Current Biology 20, R735–R745, September 14, 2010 ©2010 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2010.06.031

Hierarchical Evolution of the Bacterial Sporulation Network

Review

Michiel J.L. de Hoon[1,†], Patrick Eichenberger[2], and Dennis Vitkup[1,3]

Genome sequencing of multiple species makes it possible to understand the main principles behind the evolution of developmental regulatory networks. It is especially interesting to analyze the evolution of well-defined model systems in which conservation patterns can be directly correlated with the functional roles of various network components. Endospore formation (sporulation), extensively studied in *Bacillus subtilis*, is driven by such a model bacterial network of cellular development and differentiation. In this review, we analyze the evolution of the sporulation network in multiple endospore-forming bacteria. Importantly, the network evolution is not random but primarily follows the hierarchical organization and functional logic of the sporulation process. Specifically, the sporulation sigma factors and the master regulator of sporulation, Spo0A, are conserved in all considered spore-formers. The sequential activation of these global regulators is also strongly conserved. The feed-forward loops, which are likely used to fine-tune waves of gene expression within regulatory modules, show an intermediate level of conservation. These loops are less conserved than the sigma factors but significantly more than the structural sporulation genes, which form the lowest level in the functional and evolutionary hierarchy of the sporulation network. Interestingly, in spore-forming bacteria, gene regulation is more conserved than gene presence for sporulation genes, while the opposite is true for non-sporulation genes. The observed patterns suggest that, by understanding the functional organization of a developmental network in a model organism, it is possible to understand the logic behind the evolution of this network in multiple related species.

Introduction

Evolution is the main organizational principle of biological systems [1,2]. The emerging field of evolutionary systems biology [3,4] investigates structural and functional evolution of cellular networks. Instead of considering only the presence or absence of orthologous genes in sequenced organisms, evolutionary systems biology primarily focuses on changes in the relationships between genes and their products. A thriving area of evolutionary systems biology is the evolutionary biology of developmental networks. In metazoans, studies at the interface of evolution and development investigate the mechanistic changes leading to the organization and evolution of complex body plans. On the basis of careful analyses of several model systems, such as *Drosophila* and sea urchin, the essential role of regulatory interactions in the evolution of developmental processes has been firmly established [5,6].

Although prokaryotic organisms do not have a complex body plan, they can form multi-cellular structures, such as biofilms and fruiting bodies [7,8]. In addition, elaborate developmental processes have been characterized in many bacterial species. Endospore formation (sporulation) is the prime example of a complex bacterial developmental process. Sporulating bacteria undergo an intricate sequence of cell differentiation events leading to the formation of a highly resistant, dormant spore that can germinate when conditions improve. Initiation and progression of sporulation is controlled by a complex network of protein–protein and protein–DNA interactions, consisting of regulatory modules, signaling pathways, feed-forward network motifs, and post-translational regulation [8–10].

The sporulation process has been characterized in sufficient detail in the model organism *Bacillus subtilis* to enable fundamental evolutionary analyses from a functional perspective. Similar to developmental processes in higher organisms, bacterial sporulation is governed by a complex cascade of regulatory interactions that contains a strongly conserved regulatory kernel, i.e. core regulatory network [11]. Transcriptional regulation in the sporulation network is dominated by sigma factors — the subunit of the bacterial RNA polymerase holoenzyme that is responsible for recognizing promoter regions on the DNA [12]. The DNA-binding specificities of different sigma factors have been determined experimentally and the corresponding DNA-binding sites have been collected in DBTBS, the database of transcriptional regulation in *B. subtilis* [13].

The rapid increase in fully sequenced bacterial genomes allows us to understand the evolution of network regulation in a large number of diverged species. In this review, we first present an overview of the well-studied sequence of sporulation events in *B. subtilis*. Next, we describe the phylogenetic relationships of currently sequenced endospore-forming bacteria. We follow with a discussion of the evolution of the sporulation gene regulatory network and the properties affecting the evolvability of regulation. The functional characterization of a substantial fraction of sporulation genes in *B. subtilis* enables us to put the observed evolutionary patterns into the proper functional context. We also discuss the correlation between evolution of gene presence and regulation.

The Sporulation Process and Its Regulation in *B. subtilis*

The genetically competent, non-pathogenic soil bacterium *B. subtilis* is the prevalent model system for studies of sporulation. A significant amount of detailed molecular data has been gathered over the years to characterize the mechanism of endospore formation — in particular, the regulation of the different sporulation stages.

[1]Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA. [2]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA. [3]Department of Biomedical Informatics, Columbia University, New York, NY 10032, USA. [†]Currently at the RIKEN Omics Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan.
E-mail: pe19@nyu.edu (P.E.), vitkup@dbmi.columbia.edu (D.V.)

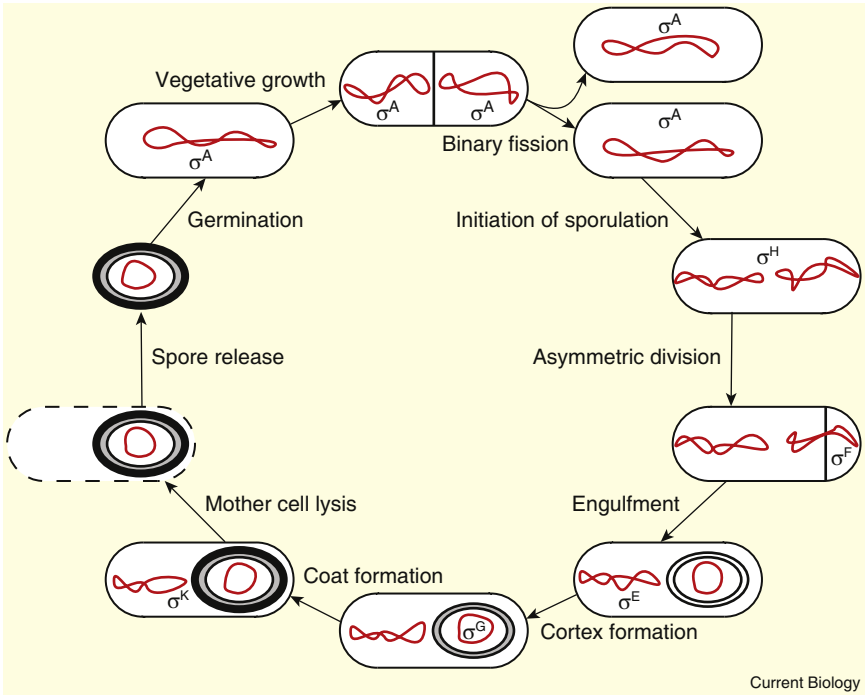

Figure 1. Morphological stages of the *B. subtilis* life cycle.

The temporal and compartment-specific activity of each sporulation sigma (σ) factor is indicated. During vegetative growth, cells divide by binary fission to generate two identical daughter cells. Sporulation is initiated in response to starvation. In the predivisional sporulating cell, the chromosomes (red) are oriented with their origin-proximal region anchored at the cell poles. During asymmetric division, two membrane-bounded compartments are generated: a small forespore and a large mother cell. After asymmetric division, the remainder of the forespore chromosome (i.e. the origin-distal region) is pulled into the forespore by translocation. Engulfment of the forespore by the mother cell results in the release of the forespore as a free protoplast in the mother cell. The cortex (composed of modified peptidoglycan, gray) is synthesized between the two membranes surrounding the forespore. The coat (black) is a complex structure made of at least 70 distinct proteins that assemble around the forespore surface. Following mother cell lysis, the mature spore is released into the environment. *B. subtilis* cells can remain in a dormant spore state for an extended period of time, but spores will germinate in response to the presence of small molecules (e.g. single amino acids, sugars or fragments of peptidoglycan) and resume vegetative growth.

Morphological Stages of Sporulation and Formation of Protective Structures

In rich medium, *B. subtilis* cells divide by binary fission approximately every 30 minutes. By contrast, deterioration of environmental conditions triggers sporulation, a developmental process that takes about 8 to 10 hours. Thus, endospore formation represents a formidable investment of time and energy and is considered to be a survival pathway of last resort, as *B. subtilis* cells only commit to sporulation after they failed to deal with starvation in other ways, such as cannibalism or establishment of a genetically competent state [14–16]. The successive morphological stages of sporulation have been defined using electron microscopy [17,18] (Figure 1). Sporulation begins with an asymmetric cell division and results in the generation of two cell types, a forespore (the smaller compartment, also called the prespore) and a mother cell. The two cells experience distinct fates, because the mother cell ultimately lyses by a programmed cell death mechanism, whereas the forespore matures as a spore. Shortly after asymmetric division, two parallel programs of gene expression are established in each compartment under the control of transcription factors that are activated in a cell-specific manner. In addition to regulatory interactions within the forespore and mother cell, precise inter-compartmental signaling is required to control the spatial and temporal progression of the developmental process.

Sporulation commences only after a round of DNA replication has been completed, in order to ensure that two chromosome copies are available in the predivisional cell [19]. The two chromosomes are oriented with their origin of replication anchored at one cell pole and their origin-distal region at mid-cell [20]. After asymmetric division, only about one-third of the forespore chromosome (i.e. the origin-proximal region) is captured in the small chamber of the dividing cell. A DNA translocase, SpoIIIE, located at the center of the polar septum, is necessary to pull the rest of this chromosome into the forespore [21–23]. The other chromosome is localized entirely inside the mother cell.

