**BMC
Systems Biology**

## RESEARCH ARTICLE

**Open Access**

# Lessons from the modular organization of the transcriptional regulatory network of Bacillus subtilis

Julio A Freyre-González[1†], Alejandra M Manjarrez-Casas[2†], Enrique Merino[2], Mario Martinez-Nuñez[2],
Ernesto Perez-Rueda[3] and Rosa-María Gutiérrez-Ríos[2*]

## Abstract

**Background:** The regulation of gene expression at the transcriptional level is a fundamental process in prokaryotes. Among the different kind of mechanisms modulating gene transcription, the one based on DNA binding transcription factors, is the most extensively studied and the results, for a great number of model organisms, have been compiled making it possible the *in silico* construction of their corresponding transcriptional regulatory networks and the analysis of the biological relationships of the components of these intricate networks, that allows to elucidate the significant aspects of their organization and evolution.

**Results:** We present a thorough review of each regulatory element that constitutes the transcriptional regulatory network of *Bacillus subtilis*. For facilitating the discussion, we organized the network in topological modules. Our study highlight the importance of σ factors, some of them acting as master regulators which characterize modules by inter- or intra-connecting them and play a key role in the cascades that define relevant cellular processes in this organism. We discussed that some particular functions were distributed in more than one module and that some modules contained more than one related function. We confirm that the presence of paralogous proteins confers advantages to *B. subtilis* to adapt and select strategies to successfully face the extreme and changing environmental conditions in which it lives.

**Conclusions:** The intricate organization is the product of a non-random network evolution that primarily follows a hierarchical organization based on the presence of transcription and σ factor, which is reflected in the connections that exist within and between modules.

**Keywords:** Master regulators, σ factors, Modularity, Hierarchy, Regulatory network, Model organisms, Paralogous proteins

## Background

*Bacillus subtilis* is the best-characterized member of the Gram-positive bacteria and represents an excellent model for the study of gene regulation and metabolism in the Firmicute phylum. This Gram-positive bacterium is a facultative aerobe that was initially classified as a soil bacterium, but its ability to grow in many diverse terrestrial and aquatic environments, from the root surface of some plants to the gastrointestinal tract of some animals, is now well known. This ability to adapt to different environments has been mainly attributed to the formation of spores, which occurs under certain conditions of stress and nutrient scarcity.

Many characteristics of this bacterium have been elucidated through the study of its complete genome. A sequence analysis of its genome revealed the presence of more than 120 transcription regulatory proteins (including 14 σ factors) that regulate the expression of 1,475 promoters [1]. This knowledge provides useful information to construct the *Bacillus subtilis* Transcriptional Regulatory Network (TRN) and make a deep and careful

* Correspondence: rmaria@ibt.unam.mx
†Equal contributors
2Departamentos de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apdo. Postal 510-3, Cuernavaca, Morelos 62250, México
Full list of author information is available at the end of the article

discussion of the intricate and elegant design that *B. subtilis* displays in the microbial world. In addition, in this bacterium, global and local regulators are integrated, which allows the TRN to be rewired in response to metabolic signals, spore formation and germination processes. This rewiring is mediated by the interplay of transcription factors (TFs) and σ factors that have fundamental roles in regulating important and well-defined metabolic and development processes, such as sporulation and germination.

An essential aspect of this work is the manual curation of the information and a thorough review of each regulatory element that constitutes the transcriptional regulatory network of *Bacillus subtilis.* As a way of presenting the vast amount of information published on *B. subtilis,* we organized this data onto network topological modules, in such a way that the relationship of different regulatory elements and the role that they play in the *B. subtilis* metabolism could be analyze from a more integrated point of view as done for other networks [2,3]. For this purpose, and based on the experimentally defined regulatory interactions of TFs and σ factors with their corresponding target gene [1], we evaluated the statistical properties of the *B. subtilis* TRN. As it has been reported for other model organisms [4-7], our results indicate that TRN of *B. subtilis* is a scale-free network with hierarchical properties consisting of nine regulatory modules that could be associated with well-defined biological processes. In addition, we discuss the evolutionary and functional implications of the topology of the TRN in this bacterial model.

## Results and discussion
### Topological organization of the *B. subtilis* regulatory network and its comparison with other TRNs
To evaluate the statistical properties and modular organization of the *B. subtilis* TRN, we obtained all of the regulatory interactions reported in DBTBS, a database of transcriptional regulation in *B. subtilis* [1]. Based on this information, we constructed a regulatory network composed of 1,626 nodes and 3,096 edges, with an average clustering coefficient of 0.538. The $P(k)$ follows a power law distribution with a power-law exponent of approximately −2.11. These results are characteristic of a large, scale-free network with a modular hierarchic organization. These properties are common in other previously described regulatory networks, such as the TRN of *Escherichia coli* K12 and *Sacharomyces cerevisiae* [6,8-10]. As a point of comparison with other reported TRNs we calculated the incoming (*Pin*) and outcoming (*Pout*) degree distributions of three well-annotated prokaryotes (*B. subtilis*, E. *coli* and *M. tuberculosis*) and an eukaryote (*S. cerevisiae*), (Figures 1 a-b respectively). We observed that the *Pin* distributions are characteristic of
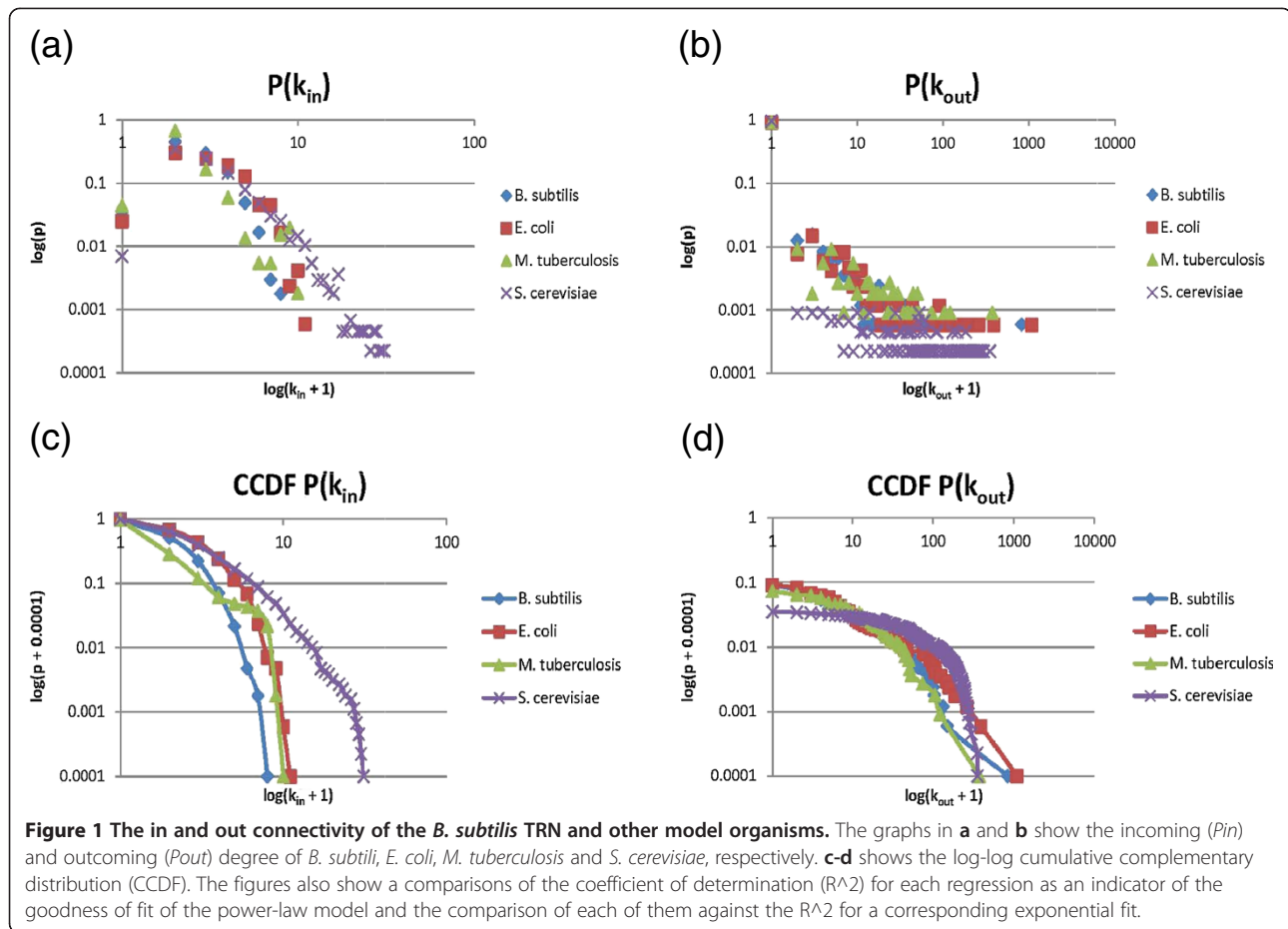
TRN, which do not show a long tail for any organism (Figure 1a). In order to estimate the exponent for each power-law we computed the log-log cumulative complementary distribution (CCDF) and then fitted a straight-line to it using least squares (Figures 1 c-d). Additionally, we computed the coefficient of determination ($R^2$) for each regression as an indicator of the goodness of fit of the power-law model, and then compared each of them against the $R^2$ for a corresponding exponential fit. We found that the *Pout* distributions for the three bacteria are better explained by a power-law than by an exponential fit. Conversely, *S. cerevisiae Pout* distribution is better explained by an exponential distribution, so we do not computed the power-law exponent.

