growth curves (A_{600}) and signatures. Time = 0 is set at ~7 h after inoculation, removing lag phase. Error bars show 1 s.d. of 60 Co colonies.

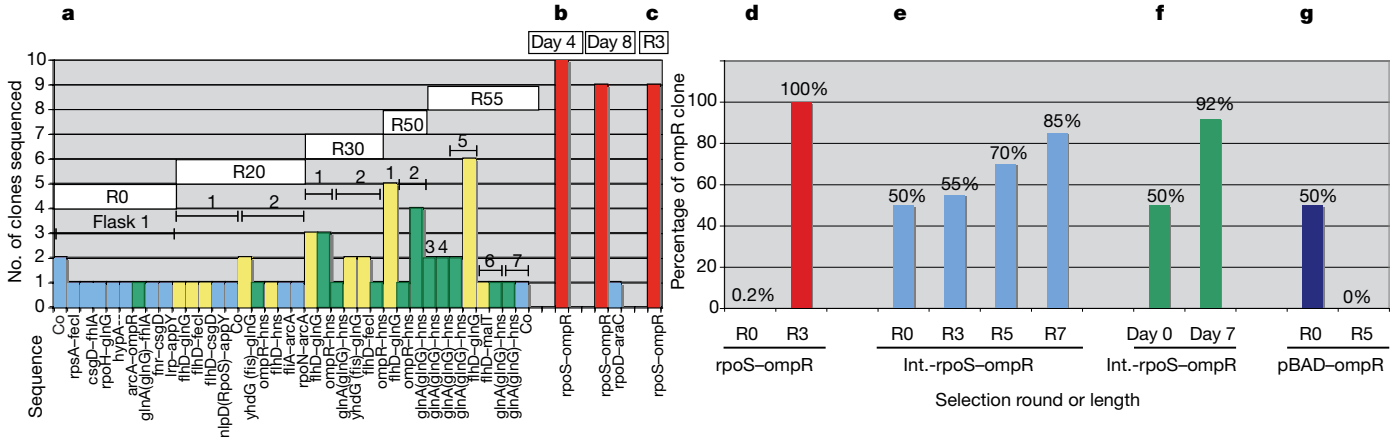

Figure 3 | Selection experiments. **a**, Serial passaging of 1 μl of culture mixture into 120 ml fresh medium, every 12–16 h. Replica flasks 1 to 7, and passage rounds R0 to R55, are indicated. Clones related to *glnA(glnG)–hns* (green) and *flhD–glnG* (yellow) were selected. **b**, Clones surviving stationary phase at 37 °C (day 4, day 8); *rpoS–ompR* is strongly selected (red). **c**, Clones obtained in heat survival (50 °C, 1 h) for three rounds (R3). **d**, *rpoS–ompR* plasmid (650-fold overexpressed transcript), heat-selected over the ~570 constructs (three rounds). **e**, Integrated *rpoS–ompR* (twofold overexpressed transcript), heat-selected over integrated Co (seven rounds). **f**, Integrated *rpoS–ompR*, selected over wild-type TOP10 cells (7 days, 37 °C). **g**, *ompR* ORF in pBAD (430-fold overexpressed), is not heat-selected over pBAD vector (five rounds). **b–d**, 10 replica flasks. **e-g**, Five replica tubes, 5 PCRs per tube.

were ranked by *P* value for different expression between samples. Family-wise error rates (FWER[22]) and false-discovery rates (FDR[23]) measured confidence in differential expression (Table 1). Comparing *rpoS–ompR* against Co control, only 13 out of ~4,000 genes were differentially expressed with high confidence, including several up-regulated chaperone and shock genes (Table 1a; FWER < 1; <1 false positive expected). Extending the list to the 23 most significant differences yields further shock genes (FDR < 10%; 2 expected false positives). After three rounds of heat selection, 39 genes changed in *rpoS–ompR* (FWER < 1), 87% being gene downregulations, including permeases (Table 1b and Supplementary Data 3). RpoS is activated in stationary phase entry, in heat stress and starvation[24], and positively regulates genes for acid, heat and salt tolerance[25,26]. OmpR controls osmoregulation[27] and is regulated by several shock pathways to control biofilm formation[28,29]. Furthermore, endogenous RpoS and OmpR are both positive regulators of *csg* genes, which are downstream of the *cpxA* shock signalling pathway[30,31]. The *rpoS–ompR* survival mechanism includes chaperone and shock gene

