

# CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers

Antonin Marchais,<sup>1</sup> Sandra Duperrier,<sup>2</sup> Sylvain Durand,<sup>2</sup> Daniel Gautheret<sup>1</sup> and Patrick Stragier<sup>2,\*</sup>

<sup>1</sup>CNRS-UMR8621; Institut de Génétique et Microbiologie; Université Paris-Sud; Orsay, France; <sup>2</sup>CNRS-UPR9073; Institut de Biologie Physico-Chimique; Université Paris-Diderot; Paris, France

**E**ndospore formation is a characteristic shared by some Bacilli and Clostridia that involves the creation of two cell types, the forespore and the mother cell. Hundreds of protein-encoding genes have been shown to be transcribed in a cell-specific fashion during this developmental process in *Bacillus subtilis*. We have used a phylogenetic profiling procedure to identify clusters of *B. subtilis* coding and non-coding sequences that co-occur in other endospore formers. One such cluster shows a strong bias for sporulation-related genes (42% among 156 genes) and is enriched in potential non-coding RNAs. We have studied one RNA candidate, encoded in the *ylbG-ylbH* interval. In vivo analysis using a transcriptional fusion to the *Escherichia coli lacZ* gene demonstrates that this region of the chromosome contains a gene, *csfG*, encoding a 147-nucleotide RNA that is transcribed only during sporulation, specifically in the forespore. *csfG* is present in many endospore formers, mostly Bacilli and some Clostridia, whereas it is absent from bacteria that do not produce endospores. All CsfG RNAs contain a strongly conserved, pyrimidine-rich, central motif that overlaps a potential stem-loop structure. The remarkable conservation of this sequence in widely divergent bacteria suggests that it plays a conserved physiological role, presumably by interacting with an unidentified target in the forespore, where it contributes to the acquisition of the spore properties.

## Introduction

Endospore formation is a primitive developmental process observed in some Bacilli and Clostridia when subjected to nutrient starvation. These bacteria belong to the phylum of Firmicutes that also contains many bacteria unable to sporulate, among which members of the Listeria, Staphylococcus, Streptococcus and Lactobacillus orders. The morphological process of sporulation appears to be extremely similar in all endospore formers, starting with a polar division that creates two cells of different sizes and different fates, the smaller forespore and the larger mother cell. Engulfment of the forespore by the mother cell is followed by the synthesis of peptidoglycan and protein layers surrounding the forespore and contributing to its resistance properties after its release in the environment by lysis of the mother cell.<sup>1</sup> Sporulation has been extensively studied in *Bacillus subtilis* because of the amenability of this bacterium to classical and reverse genetics. Transcriptome experiments have identified several hundreds of protein-encoding genes that are expressed at some stage of sporulation in one or the other cell.<sup>2-4</sup> It is now well established that the developmental process is initiated by accumulation of the Spo0A-P regulatory protein and the  $\sigma^H$  transition sigma factor, whereas a cascade of four sigma factors,  $\sigma^F$  followed by  $\sigma^G$  in the forespore, and  $\sigma^E$  followed by  $\sigma^K$  in the mother cell, is responsible for the temporal and spatial activation of sporulation-specific genes after septation.<sup>5</sup>