Following asymmetric division, the next morphological stage of sporulation is the engulfment of the forespore by the mother cell. This process is analogous to phagocytosis and is driven by mother cell proteins that facilitate membrane migration around the forespore by enzymatic removal of the peptidoglycan [24,25]. After completion of engulfment, the forespore, now entirely surrounded by its inner and outer membranes, is a free protoplast in the mother cell cytoplasm. Next, a series of protective structures is assembled around the spore core. The cortex, a modified peptidoglycan, is synthesized between the two forespore membranes [26]. Simultaneously, at least 70 individual coat proteins are synthesized in the mother cell to encase the spore in a multi-layered structure, with the crust as the outermost layer [27,28]. Finally, the mother cell lyses to release the mature spore.

Fully formed spores, recognized as the most resistant form of life on the planet [29], protect the bacterial genome against heat, desiccation, radiation, and oxidation. In addition, spore formation might be an efficient way to escape predation from higher organisms [30,31]. As soon as environmental conditions become favorable for vegetative growth, however, it is critical that *B. subtilis* quickly exits from the dormant state. This process is referred to as spore germination [32] and is triggered by the presence of nutrients in the environment. The nutrients are sensed by specific spore membrane receptors and, within minutes, the spore core rehydrates, the cortex is hydrolyzed, and the coat is shed. Ultimately, DNA replication is initiated and the first cell division soon follows.

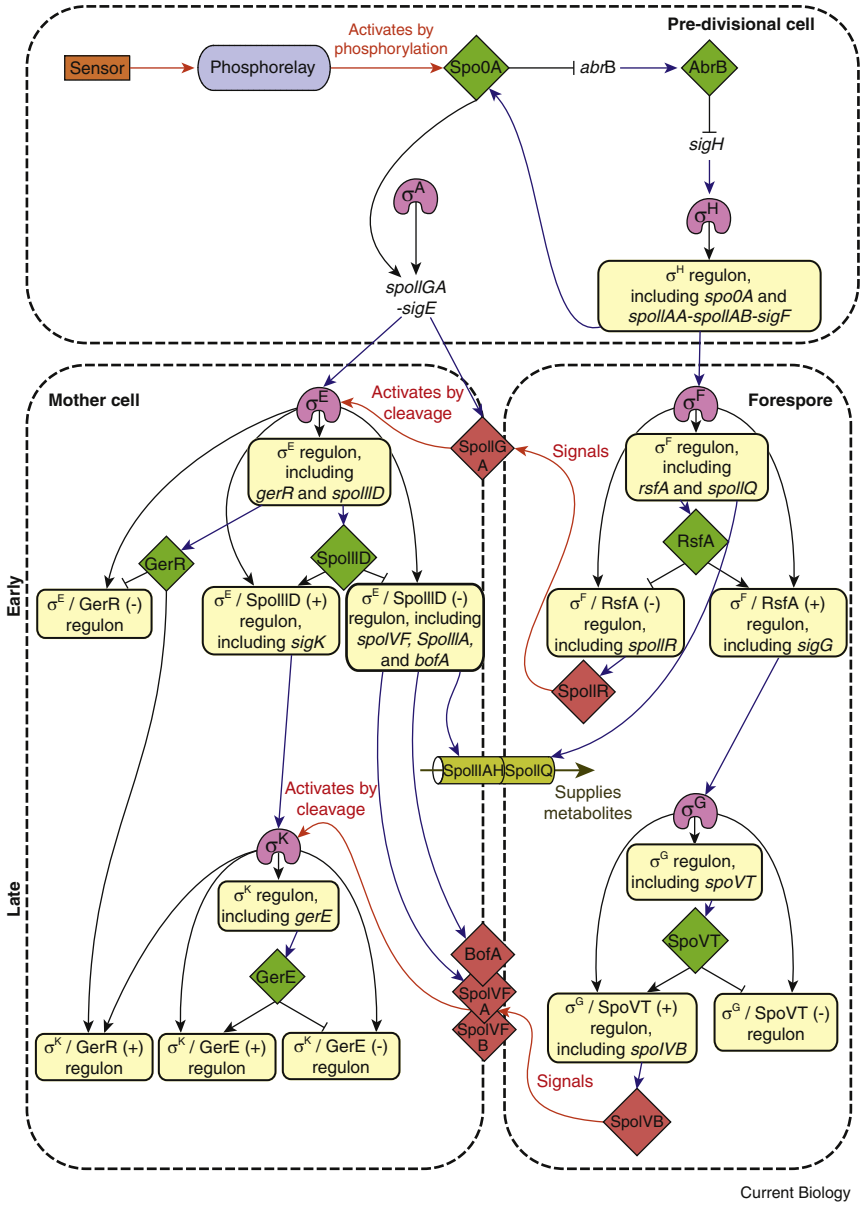

Figure 2. Modular architecture of the sporulation regulatory network in *B. subtilis*.

The temporal progression of sporulation is from the top to bottom. Each cellular compartment (pre-divisional cell, forespore and mother cell) is surrounded by dotted lines. Sigma factors are shown in pink, transcription factors in green, and regulons in yellow. Signaling proteins are shown in red, and the feeding tube in olive. Transcriptional regulation is indicated by black arrows, gene expression (protein synthesis) by blue arrows, and signaling pathways by red arrows. Coherent feed-forward loops are indicated by a plus sign, and incoherent feed-forward loops by a minus sign.

The Sporulation Transcriptional Regulatory Network in *B. subtilis*

The *B. subtilis* sporulation transcriptional regulatory network is shown schematically in Figure 2. An essential role in this network is played by sigma factors whose activity is restricted to specific compartments. The sporulation regulatory network is organized spatially and temporally into modules, with the sporulation-specific sigma factors σ^F, σ^E, σ^G, and σ^K directing gene expression in the early forespore, early mother cell, late forespore, and late mother cell, respectively [12]. The sigma factor σ^H regulates transcription in the predivisional cell [33]. Importantly, the primary sigma factor of *B. subtilis*, σ^A, remains active even during the late stages of sporulation [34].

The master regulator of sporulation, Spo0A~P, is not a sigma factor but a response regulator activated by a phosphorelay. Before initiation of sporulation, *spo0A* is transcribed at a low level by RNA polymerase containing σ^A. Sporulation is initiated by changes in environmental conditions that are sensed by a group of histidine kinases capable of auto-phosphorylation. The phosphate group is then sequentially transferred from the histidine kinases to the Spo0B and Spo0F proteins of the phosphorelay and finally to Spo0A [35]. Phosphorylation of Spo0A causes a change in the transcription of more than 500 genes, including at least 120 direct targets [36]. In the predivisional cell, σ^H increases transcription of *spo0A* using a sporulation-specific promoter. Importantly, Spo0A~P acts as an indirect positive regulator of the σ^H-encoding gene, *sigH*, by repressing transcription of its repressor AbrB. Interestingly, the Spo0A~P regulon is composed of low and high threshold genes [37], implying that Spo0A~P-dependent genes are activated with different kinetics. Following asymmetric division, Spo0A~P activity is predominantly observed in the mother cell [38].

The sigma factor σ^H also directs transcription of the *spoIIAA–spoIIAB–sigF* operon, encoding the early forespore-specific sigma factor σ^F, the anti-sigma factor SpoIIAB, and the anti-anti-sigma factor SpoIIAA. σ^F is the first compartment-specific regulator to be activated and directs transcription of about 50 genes [39,40], including *sigG*. The other transcription factor to be produced under the control of σ^F is RsfA, which activates some σ^F-transcribed genes and represses others, forming with σ^F coherent and incoherent feed-forward loops. Feed-forward loops are frequent topological motifs in cellular networks in which one gene regulates another gene, and together they also regulate a downstream gene [41]. Feed-forward loops perform various functional roles, including noise filtering, fine-tuning of expression timing, response acceleration, and pulse generation. Individual feed-forward loops in biological networks usually aggregate into larger motif generalizations [42], in which a feed-forward loop regulates many downstream genes. When RsfA acts as an activator, σ^F and RsfA form a coherent feed-forward loop, resulting in delayed and/or lengthened gene expression kinetics. On

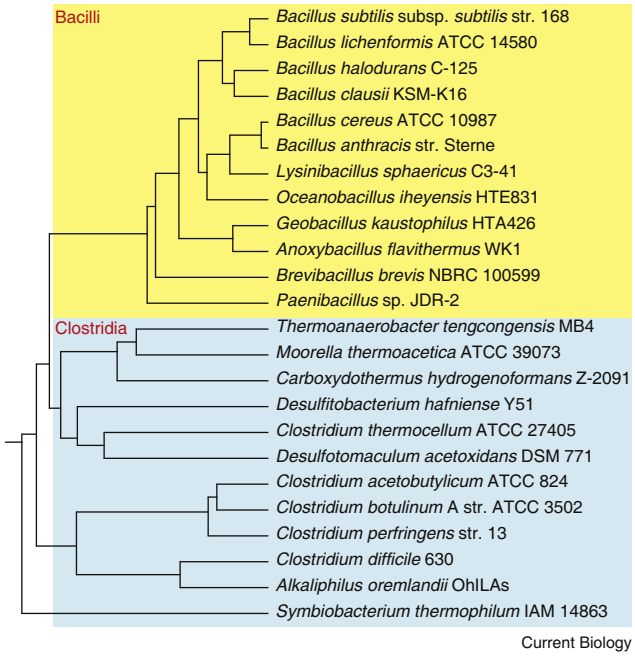

Figure 3. Phylogenetic tree of 24 representative endospore-forming species.