In a posterior step, we extracted a sub-network consisting of only the regulatory interactions of all known *B. subtilis* TFs and σ factors (54 and 16, respectively). We excluded σA interactions from this sub-network because, as a housekeeping factor, σA is tightly connected to almost every node of the network (with an outcoming connectivity of 782, connecting 46.5% of the genes in the network), generating a mega-module that encompasses all the basic physiological functions described in *B. subtilis*. Our resulting TRN was composed of 71 nodes and 81 edges and is supported by strong experimental evidence. The data were used as input to perform a hierarchical agglomerative average linkage clustering. This analysis revealed nine discrete modules (Additional file 1: Figure S1) whose genes clearly correlate with a metabolic or specific function (Figure 2), as reported for *E. coli* and *M. tuberculosis* [6,11,12].

### Granularity of the detected modules
In this work we identified modules using the hierarchical clustering method originally proposed for protein networks by Rives and Galitski [3] and applied for the first time to regulatory networks by Resendis-Antonio et al. [6]. To evaluate the identified modules, we compared our results to three alternative methods for modules detection: Girvan-Newman [2], Rosvall-Bergstrom [13], and the natural decomposition approach (NDA) [14]. In general terms, the three methods re-captured the identified modules obtained originally with the hierarchical clustering method, although showed different granularity (see Additional file 2: Table S1).

The identified by the Girvan-Newman algorithm showed the highest similarity with the ones identified with the hierarchical clustering. There were only two differences: 1) PhoP was clustered into a different module, and 2) module 8 was disaggregated into two submodules. The Rosvall-Bergstrom algorithm also clustered PhoP into a different module in addition of a more disaggregated modules. Rosvall-Bergstrom found that some elements of the modules 6, 7, 8 and 9 could

**Figure 1 The in and out connectivity of the *B. subtilis* TRN and other model organisms.** The graphs in **a** and **b** show the incoming (*Pin*) and outcoming (*Pout*) degree of *B. subtili, E. coli, M. tuberculosis* and *S. cerevisiae*, respectively. **c-d** shows the log-log cumulative complementary distribution (CCDF). The figures also show a comparisons of the coefficient of determination (R^2) for each regression as an indicator of the goodness of fit of the power-law model and the comparison of each of them against the R^2 for a corresponding exponential fit.

be disaggregated into 2, 2, 4 and 3 submodules, respectively.

Interestingly, module 6 identified by Rosvall-Bergstrom is the more transverse module, which is dispersed over three modules identified by hierarchical clustering (modules 5, 6 and 7). The NDA mathematically identifies global TFs and remove them from the network. As a consequence of this step, global TFs are not clustered into any module. This allowed the finest disaggregation into submodules of the four methods. No transverse modules were identified with this method. The transverse module 6 identified by Rosvall-Bergstrom was identified by the NDA as two modules with no functional correlation (13 and 54) and two global TFs. Despite this finer granularity, the physiological functions annotated for each module identified in this work highly correlated with the corresponding functions for the submodules identified by the NDA. We observed that the modules identified by any method are mainly subsets or supersets of the modules identified by other method. These results highlight the relevance of taking into account the previously reported *Matryoshka*-like organization of regulatory networks [14] by showing that while different methods are able to re-
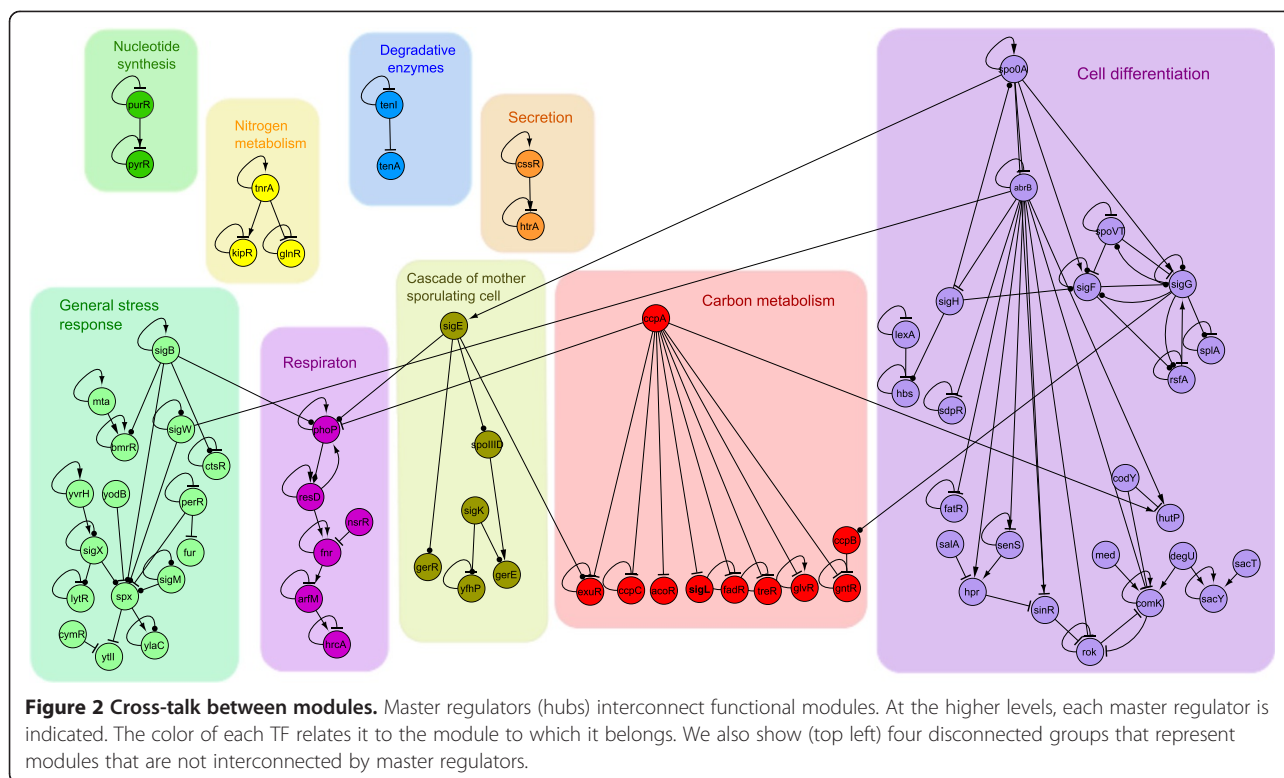
capture the identified modules, this is accomplished at different granular levels.

**The modules of the *B. subtilis* TRN clearly correlate with well-defined metabolic and physiological responses**
To characterize the metabolic and physiological responses of each of the nine modules identified in the *B. subtilis* TRN, we performed an exhaustive literature search of the experimentally validated regulatory data for each response. The complete description of each module and its relationship with their regulated genes and other modules can be viewed in the Additional file 1.

*Module 1 (M1)*
Groups the TFs TnrA, GlpR, and KipR, regulating genes involved in *Nitrogen assimilation function*s (Figure 2 and Additional file 1: Table S2). TnrA is required during nitrogen-limited growth and GlpR during growth with excess nitrogen [15]. TnrA regulates KipR, also detected in this module, whose main function is displayed during the sporulation cell fate [16] (more details are provided in the Additional file 1).

**Figure 2 Cross-talk between modules.** Master regulators (hubs) interconnect functional modules. At the higher levels, each master regulator is indicated. The color of each TF relates it to the module to which it belongs. We also show (top left) four disconnected groups that represent modules that are not interconnected by master regulators.

### Module 2 (M2)

Devoted to *Nucleotide synthesis* includes the PyrR TF, which regulates pyrimidine synthesis and metabolism by transcriptional attenuation [17], and PurR, which regulates genes involved in purine and pyrimidine synthesis and transport [18,19] (Figure 2 and Additional file 1: Table S2).

### Module 3 (M3)

Cluster the TFs CssR and HtrA, which are expressed under stressful conditions are regulate genes related to *Secretion* [20] (Figure 2) and (Additional file 1: Table S2).