**Key words:** small RNA, sporulation, germination, forespore, Bacilli, Clostridia

Submitted: 12/14/10

Revised: 01/24/11

Accepted: 01/26/11

DOI: 10.4161/rna.8.3.14998

\*Correspondence to: Patrick Stragier;  
Email: stragier@ibpc.fr

**Table 1.** A *B. subtilis* cluster enriched in sporulation genes

<b>Entry into sporulation (Spo0A~P and/or <math>\sigma^H</math>) [10]<sup>a</sup></b>	<b>Early mother cell (<math>\sigma^E</math>) [27]</b>	<b>Vegetative genes affecting sporulation [4]</b>		
<i>abrB</i> <i>kinE</i> <i>qcrB</i> <i>sigF</i> <i>spo0A</i> <i>spo0F</i> <i>spoIIAB</i> <i>spoIIIE</i> <i>ymaH</i> = <i>hfq</i> <i>ypgR</i>	<i>asnO</i> <i>cotJC</i> <i>cypA</i> <i>kamA</i> <i>mipA</i> <i>prkA</i> <i>spmB</i> <i>spoIIIAA</i> <i>spoIIID</i> <i>spoIVA</i> <i>spoIVCB</i> <i>spoVD</i> <i>sqhC</i> <i>usd</i> <i>yabP</i> <i>yhaM</i> <i>yhbH</i> <i>yhfN</i> <i>ylaK</i> <i>yloS</i> = <i>thiN</i> <i>ylxY</i> <i>yngE</i> <i>yqfY</i> = <i>ispG</i> <i>yqiQ</i> = <i>prpB</i> <i>ytpB</i> <i>ywjF</i> = <i>fadF</i> <i>ywIC</i>	<i>hpr</i> = <i>scoC</i> <i>menH</i> <i>spolIIIJ</i> <i>yIbF</i>	<i>pyrD</i> <i>radC</i> <i>rbfA</i> <i>ribT</i> = <i>ypzK</i> <i>rnc</i> <i>rnmV</i> <i>rocF</i> = <i>argI</i> <i>speA</i> <i>speD</i> <i>speE</i> <i>tagO</i> <i>topB</i> <i>udk</i> <i>yabA</i> <i>yabB</i> <i>yabN</i> <i>ybxG</i> <i>ycbL</i> <i>yceD</i> <i>yceF</i> <i>ydfO</i> <i>yfKE</i> <i>yfIG</i> = <i>mapB</i> <i>yhaO</i> <i>yjIC</i> <i>yjID</i> <i>ykfA</i> <i>ykqB</i> = <i>ktrC</i> <i>ykrT</i> = <i>mtnK</i> <i>ykrW</i> <i>ykrZ</i> = <i>mtnD</i> <i>ykul</i> <i>ykVM</i> = <i>queF</i> <i>ylxP</i> <i>yneT</i> <i>yoaS</i> <i>yobN</i> <i>ypbD</i> <i>ypvA</i> <i>yqeT</i> <i>yqfF</i>	<i>yqfU</i> <i>yqzB</i> = <i>ccpN</i> <i>yrhF</i> <i>ysbB</i> <i>ysxB</i> <i>ytjP</i> <i>ytol</i> <i>yugJ</i> <i>yusL</i> = <i>fadM</i> <i>yutE</i> <i>yutM</i> <i>yvfS</i> <i>yvfV</i> <i>yvgM</i> <i>yvoF</i> <i>yvyE</i> <i>ywhB</i> <i>ywIG</i> <i>yxdJ</i>
<b>Early forespore (<math>\sigma^F</math>) [8]</b>		<b>Other genes [90]</b>		<b>CNE [19]<sup>b</sup></b>
<i>mfd</i> <i>pgk</i> <i>sigG</i> <i>yhbA</i> <i>yjbA</i> <i>yqhH</i> <i>ytfJ</i> <i>yuiC</i>	<i>sqhC</i> <i>usd</i> <i>yabP</i> <i>yhaM</i> <i>yhbH</i> <i>yhfN</i> <i>ylaK</i> <i>yloS</i> = <i>thiN</i> <i>ylxY</i> <i>yngE</i> <i>yqfY</i> = <i>ispG</i> <i>yqiQ</i> = <i>prpB</i> <i>ytpB</i> <i>ywjF</i> = <i>fadF</i> <i>ywIC</i>	<i>ackA</i> <i>acoA</i> <i>addB</i> <i>ahpF</i> <i>ahrC</i> <i>araB</i> <i>bcd</i> <i>bkdAB</i> <i>codV</i> <i>dhbB</i> <i>fer</i> <i>ffh</i> <i>folB</i> <i>folK</i> <i>ftsX</i> <i>fur</i> <i>groES</i> <i>hemY</i> <i>hemZ</i> <i>infA</i> <i>infC</i> <i>luxS</i> <i>mntR</i> <i>motA</i> <i>mraW</i> <i>murD</i> <i>mutSB</i> <i>paiA</i> <i>paiB</i> <i>pps</i>		<i>rpoB-rpoC</i> <i>rplW-rplB</i> <i>rplB-rpsS</i> <i>yfmA-yfIT</i> <i>yjbH-yizD</i> <i>metE</i> S-box riboswitch <i>mtnKA</i> S-box riboswitch <i>yIbG-yIbH*</i> <i>polC-ylxS*</i> <i>pbpX-ymdA</i> <i>yneE-yneF</i> <i>yoaZ-penP</i> <i>alaS</i> T-box riboswitch <i>gapB-speD</i> <i>phoR-polA</i> <i>thil-sspA</i> <i>yugH-yugI</i> <i>rho-rpmEA</i> <i>yybO-yybP*</i>
<b>Late forespore (<math>\sigma^G</math>) [11]</b>	<b>Late mother cell (<math>\sigma^K</math>) [6]</b>			
<i>sIeB</i> <i>spoIVB</i> <i>spoVAE</i> <i>spoVAF</i> <i>spoVT</i> <i>sspA</i> <i>sspB</i> <i>sspC</i> <i>sspF</i> <i>sspl</i> <i>tepA</i>	<i>dapG</i> <i>gerE</i> <i>ykoY</i> <i>ypbH</i> = <i>mecB</i> <i>ytlA</i> <i>ytlD</i>			