The tree was calculated using the PHYLIP program [97] based on 16S ribosomal RNA sequences from the Ribosomal Database Project [98]. *B. subtilis* belongs to the cluster (yellow) of 12 species of aerobic bacteria from the class Bacilli. The other cluster (cyan) includes 12 species from the anaerobic class Clostridia. The anaerobic Clostridia cluster can be subdivided further into two subclusters: one corresponding to the Clostridiaceae family, which includes *C. acetobutylicum*, the other including the Thermoanaerobacteraceae and Peptococcaceae families.

the other hand, when RsfA acts a repressor, σ^F and RsfA form an incoherent feed-forward loop, producing a short pulse of gene expression. The late-forespore-specific sigma factor, σ^G, regulates about 100 genes [39,40], and forms coherent and incoherent feed-forward loops with SpoVT, the final forespore-specific transcription factor in the cascade.

A similar logic is followed in the regulation of gene expression in the mother cell. Transcription of the genes specific to the early mother cell occurs under the control of σ^E, whose synthesis depends on a σ^A-specific promoter that is activated by Spo0A~P and located upstream of the *spoIIGA–sigE* operon. σ^E directs the transcription of about 270 genes [40,43], including those encoding the transcription factors SpoIIID and GerR. Both SpoIIID and GerR form feed-forward loops with σ^E. However, while SpoIIID activates some σ^E-dependent genes and represses others, GerR only seems to act as a repressor of σ^E-dependent genes and as an activator of σ^K-dependent genes [44]. Expression of *sigK*, encoding the last sigma factor in the sporulation cascade, is dependent on both σ^E and SpoIIID [45]. σ^K regulates about 150 genes [40,43], including GerE, the final mother cell-specific transcription factor in the cascade. Like SpoIIID, GerE can act either as an activator or as a repressor of transcription depending on the promoter regulatory logic.

Non-Transcriptional Regulation and Inter-Module Communication

Although the lines of gene expression in the forespore and mother cell run in separate compartments, they are connected by intercellular signaling to ensure coordinated regulation. As the synthesis of both σ^F and σ^E is under the control of Spo0A~P, they are already present in the predivisional cell. However, these sigma factors are held inactive until asymmetric division is completed. Specifically, σ^F is sequestered in a complex with two molecules of the anti-sigma factor SpoIIAB [46], whereas σ^E is present in a pro-form (pro-σ^E) that contains an inhibitory pro-sequence of 27 amino acids [47–49]. The anti-anti-sigma factor SpoIIAA binds to the SpoIIAB$_2$–σ^F complex and induces σ^F release [50]. Importantly, in order to be able to reverse the inhibition, SpoIIAA needs to be in an unphosphorylated state. In the pre-divisional cell, SpoIIAB is the kinase that maintains the phosphorylation of SpoIIAA [51,52]. After completion of asymmetric division, the septum-associated phosphatase SpoIIE dephosphorylates SpoIIAA, facilitating σ^F release [53,54].

The processing of pro-σ^E in the mother cell requires SpoIIGA, a membrane-associated aspartate protease [47–49]. The proteolytic cleavage is triggered by SpoIIR, a signaling protein secreted from the forespore. The *spoIIR* gene is one of the first genes expressed in response to σ^F activation [55,56]. Correct timing of *spoIIR* expression, and consequently σ^E activation in the mother cell, is critical because a delay will result in a terminal phenotype, in which one forespore is formed at each pole of the sporulating cell, leaving the mother cell empty of DNA [57].

Activation of σ^G in the forespore depends on the eight proteins expressed from the σ^E-controlled *spoIIIA* operon in the mother cell. The last gene in the operon, *spoIIIAH*, encodes a protein similar to a component of a secretion system [58,59]. SpoIIIAH interacts directly with the forespore protein SpoIIQ [60] and these two proteins probably form a channel between the mother cell and the forespore [58]. The nature of the molecules that are transported through the channel is currently unclear. It is possible that the channel is used to import a specific regulatory protein from the mother cell in order to activate σ^G in the forespore, but the channel is also required for late-phase σ^F activity and maintenance of forespore integrity. Therefore, it has been suggested that the channel may in fact function as a feeding tube providing small metabolites to nurture the forespore [61,62].

Like σ^E, σ^K is synthesized in the mother cell as an inactive pro-σ^K form [45]. The proteolytic cleavage of pro-σ^K requires the membrane-embedded metalloprotease SpoIVFB [63,64], whose activity is modulated by two mother-cell membrane proteins SpoIVFA and BofA [65,66]. The *spoIVF* operon and *bofA* are transcribed from σ^E-dependent promoters [66]. The proteolysis of pro-σ^K is triggered by SpoIVB, a signaling protein transcribed under the control of σ^G in the forespore [67,68].

The Phylogeny of Endospore-Forming Bacteria

Endospore-forming bacteria belong to two classes of the Firmicutes phylum: the Bacilli (aerobic Firmicutes, shown in yellow in Figure 3) and the Clostridia (anaerobic Firmicutes, shown in cyan in Figure 3). Both classes can be divided into a number of orders, some with sporulating genera and some with non-sporulating genera. It is generally assumed that the common ancestor of all Bacilli and Clostridia was an endospore-forming organism, even though several genera (such as *Listeria*, *Staphylococcus*, *Streptococcus* and *Lactobacillus*) that presumably evolved from that common ancestor have lost the ability to sporulate. Consistent with this interpretation, loss of sporulation appears

to be advantageous for cells living in a fairly constant environment [69].

Endospore formation is an ancient process that appeared only once in the course of evolution and likely predated the rise in oxygen in the terrestrial atmosphere about 2.3 billion years ago [70]. Furthermore, given the large number of genes that are essential to endospore formation, it is unlikely that the ability to sporulate could have been gained by other phyla as a result of horizontal gene transfer. We note that certain bacterial species, such as species of *Streptomyces* or *Myxobacteria*, can form spores via processes other than endospore formation, but the corresponding sporulation mechanisms are unrelated and therefore will not be considered in this review. Recently, two species of Actinobacteria, *Mycobacterium marinum* and *M. bovis*, were described as endospore-forming species [71]. However, in the genomes of these species, we and others [72] could not find orthologs for most of the sporulation genes. Specifically, according to our analysis, all of the major regulatory genes of the sporulation network (genes shown in Figure 2) and 92% of all genes with primary function in sporulation — i.e. genes in the functional categories of regulation, sporulation, and spore coat — are missing in these organisms. The somewhat higher level of conservation in *M. marinum* and *M. bovis* (85% of genes missing) for all 307 genes regulated by the sporulation sigma factors can be explained by the high conservation (51% of genes missing) of metabolic genes with additional cellular roles beyond sporulation. These results demonstrate that it is highly unlikely that *M. marinum* and *M. bovis* are true endospore formers.

In phylogenetic analyses based on 16S ribosomal DNA sequences, the group of aerobic endospore-forming bacteria (historically, the *Bacillus* genus) has turned out to be quite heterogeneous [73]. Several clusters identified from these analyses have since been elevated to the level of genera, such as *Lysinibacillus*, *Paenibacillus*, *Brevibacillus* and *Geobacillus*. However, the majority of currently sequenced aerobic endospore formers belong to the redefined *Bacillus* genus, including *B. subtilis* [74]. The largest genomes (>5 Mb) belong to species of the *B. anthracis*/*B. cereus*/*B. thuringiensis* group, including several pathogenic species [75]. The smallest *Bacillus* genome sequenced to date (3.35 Mb) belongs to *B. coahuilensis*, a species isolated from the Chihuahan desert [76].

Similarly, what used to be defined as the *Clostridium* genus (historically, the anaerobic endospore-forming bacteria) is also extremely heterogeneous [77]. The majority of the species for which a complete genome sequence has been obtained belong to a cluster corresponding to the redefined *Clostridium* genus, including a solventogenic bacterium (*C. acetobutylicum*) and several human pathogens (*C. botulinum, C. difficile* and *C. perfringens*). Two species that do not belong to the *Clostridium* genus, *Carboxydothermus hydrogenoformans* and *Moorella thermoacetica*, are characterized by their ability to fix carbon monoxide. Importantly, *C. hydrogenoformans* has the smallest genome (2.4 Mb) of all sequenced endospore-formers. Therefore, a good estimate for the core set of sporulation genes (~70 genes) can be obtained from the comparison of this genome to the genomes of other endospore-forming species [78].

To understand the conservation of the sporulation network we selected 24 representative species (Figure 3) out of the 75 currently sequenced endospore-forming bacteria. In our selection, we attempted to cover as much phylogenetic diversity as possible, while incorporating the species that have been best characterized experimentally (i.e. *B. subtilis*, *B. anthracis*/*B. cereus*, *C. acetobutylicum*, *C. perfringens* and *C. difficile*).