### Module 4 (M4)

Entitle as *Degradative enzyme* module, is integrated by two TFs, TenA and TenI (Figure 2), that regulate the production of several extracellular enzymes [21].
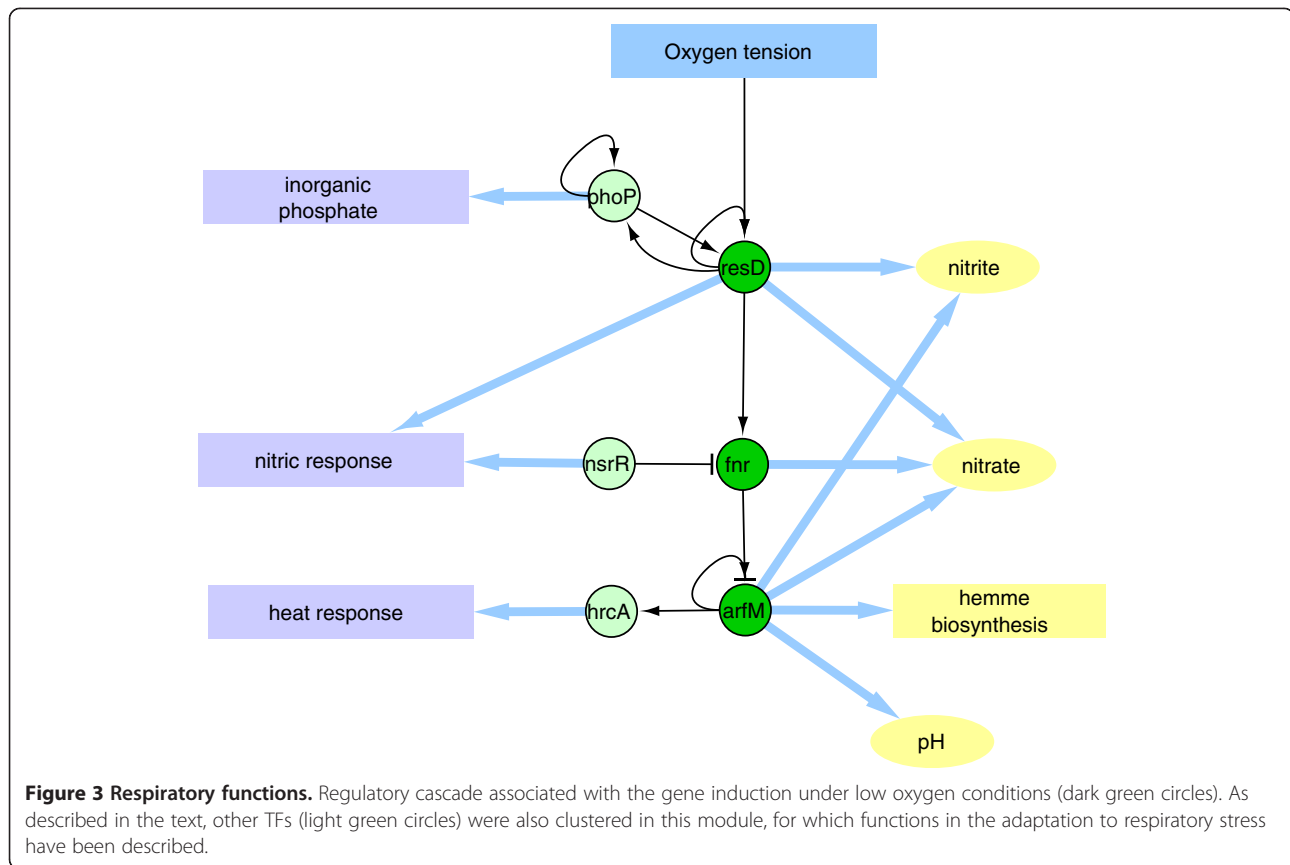
### Module 5 (M5)

Or *Respiration* module includes all the TFs that are required for switching between aerobic and anaerobic growth (see Additional file 1: Table S2). The TFs belonging to this module are ArfM, HrcA, FNR, NsrR, ResD, and PhoP, which are highly inter-regulated in a hierarchical order (see Figures 2 and 3). The complex regulation of this module correlates with the fact that *B. subtilis* grows either by fermentation or anaerobically, using nitrate or nitrite as terminal electron acceptors [22].

### Module 6 (M6)

Devoted to *Carbon metabolism*, groups the TFs CcpA (Figure 2 and Additional file 1: Figure S1). CcpA is the master regulator of sugar operons (see Additional file 1: Table S2 and a detail description in the Additional file 1), which regulates almost all the TFs in this module except for CcpB. GntR a TF that is responsible for gluconate catabolism regulation [23] is an example of this. ExuR involved in hexuronate assimilation, is regulated by CcpA and σE, which are located in the *CMCS module*. Other proteins also regulated by CcpA, are AcoR a regulatory protein that is expressed when *B. subtilis* is in the exponential growth phase and excretes diverse organic compounds, such as acetoin, TreR that coordinates the expression of different kind of genes in response to trehalose availability (Additional file 1: Table S2), GlvR involved in the maltose utilization [24], FadR involved in the fatty acid β-oxidation cycle and CcpC necessary for the catabolic repression of genes that are involved in the Krebs cycle [25,26].

### Module 7 (M7)

Cluster the TFs controlling the *Cascade of the mother cell sporulating (CMCS)*, the genes encoding the TFs of this cascade are expressed hierarchical in the following order *sigE → spoIIID* and *gerR → sigK → gerE* and *yfhP* [27,28] (see Figure 2, Additional file 1: Table S2 and annotations in Additional file 1).

**Figure 3 Respiratory functions.** Regulatory cascade associated with the gene induction under low oxygen conditions (dark green circles). As described in the text, other TFs (light green circles) were also clustered in this module, for which functions in the adaptation to respiratory stress have been described.

### Module 8 (M8)

Groups TFs involved in *Cell differentiation*, involving four master regulators: AbrB, DegU, ComK, and Spo0A [29], that coordinate in conjunction with other regulators the following well-defined cellular responses and fates: *sporulation, competence, DNA protection, matrix and extracellular protein biogenesis, cannibalism, degradative enzyme synthesis, and nutritional limitation response* (Figure 4 and Additional file 1: Table S3), that were clustered together in this module. In the Additional file 1, we discuss the direct and indirect influence of the master transcriptional regulators, on various differentiation processes and stress responses and its relationship with other TFs coordinating the above mention fates and cellular responses.

### Module 9 (M9)

In addition to cell differentiation, *B. subtilis* has other methods to face adverse growth conditions. The genes involved in these activities are regulated by different σ factors and TFs clustered in the *General stress module.* The σB regulon is one of the alternative responses, and it is activated to protect the vegetative cell during starvation or physical stress [30]. The stress responses includes TFs such as CtsR, BmrR, YtlI, CymR, PerR, YodB, LytR, YvrH.and Spx, and the σ factor YlaC, σM, σW, and σX
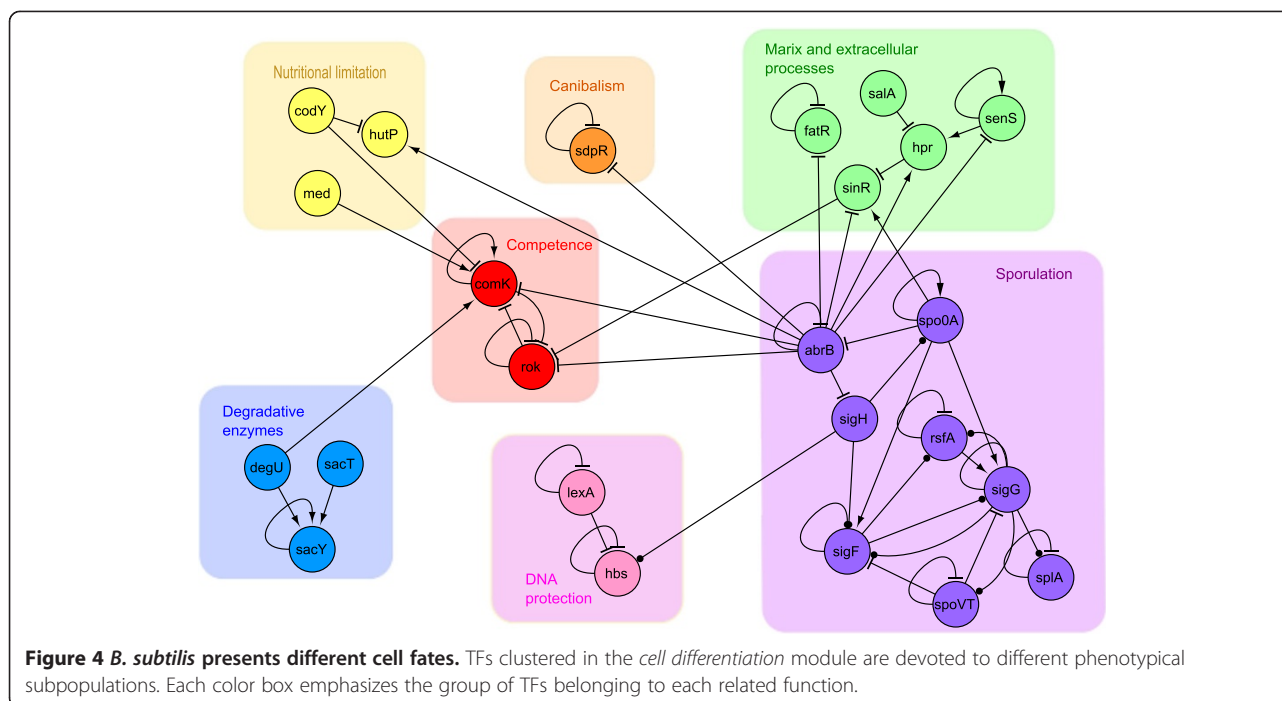
(see Figure 2), which specific function are described in the Additional file 1.

### Paralogy; an evolutionary force modifying the TRN of B. subtilis

In a previous study, we performed an exhaustive review of paralogous gene regulation in *E. coli* and *B. subtilis* based on published information [31]. In this work, we identified the paralogous TFs in the constructed TRN and briefly discussed the implications of the distribution of the TFs inside and between modules.

In our previous study, we found that TnrA and GlpR located in M1, are paralogous proteins (Figure 5) [31] that belong to the MerR family [32], and interestingly, their DNA-binding sites have the same consensus sequence [33]. A large fraction of neighboring TF binding sites have been formed by local duplications of a common sequence and might diverge as a consequence of point mutations [34]; further, these sites may have been selected for specific environmental conditions, as suggested by Singh and Hannenhalli [35]. Additional examples of interchangeable DNA-binding sites have been observed in other families of *E. coli* regulatory proteins, such as CRP and FNR [36].

