THE EVOLUTION OF GENETIC REGULATORY SYSTEMS IN BACTERIA

Harley H. McAdams, Balaji Srinivasan* and Adam P. Arkin‡*

The genomes of bacterial species show enormous plasticity in the function of individual genes, in genome organization and in regulatory organization. Over millions of years, both bacterial genes and their genomes have been extensively reorganized and adapted so that bacteria occupy virtually every environmental niche on the earth. In addition, changes have occurred in the regulatory circuitry that controls cell operations, cell-cycle progression and responses to environmental signals. The mechanisms that underlie the adaptation of the bacterial regulatory circuitry are crucial for understanding the bacterial biosphere and have important roles in the emergence of antibiotic resistance.

REGULATORY CIRCUIT
A reaction network that can involve transcription factors, promoters, enzymes, structural genes, functional RNAs and metabolites. Regulatory networks control activation of genes in development, in the cell cycle and in the activation of metabolic pathways.

**Department of Developmental Biology, Stanford University School of Medicine, B300 Beckman Center, Stanford, California 94305, USA.*
‡Physical Bioscience Division, Lawrence Berkeley Laboratories, Department of Bioengineering, University of California at Berkeley, and the Howard Hughes Medical Institute, Berkeley, California 94720, USA.
Correspondence to H.H.M.
e-mail: hmcadams@stanford.edu
doi:10.1038/nrg1292

The field of evolutionary genomics has recently devoted considerable research effort towards understanding the evolution of coding region sequences and of genome organization — research that has been boosted by the availability of a large number of complete genome sequences. Less attention, however, has been devoted to explaining the evolution of the overall genetic REGULATORY CIRCUITRY (FIG. 1) that controls cellular functions. Recent FUNCTIONAL GENOMICS studies are beginning to address this topic. The regulatory circuitry includes environmental sensors, sensors that reflect the internal state of the cell and a wide variety of signalling pathways. These signalling pathways comprise a network of protein-level reactions and genetic regulatory mechanisms that implement a type of biochemically based 'logic' — a control system — that determines how the cell responds to the sensed conditions. We are interested here in how the organization and mechanisms of this control system have evolved alongside genomic evolution. Key questions include the degree of plasticity of the regulatory network structure, how the modular organization of cell function emerges, what forces create recurrently observed CIRCUIT MOTIFS and, finally, how the complex, highly organized and biochemically based regulatory systems in cells emerged. Our ability to answer these questions is growing rapidly owing to emergence of new data sources and new experimental techniques (BOX 1).

In this review, we discuss bacterial evolution, with an emphasis on the evolution of the regulatory circuitry; however, changes in this circuitry and the organization of the genome are inextricably linked, so the discussion inevitably weaves between the two topics. We first summarize the evidence that bacterial genomes show enormous plasticity in the function of individual genes, in genome organization and in regulatory organization. Second, we discuss the evidence for the spontaneous evolutionary emergence of a hierarchical, modular functionality of increasing complexity. Third, we consider the spontaneous emergence of increasing regulatory complexity that allows cells to change their behaviour or their metabolic capabilities and therefore to survive in temporally and spatially complex environments. We then ask why the core identity of common bacterial species is conserved in the presence of seemingly highly disruptive mechanisms for genomic change. Finally, we comment on the potential for harnessing the dynamic processes of regulatory adaptation for engineering novel organisms and the implication of these processes in the emergence of antibiotic-resistant bacteria.

Although the focus in this article is the evolution of bacterial regulation, we note with interest an emerging consensus that metazoan evolution — like that of bacteria — is more strongly driven by changes in the complexity of regulation of gene expression than by changes in non-regulatory proteins[1,2].

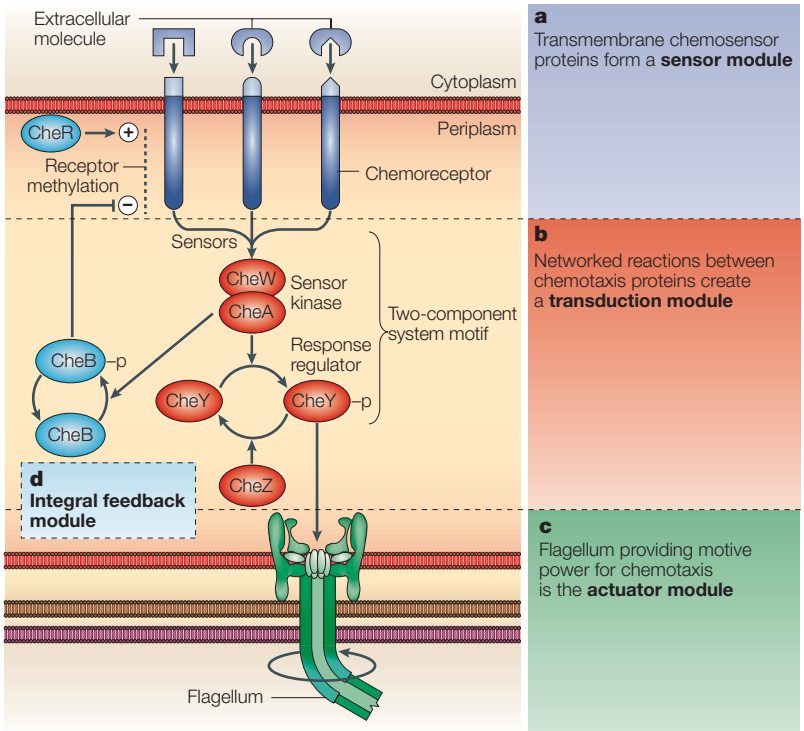


Figure 1 | **Example of a bacterial regulatory circuit.** The figure illustrates the modular nature of the *Escherichia coli* chemotaxis network. **a** | The sensor module includes several different chemoreceptors that are sensitive to different extracellular molecules. Diverse collections of hundreds or thousands of these chemoreceptors are assembled into large membrane-associated arrays that are localized predominantly at the cell poles[62]. External molecules bind to receptors on the cell surface and activate the CheW/CheA sensor kinase. **b** | The transduction module comprises biochemical reactions between different chemotaxis molecules that create a pathway that communicates a signal to the distant flagella. This signal changes the frequency of reversal of the flagella motor (the actuator module, **c**) in a manner that causes the bacteria to swim generally towards attractive chemical sources and away from hazardous sources. **d** | The feedback loop within the sensor module, which involves methylation of the receptor, allows the network to operate over wide concentration ranges of the external molecule that is being sensed. The sensor kinase–response regulator reaction, labelled as a TWO-COMPONENT SYSTEM motif, is part of the communication link that signals the status of the chemoreceptors to the motor. Because of its strong modular organization, the chemotaxis system is evolutionarily flexible, as shown by the diversity of chemical signals and response regulator functions that are found among motile bacteria.