Table 1 | DNA chip expression analysis

Rank	*P* value	FDR	FWER	+/−	Gene	Notes
a						
1	3×10^{-13}	10^{-7}%	1.1×10^{-9}	+	*ompR*	Osmotic response regulator
2	1.4×10^{-9}	3×10^{-4}%	5.6×10^{-6}	+	*nlpD*	Lipoprotein
3	3.1×10^{-5}	2.3%	0.13	+	*ycjX*	NTH domain
4	3.2×10^{-5}	2.3%	0.13	+	*ldhA*	D-lactate dehydrogenase
5	3.2×10^{-5}	2.3%	0.13	+	*dnaK*	*Chaperone Hsp70
6	3.4×10^{-5}	2.3%	0.14	−	*artM*	†A3T permease protein
7	5.1×10^{-5}	2.5%	0.21	+	*groL*	*GroEL, chaperone Hsp60
8	5.5×10^{-5}	2.5%	0.22	+	*maeB*	Predicted oxidoreductase
9	5.5×10^{-5}	2.5%	0.23	+	*dnaJ*	*Chaperone with DnaK; Hsp
10	7.1×10^{-5}	2.9%	0.29	+	*ycjF*	Conserved inner membrane protein
11	1.7×10^{-4}	6.2%	0.67	+	*clpB*	*Heat shock protein; Hsp
12	1.8×10^{-4}	6.2%	0.74	+	*htpG*	*Chaperone Hsp90, Hsp C 62.5
13	2.2×10^{-4}	6.7%	0.88	−	*artQ*	†A3T permease protein
14	2.5×10^{-4}	7.2%	1	+	*ogrK*	Prophage P2 OGR protein
15	2.7×10^{-4}	7.3%	1	−	*cspB*	CspB
16	3.1×10^{-4}	7.7%	1	+	*groS*	*GroES, Hsp60-binding chaperone
17	3.2×10^{-4}	7.7%	1	−	*yaiA*	ORF for hypothetical protein (hyp.)
18	3.6×10^{-4}	8.2%	1	+	*gnty*	Gluconate transport associated
19	4.0×10^{-4}	8.7%	1	+	*mfd*	*Mutation frequency decline
20	4.3×10^{-4}	8.7%	1	−	*cspI*	*Cold shock-like protein
21	4.9×10^{-4}	9.1%	1	+	*hmpA*	Dihydropteridine reductase
22	4.9×10^{-4}	9.1%	1		*bcsG*	*Gene involved in biofilm formation[43]
23	5.7×10^{-4}	9.8%	1	+	*hslU*	*Hsp hslVU, chaperone homology
b						
1	1.1×10^{-07}	0.04%	0.0004	−	*gadW*	Regulator for acid resistance
2	2.0×10^{-07}	0.04%	0.001	−	*artP*	†A3T component
3	1.1×10^{-06}	0.1%	0.004	−	*artQ*	†A3T permease protein
4	2.2×10^{-06}	0.2%	0.009	−	*artM*	†A3T permease protein
5	6.2×10^{-06}	0.4%	0.025	−	*yfjO*	CP4-57 prophage protein (hyp.)
6	6.3×10^{-06}	0.4%	0.025	−	*ybcM*	DLP12 prophage; AraC type TF (hyp.)
7	7.4×10^{-06}	0.4%	0.030	−	*ygiV*	DNA gyrase inhibitor paralog
8	8.0×10^{-06}	0.4%	0.032	−	*yehZ*	†Osmoprotectant (permease) (hyp.)

a, List of the 23 differentially expressed genes with lowest *P* values, for *rpoS–ompR* against the control (Co). **b**, The 8 differentially expressed genes with lowest *P* values for *rpoS–ompR* (after 3 rounds of heat selection) compared against non-heat-treated *rpoS–ompR*, highlighting downregulation of arginine transport and permeases. FDR, false discovery rate; FWER, family-wise error rate; Hsp, heat shock protein; hyp., hypothetical; NTH, nucleoside triphosphate hydrolase; A3T, arginine 3rd transport system. Up- and downregulation are respectively designated by + or − in column 5.
* Shock proteins.
† Permease proteins.

upregulation, and permease downregulation, requiring precise expression timing and refinements after multiple rounds of heat shock.