As in previous works [31,35,37,38], we observed that some duplication events does not give as a result two

**Figure 4 *B. subtilis* presents different cell fates.** TFs clustered in the *cell differentiation* module are devoted to different phenotypical subpopulations. Each color box emphasizes the group of TFs belonging to each related function.
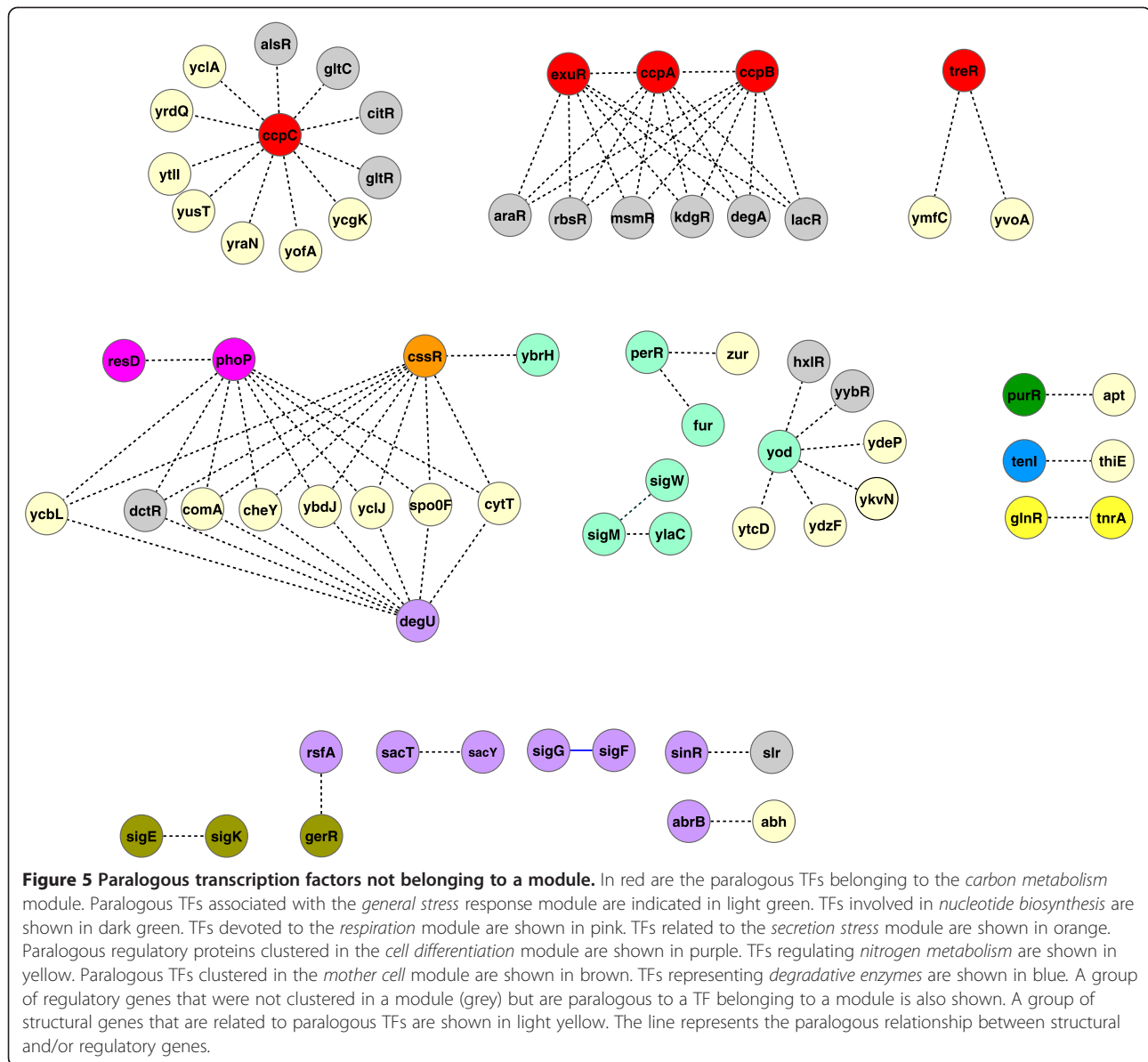
TFs. An example of this, is PurR cluster in M2, whose paralogous copy is the protein Apt, an adenine phosphoribosyltransferase [39] (Figure 5) that participates in nucleotide synthesis and is an enzyme rather than a TF. Both proteins have a PRT motif that is involved in the binding of the inducer phosphoribosylpyrophosphate. In PurR, this domain is fused to a winged-helix-turn-helix domain that is present in other DNA-binding proteins, while in Apt, the PRT domain presents catalytic activity [40]. This module, is a clear example of how the diversity of paralogous proteins and gene functions, might increase the genetic and metabolic robustness of a network. We observed something similar in M3, where the TFs CssR and HtrA are not homologous. Instead, CssR have as paralogous copy the TF YvrH, positioned in M9 that group TFs related with the *General stress* response (Figure 5) [31], As CssR, YvrH is a response regulator of a two-component system (YvrG-YvrH), but differently to CssR the regulatory function of this TF participates in the control of the homeostasis of *B. subtilis* at the cell surface level [41]. This example, illustrates that paralogous TFs do not always regulate genes that are related to functional or metabolic processing but can be part of different regulatory modules. Similar results on the plasticity and robustness of the regulatory networks of *E. coli* and *S. cerevisiae* have been described by Babu and Teichmann [38] and Conant and Wagner [42].

In M4 the only members of the module, TenA and TenI are not homologous; nevertheless, TenI has a paralogous copy with catalytic activity: the ThiE thiamine-phosphate pyrophosphorylase enzyme (Figure 5) [43].

Similar examples were found in M2 and M3, related with *Nucleotide synthesis* and *secretion stress*, this confirms that paralogous proteins with different functions might increase the plasticity and robustness of the *B. subtilis* regulatory and metabolic network.

In M5 we observed that the TFs ResD and PhoP, based on their sequence similarity, evolved from a common ancestor [44] and form part of the IIIA group of two-component systems (Figure 5). These similarity is worth noting not only because these paralogous TFs coordinate the expression of genes that are involved in the phosphate uptake, but also because they are part of a paralogous set of two-component system regulatory proteins, ResD/ResE and PhoP/PhoR, where ResE and PhoR are membrane-bound histidine kinases that sense the extracellular phosphate concentration [45]. Whitworth and Cock previously postulate, that genes regulated by two-component systems might allow rapid and robust responses to short-term changes in the environment [46].

CcpB is a paralogous copy of CcpA, the main TF in the *Carbon metabolism module* M6. These TFs share 30% sequence similarity, and as in the case of CcpA, the down-regulating activity of CcpB depends on the phosphorylated state of HPr. In parallel with CcpA, CcpB regulates the *gntR* regulatory gene and the *gnt* and *xyl* operons, which are involved in the metabolism of gluconate and xylose, respectively [32]. ExuR is another CcpA-paralogous copy in the *Carbon metabolism* module, the activity of which is down-regulated by the phosphorylated state of HPr (See Figure 5). In our previous work [31], we described many other CcpA paralogs in

**Figure 5 Paralogous transcription factors not belonging to a module.** In red are the paralogous TFs belonging to the *carbon metabolism* module. Paralogous TFs associated with the *general stress* response module are indicated in light green. TFs involved in *nucleotide biosynthesis* are shown in dark green. TFs devoted to the *respiration* module are shown in pink. TFs related to the *secretion stress* module are shown in orange. Paralogous regulatory proteins clustered in the *cell differentiation* module are shown in purple. TFs regulating *nitrogen metabolism* are shown in yellow. Paralogous TFs clustered in the *mother cell* module are shown in brown. TFs representing *degradative enzymes* are shown in blue. A group of regulatory genes that were not clustered in a module (grey) but are paralogous to a TF belonging to a module is also shown. A group of structural genes that are related to paralogous TFs are shown in light yellow. The line represents the paralogous relationship between structural and/or regulatory genes.

*B. subtilis* that are different from CcpB and ExuR. These TFs were not included in our network because no regulatory interaction different from σA, has been reported for any of them in the DBTBS database. A similar situation also exists for the paralogous copies of the TFs CcpC and TreR.

## Master regulators govern sporulation and cross-talk with other modules

A topological motif is defined as a statistically over-represented pattern of interconnected nodes and links (subgraphs) in a complex biological network [47]. Recent evidence suggests that motifs in regulatory networks could be a by-product resulting from network organization and evolution [48-51]. Two principal network motifs have

been found in TRNs: the feed forward motif (FF) and the bi-fan motif (BF) [7]. FFs are three-network motifs that comprise two regulatory genes and one target gene (A → B, B → C, A → C), and BFs involve two regulatory genes and two target genes (A → C, A → D, B → C, B → D). Some studies have suggested that FFs play important organizational [49,52] and dynamical [53,54] roles that could explain why they have been selected in TRNs, whereas other studies have shown that the overabundance of BFs does not correlate with any specific functional role [55]. Hence, we only focus our attention in this work on FFs that perform various functional roles, including noise filtering, fine tuning of expression timing, response acceleration, and pulse generation, all of which are well described in the context of sporulation in *B. subtilis* [56].