Phylogenetic Conservation of the Sporulation Network and Regulatory Interactions

Due to decades of careful experimental studies, a detailed picture of the *B. subtilis* sporulation network is already available. To what extent is the functional organization of the sporulation network in other spore-formers similar to that in *B. subtilis*? The annotations of sequenced genomes and the experimental characterization of the binding specificities of sigma factors make it possible to understand the evolutionary divergence of the sporulation network in other species. Here, we focus on the evolution of gene presence and regulation of 307 *B. subtilis* genes that, according to DBTBS [13], are directly regulated by σ^H, σ^F, σ^G, σ^E or σ^K (Figure 4; Table S1 in Supplemental Information).

The DNA-binding specificities for sporulation sigma factors and many other regulatory interactions in *B. subtilis* have been assembled in DBTBS [13]. In the database, position-specific scoring matrices (PSSM) for regulatory interactions have been calculated on the basis of experimentally characterized promoters (Figure S1). Importantly, interactions between transcription factors and their cognate DNA-binding sites can evolve in various ways. First, both the transcription factor residues involved in promoter recognition and the corresponding DNA-binding site may remain conserved. Second, the transcription factor and the DNA-binding site could diverge separately, eliminating the regulatory relationship. Third, the transcription factor and the DNA-binding site could co-evolve in such a way that the regulation is preserved, although the DNA-binding site changes significantly.

To understand the typical mode of evolution for the sporulation sigma factors, we aligned the sigma factor protein sequences from the 24 considered species and assessed the conservation of the protein residues in these factors that are important for DNA binding. Specifically, we focused on six amino acids known to be important for promoter recognition in the –10 site binding domain of sigma factors and seven amino acids involved in the –35 site binding domain (Table S2A–E in Supplemental Information). These thirteen amino acids are highly conserved in the considered bacteria, suggesting that the DNA-binding specificities of the sporulation sigma factors are also conserved, and that the absence of a sigma-factor-binding site for a particular promoter suggests the absence of the corresponding regulatory relationship.

In our analysis of the sporulation network we established conservation of gene presence using bidirectional BLAST [79] hits, requiring an E-value <0.01. Conservation of gene regulation was established by searching for putative sigma-factor-binding sites in the upstream 300 base pair region using the PSSMs available in DBTBS, requiring a P-value <0.05 (see Supplemental Information). In our searches for regulatory sites we allowed for the possibility that the sigma factor binds in front of an upstream gene in the same operon; the genome-wide operon structure was predicted by identifying transcriptional terminators [80] as well as by considering the chromosomal distances between genes (see Supplemental Information for details). To model the Spo0A-mediated regulation of the *spoIIGA–sigE* promoter,

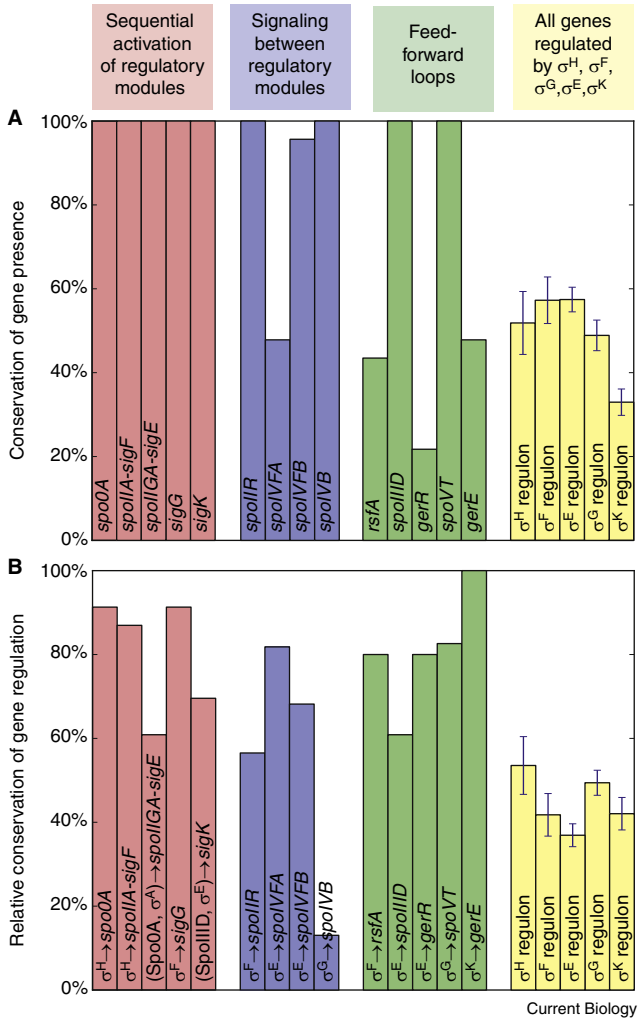

Figure 4. Conservation of gene presence and regulation in the sporulation regulatory network for the 24 representative spore-formers.

(A) The percentage of conserved genes. (B) The percentage of conserved regulations, given the presence of the regulated gene. The conservations are shown in percentages compared to *B. subtilis*. Conservation of gene presence was established using bidirectional BLAST hits, while conservation of gene regulation was established by searching for sigma-factor-binding sites using the PSSMs available in DBTBS [13]. The figure demonstrates that the observed conservation patterns follow the functional and structural hierarchy of the sporulation network: the sigma factors (red) are most conserved, followed by feed-forward loops (green) and inter-compartmental interactions (purple), followed by all downstream genes that are directly regulated by the sporulation network (yellow).

we constructed a hybrid PSSM, in which the –10 part describes a canonical σ^A-binding site, while the –35 part describes the Spo0A-binding site. Likewise, since SpoIIID binds to the –35 part of the *sigK* promoter in *B. subtilis*, we constructed a hybrid PSSM to model the activation of *sigK* by σ^E and SpoIIID; the –10 part of this hybrid PSSM describes a canonical σ^E-binding site, while the –35 part describes the SpoIIID-binding site.

Conservation of Main Regulatory Modules

The overall modular architecture of the sporulation regulatory network and the sporulation sigma factors are fully conserved in all endospore-forming bacteria (Figure 4A, red). Here, we define sporulation modules as collections of genes controlled by specific σ factors. The conservation of sporulation modules was observed previously by Stragier, who analyzed the genomes of six endospore-forming bacteria [11]. The clustering of the genes encoding the sporulation sigma factors is also fully conserved. The *spoIIAA–spoIIAB–sigF* operon, as well as *spoIIE*, is present in all 24 organisms shown in Figure 3. This suggests that the bistable switch that restricts σ^F activity to the forespore is probably conserved in all spore-formers. Similarly, the gene cluster *spoIIGA–sigE–sigG* can be found in all considered species. Although a *sigK* gene could be identified in every species analyzed, σ^K activity could not be unambiguously detected in *C. acetobutylicum*, suggesting that this species might be lacking a functional σ^K [81].

The master regulator Spo0A is also present in all 24 species. Importantly, in contrast to the *Bacillus* species, in all *Clostridium* species the phosphorelay responsible for the post-translational activation of Spo0A is not conserved. In those species, Spo0A is probably phosphorylated directly by an orphan histidine kinase [82]. Nonetheless, although the signals for sporulation initiation may vary from species to species, phosphorylation of Spo0A remains the principal trigger for sporulation initiation in all endospore-forming bacteria.

Conservation of the Sequential Activation of Regulatory Modules

As we describe above, the available PSSMs [13] can be used to investigate the conservation of the regulatory interactions essential for the sequential activation of the sporulation modules (Figure 4B; Table S1). When sporulation commences, *spo0A* is transcribed in large quantities from a sporulation-specific promoter recognized by σ^H (Figure 2). This promoter can be identified in all but two of the considered bacteria. The Spo0A-binding site upstream of *abrB*, used to repress its transcription, is present in 20 out of 23 organisms containing an *abrB* ortholog. No other regulation by Spo0A is conserved more strongly.

We found the σ^H promoter upstream of *sigF* in 21 of the considered organisms and the σ^F promoter upstream of *sigG* in 22. The σ^A-dependent promoter for *sigE* is activated by Spo0A~P binding to the –35 promoter region [37]. The hybrid σ^A–Spo0A-binding site upstream of the *spoIIGA–sigE* operon can be identified in 15 out of the 24 species. Similarly, the sigma factor σ^E and the transcription factor SpoIIID activate transcription of *sigK*, and the hybrid σ^E–SpoIIID-binding site upstream of *sigK* can be identified in 17 out of the 24 species.

Overall, about 80% of the regulatory interactions involved in the sequential activation of global regulatory modules can be identified in the spore-forming bacteria selected for the analysis (Figure 4B, red; Table S1). The actual level of conservation may be even higher, as some promoters may have been missed by the PSSM search due to its limited sensitivity. This indicates that the sequential activation of main regulatory modules is highly conserved across all sporulating species and, consequently, constitutes a fundamental, although possibly not essential, part of this developmental network.

Conservation of Signaling Between Regulatory Modules

As described above, one of the important challenges faced by the sporulating cell is the maintenance of reliable

communication between the mother cell and the forespore. This communication is necessary to ensure the appropriate timing of the transcriptional program in both compartments. Upon completion of the asymmetric division and σ^F activation, the forespore produces a signaling protein, SpoIIR, which activates σ^E in the mother cell. Orthologs of *spoIIR* can be identified in all 24 species (Figure 4A, purple). Similarly, after completion of engulfment and σ^G activation, another signaling molecule, SpoIVB, is produced in the forespore in order to activate σ^K in the mother cell. Orthologs of *spoIVB* can also be identified in all 24 species. These strong conservation patterns suggest that the signaling pathways mediated by SpoIIR and SpoIVB are essential for the sporulation process. As the σ^G-dependent promoter of *spoIVB* strongly depends on the activating transcription factor SpoVT, for which no consensus DNA-binding sequence is available, we were not able to use a PSSM to reliably detect σ^G-dependent transcription of *spoIVB*, even in *B. subtilis* (Figure 4B, blue).