<sup>a</sup>Protein encoding genes are grouped according to their transcription dependencies and their number is shown for each subset. Alternate names are indicated. <sup>b</sup>The conserved noncoding elements (CNE) of unknown function are labeled by their bracketing genes. Asterisks point to previously identified CNEs.

Analysis of the complete genomes of more than 50 endospore formers, including many Clostridia,<sup>6</sup> indicates the conserved presence of the major regulatory players identified in *B. subtilis*, in accordance with an early suggestion that all modern endospore formers derive from a single sporulating ancestor.<sup>7</sup>

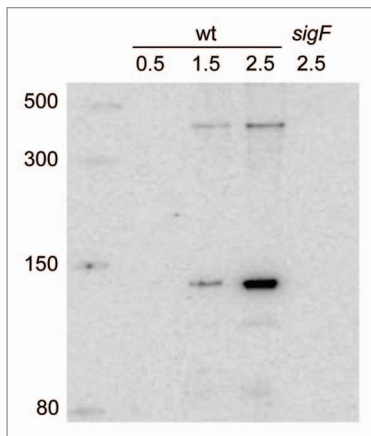
Until recently, only proteins were considered to be involved in the control of sporulation. However, with the identification of many small non-coding RNAs (sRNAs) in a wide variety of bacteria, where they modulate gene expression of multiple regulons, including complex processes such as biofilm formation,<sup>8</sup> morphological development<sup>9</sup> or host-pathogen

interaction,<sup>10</sup> it becomes essential to investigate the presence and the role of sRNAs during sporulation. This question has been addressed by Silvaggi et al. who combined microarray transcriptional profiling of intergenic regions in *B. subtilis* with a computational screen for potential RNA secondary structures conserved in other endospore formers.<sup>11</sup> Importantly for what follows, Silvaggi and co-workers eliminated the candidates for which a role as a cis-regulatory RNA element had been previously reported. They identified one sRNA that accumulates early during sporulation before septation and for which no ortholog was indicated, a second sRNA synthesized under the control

of  $\sigma^K$  in the mother cell and also present in *Bacillus anthracis*, and finally a complex pattern of sRNAs synthesized from an intergenic region under the control of  $\sigma^G$  in the forespore and  $\sigma^K$  in the mother cell and conserved in *Bacillus halodurans*. Recently, an additional sRNA, identified only in *B. subtilis*, was reported to be expressed under the control of  $\sigma^E$  in the mother cell.<sup>12</sup> Disrupting the intergenic regions encoding these various sRNAs did not lead to any obvious sporulation defect.

Another study<sup>13</sup> focused exclusively on *B. subtilis* intergenic regions larger than 500 bp and did not detect any sporulation related sRNA. A comprehensive,





**Figure 2.** Northern blot detection of CsfG. RNA was extracted from *B. subtilis* cells (either wild type or deleted for the *sigF* gene) grown in DSM medium and harvested at various times after entry into sporulation. Hours of sporulation are indicated. Total RNA was separated by electrophoresis on an 8% polyacrylamide gel, transferred to a nylon membrane, and CsfG RNA was detected by hybridization with a complementary oligonucleotide. Samples were run alongside ssRNA markers (NEB) with the indicated sizes.

(for *alaS*, *metE* and the *mtnKA* operon), two genetic intervals previously reported as potentially expressing a non-coding RNA (next to *ylbH* and to *yybP*),<sup>16</sup> and one region already identified as producing several sRNAs during sporulation (the *polC-ylxS* interval).<sup>11</sup> The latter result confirms the existence of a specific functional bias among the 175 members of the cluster and suggests that sporulation related sRNAs might be encoded by some of the other CNEs.