We used the *CMCS* to exemplify the relevance of FFs and to highlight the possibility that FF regulatory genes perform cross-talk regulation between this module and others. We searched for the entire set of three-node network motifs in the *B. subtilis* TRN, as described in the methods section. Based on this analysis, two motifs were identified: the FF and an alternative version consisting of a two-node feedback circuit between the regulatory nodes. The latter, herein called the complex feed forward motif (CFF), has also been identified in the *E. coli* TRN, which highlights the importance of feedback circuits for TRN organization [47,49]. At the global scale, 89% of the FFs and 100% of the CFFs in the *B. subtilis* TRN are embedded within specific modules, while the remaining FFs enable, at the level of regulatory genes, cross-talk between modules (Table 1).

The two regulatory genes involved in a FF could be classified into master and local regulators. The master regulator governs the expression of the local regulator and the target gene, while the local regulator only governs the expression of the target gene (Table 1). Due to the hierarchical nature of the *B. subtilis* TRN, a gene that is considered to be a master regulator in a given FF could appear as a local regulator in another FF [14,49,57]. However, by considering the number of FFs in which a TF is a master or local regulator, it is possible to infer its role in the entire TFN.

### Applying these criteria to the FFs in the *CMCS* module, we found that SigE and SpoIIID could be classified as master regulators

The gene targets of SpoIIID regulation are also targets of GerE or SigK regulation. In contrast, the gene targets of SigE regulation are also commonly targets of SigK, SpoIIID, or GerR TFs. In some cases, SigE (*CMCS* module), ExuR (carbon metabolism module), and PhoP (respiration module) regulate the same set of genes, thus enabling regulatory cross-talk between these three modules.

A similar analysis of the CFFs indicates that GerE and SigK, which are both involved in a two-node feedback loop, coregulate a large set of common target genes (Figure 6). As discussed in a recent paper published by our group [14], the feedback loops, are not over represented structures but no for that less important. As in the presented circuit formed by the TFs GerE and SigK for which an experimental study has provided evidence showing that this FBL plays a key role in enhancing the robustness of the mother cell network and optimizing the expression of target genes [58].

### The role of σ factors in the TRN

In our initial attempt, we performed an analysis of the *B. subtilis* TRN considering only TFs in the absence of any σ factor (data not shown) and found that the resulting network was biologically meaningless because it was decomposed into a very large number of small modules, many of which shared the same function or metabolic process. For example, we found multiple *Respiratory*, *Sporulation*, and *Carbon compound* modules. In addition, we found that the number of TFs in this TRN was reduced when the σ factors were not considered because they were the only connections of many TFs to the regulatory network. In some cases, the loss of TFs from the TRN led to an absence of specific functional descriptions in the resulting modules. In contrast, the inclusion of σ factors in the *B. subtilis* TRN generated cohesive modules associated with well-defined physiological functions and cell processes that are characteristic of this model organism. Furthermore, the 11 *B. subtilis* σ factors included in the analysis were also grouped into functional and characteristic modules (see Figure 2). For example, the presence of σL in the *Carbon metabolism* module established regulatory links with TFs that regulate genes involved in the metabolism of fructose, levanase, arginine, acetoin, isoleucine, leucine, and valine. In addition, σK and σE participate in a regulatory cascade that is required for sporulation and were clustered in the *CMCS* module, while σH, σF, and σG were clustered in the *Cell differentiation* module and are responsible for different stages of the spore formation: initiation of sporulation, early spore formation, and late spore formation, respectively. It is also important to observe that σE and σF were assigned to different modules and are instrumental in preparing for cell-specific programs after the septum formation. The remaining five σ factors, σB, σM, σW, σX, and YlaC, were organized into one module whose function is associated with general stress response. The σB factor is considered to be the master regulator of the stress response in Gram-positive bacteria, while the other four factors belong to the extracytoplasmic function σ factor family, which is characterized by a response to various stress factors [59]. The aforementioned examples suggest that the inclusion of σ factors in the construction of the TRN provides relevant information regarding their important regulatory

**Table 1 Motif distribution in the *Cascade of the mother cell sporulating* module in the *B. subtilis* TRN**

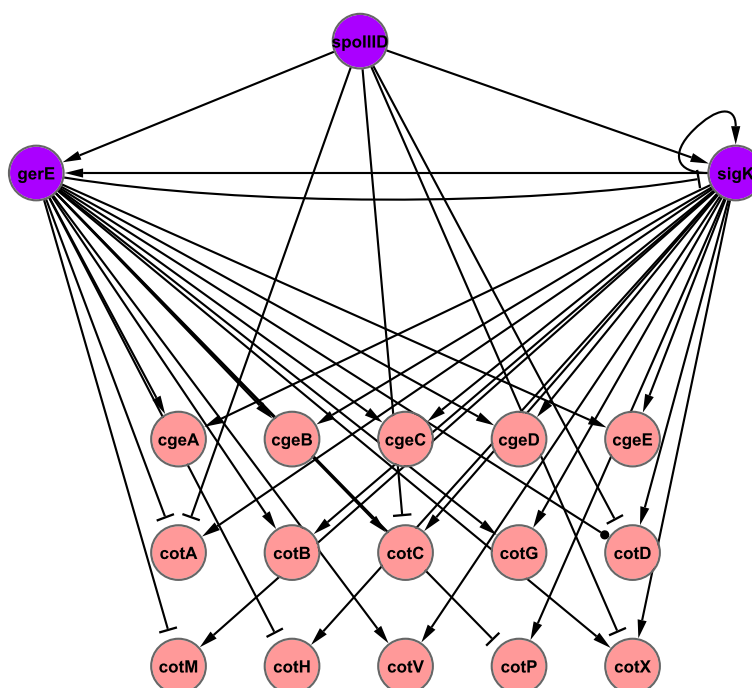| Master regulator (module) | Local regulator (module) | Number of target genes |
| --- | --- | --- |
| SigE (sporulation) | SigK (sporulation) | 11 |
| SigE (sporulation) | SpoIIID (sporulation) | 48 |
| SigE (sporulation) | YlbO (sporulation) | 11 |
| SigE (sporulation) | ExuR (carbon metabolism) | 7 |
| SigE (sporulation) | PhoP (respiration) | 4 |
| SpoIIID (sporulation) | GerE (sporulation) | 4 |
| SpoIIID (sporulation) | SigK (sporulation) | 7 |

**Figure 6 Motif over-representation in the *Cascade of the mother cell sporulating* module.** GerE and SigK TFs are involved in a two-node feedback circuit that coregulates a large set of target genes.

roles in the metabolic and cellular processes that take place in *B. subtilis.*

### Different functions in one module

A meticulous analysis of the regulatory roles of the TFs showed that each module typically has one function or is related to a specific metabolic process. However, in some modules where more than one function was identified, meaningful biological relationships between these functions were discovered. In this section, we discuss these cases.

*The respiration module* cluster has six TFs, ArfM, Fnr, NsrR, ResD, PhoP and HrcA. The first four TFs are part of a regulatory cascade that is triggered in response to changes in oxygen availability in the environment and regulate the expression of genes required for the cell adaptation from aerobic to anaerobic environments, and *vice versa.* PhoP is a TF whose expression is induced when decreasing the presence of phosphate in the medium and shares common regulated gene targets with ResD, that are involved in the first steps of the respiratory process. This concurrent regulation reflects the dependent relationship of the bacterial respiratory process with the phosphate availability in the media. Furthermore, transcription induced by ResD under limited phosphate conditions provides essential components for the transport of electrons, required for the assimilation of inorganic phosphate into ATP [45].

The sixth TF clustered in the *respiration module* is HrcA that participates in heat-shock stress response regulation [60] and is encoded in the *lepA-hemN-hrcA-grpE-dnaKJ-yqeTUV* operon. The heat-shock-related genes of this operon include the *grpE-*, *dnaK-*, and *dnaJ-*encoding chaperons. This operon also encodes for genes required in the respiratory process, such as the GTP-binding protein LepA and the oxygen-independent coproporphyrinogen III oxidase HemN, which participates in heme biosynthesis and anaerobic respiratory energy metabolism [61,62]. The *yqeT* gene codes for a protein that is homologous to the L11 methyltransferase ribosomal protein, while *yqeU* and *yqeV* code for proteins with unknown function [61,62]. Transcription of the *lepA-hemN-hrcA-grpE-dnaKJ-yqeTUV* operon is controlled by at least four promoters that depend on σA, three terminators, and ArfM and is autoregulated by HrcA [62,63]. One start site that responds to aerobic/anaerobic conditions is located 37 bp upstream of the *lepA* translational start codon [63]. The remaining sites are located downstream of *hemN* and respond to heat-shock stress [64]. The means by which ArfM controls the expression of this operon in anaerobic conditions remains unclear, although a cascade of events in which ResD favors the expression of FNR, a mediator of the anaerobic induction of ArfM, may be responsible for the activation of the *lepA-hemN-hrcA-grpE-dnaKJ-yqeTUV* operon [65]. However, more experiments need to be performed to confirm this hypothesis.