Plasticity of bacterial genomes

Many lines of evidence indicate that the content and organization of the bacterial genome is highly changeable — that is, highly plastic. Genes and parts of genes can move within and between genomes, and contiguous pieces of DNA that encode many genes are transferred within genomes and among cells by mobile genetic elements[3–5].

The processes that lead to plastic bacterial genomes occur randomly and usually at such a slow rate that almost all bacterial cell divisions yield progeny with genomes that are identical to the parent. (The rate of spontaneous point mutations in bacteria is in the range of 10^{-9}–10^{-10} mutations/cell/generation[6].) However, owing to the vast numbers of bacteria and their short generation times, there is extensive exploration, over long time spans, of possible bacterial genomic arrangements. As a result, most observed bacterial species are in rough dynamic balance (that is, they are more or less optimized) for the range of environmental conditions that they encounter in their target niches. In this situation, point mutations or other genomic changes that occur will almost always reduce fitness and therefore tend to disappear from the population[7]. However, any subpopulation of bacteria that encounters a persistently different environment can undergo a wide range of adaptive changes owing to selection from naturally occurring genetic variants.

For large-scale and rapid bacterial evolution, new functions and new pathways arise from horizontal gene transfer (HGT), and by large-scale internal recombination processes such as duplication, deletion and inversion. Point mutation is probably the main means by which bacteria achieve fine-tuning — that is, adjustment of their KINETIC PARAMETERS and individual DNA-binding sites. A phylogenetic tree derived from the entire genomes of 41 bacteria and 10 ARCHAEBACTERIA shows that changes from gene loss and gene birth are several times more frequent than gene changes due to HGT[8]. A comparison of protein structural domains and their reuse in a number of pathways across a wide array of organisms indicates that the dominant mechanisms for expansion of the protein repertoire are gene duplication, divergence and recombination[9]. Importantly, however, the genetic variation for bacterial evolution comes not only from internal reorganization of the genome, but also from the vast genetic resource (the 'metagenome') that is available in the biosphere, which cells can access through HGT[10]. Striking examples of rapid adaptation by means of HGT include the emergence, within time spans of a few years, of drug-resistant pathogenic strains[11] and of bacteria that are capable of breaking down newly introduced XENOBIOTIC compounds[12,13] (BOX 2).

An exhaustive analysis of the dynamics of evolutionary change in the TRYPTOPHAN OPERON, for example, shows that an impressive degree of operon reorganization has occurred in different bacteria, with little effect on essential tryptophan metabolic functions[14]. In general, the fitness advantages, if there are any, of the operon organizations observed in different bacteria are unknown, but two cases have been analysed that are pertinent to regulatory evolution. First, both tryptophan (Trp) transport and Trp synthesis rates vary over a much broader range in *Bacillus subtilis* than in related *Cyanobacter* species[14]. The difference is thought to relate to the differences in dynamic requirements imposed by their environments: *B. subtilis* lives in the soil where conditions can change rapidly, whereas *Cyanobacter* lives in aquatic environments, which are presumably more stable. Second, *Buchnera aphidicola*, an ENDOSYMBIONT that must overproduce Trp to meet the needs of its host, has reorganized the enzymes for the rate-limiting first step of Trp synthesis to a plasmid, which results in a 16-fold amplification of that step[14]. These cases illustrate how the long-term optimization of organisms that are now in equilibrium with their average environment can include a rearrangement of the control system and the genome to match the needs of their individual fitness strategies.

FUNCTIONAL GENOMICS
The use of genome-wide or system-wide experimental approaches to assess gene function. It also refers to the analysis of gene function within the context of the overall design and behaviour of the organism.

CIRCUIT MOTIFS
Elements of circuit organization that are found repeatedly in regulatory circuits of different organisms and even in different regulatory subcircuits in the same organism.

Box 1 | New data sources and experimental techniques

Many new data sources and new types of experimental techniques relating to bacterial evolution have become available in recent years. These include:

- Genome-wide studies that have identified the organization of bacterial regulatory networks comprising hundreds of genes[23,34].
- Cross-genomic analyses stimulated by nearly 150 completed microbial genome sequences[31,38,57].
- Bioinformatic, experimental and theoretical studies of genome organization showing pervasive and persistent HORIZONTAL GENE TRANSFER (HGT) and genome reorganization[3,5,7,58,59].
- Experimental studies of bacterial evolution that have tracked the molecular-level adaptive changes that occur over tens of thousands of generations[60].
- *In silico* experiments to test evolutionary theories[42].

The evolutionary optimization of a bacterial species or of a specialized strain necessarily applies to the system as a whole: components or subsystems cannot be truly optimized in isolation. Each subsystem of an organism must make a fitness contribution that is related to the fitness strategy of the organism for succeeding in a particular environmental niche. Every particular fitness strategy requires specific capabilities — for example, specialized metabolic pathways, unique motility mechanisms, offence/defence methods against other organisms and a reproduction strategy. These capabilities are embodied in morphological adaptations, in metabolic adaptations and in the sensor–decision–response network that coordinates the whole system.

This 'system-level optimization' extends beyond individual species. If one species in a stable community achieves a notable improvement in fitness through mutation or HGT, it causes a wholesale reoptimization of the local biosphere. Such events are random, rare and unpredictable, but are of wide significance when they occur. One example is the appearance in 1992 of *Vibrio cholerae* O139, a new virulent strain, in Madras (India), apparently as the result of a HGT event[15]. Within a year, the resulting cholera epidemic spread across the Indian subcontinent, affecting millions of people.

HGT facilitates bacterial adaptation. It has long been speculated that HGT accelerates bacterial evolution by the sporadic introduction of novel gene repertoires. Experimental investigations of this hypothesis were relatively infrequent until an analysis of whole genome sequences showed that HGT occurs frequently and ubiquitously in many bacteria (for example, HGT is arguably the source of 18% of the *E. coli* genome[16]). Furthermore, the role of transposons and INTEGRONS in interspecies transfer of antibiotic-resistance genes is well known[11,17].

Many examples support the role of HGT in facilitating evolutionary innovation and adaptation (BOX 2). By rapidly introducing new genes into existing genomes, HGT circumvents the slow process of creating novel genes and accelerates genome innovation. HGT is more common between organisms that share similar characteristics — in particular, genome size, genome G+C composition, carbon utilization and oxygen tolerance[18].

Although some of the mechanisms by which bacterial cells acquire foreign DNA or experience chromosome rearrangement involve external agents (such as phage infection and CONJUGATIVE TRANSPOSONS[19]), others are innate features of bacterial genetic design. For example, *Bacillus subtilis* has a complex control system[20] for activating the competence pathway for DNA uptake; this indicates that the mechanism that confers competence for DNA importation must be considered to be part of the larger 'system' for genetic adaptation.