The 'promoter-only' constructs are interesting, because high-copy promoters could titre out factors that bind to the endogenous promoter, changing overall transcription. However, comparing the *rpoS* promoter (with GFP ORF) to Co, only three high-confidence changes were seen: *nlpD(rpoS)*, *cspB* and *ilvC* (FWER < 1). In this case, the promoter *per se* has rather little influence on the transcriptome of the cell. Unlike the relatively few changes between Co, *rpoS*, untreated *rpoS–ompR* and heat-selected *rpoS–ompR*, 359 genes were differentially expressed in *malT–fliA* versus Co (FWER < 1; Supplementary Information). This clone has a high GFP level like *rpoS–ompR* but a contrastingly altered growth signature (Fig. 2d). Therefore rewiring perturbs ~10% of genes, yet the cell remains viable. Interestingly, integrated *malT–fliA* has lower transcript expression, relative to Co (11-fold, compared with 67-fold in the plasmid), but it also has perturbed growth, albeit less pronounced (Supplementary Data 2). As >80% of constructs have near-wild-type growth characteristics, they may be much closer to the *rpoS–ompR* situation than to *malT–fliA*, with very few differentially expressed genes. In that case, the reconnected gene network, even when highly expressed, does not appear to propagate changes across the whole network.

Synthetic biology and gene networks

To understand the forces, hurdles and design principles moulding gene network evolution[14,32–35], we need to test our understanding by constructing synthetic model systems[36–40]. The observations described here show that bacteria can both tolerate and exploit radical changes in their circuitry. This raises the possibility that similar experiments could be tried in other organisms, from yeast to mammals, to ascertain whether tolerance towards rewiring is a general feature of evolved biological networks. It is interesting to compare our rewiring results to those of ref. 41, when 5,280 *Saccharomyces cerevisiae* genes were overexpressed and only 15% were found to cause growth defects. The effects of rewiring (~16% growth phenotypes) include an element of dosage dependency (overexpression), but also altered timing of expression, and potentially subverting elements in more than one pathway. For *E. coli*, it is surprising that rewired clones can have such limited genome-wide transcriptional changes, indicating that bacterial networks have an in-built predisposition to dampen change. *E. coli* is a complex, tightly coordinated biological system regulated by multiple layers of molecular networks: in tampering with the transcription regulatory network alone, we learn that the static network view provides a map of poor quality to predict the result of genetic perturbations. However, some general trends are ascertainable, such as network hierarchy correlating with expression. Also, our results indicate that partition of a network into small modules (negative feedback, feed-forward, and so on) could in some cases be misleading, as the behaviour of these modules is affected to a large extent by the rest of the network in which they are embedded. The vast majority of added network connections gave no evidence of new phenotypes, even for highly connected hub genes, yet a few gave selective advantages. This pays tribute to the evolutionary potential provided to the cell by the plasticity of its genome.

METHODS SUMMARY

Cloning. The 26 promoter regions (defined as including all upstream TF binding sites annotated in Ecocyc[15]) were cloned with the 22 associated ORFs into pGLOW-TOPO (Invitrogen). Each construct also contained a downstream GFP ORF (with separate Shine-Dalgarno sequence). Full sequences are in Supplementary Information. For ORFs with more than one annotated promoter, both were cloned separately (for example, *rpoS* and *nlpD* promoters for rpoS transcription[42]; denoted here by *nlpD(rpoS)*). *rpoD* has two promoters, and *dnaG*(*rpoD*) did not clone successfully. The non-expressing control plasmid (Co) contained a 66 bp non-regulatory DNA sequence upstream of the promoter-less GFP ORF.

GFP measurements. Bacterial cultures (200 µl; 16 h growth) were diluted 20 µl:180 µl PBS in 96-well plates. Six independent sample readings (excitation, 485 nm; emission, 520 nm) had Co-background subtracted and were normalized for A_{600}, with a threshold to remove very low A readings (background-corrected $A_{600} < 0.03$).

Growth signatures. A quantity (0.2 µl) of 1:200-diluted overnight bacterial culture was added to 120 µl LB medium (with 100 µg ml^{-1} ampicillin and 50 µg ml^{-1} streptomycin) in 96-well plates. A Tecan Genios plate reader measured A_{600} (XFLUOR4 software; 37 °C; 595 nm absorbance; 3 flashes; interval 190 s; shake duration (orbital low) 130 s; 1,000 cycles, ~20 h; lids on). To avoid edge-effects, only the plates' central 60 wells were used (outer wells contained sterile medium). The assay is sensitive to volume, evaporation and lid condensation; the Tecan machine was optimal (other machines had lid effects). The slope of linear regression of the A_{600} readings, over a sliding window of nine sequential time-points, gave the growth signatures.