The *Cell differentiation* module is another example where different but related functions are observed together. This module contains the most TFs, and although most have a direct implication in the phenotype expressed in each cell fate, there are some genes whose presence warrants further discussion. Two of them code for the LexA and Hbs TFs, which are associated with the care, protection, organization, and structuration of DNA. Hbs has an important effect on gene expression and growth; thus, it is required not only during sporulation, but also during vegetative growth [66]. Mutants of this gene show reduced sporulation efficiency [66]. In contrast, LexA is a master regulator of genes that is involved in DNA damage and the SOS response, and it has a role in coordinating the initiation of sporulation when the cell is exposed to DNA damage. Finally, HutP is directly involved in the use of alternative carbon and nitrogen sources [67], but was clustered in the *Cell differentiation* module because it is regulated by the master regulator AbrB and the CodY TF, which regulate gene expression during the competence state. Additionally, HutP is regulated by CcpA, the master TF for carbon metabolism. Sporulation is a critical cell fate that allows *B. subtilis* to adopt a resistant structure, thereby allowing its survival in extreme, unfavorable conditions. The decision to begin the sporulation process is critical because once the process is started, it will continue until it has been completed, committing these cells to a latent state of viability. For this reason, the transcription regulation of genes involved in sporulation must be concurrently regulated by TFs like HutP and CcpA in response to the availability of nitrogen and carbon sources.

One cluster that certainly represents a module with more than one function is the *General stress response* module. This module is coordinated by the master regulator σB, which controls the expression of more than 150 genes that are involved in *B. subtilis* adaptation to different types of stresses and starvation stimuli typical of its natural ecosystems [68]. A cascade of regulatory events induces the activity of σB, which is a consequence of a switch mechanism that phosphorylates the proteins RbsV and RvsW as the final intermediates of the cascade. RvsW captures σB in a stable complex that prevents the union of σB with RNApol, a condition prevailing in exponential growth. This condition is reverted when *B. subtilis* is exposed to stress, for which two modes of action have been described [68]. The first response is induced by environmental stresses such as heat-shock, salt, acid, ethanol, Mn2+, and blue light [68]. The second response is induced by energy depletion and requires the detection of glucose, oxygen, and phosphate starvation and exposure to agents such as NO, azide, CCCP, and mycophenolic acid [68]. As a consequence of these stresses, a portion of the cells commits to sporulation, but the rest rely on alternative survival strategies. One such strategy is provided by

σB or the structural protein RelA and consists of coordinating a stringent response that can "lead to a vegetative dormancy characterized by drastically reduced anabolic reactions and a prospective protection by a multiple stress resistance machine directed mainly by different sets of TF," as described by Hecker and colleagues [68].

## Functions present in more than one module

In our network analysis, we were able to define regulatory modules with specific metabolic or cellular function; nevertheless, we found two cases in which one function was redundantly regulated by TFs belonging to different modules of the network. These cases were associated with the regulation of heat-shock proteins and the regulation of degradative enzymes synthesis. These cases are discussed below.

One hardship that bacteria must face is growth adaptation to different temperature changes, for which they have developed diverse programs of gene regulation. The response to heat-shock stress is one of the best-known systems and involves the so-called heat-shock proteins (HSPs). In *B. subtilis*, more than 200 genes are induced in response to a heat shock. These genes are regulated through different mechanisms that provide a method of classification. Based on their regulatory elements, HSPs have been classified into six main classes or regulons [69]. Although all of the genes of these regulons are classified as being involved in heat shock, many of them also respond to a variety of other stress stimuli; therefore, their corresponding TFs are included in more than one module. For example, class III is regulated in response to heat-shock stress and oxidative stress by CtsR, which belongs to the *General stress module* [70], while class V is regulated in response to heat-shock and secretion stress by the CssRS system, which belongs to the *Secretion* module.

In contrast, Class II and Class III heat-shock proteins were grouped into the *General stress module* and are regulated by σB in response to heat-shock and a diversity of other stresses. The activity of σB is controlled by a complex signal transduction network that includes at least seven other gene products encoded by the *sigB* operon [71]. This complex regulation might allow the efficient adaptation of *B. subtilis* to a broad diversity of stresses and starvation in non-sporulating conditions [72] (Figure 2). Finally, Class V HSPs were grouped in the *Secretion stress* module that is regulated by the two-component system CssR-CssS, with a possible dual role of CssRS in the heat-shock and secretion stress responses [73]. Our above description of the heat-shock response in *B. subtilis* demonstrates that modular analysis is a tool that can be used to understand potential relationships among the different regulons of the heat-shock stress proteins. Furthermore, this type of analysis could

be expanded by comparing the regulatory networks of different organisms. For example, in the proteobacteria *E. coli*, only one heat-shock regulator, the σ32 has been described at the transcription level, while in *B. subtilis* and others Gram-positive organisms, several mechanisms regulate the heat-shock response. In this regard, the complex regulation of *B. subtilis* may allow this organism to tolerate the extensive stresses that are present in the different and changing environments where it grows [69].

The presence of the same function in different modules was also observed in the transcription regulation of degradative enzymes, which clustered in the *Degradative enzymes* and *Cell differentiation* modules. In the first module, the transcription of the genes coding for degradative enzymes are regulated by TenA and TenI, while genes in the second module are regulated by SenS and DegU, which also regulate the expression of genes responsible for biofilm formation. TenA and TenI seem to be unessential for the production of degradative enzymes in *B. subtilis*. These TFs might compensate for the function of SenS and DegU in the case of mutation. In this and other instances, redundancy in regulatory elements might increase the robustness of biological systems [74,75].

Sporulation is another cellular process that appears in more than one module. In our analysis, we found that different sets of regulatory elements were associated with this process in the *CMCS cell* and *Cell differentiation* modules. Sporulation is primarily controlled by Spo0A and σH, which are both involved in the initiation of this process, and by genes that are expressed in the regulatory cascade that occurs in the forespore. The separation of both groups of genes into two modules originates from the compartmentalization of the mother cell and forespore gene expression during endospore formation, which is triggered by the phosphorylated form of Spo0A.

## Conclusions

The work here presented reviewed the TRN of *B. subtilis* by an exhaustive manual functional annotation of the identified regulatory modules and motif distributions.

In good agreement with previous works, we found that the *B. subtilis* regulatory network displays the typical characteristics of a scale-free network with modular and hierarchical organization. The determination of these properties allowed us to classify the TFs into nine discrete modules that are highly connected in an intra-modular fashion and show a hierarchically organized inter-modular structure (See Figure 2). The detailed literature analysis demonstrated that each module was associated with well-defined physiological functions. Although the modules were not entirely homogeneous in some cases, their components respond to common conditions or stimuli and were mainly regulated by global or pleiotropic TFs as

describe previously for *Saccharomyces cerevisiae* and *Escherichia coli* [76]. In addition to the TFs, we showed that an important element of our analysis was the inclusion of the σ factors. This addition allows the clustering of TFs into larger and more biologically meaningful modules than those obtained in our previous study in *E. coli*, where the σ factors were not considered [6]. The advantages mentioned above seem to be more evident in organisms with many σ factors, such as *B. subtilis*, which has 16 σ factors [1], compared with organisms with fewer σ factors, such as *E. coli*, which has only seven σ factors in its regulatory network [77]. In the particular case of *B. subtilis*, important cell fates, such as sporulation and the general stress response, depend on cascades of σ factors. These regulatory relationships were observed in our analysis of regulatory modules, in which some σ factors can play the role of pleiotropic regulators [14], governing different cell fates and cross-talking; a classification that can be done though the analysis of motifs like FFs, and FBLs and the previously proposed intermodular genes [14]. In good agreement with previous studies on the evolution of cellular networks by duplication events [31,35,37,38], we identified paralogous TFs playing important roles in the TRN of *B. subtilis*. In summary, the analytical methodology utilized in our study provides an excellent approach to integrate and understand the complex regulatory network of *B. subtilis*, a network that modulates the expression of the different mechanisms of adaptation and cell differentiation that this bacterium has evolved in response to the environmental changes that it must constantly face.

## Methods
### Network construction

From DBTBS version 2010 [1], we extracted all the regulatory interactions between TFs, σ factors, and target genes reported in this database. We selected regulatory interactions with strong evidence, using as a parameter the strong and weak evidence classification performed in the RegulonDB database [77]. The full TRN of *B. subitilis* and the TFs and σ factors can be consulted and download from the Additional files 3 and 4 respectively and or the address provided by the journal.

### Statistical analysis of the regulatory networks

The connectivity ($P(k)$), and the clustering coefficient ($C(k)$) distributions of the *B. subtilis* network were obtained as described in Resendis et al. [6]. Our analysis shows that the transcriptional regulatory network of *B. subtilis* follows a scale-free distribution with hierarchical modularity. We also compute (see the methods section) the incoming ($Pin$) and outcoming ($Pout$) degree distributions for three bacteria (*B. subtilis* [14], *E. coli* [78] and *M. tuberculosis* [79]) and one eukaryote (*S. cerevisiae* [80]) to visualize and understand generic

characteristics of the regulatory network topology. Additionally, we computed the coefficient of determination ($R^2$) for each regression as an indicator of the goodness of fit of the power-law model, and then compared each of them against the $R^2$ for a corresponding exponential fit.

In a posterior step, we extracted a sub-network considering only the regulatory interactions of all known *B. subtilis* TFs and σ factors. Over this sub-network, we calculated the shortest path length between every pair of genes ($d_{ij}$ is the shortest path length between gene i and gene j). Next, we calculated the association function ($1/d_{ij}^2$) of these shortest path lengths [3,6]. This function gives a measure of the closeness among genes by amplifying close relationships and minimizing remote distances. The data were used as input to perform a hierarchical agglomerative average linkage clustering using the programs Cluster 3 [81] and TreeView [56] for cluster visualization.

## Modules homogeneity

With the aim to corroborate the homogeneity of the modules obtained by the hierarchical clustering method, we performed three extra clustering algorithms: the Girvan-Newman algorithm [2], the Rosvall-Bergstrom algorithm [13] and the Natural decomposition approach [14].