Origin of modularity in regulatory networks

A 'modular' organization of cellular functions appears in spatial, temporal, chemical and genetic contexts[21]. Although there is no general agreement on the definition of a regulatory module[22], here we define modules as groups of proteins that work together to execute a function (for example, a metabolic pathway or the

TWO-COMPONENT SYSTEMS
Signal-transduction systems that enable bacteria to regulate cellular functions in response to changing environmental conditions. They are composed of a histidine kinase sensor protein and a response regulator that frequently acts as a transcription factor.

KINETIC PARAMETERS
The rate constants of chemical reactions that describe how fast the reaction takes place.

ARCHAEBACTERIA
An ancient kingdom of unicellular microorganisms that are phylogenetically distinct from bacteria and eukaryotes. They are often found in extreme environments, such as near deep-sea vents.

XENOBIOTIC
A compound that is foreign to biological systems, often referring to human-made compounds that are resistant to biodegradation.

TRYPTOPHAN OPERON
The group of genes that control the biosynthesis of tryptophan.

ENDOSYMBIONT
An organism that grows inside another organism. The relationship can be either mutualistic (both species benefit) or commensalistic (one species benefits, whereas the other is not affected).

HORIZONTAL GENE TRANSFER
The transfer of genetic material among cells that belong to different strains, species or genera.

Box 2 | Examples of the role of HGT in facilitating evolutionary innovation and adaptation

- Analyses of how bacterial communities have adapted to the recent introduction of xenobiotic compounds revealed the importance of horizontal gene transfer (HGT) by mobile genetic elements in the worldwide spread of catabolic pathways and in the formation of novel pathways[13]. Four findings are of particular interest[13]:

  Evolutionarily related catabolic genes are found in gene clusters in bacteria originating from widely separated locations.

  The phylogeny of the catabolic genes is not congruent with that of the 16S rRNA GENES of the corresponding host.

  Genes for the degradation of organic pollutants are often associated with mobile genetic elements such as plasmids and transposons.

  Evolutionarily related catabolic genes and entire gene modules are involved in the degradation of different, but structurally similar, xenobiotic compounds.
- Analysis of GENE CASSETTES identified in environmental soil samples found an eclectic collection of genes and noncoding DNA. This indicates that there is a vast library of disparate genes (the 'metagenome') available in the bacterial biosphere that can be imported into cells and integrated into the host regulatory network to yield new phenotypes[10].
- The *Salmonella* PhoP–PhoQ system senses environmental Mg^{2+} to determine whether the bacterium is inside a host cell and, if so, to activate the *mgtC* virulence locus via PhoP signalling. This pathway facilitates intracellular survival as well as mediating other physiological responses to low Mg (REF. 61). All PhoP-regulated genes that mediate *Salmonella* virulence have been acquired by HGT[61]. These externally acquired virulence gene cassettes were then integrated into a pre-existing *Salmonella* regulatory system so that they are activated exactly as needed to achieve host infection.

chemotaxis control network) or that create a multiprotein machine (for example, the replisome that replicates the chromosome or the complicated flagellum structure). Numerous gene-expression studies indicate that coherent sets of genes in many organisms are regulated together as a unit. For example, in the *C. crescentus* cell cycle, functional groups are not only co-regulated, but are turned on just when they are needed[23]; the regulation of many genes in the yeast *Saccharomyces cerevisiae* are similarly organized[24]. We will consider how modularity relates to the organization of the regulatory network and how the fitness advantages of a modular regulatory organization might lead to selection for modularity.

Several explanations for the emergence of modularity have been proposed. Groups of proteins that perform a useful function might simply tend to survive the scrambling and selection processes of bacterial adaptation cited above and be 'tuned up' by successive point mutations. These optimized collections of proteins would then survive to be dubbed 'modules'. Another proposed idea is that a modular regulatory organization has superior EVOLVABILITY. This is because selection on point mutations within modular functions can produce localized adaptive responses while having limited impact on the overall fitness of the organism in a target niche[25]. In addition, the modular organization of the regulatory circuitry enhances evolvability, because a simple change in the wiring of the regulatory circuitry can cause large changes in the organism's response to a signal. For example, a mutation in the promoter of a master regulator gene that changes the regulatory protein controlling its activity could introduce radical changes in either the timing of expression of the master regulator or the conditions leading to its expression. This would in turn change the pattern of expression of many downstream genes regulated by the master regulator. On rare occasions, such regulatory circuitry mutations will lead to significant enhancements to the fitness of an organism in the current niche or enable entry into a new niche; otherwise, of course, they will be removed from the population by selection.

Another reason for the emergence of modularity could be that modular organization facilitates the transfer of useful complex functions between organisms by HGT (a drug-resistance cassette, for example). The fact that beneficial gene collections are co-localized so that they can all be transferred together between species has been thought to benefit the genes themselves, an idea known as the 'selfish operon model'[26]. Cumulatively, the evidence described above indicates that HGT and modular genome organization are mutually reinforcing phenomena and that they have co-evolved.

Finally, modular organization simplifies the circuitry that is needed for complex responses. For example, the logic whereby incoming environmental signals determine whether to activate a complex pathway can be focused at the promoter site of a master regulator. If the (combinational) signal conditions that are appropriate for activating the master regulator are satisfied, then the downstream proteins in the regulon it controls will be activated. (This advantage is similar to the advantages of organizing computer programs into subroutine structures.)

Although the independent evolution of complex functions (for example, several vitamin-B6-dependent enzymes[27] or compounds for arsenic resistance[28]) is widely observed, the more complex the functionality, the longer independent reinvention will take. A quicker solution to evolve complex functionalities might therefore be to move pre-existing modular functions (such as drug-resistance cassettes) between bacterial species by HGT. The stressed bacterium that by chance finds itself in receipt of a ready-made solution to its difficulties will be advantaged, and, on rare occasions, the imported functional complex of genes will enable domination of a niche so that both the imported modular complex and its new host lineage survive.

Evolution of the regulatory circuitry

As discussed above, large sections of DNA containing many genes can be imported into bacterial cells and these imported genes might even comprise potentially functional circuits or modules. For the cell to benefit, the newly acquired genes have to be successfully integrated into the cellular regulatory system so that they are turned on and off at appropriate times. Interestingly, this ability to adapt potentially functional, but unused, modules to productive use in the cell has been shown to occur spontaneously. In experiments with cells constructed without an essential metabolic pathway, but containing a promoterless rescue operon, mutations in the upstream operator region of the rescue operon readily generated active promoter sites and created viable mutant strains[29,30].

The idea that the 'wiring' of a cell's regulatory networks is as susceptible to change as its genome organization is reinforced by an extensive analysis of the domain architecture of *E. coli* transcription factors[31]. This study showed that the many differentiated classes of transcription factors have evolved by associating DNA-binding domains with different regulatory domains through extensive recombination and by the widespread duplication of particular architectures followed by adaptive divergence (this allows them to bind regulatory regions for distinct genes or operons). In addition, an analysis of genetic circuit motifs in *E. coli* and in *S. cerevisiae* found that duplicate regulatory genes are randomly distributed across different types of gene circuits, implying that duplicated transcriptional regulators can readily evolve new interactions[32].