Integrations. About 40 representative pGLOW constructs, including Co, were integrated into the *E. coli* chromosome using *manX* locus site-directed integration (Gene Bridges Kit K006).

Selection experiments. After serial passaging, 37 °C-longevity or 50 °C-heat shock assays, samples were plated onto selective agar media and colonies were picked at random and sequenced or PCR-verified.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 December 2007; accepted 22 February 2008.

1. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1474 (1997).
2. Perez-Rueda, E. & Collado-Vides, J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **28,** 1838–1847 (2000).
3. Salgado, H. *et al.* RegulonDB (version 3.2): Transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res.* **29,** 72–74 (2001).
4. Madan Babu, M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* **31,** 1234–1244 (2003).
5. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31,** 64–68 (2002).
6. Martinez-Antonio, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.* **6,** 482–489 (2003).
7. Salgado, H. *et al.* RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.* **34,** D394–D397 (2006).
8. Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286,** 509–512 (1999).
9. Guelzim, N., Bottani, S., Bourgine, P. & Kepes, F. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genet.* **31,** 60–63 (2002).
10. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. *Nature* **406,** 378–382 (2000).
11. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290,** 1151–1155 (2000).
12. Hooper, S. D. & Berg, O. G. On the nature of gene innovation: Duplication patterns in microbial genomes. *Mol. Biol. Evol.* **20,** 945–954 (2003).
13. Teichmann, S. A., Park, J. & Chothia, C. Structural assignments to the *Mycoplasma genitalium* proteins show extensive gene duplications and domain rearrangements. *Proc. Natl Acad. Sci. USA* **95,** 14658–14663 (1998).
14. Teichmann, S. A. & Babu, M. M. Gene regulatory network growth by duplication. *Nature Genet.* **36,** 492–496 (2004).
15. Keseler, I. M. *et al.* EcoCyc: A comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* **33,** D334–D337 (2005).
16. Mangan, M. W. *et al.* The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.* **59,** 1831–1847 (2006).
17. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature Methods* **3,** 623–628 (2006).
18. Pruss, B. M. & Matsumura, P. A regulator of the flagellar regulon of *Escherichia coli*, flhD, also affects cell division. *J. Bacteriol.* **178,** 668–674 (1996).
19. Fong, S. S., Joyce, A. R. & Palsson, B. O. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* **15,** 1365–1372 (2005).
20. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436,** 588–592 (2005).
21. Delaney, J. M., Ang, D. & Georgopoulos, C. Isolation and characterization of the *Escherichia coli htrD* gene, whose product is required for growth at high temperatures. *J. Bacteriol.* **174,** 1240–1247 (1992).
22. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6,** 65–70 (1979).

23. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
24. Loewen, P. C., Hu, B., Strutinsky, J. & Sparling, R. Regulation in the rpoS regulon of *Escherichia coli*. *Can. J. Microbiol.* **44,** 707–717 (1998).
25. Cheville, A. M., Arnold, K. W., Buchrieser, C., Cheng, C. M. & Kaspar, C. W. rpoS regulation of acid, heat, and salt tolerance in *Escherichia coli* O157:H7. *Appl. Environ. Microbiol.* **62,** 1822–1824 (1996).
26. Foster, J. W. & Moreno, M. Inducible acid tolerance mechanisms in enteric bacteria. *Novartis Found. Symp.* 221, 55–69; discussion 70–74 (1999).
27. Pratt, L. A., Hsing, W., Gibson, K. E. & Silhavy, T. J. From acids to osmZ: Multiple factors influence synthesis of the OmpF and OmpC porins in *Escherichia coli*. *Mol. Microbiol.* **20,** 911–917 (1996).
28. Vidal, O. *et al.* Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: Involvement of a new ompR allele that increases curli expression. *J. Bacteriol.* **180,** 2442–2449 (1998).
29. Prigent-Combaret, C. *et al.* Complex regulatory network controls initial adhesion and biofilm formation in *Escherichia coli* via regulation of the *csgD* gene. *J. Bacteriol.* **183,** 7213–7223 (2001).
30. Cosma, C. L., Danese, P. N., Carlson, J. H., Silhavy, T. J. & Snyder, W. B. Mutational activation of the Cpx signal transduction pathway of *Escherichia coli* suppresses the toxicity conferred by certain envelope-associated stresses. *Mol. Microbiol.* **18,** 491–505 (1995).
31. Romling, U., Bian, Z., Hammar, M., Sierralta, W. D. & Normark, S. Curli fibers are highly conserved between *Salmonella typhimurium* and *Escherichia coli* with respect to operon structure and regulation. *J. Bacteriol.* **180,** 722–731 (1998).
32. Madan Babu, M., Teichmann, S. A. & Aravind, L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J. Mol. Biol.* **358,** 614–633 (2006).
33. Lozada-Chavez, I., Janga, S. C. & Collado-Vides, J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res.* **34,** 3434–3445 (2006).
34. Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **14,** 283–291 (2004).
35. Poelwijk, F. J., Kiviet, D. J. & Tans, S. J. Evolutionary potential of a duplicated repressor–operator pair: Simulating pathways using mutation data. *PLoS Comput. Biol.* **2,** e58 (2006).
36. Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420,** 224–230 (2002).
37. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).
38. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403,** 339–342 (2000).
39. Isalan, M., Lemerle, C. & Serrano, L. Engineering gene networks to emulate *Drosophila* embryonic pattern formation. *PLoS Biol.* **3,** e64 (2005).
40. Isalan, M., Santori, M. I., Gonzalez, C. & Serrano, L. Localized transfection on arrays of magnetic beads coated with PCR products. *Nature Methods* **2,** 113–118 (2005).
41. Sopko, R. *et al.* Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21,** 319–330 (2006).
42. Lange, R. & Hengge-Aronis, R. The *nlpD* gene is located in an operon with rpoS on the *Escherichia coli* chromosome and encodes a novel lipoprotein with a potential function in cell wall formation. *Mol. Microbiol.* **13,** 733–743 (1994).
43. Solano, C. *et al.* Genetic analysis of *Salmonella enteritidis* biofilm formation: Critical role of cellulose. *Mol. Microbiol.* **43,** 793–808 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Martinez Arias, J. Sharpe, M. Babu, P. Bork and B. Schoenwetter for critical reading of the manuscript; P. Ribeca for RegulonDB analysis; and B. Di Ventura and S. Martinez de Pablo for cloning assistance. C.H. and E.R. were funded by European Commission FP6 Netsensor Grant 012948.

Author Contributions M.I., C.L., K.M., P.B., C.H. and M.G.-C. carried out experiments. E.R. and C.L. did computational analysis. M.I., C.L. and L.S. conceived experiments. M.I. and L.S. supervised experiments.

Author Information Microarray data are MIAME-compliant and have been deposited at ArrayExpress http://www.ebi.ac.uk/microarray-as/aer/entry, accession E-MEXP-732. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.I. (isalan@crg.es).

METHODS

Cloning. The 27 promoter regions and 22 ORFs were cloned from *Escherichia coli* strain Top10 (Invitrogen), using genomic PCR (DNA template extracted with a Genomic-tip 500/G (Qiagen); 1 μg per PCR). PCR conditions were typically: 97 °C, 3 min; (97 °C, 30 s; 50 °C, 30 s; 72 °C, 2 min 30 s) × 10 cycles; (97 °C, 30 s; 72 °C, 2 min 30 s) × 20 cycles. PCR products were cloned via topoisomerase then *Fse*I-*Pac*I cloning. Full sequences and maps are in Supplementary Information.

The promoter-less control (Co) plasmid was pGLOW-TOPO with a 66 bp DNA fragment in the TOPO cloning site: CGTCGACGTGGCGCCGCCGGATAAGGCGTTTACGTGACGGCCGGCCCGGGTTCTGGCTTAATTAAA. This sequence was chosen empirically because it was obtained as an oligonucleotide by-product when cloning promoters into pGLOW-TOPO. The sequence contains no regulatory motifs and has no GFP-inducing activity in bacteria, as measured by fluorimetry and western blot with anti-GFP antibody.

Site-directed integration into the *E. coli* chromosome. Approximately 40 of the pGLOW constructs (representing growth phenotypes, selections and Co) were stably integrated into the *E. coli* chromosome at the *manX* locus (Gene Bridges Kit K006). The kit uses PCR to provide two 50 bp homology arms, matching the *manX* locus. Transient expression of the Red/ET recombination proteins provides a highly specific site-directed integration, which is verified by genomic PCR. The following generic primers (gel pure) amplify approximately 2.3 kb of plasmid backbone from the original pGLOW constructs (including the ampicillin resistance gene) and add *manX* homology arms: pGLOW_Amp_F, GTTGATACATGGGGAGGCAGCCCGTTCAATGCTGCCAGCCGCATTGTCGTCGCTCAGTGGAACGAAAACTC; pGLOW_polyA_R, CGAGCATTGGAATGTTAACGCCTGCAATGACTTCATAATGCTCTTTGTCGAGCTGGTTCTTTCCGCCTCA.