In *B. subtilis*, σ^E is activated by proteolytic cleavage (Figure 2) of pro-σ^E [47,49]. The gene encoding the corresponding protease SpoIIGA is present, immediately upstream of *sigE*, in all investigated species. Thus, conservation of the signaling protein, the protease, their transcriptional regulation, and the presence of a prosequence suggest that this signaling pathway is also highly conserved.

Similarly, σ^K in *B. subtilis* is produced with a pro-sequence [45], which is removed by the protease SpoIVFB in order to activate σ^K. A similar prosequence can be identified in all considered organisms, except *C. difficile* [83]. In *B. subtilis*, the gene *spoIVFB* encoding the σ^K protease is transcribed from a σ^E-dependent promoter, effectively avoiding premature σ^K activation. Orthologs of *spoIVFB* can be identified in all 24 spore-forming bacteria except *C. difficile*, while the corresponding σ^E-dependent promoter is found in 16 species. On the other hand, we found orthologs of the second operon gene, *spoIVFA*, only in half of the considered species. We note that, because SpoIVFA is simply a modulator of proteolysis, it may not be essential for σ^K activation in some bacteria.

Conservation of Feed-Forward Network Motifs

Five feed-forward motif generalizations, consisting of a sigma factor and a transcription factor, are found in *B. subtilis:* σ^F, RsfA; σ^G, SpoVT; σ^E, GerR; σ^E, SpoIIID; and σ^K, GerE. These motifs fine-tune the timing and duration of gene expression in the sporulation network. In each of these loops, a sigma factor regulates a transcription factor, and the sigma factor and transcription factor jointly regulate expression of downstream sporulation genes. In terms of gene presence, *spoIIID* and *spoVT* can be identified in all species, while *rsfA*, *gerR*, and *gerE* are less conserved (Figure 4A, green). Overall, the transcription factors involved in the feed-forward loops are not as strongly conserved as the sporulation sigma factors.

In contrast, the regulation of feed-forward loops, when the corresponding genes are present, is strongly conserved in the sporulation network (Figure 4B, green). For example, the regulation of *rsfA* by σ^F is conserved in 9 of the 11 species in which *rsfA* is present, the regulation of *gerE* by σ^K is conserved in all organisms in which *gerE* is present, and the regulation of *spoVT* by σ^G is conserved in 20 out of 24 species. Surprisingly, regulation of *spoIIID* by σ^E is conserved in only 15 organisms, although this is consistent with a recent report that *spoIIID* transcription is independent of σ^E in *C. perfringens* [84]. The observed conservation pattern suggests that feed-forward loops, if present, are likely to perform similar roles in fine-tuning the timing of gene expression in diverse sporulating bacteria.

What is the logic behind the difference in conservation of various feed-forward motifs involved in sporulation? In *B. subtilis*, relatively few genes are regulated by the RsfA and GerR motifs in comparison with motifs involving SpoVT, SpoIIID, and GerE. This suggests that the conservation of feed-forward loops formed by various transcription factors largely correlates with their functional importance in the sporulation network. The relatively low degree of the GerE conservation, in comparison with the other feed-forward transcription factors, is somewhat surprising. However, as GerE is the last transcription factor in the cascade, it is more dispensable as it is unlikely to significantly affect other sporulation stages.

Conservation of Downstream Genes and Their Regulation

The vast majority of 276 downstream components of the sporulation network are non-regulatory genes that are involved in various molecular and cellular functions related to the sporulation process. Overall, the presence of downstream genes (Figure 4A, yellow) and their regulation (Figure 4B, yellow) are significantly less conserved than those of the main regulatory kernel of the sporulation network. It is likely that, although the regulatory logic of the network is mostly conserved, the regulated components vary in different bacteria to accommodate specific functions and environmental conditions. We estimate that, while the conservation of gene presence is about 50%, the conservation of a gene regulation, given the gene is present, is close to 70%, after correcting for the limited sensitivity of detecting sigma-factor-binding sites. Consequently, if a particular gene or a function is present, its regulation tends to be conserved.

Conservation of Genes and Regulation in Different Functional Categories

The evolutionary patterns discussed above suggest that gene conservation depends on its network role. To understand how the evolution of gene presence, gene regulation, and their interplay depend on their functional role in sporulation, we used the annotations available in SubtiList [85], SubtiWiki [86], and GenoList [87] to divide genes regulated by sporulation sigma factors into several broad non-overlapping functional categories: sporulation; metabolism; coat; regulation; and cell wall (see the Figure 5 legend for the number of genes in each category).

The degree of gene conservation for different functional categories across the considered organisms is shown in Figure 5A. In terms of gene presence, we do not find large differences between the functional categories except for coat proteins, which are significantly less conserved. Similar observations have been reported previously [28]. It is also known that the outer spore structure varies significantly, even between closely related organisms [72]. It is likely that coat proteins are less conserved because they are surface-exposed and, consequently, specifically adapted to diverse environments.

Complementary to the conservation of gene presence, Figure 5B shows the estimated conservation of gene regulation given gene presence and corrected for the limited

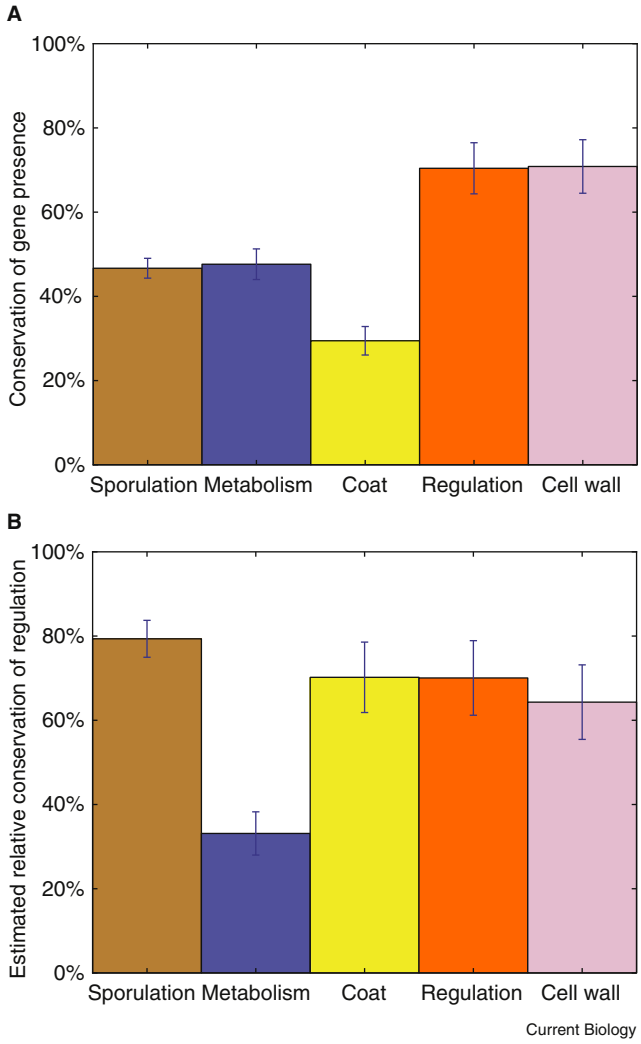

Figure 5. Conservation of gene presence and regulation in different functional categories of the sporulation network.

Conservation of gene presence is shown in (A) and regulation in (B). The percentages are shown for genes regulated by the sporulation sigma factors (σ^F, σ^G, σ^E, σ^K, σ^H) in the 24 representative organisms. The following functional categories are considered: sporulation (179 genes, brown), genes encoding proteins specifically involved in sporulation, except for spore coat proteins, regulation, cell wall biosynthesis, and metabolism; metabolism (45 genes, purple), genes encoding metabolic enzymes; coat (38 genes, yellow), genes encoding spore coat proteins; regulation (31 genes, orange), genes encoding proteins with regulatory functions (e.g. transcription factors and proteins involved in signaling); and cell wall or spore cortex (14 genes, pink), genes whose products are involved in cell wall biosynthesis or hydrolysis. Interestingly, while the conservation of gene presence is weak for coat proteins, conservation of its regulation is relatively strong.

sensitivity of predicting sigma-factor-binding sites using PSSMs (see Supplemental Experimental Procedures for details). Interestingly, the conservation of regulatory relationships for the coat category is relatively strong, whereas the regulatory conservation for metabolism is weak. The conserved regulation for coat proteins is not surprising, as genes encoding coat proteins, if present, are likely to be transcribed in the same compartment and at the same time in different bacteria. On the other hand, the weak conservation of metabolic regulation by sporulation-specific sigma factors is probably due to the fact that, although metabolic pathways are used during sporulation to produce specific metabolites, such as spore components, the necessary compounds probably vary significantly between different bacteria. For example, we recently described the leucine and fatty acid degradation pathway (*yng* gene cluster) regulated by σ^E and active in *B. subtilis* during sporulation [88]. Here, we found that the regulation of this pathway by σ^E is conserved in only a few of the considered species.