Conservation of internal organization of regulatory circuits. The interfaces within conserved modular functions is frequently more highly conserved among organisms than the interfaces with the regulatory network of the different organisms. For example, the internal organization and molecular structure of the complex machinery of the bacterial flagellum are more similar in distantly related bacteria than is the top-level regulatory interface that determines when and where the flagella are constructed and even their function within the cell[33]. The

INTEGRON
A genetic unit that, among others, encodes proteins that splice gene cassettes into chromosomes, where the cassettes can become functional.

CONJUGATIVE TRANSPOSONS
Discrete DNA elements that can transfer themselves from donor to recipient while the two are in direct physical contact. Their broad host range makes them important in horizontal transfer and bacterial evolution.

EVOLVABILITY
The ability of random genetic variation to produce phenotypic changes that can increase fitness (intrinsic evolvability) or the ability of a population to respond to selection (extrinsic evolvability).

16S rRNA GENES
Genes that are transcribed into the 16S rRNA molecule, a major component of the bacterial small ribosomal subunit. The strong sequence conservation of this molecule makes it ideal for detecting large evolutionary distances between two organisms.

GENE CASSETTES
Small mobile DNA elements that typically consist of a promoterless open reading frame and a recombination site. Gene cassettes are ubiquitous in environmental DNA samples.

α-PROTEOBACTERIA
A class of primarily oligotrophic bacteria within the proteobacteria that have high morphological and ecological diversity.

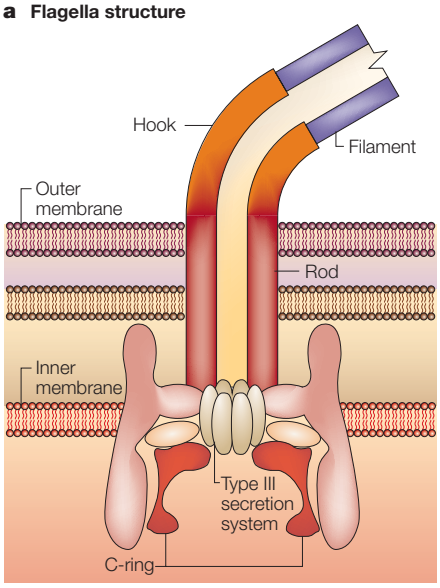
examples in FIG. 2 illustrate the adaptability of bacterial regulatory networks by showing the distinctly different positioning and means of control of flagella in three different bacteria. This adaptability enables opportunistic use of an essentially conserved flagellar function to enhance the fitness of each organism in its particular environmental niche.

Another perspective on the flexibility of regulatory networks comes from considering how the functions of conserved regulatory proteins and conserved regulatory motifs differ among bacterial species. One example of this is the CtrA protein, a DNA-binding response regulator that controls many cell-cycle functions in *Caulobacter crescentus*[23,34]. The CtrA protein is evolutionary conserved, as CtrA homologues are found in several α-PROTEOBACTERIA: *Sinorhizobium meliloti, B. abortus, Rhodobacter capsulatus, Agrobacterium tumefaciens* and *Rickettsia prowazekii*. There is also intriguing evidence that several elements of the complex regulatory circuitry that controls the timing of CtrA expression in *C. crescentus* might also be conserved in *S. meliloti* and *B. abortus*. The *Brucella abortus* homologue of *C. crescentus* CtrA is also an essential master regulator, but it controls a different portfolio of functions from those in *C. crescentus*[35] (BOX 3). This CtrA case shows the conservation of key internal design elements of a complex cell-cycle regulatory subsystem, accompanied by wholesale changes in the functions that the subsystem regulates to meet the control needs of each different organism.

A similar example of a regulatory subsystem that has been conserved together with changes in its cellular function is the chemotaxis control system that regulates the operation of flagella in many bacteria and that has, interestingly, been adapted to carry out different functions in *Myxococcus xanthus*. *M. xanthus* has several distinct regulatory subsystems, each comprised of proteins that are homologous to chemotaxis proteins in enteric bacteria. Two of these subsystem/circuits, involving the *dif* and *frz* genes, respectively, control coordinated social motility and the chemotactic capability that directs cells into aggregation centres[36]. However, motility in *M. xanthus* involves a gliding mechanism rather than chemotaxis using flagella. A third *M. xanthus* chemotaxis-derived system, involving the *che3* gene cluster, has nothing to do with motility, but instead controls developmental gene expression by regulating a σ_{54} TRANSCRIPTIONAL ACTIVATOR[37].

Two-component systems (see FIG. 1) are regulatory subsystems that provide a versatile means of signalling between events that occur at two different positions in a cell. They are used in pathways that detect and respond to environmental signals and also in other situations in which information or status at one place in the cell must be rapidly transmitted to cause an action elsewhere. This is a widespread problem within cells, so it is not surprising that this signalling component has been adapted for use in many applications. Cross-genomic analysis indicates that two-component systems emerged and became widespread during early bacterial evolution through HGT[38]. The adaptive link between bacterial two-component signalling and new cellular applications apparently involves the rapid evolution of signalling domains while conserving the protein–protein and protein–DNA interfaces in the internal phospho-relay reaction of the link. A high degree of variation has been found in the sensor and effector domains of two-component-system proteins among *Bacillus subtilis*, *B. halodurans*, *B. anthracis* and *B. stearothermophilus*. By contrast, the protein–protein phospho-reaction domains were much more highly conserved[39]. This is another example of a useful regulatory element (a two-protein mechanism for fast point-to-point signalling) that has maintained its internal interfaces while showing much plasticity in its interconnections to the different regulatory networks of its host.

a **Flagella structure**



Common elements of flagella design

- Overall design
- Assembly strategy, after gene activation
- Hierarchical, multi-level cascaded regulation of assembly involving three different promoter classes
- Driven by proton pumps that drive the C-ring

b

	Organization	Promoter	Master regulator
Caulobacter	Polar monotrichous	σ_{73}	CtrA
Vibrio	Polar monotrichous	σ_{54}	FlaK
Salmonella	Peritrichous (6-8 flagella)	σ_{70}	FlhDC

Polar monotrichous flagellum

Peritrichous flagella

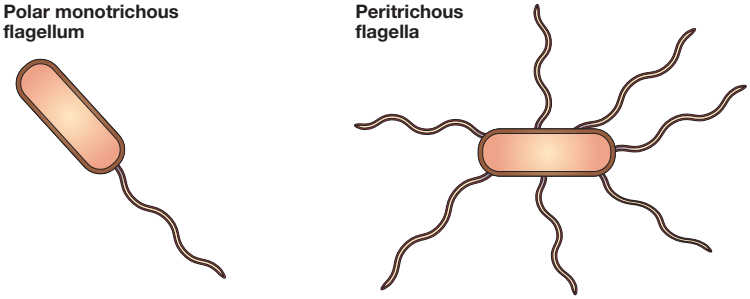


Figure 2 | **Flagella organization in *Caulobacter*, *Vibrio* and *Salmonella*.** The overall design and assembly processes of flagella are much more highly conserved than are the control interfaces that integrate the flagella into their respective host species or the number and distribution of flagella on the cell surface. **a** | The structure, the organization of assembly, and the motive power source of the individual flagella in all three bacteria are essentially the same. The diagram on the left illustrates the common structure of flagella. **b** | However, differences exist among the three species in surface organization of flagella and top-level regulation. *Caulobacter* and *Vibrio* have one polar flagellum (that is, have monotrichous flagella); *Salmonella* has many flagella distributed over the cell surface (it has peritrichous flagella). Each bacterium activates flagella construction using a different class of promoter and a different master regulator protein. Adapted from REF. 33.