Library mixtures for selections. The ~570 cloned constructs, plus 23 control (Co) samples, were inoculated from the frozen glycerol stock archive and grown in individual 200 μl wells for 16 h, as described above. The cultures were then pooled to make a library mixture. Library glycerol stocks were made by adding 4 ml of 50% glycerol to 5 ml of library culture, and storing at −80 °C.

Selection by serial passaging. 100 μl of library glycerol stock was grown in 120 ml LB medium (with 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} streptomycin; 37 °C and orbital shaking, 300 r.p.m). Culture samples were passaged into fresh medium every 12–16 h, in seven replica flasks; for rounds 2 and 3, 100 μl and 10 μl of culture were passaged into fresh medium, respectively; for round 4 onwards, 1 μl samples were passaged (10^5–10^6 c.f.u). After 20, 30, 50 and 55 rounds of passaging, samples were plated onto Petri dishes (containing 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} streptomycin) in order to get single colonies for DNA sequencing (sequencing primers for pGLOW vector: pGLOW_TOPO_F, TGGCTAGCGTTTAAACTTAAGC; pGLOW_TOPO_R, GAATTGGGACAACTCCAGTG).

Selection by longevity in stationary phase. 1 μl of glycerol stock library mixture inoculated 2 ml LB medium (supplemented with 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} streptomycin); 10 replica tubes; 37 °C, 300 r.p.m. After 24 h, 4 days and 8 days, 1 μl samples (diluted in 200 μl LB) were plated onto Petri dishes to get single colonies for DNA sequencing, as above.

Selection by heat shock at 50 °C. 2 ml library cultures were grown as above. 100 μl samples were then transferred to 0.2 ml PCR tubes on a PCR block, programmed to incubate at 42 °C, 15 min, 50 °C for 1 h, and then 4 °C, 5 min. 1 μl samples were immediately diluted and plated out onto Petri dishes, as above. Surviving colonies were harvested, grown to stationary phase and the entire procedure was repeated for 3 rounds. One colony was then sequenced per plate, from 10 independent selection tubes.

Affymetrix chip analysis. Sample preparation and treatment. Top10 cells containing pGLOW constructs were grown for 16 h at 37 °C (as above), diluted 1,000-fold and then grown in 5 ml (pre-warmed) LB medium for 6 h (note that cells should not be in stationary phase). 400 μl of culture was used per RNA extraction (Qiagen kit RNeasy). Absorbance measurements were taken to quantitate the RNA. Aliquots were checked for RNA degradation by capillary electrophoresis. All subsequent microarray handling was carried out following Affymetrix recommendations. 15 chips (3 per sample) with MG1655 (K12) probesets (Affymetrix *E. coli* Genome 2.0 Array) were used to test the 5 bacterial populations: Co, *rpoS*–(no ORF except GFP), *rpoS–ompR*, *rpoS–ompR* after heat selection and *malT–fliA*. To analyse differential expression, a linear model was used through Limma software[44] and lists of probabilities of individual genes being differentially expressed were compiled. Holm family-wise error rate[45] was used to determine differential expression (FWER < 1 as cut-off: fewer than 1 false positive expected in the list of differentially expressed genes). False discovery rates[46] were used as an alternative to calculate the expected number of false positives (fewer than 1 false positive was used as a confidence cut-off to determine significant differential expression).

RT–qPCR (reverse transcription real time quantitative PCR). RNA was extracted from bacterial cultures with an RNeasy Protect Mini Kit (Qiagen). Complementary DNA was made from 500 ng total RNA, with primer p(dT)15 (Roche) and SuperScript II Reverse Transcriptase (Invitrogen). 0.2 μl cDNA, 0.3 pmol of each primer and 5 μl LightCycler 480 SYBR Green I Master mix (Roche) and a Roche Applied Science LightCycler 480 Instrument (384 wells) were used (10 μl reactions). Samples were normalized for *gnd* housekeeping gene mRNA and compared to Co expression for fold-difference calculations.

44. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, article–3 (2004).
45. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
46. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).