## Manual annotation of identified modules

We annotated each module as follows. First, we compiled a list of cellular processes in which the genes composing each module were involved. We obtained this information from the DBTBS (version 2010) database [1] and a review of pertinent literature. Then, taking into account the regulatory and physiological context, this information was analyzed using expert biological knowledge to make human inferences about the physiological function of each module.

## Motif identification

We searched for the entire set of three-node network motifs in the *B. subtilis* TRN by using the *mfinder* program [47]. To determine the statistical over-representation of three-node subgraphs, a switching algorithm was used to generate 1,000 random networks that conserved the number of nodes, links, and the degree sequence of the real network. All three-node subgraphs showing a *p*-value < 0.01 were selected as network motifs.

## Availability of supporting data

The data sets and the following information supporting the results of this article are included within a Additional files 1 and 2. An address with down load files of the full TRN of B. *subtilis* and the TFs and σ factors TRN of *B. subtilis* can be download from Additional files 3 and 4 respectively.

## Additional files

**Additional file 1: Lessons from the modular organization of the transcriptional regulatory network of *Bacillus subtilis*.**

**Additional file 2: Table S1.** Comparison between different clustering methods.

**Additional file 3: B. subtilis full TRN.**

**Additional file 4: Transcriptional regulatory network of TFs of Bacillus subtilis.**

#### Abbreviations
TF: DNA-binding transcription factor; TRN: Transcriptional regulatory network; CMCS: Cascade of the mother cell sporulating module; FF: Feed forward motif; CFF: Complex feed forward motif; CCR: Catabolite repression; NDA: Natural decomposition approach; M: Module; Pin: Incoming degree distribution; Pout: Outcoming degree distribution; CCDF: Cumulative complementary distribution.

#### Competing interests
The authors declare that they have no competing interests.

#### Authors' contributions
JAFG reconstructed the TRN, performed topological, module identification and motifs analyses and contribute to writing. AMMC designed the study, collected the data and analyzed the results. EM was involved in draft and revising the manuscript. MMN contributed with the paralogous assignations. EPR was involved in revising the manuscript. RMGR designed the study, collected and analyzed the data, and wrote the paper. All authors read and approved the final manuscript.

#### Author details
[1]Programa de Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Av. Universidad s/n,Col. Chamilpa, Cuernavaca, Morelos 62210, México. [2]Departamentos de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apdo. Postal 510-3, Cuernavaca, Morelos 62250, México. [3]Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apdo. Postal 510-3, Cuernavaca, Morelos 62250, México.

#### References
1. Sierro N, Makita Y, de Hoon M, Nakai K: **DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information.** *Nucleic Acids Res* 2008, **36**:D93–D96.
2. Girvan M, Newman ME: **Community structure in social and biological networks.** *Proc Natl Acad Sci USA* 2002, **99**:7821–7826.
3. Rives AW, Galitski T: **Modular organization of cellular networks.** *Proc Natl Acad Sci USA* 2003, **100**:1128–1133.
4. Guelzim N, Bottani S, Bourgine P, Kepes F: **Topological and causal structure of the yeast transcriptional regulatory network.** *Nat Genet* 2002, **31**:60–63.
5. Jothi R, Balaji S, Wuster A, Grochow JA, Gsponer J, Przytycka TM, Aravind L, Babu MM: **Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture.** *Mol Syst Biol* 2009, **5**:294.
6. Resendis-Antonio O, Freyre-Gonzalez JA, Menchaca-Mendez R, Gutierrez-Rios RM, Martinez-Antonio A, Avila-Sanchez C, Collado-Vides J: **Modular analysis of the transcriptional regulatory network of E. coli.** *Trends Genet* 2005, **21**:16–20.

7. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of Escherichia coli.** *Nat Genet* 2002, **31**:64–68.

8. Albert R: **Scale-free networks in cell biology.** *J Cell Sci* 2005, **118**:4947–4957.

9. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA: **Transcriptional regulatory code of a eukaryotic genome.** *Nature* 2004, **431**:99–104.

10. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M: **Genomic analysis of regulatory network dynamics reveals large topological changes.** *Nature* 2004, **431**:308–312.

11. Balazsi G, Barabasi AL, Oltvai ZN: **Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli.** *Proc Natl Acad Sci USA* 2005, **102**:7841–7846.

12. Balazsi G, Heath AP, Shi L, Gennaro ML: **The temporal response of the Mycobacterium tuberculosis gene regulatory network during growth arrest.** *Mol Syst Biol* 2008, **4**:225.

13. Rosvall M, Bergstrom CT: **Maps of random walks on complex networks reveal community structure.** *Proc Natl Acad Sci USA* 2008, **105**:1118–1123.

14. Freyre-Gonzalez JA, Trevino-Quintanilla LG, Valtierra-Gutierrez IA, Gutierrez-Rios RM, Alonso-Pavon JA: **Prokaryotic regulatory systems biology: common principles governing the functional architectures of bacillus subtilis and escherichia coli unveiled by the natural decomposition approach.** *J Biotechnol* 2012, **161**:278–286.

15. Doroshchuk NA, Gel'fand MS, Rodionov DA: **Regulation of nitrogen metabolism in gram-positive bacteria.** *Mol Biol (Mosk)* 2006, **40**:919–926.

16. Wang L, Grau R, Perego M, Hoch JA: **A novel histidine kinase inhibitor regulating development in Bacillus subtilis.** *Genes Dev* 1997, **11**:2569–2579.

17. Turner RJ, Lu Y, Switzer RL: **Regulation of the Bacillus subtilis pyrimidine biosynthetic (pyr) gene cluster by an autogenous transcriptional attenuation mechanism.** *J Bacteriol* 1994, **176**:3708–3722.

18. Bera AK, Zhu J, Zalkin H, Smith JL: **Functional dissection of the Bacillus subtilis pur operator site.** *J Bacteriol* 2003, **185**:4099–4109.

19. Smith E, Morowitz HJ: **Universality in intermediary metabolism.** *Proc Natl Acad Sci USA* 2004, **101**:13168–13173.

20. Hyyrylainen HL, Bolhuis A, Darmon E, Muukkonen L, Koski P, Vitikainen M, Sarvas M, Pragai Z, Bron S, van Dijl JM, Kontinen VP: **A novel two-component regulatory system in Bacillus subtilis for the survival of severe secretion stress.** *Mol Microbiol* 2001, **41**:1159–1172.

21. Pang AS, Nathoo S, Wong SL: **Cloning and characterization of a pair of novel genes that regulate production of extracellular enzymes in Bacillus subtilis.** *J Bacteriol* 1991, **173**:46–54.

22. Nakano MM, Zuber P: **Anaerobic growth of a "strict aerobe" (Bacillus subtilis).** *Annu Rev Microbiol* 1998, **52**:165–190.

23. Fujita Y, Fujita T, Miwa Y, Nihashi J, Aratani Y: **Organization and transcription of the gluconate operon, gnt, of Bacillus subtilis.** *J Biol Chem* 1986, **261**:13744–13753.

24. Yamamoto H, Serizawa M, Thompson J, Sekiguchi J: **Regulation of the glv operon in bacillus subtilis: YfiA (GlvR) is a positive regulator of the operon that is repressed through CcpA and cre.** *J Bacteriol* 2001, **183**:5110–5121.

25. Kallio PT, Fagelson JE, Hoch JA, Strauch MA: **The transition state regulator Hpr of Bacillus subtilis is a DNA-binding protein.** *J Biol Chem* 1991, **266**:13411–13417.

26. Kim HJ, Jourlin-Castelli C, Kim SI, Sonenshein AL: **Regulation of the bacillus subtilis ccpC gene by ccpA and ccpC.** *Mol Microbiol* 2002, **43**:399–410.

27. Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, Losick R: **The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis.** *PLoS Biol* 2004, **2**:e328.

28. Kroos L, Zhang B, Ichikawa H, Yu YT: **Control of sigma factor activity during Bacillus subtilis sporulation.** *Mol Microbiol* 1999, **31**:1285–1294.

29. Lopez D, Vlamakis H, Kolter R: **Generation of multiple cell types in Bacillus subtilis.** *FEMS Microbiol Rev* 2009, **33**:152–163.

30. Petersohn A, Brigulla M, Haas S, Hoheisel JD, Volker U, Hecker M: **Global analysis of the general stress response of Bacillus subtilis.** *J Bacteriol* 2001, **183**:5617–5631.

31. Martinez-Nunez MA, Perez-Rueda E, Gutierrez-Rios RM, Merino E: **New insights into the regulatory networks of paralogous genes in bacteria.** *Microbiology* 2010, **156**:14–22.

32. Brown NL, Stoyanov JV, Kidd SP, Hobman JL: **The MerR family of transcriptional regulators.** *FEMS Microbiol Rev* 2003, **27**:145–163.

33. Zalieckas JM, Wray LV Jr, Fisher SH: **Cross-regulation of the Bacillus subtilis glnRA and tnrA genes provides evidence for DNA binding site discrimination by GlnR and TnrA.** *J Bacteriol* 2006, **188**:2578–2585.

34. Nourmohammad A, Lassig M: **Formation of regulatory modules by local sequence duplication.** *PLoS Comput Biol* 2011, **7**:e1002167.

35. Singh LN, Hannenhalli S: **Correlated changes between regulatory cis elements and condition-specific expression in paralogous gene families.** *Nucleic Acids Res* 2010, **38**:738–749.