Box 3 | **Conserved CtrA functions in *Caulobacter crescentus* and *Brucella abortus*[35]**

The CtrA protein is a master regulator that is important in controlling cell-cycle progression in several α-proteobacteria. Intriguing data indicate that master regulators that control bacterial cell-cycle progression are conserved, but that the functions that they each control vary widely among species[35]. For example, comparison of the homologous CtrA proteins in *Caulobacter crescentus* and *Brucella abortus* shows many common features:

- **Both *C. crescentus* and *B. abortus* CtrA proteins are essential and regulate the expression of many other proteins.**
- **There is 81% homology between the two CtrA proteins.**
- **Both are activated by phosphorylation of a conserved aspartate residue.**
- ***C. crescentus* CtrA can bind to *B. abortus* CtrA binding sites.**
- **Both CtrA proteins are similarly regulated:**

  They have two promoter sites (but with different spacing).

  They autoregulate their own expression.

  Each organism regulates CtrA through the methylation of a *ctrA* promoter site by homologous CcrM METHYLTRANSFERASE proteins and, in both cases, the *ccrM* genes are regulated by CtrA.
- **In contrast to the similarity of the two *ctrA* genes and their regulation, the two genes have significantly different functions in the regulatory systems of the two organisms. For example:**

  A principal function of *C. crescentus* CtrA is to control the initiation of chromosome replication. This function is not performed by CtrA in *B. abortus*.

  Although CtrA proteins are master regulators in both organisms, the genes and functions that they control are quite different.

Emergence of complex regulatory networks. The high frequency of wiring rearrangements that have occurred in bacteria was also shown by analysing the extent of protein homologies across many small genetic circuit motifs in *E. coli* and in *S. cerevisiae*. These studies indicate that most motifs have arisen by CONVERGENT EVOLUTION and not by duplication of ancestral circuits[32]. Therefore, small circuit motifs can arise spontaneously and be preserved by selection. Regulatory circuits with higher functionality can be built up by combining a relatively small number of common circuit motifs in different ways[40,41] in the same manner that electronic circuits with complex behaviour are created by interconnecting simpler circuit elements. But how do evolutionary processes create the larger circuits that regulate most bacterial-cell responses and that are constructed from many smaller motifs? This is a long-standing question in evolutionary theory.

Recent *in silico* experiments by Lenski *et al.*[42] have addressed this question. These simulation experiments illustrate how the processes of mutation and selection can lead to evolution towards organisms of increasing functionality when the enhanced functionality confers a fitness advantage (BOX 4). Although the digital creatures in these experiments embody an abstract view of life as an information-processing phenomenon, it is highly informative to observe, in a laboratory microcosm, the ability of ever-more fit organisms to emerge while less fit variants disappear from the population.

In these experiments, complex functions originated spontaneously from combinations of simpler functions. This result is consistent with Darwin's hypothesis that complex features evolved by modification of existing structures and functions. In addition, in every case the evolutionary pathway by which the eventually successful organisms evolved the ability to perform complex functions always involved a highly improbable succession of mutation events. As in the organic world, all traces of large numbers of more probable (but less competitive) phenotypes were eliminated by the success of the improbable (but more competitive) winner. These experiments are particularly valuable as they show how straightforward evolutionary mechanisms of mutation and selection can produce steady increases in organism complexity without invoking 'intelligent design'[43].

What are bacterial 'species'?

So far, we have emphasized the plasticity of the bacterial genome and the selective pressures that favour regulatory circuit innovation. Given these pressures and the evidence that bacteria can, in fact, rapidly adapt to changing environmental conditions, it is perhaps surprising that common bacterial species are found in essentially the same form around the globe. There is no natural force holding the properties that are characteristic of a species together, other than the pressure of selection[44]. The selective pressure is not only exerted to maintain particular enzymes, pathways and structures (such as flagella, pili or secretion systems), but also to maintain the sensors, signalling networks and the decision-logic circuits that coordinate all processes of the cell to implement the fitness strategy of the species.

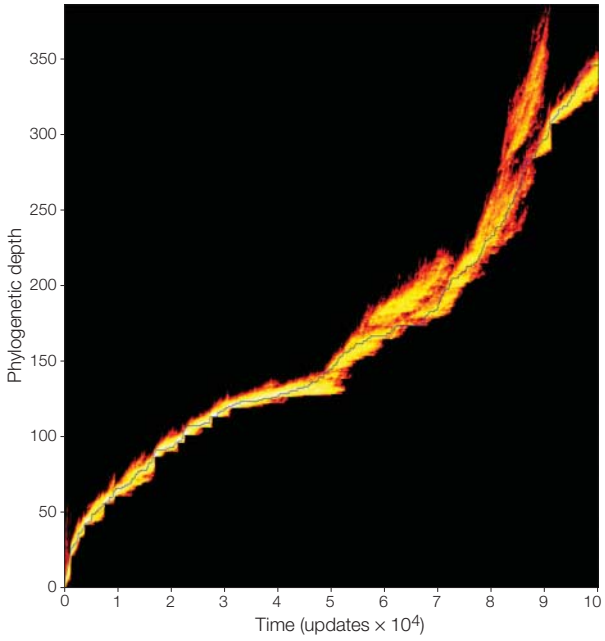
There is no consensus on the properties that define a bacterial species. The groupings of bacteria commonly called a species are ecotypes — that is, populations of organisms occupying the same ecological niche and whose divergence is preserved by natural selection[44]. Therefore, a bacterial 'species' represents a collection of capabilities that lead to dominance in a niche. Although the bacterial genome is extraordinarily changeable, the species identity can be maintained if the combined rate

METHYLTRANSFERASE
An enzyme that catalyses the addition of a methyl group, often to adenine or cytosine molecules in DNA.

σ_{54} TRANSCRIPTIONAL ACTIVATOR
Sigma factors are variable protein components of the bacterial RNA polymerase that have great influence on where the polymerase binds to DNA. The ability of σ_{54}, in particular, to initiate transcription by the polymerase might be affected by activators that bind at distant sites on the DNA.