Conservation of Gene Presence Relative to Conservation of Regulation

In Figure 6A we contrast the evolutionary conservation of regulatory interaction for sporulation (σ^F, σ^G, σ^E, σ^K, σ^H) and non-sporulation (σ^B, σ^D, σ^W, σ^X) sigma factors. This comparison demonstrates a significantly (two to three times) higher conservation of regulatory relationships involved in sporulation. It is also interesting to compare the conservation of gene presence with conservation of regulatory relationships (Figure 6B). In eukaryotic organisms, especially at short evolutionary distances, regulatory changes usually serve as the main driving force behind evolution and functional adaptation [89,90]. In bacteria, gene regulatory networks have also been found to be extremely flexible, with only a small fraction of transcriptional regulations conserved across diverged phyla. In contrast, while the estimated conservation of regulation by sporulation sigma factors is generally proportional to the conservation in gene content (Figure 6B, red), regulatory relationships involved in sporulation are relatively more conserved. On average, for a bacterium with 50% gene content conservation, about 70% of regulatory interactions involved in sporulation are conserved.

The observed pattern, indicating relatively faster changes in gene content, is likely to be a consequence of the extreme physiological importance of the regulations involved in sporulation. In contrast, if we consider the changes in regulatory relationships for sigma factors not involved in sporulation, a more conventional picture emerges (Figure 6B, green) in which regulatory changes are diverging faster than changes in gene content.

Conclusions

The evolutionary patterns discussed in this review allow us to draw several important conclusions about the evolution of a bacterial developmental network and its regulation. Previously, several eukaryotic developmental systems were considered in detail, notably the sea urchin developmental network [91]. In developmental regulatory networks of higher organisms, conservation of regulatory interactions reflects the hierarchy inherent to the formation of a new body part. In the early stage, the domain that will develop into a body part is specified, followed by the middle stage in which the morphology of the body part is determined, and the late stage, specifying the details of the body part. As each stage in this hierarchy builds on the previous stage, gene regulatory interactions in the earlier stages of development have more widespread consequences than those in later stages, and therefore tend to be evolutionarily conserved [92]. Interestingly, no major changes in developmental organization of the animal body plan have appeared since the Cambrian Explosion about 500 million years ago. The likely reason for this conservation is a high interdependency between development modules. In other words,

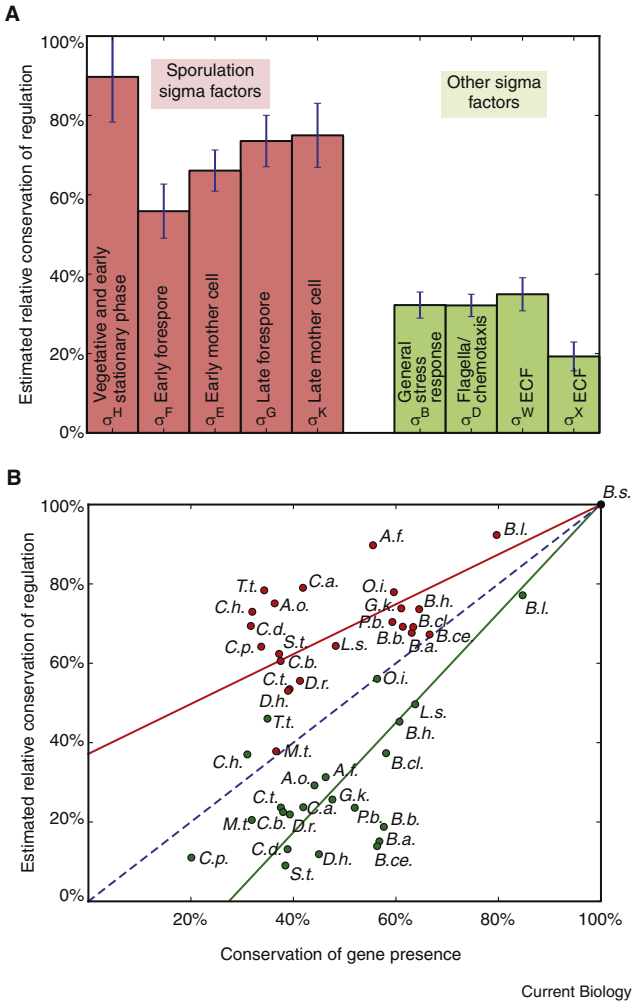

Figure 6. The estimated conservation of regulatory interactions for genes regulated by sporulation (red) and non-sporulation (green) sigma factors.

The conservation is shown in percentages compared to *B. subtilis*. The conservation results were corrected for the limited sensitivity of predicting sigma-factor-binding sites using PSSMs. (A) The estimated conservation of regulation, given the presence of a regulated gene, for the 307 genes regulated by the sporulation sigma factors, and the 211 genes regulated by the non-sporulation sigma factors (other). ECF, extracytoplasmic function. (B) The conservation of gene presence versus the estimated conservation of regulation for the 24 representative organisms. The conservation is shown in red for regulations by the sporulation sigma factors (σ^F, σ^G, σ^E, σ^K, σ^H), and in green for regulations by the non-sporulation sigma factors (σ^B, σ^D, σ^W, σ^X). The initials of each organism name are shown (see phylogenetic tree in Figure 3 for full names), with *B. subtilis* appearing in the upper-right corner. The solid red and green lines represent a linear fit between the conservation of gene presence and the estimated conservation of gene regulation for sporulation and non-sporulation sigma factors, respectively. The dashed diagonal line (in blue) corresponds to the equal conservation of gene presence and gene regulation. Interestingly, for the sporulation sigma factors, gene presence evolves faster than gene regulation, while the reverse is true for the non-sporulation sigma factors.

once the basic development plan has been established, it is unlikely to change in evolution.

Although spore-forming bacteria do not have an elaborate body plan, similar evolutionary patterns to the ones observed in higher organisms [3,93] are also present in the bacterial model system that was discussed here. Importantly, the observed conservation patterns clearly demonstrate that the evolution of the sporulation network is not random, but, to a large extent, follows the functional and hierarchical logic of the sporulation network (Figure 2). Specifically, the regulatory modules governed by each sigma factor are conserved in all spore-forming bacteria that we considered. The sequential activation of these global regulatory modules is also strongly conserved, as well as signaling interactions between the modules. The feed-forward motifs show an intermediate level of conservation, i.e. they are less conserved than the sigma factors but significantly more than the other sporulation genes, which occupy the lowest level in the functional and evolutionary hierarchy.

Evolvability is the concept conjugate to evolutionary conservation [94,95], encapsulating the capacity of an organism to generate heritable phenotypic variations, and consequently the ability to adapt efficiently to new environments. Several hallmarks of evolvable systems discussed previously [96], such as modularity and weak linkage, i.e. small number of interactions between functional modules, are clearly present in the structure of the sporulation network. For example, the coat proteins are restricted in their expression to the mother cell. This allows the efficient adaptation of the coat structure, probably through production of different metabolic building blocks in the mother cell, to the necessary environment. Similarly, the fine-tuning of expression within each regulatory module using feed-forward loops can occur without significant perturbation of gene expression in the other modules. The ability to efficiently adapt the sporulation process to various environmental niches may have contributed to the conservation of the sporulation network structure over more than a billion years of evolution.

Supplemental Information

Supplemental Information includes three tables, one figure and supplemental methods and can be found with this article online at doi:10.1016/j.cub.2010.06.031.

Acknowledgments

The authors wish to thank Rich Losick for advice and helpful discussions, and members of D.V.'s and P.E.'s laboratories for comments on the manuscript. We sincerely apologize to colleagues whose important work has not been cited due to the limit on the number of references. Research was funded in part by NIH grants GM081571 and GM092616 to P.E., NIH grant GM079759 and GC207272NGC to D.V., and National Centers for Biomedical Computing (MAGNet) grant U54CA121852 to Columbia University.

References

1. Dobzhansky, T.G. (1937). Genetics and the Origin of Species (New York: Columbia University Press).
2. Fisher, R.A. (1930). The Genetical Theory of Natural Selection (Oxford: The Clarendon Press).
3. Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. Nat. Rev. Genet. *10*, 141–148.
4. Lynch, M. (2007). The evolution of genetic networks by non-adaptive processes. Nat. Rev. Genet. *8*, 803–813.
5. Davidson, E.H., and Levine, M.S. (2008). Properties of developmental gene regulatory networks. Proc. Natl. Acad. Sci. USA *105*, 20063–20066.
6. Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N., *et al.* (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature *450*, 219–232.
7. Vlamakis, H., Aguilar, C., Losick, R., and Kolter, R. (2008). Control of cell fate by the formation of an architecturally complex bacterial community. Genes Dev. *22*, 945–953.
8. Kroos, L. (2007). The Bacillus and Myxococcus developmental networks and their transcriptional regulators. Annu. Rev. Genet. *41*, 13–39.