36. Sawers G, Kaiser M, Sirko A, Freundlich M: **Transcriptional activation by FNR and CRP: reciprocity of binding-site recognition.** *Mol Microbiol* 1997, **23**:835–845.

37. Tanay A, Regev A, Shamir R: **Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast.** *Proc Natl Acad Sci USA* 2005, **102**:7203–7208.

38. Teichmann SA, Babu MM: **Gene regulatory network growth by duplication.** *Nat Genet* 2004, **36**:492–496.

39. Saxild HH, Nygaard P: **Genetic and physiological characterization of Bacillus subtilis mutants resistant to purine analogs.** *J Bacteriol* 1987, **169**:2977–2983.

40. Sinha SC, Krahn J, Shin BS, Tomchick DR, Zalkin H, Smith JL: **The purine repressor of Bacillus subtilis: a novel combination of domains adapted for transcription regulation.** *J Bacteriol* 2003, **185**:4087–4098.

41. Serizawa M, Kodama K, Yamamoto H, Kobayashi K, Ogasawara N, Sekiguchi J: **Functional analysis of the YvrGHb two-component system of Bacillus subtilis: identification of the regulated genes by DNA microarray and northern blot analyses.** *Biosci Biotechnol Biochem* 2005, **69**:2155–2169.

42. Conant GC, Wagner A: **Convergent evolution of gene circuits.** *Nat Genet* 2003, **34**:264–266.

43. Lawhorn BG, Gerdes SY, Begley TP: **A genetic screen for the identification of thiamin metabolic genes.** *J Biol Chem* 2004, **279**:43555–43559.

44. Fabret C, Feher VA, Hoch JA: **Two-component signal transduction in Bacillus subtilis: how one organism sees its world.** *J Bacteriol* 1999, **181**:1975–1983.

45. Birkey SM, Liu W, Zhang X, Duggan MF, Hulett FM: **Pho signal transduction network reveals direct transcriptional regulation of one two-component system by another two-component regulator: bacillus subtilis PhoP directly regulates production of ResD.** *Mol Microbiol* 1998, **30**:943–953.

46. Whitworth DE, Cock PJ: **Evolution of prokaryotic two-component systems: insights from comparative genomics.** *Amino Acids* 2009, **37**:459–466.

47. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks.** *Science* 2002, **298**:824–827.

48. Cordero OX, Hogeweg P: **Feed-forward loop circuits as a side effect of genome evolution.** *Mol Biol Evol* 2006, **23**:1931–1936.

49. Freyre-Gonzalez JA, Alonso-Pavon JA, Trevino-Quintanilla LG, Collado-Vides J: **Functional architecture of Escherichia coli: new insights provided by a natural decomposition approach.** *Genome Biol* 2008, **9**:R154.

50. Mazurie A, Bottani S, Vergassola M: **An evolutionary and functional assessment of regulatory network motifs.** *Genome Biol* 2005, **6**:R35.

51. Sole RV, Valverde S: **Are network motifs the spandrels of cellular complexity?** *Trends Ecol Evol* 2006, **21**:419–422.

52. Macia J, Widder S, Sole R: **Specialized or flexible feed-forward loop motifs: a question of topology.** *BMC Syst Biol* 2009, **3**:84.

53. Mangan S, Alon U: **Structure and function of the feed-forward loop network motif.** *Proc Natl Acad Sci USA* 2003, **100**:11980–11985.

54. Wall ME, Dunlop MJ, Hlavacek WS: **Multiple functions of a feed-forward-loop gene circuit.** *J Mol Biol* 2005, **349**:501–514.

55. Ingram PJ, Stumpf MP, Stark J: **Network motifs: structure does not determine function.** *BMC Genomics* 2006, **7**:108.

56. de Hoon MJ, Eichenberger P, Vitkup D: **Hierarchical evolution of the bacterial sporulation network.** *Curr Biol* 2010, **20**:R735–R745.

57. Dobrin R, Beg QK, Barabasi AL, Oltvai ZN: **Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network.** *BMC Bioinformatics* 2004, **5**:10.

58. Wang L, Perpich J, Driks A, Kroos L: **One perturbation of the mother cell gene regulatory network suppresses the effects of another during sporulation of Bacillus subtilis.** *J Bacteriol* 2007, **189**:8467–8473.

59. Ryu HB, Shin I, Yim HS, Kang SO: **YlaC is an extracytoplasmic function (ECF) sigma factor contributing to hydrogen peroxide resistance in Bacillus subtilis.** *J Microbiol* 2006, **44**:206–216.

60. Schulz A, Schumann W: **hrcA, the first gene of the Bacillus subtilis dnaK operon encodes a negative regulator of class I heat shock genes.** *J Bacteriol* 1996, **178**:1088–1093.

61. Homuth G, Heinemann M, Zuber U, Schumann W: **The genes of lepA and hemN form a bicistronic operon in Bacillus subtilis.** *Microbiology* 1996, **142**(Pt 7):1641–1649.

62. Homuth G, Masuda S, Mogk A, Kobayashi Y, Schumann W: **The dnaK operon of Bacillus subtilis is heptacistronic.** *J Bacteriol* 1997, **179**:1153–1164.

63. Hippler B, Homuth G, Hoffmann T, Hungerer C, Schumann W, Jahn D: **Characterization of Bacillus subtilis hemN.** *J Bacteriol* 1997, **179**:7181–7185.

64. Magos L, Kovacs G: **Compensatory changes in respiration against resistance.** *Acta Physiol Hung* 1956, **9**:223–230.

65. Homuth G, Rompf A, Schumann W, Jahn D: **Transcriptional control of Bacillus subtilis hemN and hemZ.** *J Bacteriol* 1999, **181**:5922–5929.

66. Micka B, Marahiel MA: **The DNA-binding protein HBsu is essential for normal growth and development in Bacillus subtilis.** *Biochimie* 1992, **74**:641–650.

67. Kumarevel T: **Structural insights of HutP-mediated regulation of transcription of the hut operon in Bacillus subtilis.** *Biophys Chem* 2007, **128**:1–12.

68. Hecker M, Pane-Farre J, Volker U: **SigB-dependent general stress response in Bacillus subtilis and related gram-positive bacteria.** *Annu Rev Microbiol* 2007, **61**:215–236.

69. Schumann W: **The Bacillus subtilis heat shock stimulon.** *Cell Stress Chaperones* 2003, **8**:207–217.

70. Derre I, Rapoport G, Msadek T: **CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in gram-positive bacteria.** *Mol Microbiol* 1999, **31**:117–131.

71. Kruger E, Hecker M: **The first gene of the Bacillus subtilis clpC operon, ctsR, encodes a negative regulator of its own operon and other class III heat shock genes.** *J Bacteriol* 1998, **180**:6681–6688.

72. Hecker M, Schumann W, Volker U: **Heat-shock and general stress response in Bacillus subtilis.** *Mol Microbiol* 1996, **19**:417–428.

73. Darmon E, Noone D, Masson A, Bron S, Kuipers OP, Devine KM, van Dijl JM: **A novel class of heat and secretion stress-responsive genes is controlled by the autoregulated CssRS two-component system of Bacillus subtilis.** *J Bacteriol* 2002, **184**:5661–5671.

74. Babu MM: **Early Career Research Award Lecture. Structure, evolution and dynamics of transcriptional regulatory networks.** *Biochem Soc Trans* 2010, **38**:1155–1178.

75. Gutierrez-Rios RM, Rosenblueth DA, Loza JA, Huerta AM, Glasner JD, Blattner FR, Collado-Vides J: **Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles.** *Genome Res* 2003, **13**:2435–2443.

76. Yu H, Gerstein M: **Genomic analysis of the hierarchical structure of regulatory networks.** *Proc Natl Acad Sci USA* 2006, **103**:14724–14731.

77. Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavaleta A, Muniz-Rascado L, Solano-Lira H, Jimenez-Jacinto V, Weiss V, Garcia-Sotelo JS, Lopez-Fuentes A, Porron-Sotelo L, Alquicira-Hernandez S, Medina-Rivera A, Martinez-Flores I, Alquicira-Hernandez K, Martinez-Adame R, Bonavides-Martinez C, Miranda-Rios J, Huerta AM, Mendoza-Vargas A, Collado-Torres L, Taboada B, Vega-Alvarado L, Olvera M, Olvera L, Grande R, Morett E, Collado-Vides J: **RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units).** *Nucleic Acids Res* 2011, **39**:D98–105.

78. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muniz-Rascado L, Garcia-Sotelo JS, Weiss V, Solano-Lira H, Martinez-Flores I, Medina-Rivera A, Salgado-Osorio G, Alquicira-Hernandez S, Alquicira-Hernandez K, Lopez-Fuentes A, Porron-Sotelo L, Huerta AM, Bonavides-Martinez C, Balderas-Martinez YI, Pannier L, Olvera M, Labastida A, Jimenez-Jacinto V, Vega-Alvarado L, Del Moral-Chavez V, Hernandez-Alvarez A, Morett E, Collado-Vides J: **RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more.** *Nucleic Acids Res* 2013, **41**:D203–D213.

79. Rohde KH, Veiga DF, Caldwell S, Balazsi G, Russell DG: **Linking the transcriptional profiles and the physiological states of Mycobacterium tuberculosis during an extended intracellular infection.** *PLoS Pathog* 2012, **8**:e1002769.

80. Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L: **Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast.** *J Mol Biol* 2006, **360**:213–227.

81. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863–14868.