CONVERGENT EVOLUTION
Two items are said to be the result of convergent evolution if their similarities arose by independent processes without common ancestry. This usually reflects evolutionary adaptation to similar environmental conditions.

Box 4 | **Evolutionary origin of complex features**



Computer models were used by Lenski *et al.*[42] to address the long-standing question of how complex organismal features are generated. Computer programs that self-replicate, mutate, compete and evolve provide a laboratory simulation of genetic evolution that demonstrates how the spontaneous emergence of increasing complexity can occur. The model genome of these digital creatures is a series of instructions for different computational operations. The set of instructions is cleverly designed so that various instruction sequences can perform other logical functions of varying complexity (such as NOT, AND, OR, and others, including EQUALS).

Mutations in the form of defective copy operations can either replace an instruction with another or perform an insertion or deletion operation that changes the genome's length. During the simulation, each organism executes its current instructions, leading to replication with a small probability of a mutation in the next generation. Each experiment starts with 3,600 identical organisms. At first, each organism has no functionality other than the ability to self-replicate and perform a single elemental logic function, a NAND (not-AND) operation on two numbers. The organism 'feeds' on a ration of 'single-instruction processing' units (SIPs). One SIP enables one instruction to be carried out. When, in a mutation event, an evolving creature acquires the ability to perform another function, it is rewarded with a larger ration of SIPs, so that it processes its instructions faster and therefore replicates sooner. The ability to perform more complex functions is rewarded with more SIPs than are simpler functions.

In this world, 'fitness' is defined as 'knowing' how to do more logic functions. Greater fitness produces greater success in reproductive competition. As the total population is constrained, a faster growing creature — that is, one that is more complex and harvests more SIPs — will eventually take over the population.

Full details of the evolution of the population can be stored and analysed. For example, the figure shows phylogenetic depth (the cumulative number of generations in which an organism's genotype differs from the initial genotype) versus time in one case-study population. Colours indicate the relative number of genotypes at any time (yellow equals more, red equals fewer). The blue line is the line of descent that led to the most abundant final genotype. One strength of these digital models is that predictions of evolutionary theory can be tested against the simulation results that are recorded in these stored population histories. The ability to run many such simulations with identical starting conditions provides an evolution laboratory that is not available in experiments with biological organisms.

Figure reproduced with permission from REF. 42 and kindly provided by C. Ofria, Michigan State University, USA.

of all genomic change processes is sufficiently slow that the portfolio of characteristics defining the species is maintained by selection[44].

For example, a comparison of two sequenced clinical isolates of *Helicobacter pylori* showed that 6% of all genes were specific to one strain and absent from the other[45]. Even greater variability is found in *E. coli* strains, in which comparison of natural isolates illustrates the heterogeneity of ecotypes. The genomes of 26 different human pathogenic, commensal, extraintestinal and laboratory *E. coli* isolates were heterogeneous; up to 10% of the open reading frames that are specific to *E. coli* K-12 were not detectable in the other strains, and genome sizes also differed by up to 1 Mb[46]. A comparison of 202 *E. coli* strains isolated from a geographically and species-diverse group of wild animals found high genetic diversity. Differences between geographical location and host taxonomic group of each animal were the primary causes of the genomic divergence[47].

The common theme in these studies is that the bacterial genome is a patchwork of laterally acquired genes inserted in a common backbone. The genes in the common backbone seem necessary to encode the core cellular processes. The variable component of the genome must confer specific selective advantages that allow the bacterium to thrive within particular hosts. In addition, genome organization seems to be of secondary importance to gene content, because the organization of genes either in the genome or as part of individual regulatory systems is not strongly conserved.

Regulation and environmental complexity

The complexity of an organism's regulatory circuitry is affected by the complexity of its environment. Regulatory complexity is an adaptation to the dynamic characteristics of the changing environmental stresses that are characteristic of an organism's target niche. Environmental changes can be regarded as random fluctuations over

some frequency range, frequently overlaid on a trend (such as global warming) or on top of periodic patterns (for example, seasons or day/night cycles). Enteric bacteria, soil bacteria and other free-living bacteria live in complex environments and have correspondingly complex sensor–response–control subsystems[48]. Surface-dwelling bacteria experience a broad distribution of timescales in the fluctuations in environmental parameters, such as broad annual temperature swings and the rapid changes in osmolarity that can occur within minutes following a sudden rainstorm. These bacteria have many environmental sensors and corresponding control circuits to invoke a wide variety of contingent responses. By contrast, obligate symbiotic bacteria live in a more constant host environment, and generally have both fewer genes and a simplified regulatory structure[49,50].

When there is a persistent environmental change or if an important fitness-enhancing innovation emerges in one of the co-occupants of a niche, then all occupants of the niche are stressed and begin to adapt towards a new collective equilibrium. For selective pressures that persist for many bacterial generations, mutation and HGT are suitable adaptive response mechanisms, but lethal stresses with onset time comparable to, or less than, bacterial generation times require a faster response mechanism. Pre-existing pathways in the genome that can be quickly activated by sensor–response control systems provide contingent responses for these emergencies. The HEAT-SHOCK RESPONSE and various metabolic-pathway activating responses are examples.

Particularly interesting cases, both practically and theoretically, arise when the information the bacteria can glean from its environment does not determine which of several possible responses is best. Bacteria, after all, can only sense what is happening at the present time and in their immediate environment. Depending on particular circumstances, several regulatory options are observed: *C. crescentus* responds to environmental xylose by activating several transporters and degradative exoenzymes; it seems that, in addition to activating xylose-specific pathways, *C. crescentus* also activates pathways that would enable it to metabolize environmental cellulose. This is thought to be a 'guess' on the part of *C. crescentus* that if xylose is present, other plant products are probably present as well[51]. In another well-studied case, a stochastic switch in the phage lambda regulatory network causes a small fraction of infected *E. coli* cells to enter the LYSOGENIC STATE so that these cells (and the embedded phage genome) survive the phage attack[52]. In this way, the phage population hedges against the possibility that the infection is so severe that all the *E. coli* cells in the vicinity will be killed, which would also lead to death of all the phages. Other bacterial regulatory systems implement even more complex strategies that can be analysed as hedging bets or competitive games[22].

HEAT-SHOCK RESPONSE
A mechanism that involves activation of many genes that cells use to maintain stability when subjected to thermal stress.

LYSOGENIC STATE
A phage integrated into a bacterial cell's chromosome is, in a latent form, called a 'lysogenic state'. Environmental stress can cause the lysogenic phage to leave the chromosome and produce infectious phage particles followed by bursting of the host cell.

Implications

Engineering of genetic circuits. The preceding sections have reviewed the adaptive mechanisms that affect bacterial regulatory circuitry and that have produced the complex bacterial regulatory systems that we now observe. As we have discussed, the continuous testing and optimization process that occurs in Nature ensures that the regulatory circuitry of each species is exquisitely adapted to the exigencies of living in its target niche. Biological engineers have recently started to construct artificial regulatory circuits in bacteria with the eventual objective of creating organisms with new behaviours[53]. However, even for simple functions such as switches and oscillators, the human-engineered circuits are much more noisy and unreliable than evolved regulatory circuits[54]. By contrast, natural 'engineering' by mutation and selection produces robust designs for controlling important cell functions, simply because the fitness penalty of unreliable cell operation clears unreliable designs from the population.