9. Errington, J. (2003). Regulation of endospore formation in Bacillus subtilis. Nat. Rev. Microbiol. *1*, 117–126.
10. Piggot, P.J., and Losick, R. (2002). Sporulation genes and intercompartmental regulation. In Bacillus subtilis and Its Closest Relatives: From Genes to Cells, A.L. Sonenshein, J.A. Hoch, and R. Losick, eds. (Washington, DC: American Society for Microbiology), pp. 483–518.
11. Stragier, P. (2002). A gene odyssey: Exploring the genomes of endospore-forming bacteria. In Bacillus subtilis and Its Closest Relatives: From Genes to Cells, A.L. Sonenshein, J.A. Hoch, and R. Losick, eds. (Washington, DC: American Society for Microbiology), pp. 519–526.
12. Losick, R., and Stragier, P. (1992). Crisscross regulation of cell-type-specific gene expression during development in B. subtilis. Nature *355*, 601–604.
13. Sierro, N., Makita, Y., De Hoon, M., and Nakai, K. (2008). DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. Nucleic Acids Res. *36*, D93–D96.
14. Suel, G.M., Garcia-Ojalvo, J., Liberman, L.M., and Elowitz, M.B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. Nature *440*, 545–550.
15. Gonzalez-Pastor, J.E., Hobbs, E.C., and Losick, R. (2003). Cannibalism by sporulating bacteria. Science *301*, 510–513.
16. Veening, J.W., Smits, W.K., Hamoen, L.W., and Kuipers, O.P. (2006). Single cell analysis of gene expression patterns of competence development and initiation of sporulation in Bacillus subtilis grown on chemically defined media. J. Appl. Microbiol. *101*, 531–541.
17. Ryter, A., Schaeffer, P., and Ionesco, H. (1966). [Cytologic classification, by their blockage stage, of sporulation mutants of Bacillus subtilis Marburg]. Ann. Inst. Pasteur (Paris) *110*, 305–315.
18. Piggot, P.J., and Coote, J.G. (1976). Genetic aspects of bacterial endospore formation. Bacteriol. Rev. *40*, 908–962.
19. Veening, J.W., Murray, H., and Errington, J. (2009). A mechanism for cell cycle regulation of sporulation initiation in Bacillus subtilis. Genes Dev. *23*, 1959–1970.
20. Teleman, A.A., Graumann, P.L., Lin, D.C., Grossman, A.D., and Losick, R. (1998). Chromosome arrangement within a bacterium. Curr. Biol. *8*, 1102–1109.
21. Ptacin, J.L., Nollmann, M., Becker, E.C., Cozzarelli, N.R., Pogliano, K., and Bustamante, C. (2008). Sequence-directed DNA export guides chromosome translocation during sporulation in Bacillus subtilis. Nat. Struct. Mol. Biol. *15*, 485–493.
22. Burton, B.M., Marquis, K.A., Sullivan, N.L., Rapoport, T.A., and Rudner, D.Z. (2007). The ATPase SpoIIIE transports DNA across fused septal membranes during sporulation in Bacillus subtilis. Cell *131*, 1301–1312.
23. Wu, L.J., and Errington, J. (1994). Bacillus subtilis SpoIIIE protein required for DNA segregation during asymmetric cell division. Science *264*, 572–575.
24. Morlot, C., Uehara, T., Marquis, K.A., Bernhardt, T.G., and Rudner, D.Z. (2010). A highly coordinated cell wall degradation machine governs spore morphogenesis in Bacillus subtilis. Genes Dev. *24*, 411–422.
25. Pogliano, J., Osborne, N., Sharp, M.D., Abanes-De Mello, A., Perez, A., Sun, Y.L., and Pogliano, K. (1999). A vital stain for studying membrane dynamics in bacteria: a novel mechanism controlling septation during Bacillus subtilis sporulation. Mol. Microbiol. *31*, 1149–1159.
26. Popham, D.L. (2002). Specialized peptidoglycan of the bacterial endospore: the inner wall of the lockbox. Cell Mol. Life Sci. *59*, 426–433.
27. McKenney, P.T., Driks, A., Eskandarian, H.A., Grabowski, P., Guberman, J., Wang, K.H., Gitai, Z., and Eichenberger, P. (2010). A distance-weighted interaction map reveals a previously uncharacterized layer of the *B. subtilis* spore coat. Curr. Biol. *20*, 934–938.
28. Henriques, A.O., and Moran, C.P., Jr. (2007). Structure, assembly, and function of the spore surface layers. Annu. Rev. Microbiol. *61*, 555–588.
29. Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J., and Setlow, P. (2000). Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments. Microbiol. Mol. Biol. Rev. *64*, 548–572.
30. Klobutcher, L.A., Ragkousi, K., and Setlow, P. (2006). The Bacillus subtilis spore coat provides "eat resistance" during phagocytic predation by the protozoan Tetrahymena thermophila. Proc. Natl. Acad. Sci. USA *103*, 165–170.
31. Laaberki, M.H., and Dworkin, J. (2008). Role of spore coat proteins in the resistance of Bacillus subtilis spores to Caenorhabditis elegans predation. J. Bacteriol. *190*, 6197–6203.
32. Setlow, P. (2003). Spore germination. Curr. Opin. Microbiol. *6*, 550–556.
33. Carter, H.L., 3rd, and Moran, C.P., Jr. (1986). New RNA polymerase sigma factor under spo0 control in Bacillus subtilis. Proc. Natl. Acad. Sci. USA *83*, 9438–9442.
34. Li, Z., and Piggot, P.J. (2001). Development of a two-part transcription probe to determine the completeness of temporal and spatial compartmentalization of gene expression during bacterial development. Proc. Natl. Acad. Sci. USA *98*, 12538–12543.
35. Burbulys, D., Trach, K.A., and Hoch, J.A. (1991). Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay. Cell *64*, 545–552.
36. Molle, V., Fujita, M., Jensen, S.T., Eichenberger, P., Gonzalez-Pastor, J.E., Liu, J.S., and Losick, R. (2003). The Spo0A regulon of Bacillus subtilis. Mol. Microbiol. *50*, 1683–1701.
37. Fujita, M., Gonzalez-Pastor, J.E., and Losick, R. (2005). High- and low-threshold genes in the Spo0A regulon of Bacillus subtilis. J. Bacteriol. *187*, 1357–1368.
38. Fujita, M., and Losick, R. (2005). Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A. Genes Dev. *19*, 2236–2244.
39. Wang, S.T., Setlow, B., Conlon, E.M., Lyon, J.L., Imamura, D., Sato, T., Setlow, P., Losick, R., and Eichenberger, P. (2006). The forespore line of gene expression in Bacillus subtilis. J. Mol. Biol. *358*, 16–37.
40. Steil, L., Serrano, M., Henriques, A.O., and Volker, U. (2005). Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of Bacillus subtilis. Microbiology *151*, 399–420.
41. Alon, U. (2006). An Introduction to Systems Biology: Design Principles of Biological Circuits (Boca Raton: Chapman & Hall/CRC).
42. Kashtan, N., Itzkovitz, S., Milo, R., and Alon, U. (2004). Topological generalizations of network motifs. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *70*, 031909.
43. Eichenberger, P., Fujita, M., Jensen, S.T., Conlon, E.M., Rudner, D.Z., Wang, S.T., Ferguson, C., Haga, K., Sato, T., Liu, J.S., *et al.* (2004). The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis. PLoS Biol. *2*, e328.
44. Kuwana, R., Okumura, T., Takamatsu, H., and Watabe, K. (2005). The ylbO gene product of Bacillus subtilis is involved in the coat development and lysozyme resistance of spore. FEMS Microbiol. Lett. *242*, 51–57.
45. Kroos, L., Kunkel, B., and Losick, R. (1989). Switch protein alters specificity of RNA polymerase containing a compartment-specific sigma factor. Science *243*, 526–529.
46. Campbell, E.A., Masuda, S., Sun, J.L., Muzzin, O., Olson, C.A., Wang, S., and Darst, S.A. (2002). Crystal structure of the Bacillus stearothermophilus anti-sigma factor SpoIIAB with the sporulation sigma factor sigmaF. Cell *108*, 795–807.
47. LaBell, T.L., Trempy, J.E., and Haldenwang, W.G. (1987). Sporulation-specific sigma factor sigma 29 of Bacillus subtilis is synthesized from a precursor protein, P31. Proc. Natl. Acad. Sci. USA *84*, 1784–1788.
48. Stragier, P., Bonamy, C., and Karmazyn-Campelli, C. (1988). Processing of a sporulation sigma factor in Bacillus subtilis: how morphological structure could control gene expression. Cell *52*, 697–704.
49. Imamura, D., Zhou, R., Feig, M., and Kroos, L. (2008). Evidence that the Bacillus subtilis SpoIIGA protein is a novel type of signal-transducing aspartic protease. J. Biol. Chem. *283*, 15287–15299.
50. Igoshin, O.A., Price, C.W., and Savageau, M.A. (2006). Signalling network with a bistable hysteretic switch controls developmental activation of the sigma transcription factor in Bacillus subtilis. Mol. Microbiol. *61*, 165–184.
51. Duncan, L., and Losick, R. (1993). SpoIIAB is an anti-sigma factor that binds to and inhibits transcription by regulatory protein sigma F from Bacillus subtilis. Proc. Natl. Acad. Sci. USA *90*, 2325–2329.
52. Min, K.T., Hilditch, C.M., Diederich, B., Errington, J., and Yudkin, M.D. (1993). Sigma F, the first compartment-specific transcription factor of B. subtilis, is regulated by an anti-sigma factor that is also a protein kinase. Cell *74*, 735–742.
53. Arigoni, F., Pogliano, K., Webb, C.D., Stragier, P., and Losick, R. (1995). Localization of protein implicated in establishment of cell type to sites of asymmetric division. Science *270*, 637–640.
54. Duncan, L., Alper, S., Arigoni, F., Losick, R., and Stragier, P. (1995). Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. Science *270*, 641–644.
55. Karow, M.L., Glaser, P., and Piggot, P.J. (1995). Identification of a gene, spoIIR, that links the activation of sigma E to the transcriptional activity of sigma F during sporulation in Bacillus subtilis. Proc. Natl. Acad. Sci. USA *92*, 2012–2016.
56. Londono-Vallejo, J.A., and Stragier, P. (1995). Cell-cell signaling pathway activating a developmental transcription factor in Bacillus subtilis. Genes Dev. *9*, 503–508.
57. Eldar, A., Chary, V.K., Xenopoulos, P., Fontes, M.E., Loson, O.C., Dworkin, J., Piggot, P.J., and Elowitz, M.B. (2009). Partial penetrance facilitates developmental evolution in bacteria. Nature *460*, 510–514.
58. Meisner, J., Wang, X., Serrano, M., Henriques, A.O., and Moran, C.P., Jr. (2008). A channel connecting the mother cell and forespore during bacterial endospore formation. Proc. Natl. Acad. Sci. USA *105*, 15100–15105.
59. Camp, A.H., and Losick, R. (2008). A novel pathway of intercellular signalling in Bacillus subtilis involves a protein with similarity to a component of type III secretion channels. Mol. Microbiol. *69*, 402–417.
60. Blaylock, B., Jiang, X., Rubio, A., Moran, C.P., Jr., and Pogliano, K. (2004). Zipper-like interaction between proteins in adjacent daughter cells mediates protein localization. Genes Dev. *18*, 2916–2928.
61. Camp, A.H., and Losick, R. (2009). A feeding tube model for activation of a cell-specific transcription factor during sporulation in Bacillus subtilis. Genes Dev. *23*, 1014–1024.
62. Doan, T., Morlot, C., Meisner, J., Serrano, M., Henriques, A.O., Moran, C.P., Jr., and Rudner, D.Z. (2009). Novel secretion apparatus maintains spore integrity and developmental gene expression in Bacillus subtilis. PLoS Genet. *5*, e1000566.

63. Rudner, D.Z., Fawcett, P., and Losick, R. (1999). A family of membrane-embedded metalloproteases involved in regulated proteolysis of membrane-associated transcription factors. Proc. Natl. Acad. Sci. USA *96*, 14765–14770.
64. Zhou, R., Cusumano, C., Sui, D., Garavito, R.M., and Kroos, L. (2009). Intramembrane proteolytic cleavage of a membrane-tethered transcription factor by a metalloprotease depends on ATP. Proc. Natl. Acad. Sci. USA *106*, 16174–16179.
65. Cutting, S., Oke, V., Driks, A., Losick, R., Lu, S., and Kroos, L. (1990). A forespore checkpoint for mother cell gene expression during development in B. subtilis. Cell *62*, 239–250.
66. Ricca, E., Cutting, S., and Losick, R. (1992). Characterization of bofA, a gene involved in intercompartmental regulation of pro-sigma K processing during sporulation in Bacillus subtilis. J. Bacteriol. *174*, 3177–3184.
67. Cutting, S., Driks, A., Schmidt, R., Kunkel, B., and Losick, R. (1991). Forespore-specific transcription of a gene in the signal transduction pathway that governs Pro-sigma K processing in Bacillus subtilis. Genes Dev. *5*, 456–466.
68. Lu, S., Halberg, R., and Kroos, L. (1990). Processing of the mother-cell sigma factor, sigma K, may depend on events occurring in the forespore during Bacillus subtilis development. Proc. Natl. Acad. Sci. USA *87*, 9722–9726.
69. Maughan, H., Birky, C.W., Jr., and Nicholson, W.L. (2009). Transcriptome divergence and the loss of plasticity in Bacillus subtilis after 6,000 generations of evolution under relaxed selection for sporulation. J. Bacteriol. *191*, 428–433.
70. Paredes, C.J., Alsaker, K.V., and Papoutsakis, E.T. (2005). A comparative genomic view of clostridial sporulation and physiology. Nat. Rev. Microbiol. *3*, 969–978.
71. Ghosh, J., Larsson, P., Singh, B., Pettersson, B.M., Islam, N.M., Sarkar, S.N., Dasgupta, S., and Kirsebom, L.A. (2009). Sporulation in mycobacteria. Proc. Natl. Acad. Sci. USA *106*, 10781–10786.
72. Traag, B.A., Driks, A., Stragier, P., Bitter, W., Broussard, G., Hatfull, G., Chu, F., Adams, K.N., Ramakrishnan, L., and Losick, R. (2010). Do mycobacteria produce endospores? Proc. Natl. Acad. Sci. USA *107*, 878–881.
73. Priest, F.G. (1993). Systematics and ecology of Bacillus. In Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics, A.L. Sonenshein, J. Hoch, and R. Losick, eds. (Washington DC: ASM Press), pp. 3–16.
74. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.*, et al.* (1997). The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature *390*, 249–256.
75. Rasko, D.A., Altherr, M.R., Han, C.S., and Ravel, J. (2005). Genomics of the Bacillus cereus group of organisms. FEMS Microbiol. Rev. *29*, 303–329.
76. Alcaraz, L.D., Olmedo, G., Bonilla, G., Cerritos, R., Hernandez, G., Cruz, A., Ramirez, E., Putonti, C., Jimenez, B., Martinez, E.*, et al.* (2008). The genome of Bacillus coahuilensis reveals adaptations essential for survival in the relic of an ancient marine environment. Proc. Natl. Acad. Sci. USA *105*, 5803–5808.
77. Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., and Farrow, J.A. (1994). The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. *44*, 812–826.
78. Wu, M., Ren, Q., Durkin, A.S., Daugherty, S.C., Brinkac, L.M., Dodson, R.J., Madupu, R., Sullivan, S.A., Kolonay, J.F., Nelson, W.C.*, et al.* (2005). Life in hot carbon monoxide: The complete genome sequence of Carboxydothermus hydrogenoformans Z-2901. PLoS Genet. *1*, e65.
79. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
80. De Hoon, M.J., Makita, Y., Nakai, K., and Miyano, S. (2005). Prediction of transcriptional terminators in Bacillus subtilis and related species. PLoS Comput. Biol. *1*, e25.
81. Jones, S.W., Paredes, C.J., Tracy, B., Cheng, N., Sillers, R., Senger, R.S., and Papoutsakis, E.T. (2008). The transcriptional program underlying the physiology of clostridial sporulation. Genome Biol. *9*, R114.
82. Stephenson, K., and Hoch, J.A. (2002). Evolution of signalling in the sporulation phosphorelay. Mol. Microbiol. *46*, 297–304.
83. Haraldsen, J.D., and Sonenshein, A.L. (2003). Efficient sporulation in Clostridium difficile requires disruption of the sigmaK gene. Mol. Microbiol. *48*, 811–821.
84. Harry, K.H., Zhou, R., Kroos, L., and Melville, S.B. (2009). Sporulation and enterotoxin (CPE) synthesis are controlled by the sporulation-specific sigma factors SigE and SigK in Clostridium perfringens. J. Bacteriol. *191*, 2728–2742.
85. Moszer, I., Glaser, P., and Danchin, A. (1995). SubtiList: a relational database for the Bacillus subtilis genome. Microbiology *141*, 261–268.
86. Florez, L.A., Roppel, S.F., Schmeisky, A.G., Lammers, C.R., and Stulke, J. (2009). A community-curated consensual annotation that is continuously updated: the Bacillus subtilis centred wiki SubtiWiki. Database (Oxford) *2009*, bap012.
87. Lechat, P., Hummel, L., Rousseau, S., and Moszer, I. (2008). GenoList: an integrated environment for comparative analysis of microbial genomes. Nucleic Acids Res. *36*, D469–D474.
88. Hsiao, T.L., Revelles, O., Chen, L., Sauer, U., and Vitkup, D. (2010). Automatic policing of biochemical annotations using genomic correlations. Nat. Chem. Biol. *6*, 34–40.
89. Borneman, A.R., Gianoulis, T.A., Zhang, Z.D., Yu, H., Rozowsky, J., Seringhaus, M.R., Wang, L.Y., Gerstein, M., and Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species. Science *317*, 815–819.
90. Prud'homme, B., Gompel, N., and Carroll, S.B. (2007). Emerging principles of regulatory evolution. Proc. Natl. Acad. Sci. USA *104* (Suppl 1), 8605–8612.
91. Davidson, E.H. (2009). Network design principles from the sea urchin embryo. Curr. Opin. Genet. Dev. *19*, 535–540.
92. Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science *311*, 796–800.
93. Davidson, E.H., and Erwin, D.H. (2010). Evolutionary innovation and stability in animal gene networks. J. Exp. Zool. B Mol. Dev. Evol. *314*, 182–186.
94. Kirschner, M., and Gerhart, J. (1998). Evolvability. Proc. Natl. Acad. Sci. USA *95*, 8420–8427.
95. Wagner, A. (2005). Robustness and Evolvability in Living Systems (Princeton: Princeton University Press).
96. Kirschner, M.W., and Gerhart, J.C. (2005). The Plausibility of Life: Resolving Darwin's Dilemma, 1st edition (New Haven, CT: Yale University Press).
97. Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics *5*, 164–166.
98. Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M.*, et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. *37*, D141–D145.