This situation has motivated an emerging 'design, then evolve' strategy for genetic circuit development[54–56]. With this approach, the engineers develop cells with an approximation of the desired behaviour and then seek to optimize the design by successive mutation–selection cycles. The central challenge in this approach to producing *in vivo* circuitry with novel and complex behaviour lies in the design of screens and penalties that cause the bacteria to evolve rapidly in the laboratory towards the desired functionality.

Our review of the evolution of bacterial regulatory circuits allows us to draw several conclusions that are relevant to the 'design then evolve' circuit design strategy. First, it seems that obtaining results reasonably quickly (that is, in weeks or months, rather than in decades or millennia) requires some combination of using mutator strains, facilitating HGT and genome shuffling, plus (ideally) tapping the resources of the 'metagenome'. Second, if we want the resulting evolved design to be robust — that is to operate predictably over a wide range of conditions — then it is necessary to select for the desired behaviour over this range of conditions and to impose strong penalties for behaviour that is inconsistent with the desired behaviour. Third, given the plasticity of the genetic wiring diagram, the starting circuit design is probably considerably less important to eventual success than the design of the selective screen. Generally, this 'mutate and select' approach is significantly different from most current engineering efforts that aim to design novel artificial behaviour into bacteria.

Combating antibiotic resistance. We have seen that bacteria have enormous capability to overcome obstacles to their invasion of fertile environmental niches by internal genome mutations and by exploiting the vast genetic resources within the bacterial biosphere by HGT. An extraordinarily important consequence has been the rapid emergence of drug-resistant bacteria. Now we know that it is a certainty that bacteria will develop resistance to any new drugs that we discover; the only issue is how long it will take and whether we can delay it. This is a public-health challenge that is not going to go away, and it will inevitably increase in severity with time.

What should be done to respond to this threat? Perhaps the mechanisms that bacteria use to import and

integrate foreign DNA should be ancillary targets for inclusion in combined drug formulations to reduce the effectiveness of this adaptation mechanism. Continued expansion of the number of bacterial sequences available is one of our most effective investments for understanding bacterial evolutionary dynamics and the mechanisms of bacterial adaptation. In this regard, it is not only important to continue sequencing a broad sampling of microbial species, but also to sequence many isolates of a few model systems. This will facilitate detailed cross-genomic analysis and rapid progress in understanding of bacterial adaptive mechanisms.

Conclusions

The bacterial kingdom provides many examples of the remarkable adaptability that results from evolutionary selection. These remarkable organisms have a wonderful diversity of behaviours, morphologies and natural habitats, and their small size and rapid growth rates allow direct observation of evolutionary processes in the laboratory. These experiments, together with the availability of many bacterial sequences and new genomic analysis techniques, produce new and surprising discoveries every year that deepen our understanding of evolutionary mechanisms.

1. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151 (2003).
2. Carroll, S. B. Genetics and the making of *Homo sapiens*. *Nature* **422**, 849–857 (2003).
3. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
4. Brown, J. R. Ancient horizontal gene transfer. *Nature Rev. Genet.* **4**, 121–132 (2003).
A thorough overview of the methods for detecting HGT, its implications for taxonomy, and the controversies that exist in the field.
5. Omelchenko, M. V., Makarova, K. S., Wolf, Y. I., Rogozin, I. B. & Koonin, E. V. Evolution of mosaic operons by horizontal gene transfer and gene displacement *in situ*. *Genome Biol.* **4**, R55 (2003).
6. Hudson, R. E., Bergthorsson, U. & Ochman, H. Transcription increases multiple spontaneous point mutations in *Salmonella enterica*. *Nucl. Acids Res.* **31**, 4517–4522 (2003).
7. Kurland, C. G., Canback, B. & Berg, O. G. Horizontal gene transfer: a critical view. *Proc. Natl Acad. Sci. USA* **100**, 9658–9662 (2003).
Presents a critique of the 'rampant HGT' hypothesis and stresses the continuing usefulness of tree-like phylogenies.
8. Kunin, V. & Ouzounis, C. A. The balance of driving forces during genome evolution in prokaryotes. *Genome Res.* **13**, 1589–1594 (2003).
9. Chothia, C., Gough, J., Vogel, C. & Teichmann, S. A. Evolution of the protein repertoire. *Science* **300**, 1701–1703 (2003).
10. Holmes, A. J. *et al.* The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environ. Microbiol.* **5**, 383–394 (2003).
11. Rice, L. B. Association of different mobile elements to generate novel integrative elements. *Cell. Mol. Life Sci.* **59**, 2023–2032 (2002).
12. Muller, T. A., Werlen, C., Spain, J. & Van Der Meer, J. R. Evolution of a chlorobenzene degradative pathway among bacteria in a contaminated groundwater mediated by a genomic island in *Ralstonia*. *Environ. Microbiol.* **5**, 163–173 (2003).
13. Top, E. M. & Springael, D. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr. Opin. Biotechnol.* **14**, 262–269 (2003).
An experimental demonstration of how novel catabolic pathways have been created through forward evolution in a high HGT environment.
14. Xie, G., Keyhani, N. O., Bonner, C. A. & Jensen, R. A. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol. Mol. Biol. Rev.* **67**, 303–342 (2003).
Detailed, thorough paper that traces the molecular evolution of the *trp* operon through several bacterial species.
15. Garg, P. *et al.* Molecular epidemiology of O139 *Vibrio cholerae*: mutation, lateral gene transfer, and founder flush. *Emerg. Infect. Dis.* **9**, 810–814 (2003).
16. Lawrence, J. G. & Ochman, H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl Acad. Sci. USA* **95**, 9413–9417 (1998).
17. Hacker, J., Hentschel, U. & Dobrindt, U. Prokaryotic chromosomes and disease. *Science* **301**, 790–793 (2003).
18. Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**, 1598–1602 (2003).
19. Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.* **46**, 601–610 (2002).
20. Hamoen, L. W., Venema, G. & Kuipers, O. P. Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* **149**, 9–17 (2003).
21. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
22. Wolf, D. M. & Arkin, A. P. Motifs, modules and games in bacteria. *Curr. Opin. Microbiol.* **6**, 125–134 (2003).
23. Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M. & Shapiro, L. Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290**, 2144–2148 (2000).
24. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genet.* **34**, 166–176 (2003).
25. Hansen, T. F. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *Biosystems* **69**, 83–94 (2003).
Provides quantitative definitions of evolvability and modularity in population genetics terms.
26. Lawrence, J. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr. Opin. Genet. Dev.* **9**, 642–648 (1999).
27. Paiardini, A., Contestabile, R., D'Aguanno, S., Pascarella, S. & Bossa, F. Threonine aldolase and alanine racemase: novel examples of convergent evolution in the superfamily of vitamin B6-dependent enzymes. *Biochim. Biophys. Acta* **1647**, 214–219 (2003).
28. Mukhopadhyay, R., Rosen, B. P., Phung le, T. & Silver, S. Microbial arsenic: from geocycles to genes and enzymes. *FEMS Microbiol. Rev.* **26**, 311–325 (2002).
29. Dabizzi, S., Ammannato, S. & Fani, R. Expression of horizontally transferred gene clusters: activation by promoter-generating mutations. *Res. Microbiol.* **152**, 539–549 (2001).
Shows that horizontally transferred elements can be incorporated into an extant cellular circuitry with comparatively few mutations.
30. Kasak, L., Horak, R. & Kivisaar, M. Promoter-creating mutations in *Pseudomonas putida*: a model system for the study of mutation in starving bacteria. *Proc. Natl Acad. Sci. USA* **94**, 3134–3139 (1997).
31. Madan Babu, M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucl. Acids Res.* **31**, 1234–1244 (2003).
This analysis classifies *E. coli* transcription factors on the basis of structural homology and identifies global regulators that control cellular circuitry.
32. Conant, G. C. & Wagner, A. Convergent evolution of gene circuits. *Nature Genet.* **34**, 264–266 (2003).
33. Aldridge, P. & Hughes, K. T. Regulation of flagellar assembly. *Curr. Opin. Microbiol.* **5**, 160–165 (2002).
34. Laub, M. T., Chen, S. L., Shapiro, L. & McAdams, H. H. Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc. Natl Acad. Sci. USA* **2**, 4632–4637 (2002).
35. Bellefontaine, A. F. *et al.* Plasticity of a transcriptional regulation network among alpha-proteobacteria is supported by the identification of CtrA targets in *Brucella abortus*. *Mol. Microbiol.* **43**, 945–960 (2002).
36. Shi, W., Yang, Z., Sun, H., Lancero, H. & Tong, L. Phenotypic analyses of *frz* and *dif* double mutants of *Myxococcus xanthus*. *FEMS Microbiol. Lett.* **192**, 211–215 (2000).
37. Kirby, J. R. & Zusman, D. R. Chemosensory regulation of developmental gene expression in *Myxococcus xanthus*. *Proc. Natl Acad. Sci. USA* **100**, 2008–2013 (2003).
38. Galperin, M. Y., Nikolskaya, A. N. & Koonin, E. V. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett.* **202**, 11–21 (2001).
39. Stephenson, K. & Hoch, J. A. Evolution of signalling in the sporulation phosphorelay. *Mol. Microbiol.* **46**, 297–304 (2002).
40. Tyson, J. J., Chen, K. C. & Novak, B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231 (2003).
41. 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).
42. Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003).
Describes simulation experiments that model the evolution of digital 'creatures' and illustrates how an increasingly complex functionality can arise by mutation and selection.
43. Pennock, R. T. Creationism and intelligent design. *Annu. Rev. Genom. Hum. Genet.* **4**, 143–163 (2003).
44. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487 (2002).
Reviews the history of bacterial systematics and presents a definition of species that are based on molecular metrics of divergence and cohesion.
45. Joyce, E. A., Chan, K., Salama, N. R. & Falkow, S. Redefining bacterial populations: a post-genomic reformation. *Nature Rev. Genet.* **3**, 462–473 (2002).
46. Dobrindt, U. *et al.* Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J. Bacteriol.* **185**, 1831–1840 (2003).
47. Souza, V., Rocha, M., Valera, A. & Eguiarte, L. E. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. *Appl. Environ. Microbiol.* **65**, 3373–3385 (1999).
48. Cases, I., de Lorenzo, V. & Ouzounis, C. A. Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol.* **11**, 248–253 (2003).
49. Wilcox, J. L., Dunbar, H. E., Wolfinger, R. D. & Moran, N. A. Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol. Microbiol.* **48**, 1491–1500 (2003).
50. van Ham, R. C. *et al.* Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl Acad. Sci. USA* **100**, 581–586 (2003).
51. Hottes, A. K. *et al.* Transcriptional profiling of *Caulobacter crescentus* during growth on complex and minimal media. *J. Bacteriol.* (in the press).
52. Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda infected *E. coli* cells. *Genetics* **149**, 1633–1648 (1998).
53. Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
54. McAdams, H. H. & Arkin, A. Towards a circuit engineering discipline. *Curr. Biol.* **10**, R318–R320 (2000).
55. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99**, 16587–16591 (2002).
56. Hasty, J. Design then mutate. *Proc. Natl Acad. Sci. USA* **99**, 16516–16518 (2002).

57. National Center for Biotechnology Information (NCBI). *Completed microbial genomes* [online], (cited 16 Jan 04), <http://www.ncbi.nlm.nih.gov/genomes/MICROBES/Complete.html>.
58. Lawrence, J. G. Selfish operons and speciation by gene transfer. *Trends Microbiol.* **5**, 355–359 (1997).
59. Doolittle, W. F. *et al.* How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos. Trans. R. Soc. Lond. B* **358**, 39–57; discussion 57–58 (2003).
60. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Rev. Genet.* **4**, 457–469 (2003). **A clear description of bacterial evolutionary experiments and the insights derived from the field.**
61. Groisman, E. A. The pleiotropic two-component regulatory system PhoP-PhoQ. *J. Bacteriol.* **183**, 1835–1842 (2001).
62. Ames, P., Studdert, C. A., Reiser, R. H. & Parkinson, J. S. Collaborative signaling by mixed chemoreceptor teams in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 7060–7065 (2002).

Acknowledgements

H.H.M. is supported by the Office of Science of the U.S. Department of Energy (DOE), by the Office of Naval Research and by the DARPA Defense Sciences Office. B.S. is supported by a National Defense Science and Engineering fellowship from the Army Research Office. A.P.A. is supported by the Office of Science of the U.S. DOE, the DARPA Information Processing Technology Office and by the Howard Hughes Medical Institute. We thank L. Shapiro and R. Losick for helpful comments on the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

FURTHER INFORMATION
Harvard University, Dept of Molecular & Cellular Biology's evolution links:
http://www.mcb.harvard.edu/BioLinks/Evolution.html
HGT database: http://www.fut.es/~debb/HGT/
Information about bacteria:
http://www.bact.wisc.edu/MicrotextBook/
Islander prokaryotic island database:
http://129.79.232.60/cgi-bin/islander/islander.cgi
Microbiology current issues: Antibiotic resistance:
http://www.microbeworld.org/htm/cissues/resist/resist_0.htm
Microsatellite analysis server:
http://210.212.212.7/MIC/index.html
Public Broadcasting Evolution site:
http://www.pbs.org/wgbh/evolution/
Access to this interactive links box is free online.