**A forespore-specific sRNA.** Our attention was drawn to the CNE corresponding to the *ylbG-ylbH* interval (Fig. 1A). This CNE overlaps an annotated cis-acting RNA named “*ylbH* leader” (RFAM entry RF00516),<sup>16</sup> and was therefore excluded from the analysis of Silvaggi et al. However, we observed that this CNE is most likely expressed from the reverse strand as a trans-acting RNA, since it is followed on this strand by a run of Ts that could be part of a strong rho-independent transcription terminator. Moreover, aligning this region with the orthologous regions from three closely related Bacilli reveals an impressive sequence conservation that starts immediately downstream of a putative  $\sigma^F$ -controlled promoter and

stops immediately after the putative transcription terminator (Fig. 1B). A consensus, fully-conserved -35 recognition motif (GTATA) is followed at the optimal 15 bp distance by a sequence closely related to the consensus -10 recognition motif, Gg--A-AHTR, where H is A or C or T and R is A or G.<sup>4</sup> Some  $\sigma^F$ -controlled promoters are also recognized by  $\sigma^G$  because of an overlap of sequence recognition by the two sigma factors. This is apparently the case here since the conserved putative promoter motifs fit well with the consensus for  $\sigma^G$ -dependent promoters, tGHATA (17–18 bp) MAWAMTA, where M is A or C, and W is A or T.<sup>4</sup>

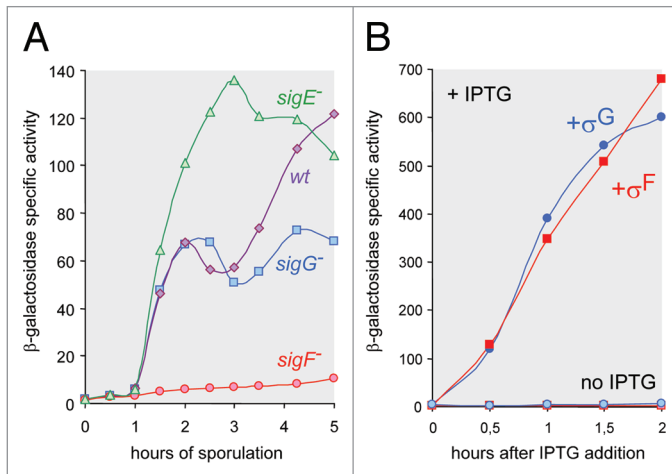
If this promoter is actually functional it would lead to the synthesis of a 147-nt RNA starting with an A residue at position 1569347 in the revised *B. subtilis* genome sequence (NCBI accession number NC\_000964) and ending with a run of seven Us at position 15692001, in the opposite orientation to the adjacent *ylbG* and *ylbH* genes (Fig. 1A). To confirm the existence of this RNA we carried out a northern blot analysis on total RNA extracted from cells harvested at early stages of sporulation. In our growth conditions polar septation takes place one hour after the entry into stationary phase. As shown in Figure 2 an RNA of the expected size is present 30 minutes after septation (lane 1.5), but absent 30 minutes earlier (lane 0.5). This small RNA accumulates during the next hour (lane 2.5) but is not synthesized in the absence of  $\sigma^F$ . The probe hybridizes with a less abundant, larger RNA (ca. 400 nt), that follows the same temporal pattern. Since no promoter sequence can be identified at the corresponding position upstream, on the strand complementary to *ylbH*, it is likely that this larger RNA results from readthrough transcription and therefore ends around the beginning of the *ylbG* gene. In any case it is a very minor species as compared to the smaller RNA.

A 11-codon reading frame is present in the four RNAs encoded by the orthologous DNA regions shown in Figure 1B, starting with a UUG triplet at position 22 in the *B. subtilis* RNA. However, the corresponding ribosome binding site (GAGG) is conspicuously absent in the sequences from the three other Bacilli, making it unlikely that

this small open reading frame is actually translated. Moreover, this reading frame is missing in orthologs of this RNA from more distant organisms discussed hereafter. Potential base pairing is identical in these Bacilli RNAs and comprises four stem-loop structures identified by specific colors in Figure 1B. We note that the second loop is 100% conserved between the four Bacilli, whereas the three other loops are highly variable. Also, 14 out of 15 nucleotides between the second and the third stem-loop are fully conserved.

In order to assess quantitatively the synthesis of this sRNA in *B. subtilis* we fused its promoter region to the *Escherichia coli lacZ* gene and integrated the resulting transcriptional fusion into the *B. subtilis* chromosome at the non-essential *amyE* locus. The cloned fragment contains 132 bp upstream of the predicted transcription start as well as the first 35 bp from the transcribed region. As shown in Figure 3A,  $\beta$ -galactosidase synthesis starts 1-hour after the initiation of the sporulation process and is abolished in the absence of  $\sigma^F$ . A second wave of  $\beta$ -galactosidase synthesis is observed 2 hours later in a wild-type strain but is missing in the absence of  $\sigma^G$ . Finally, in a mutant strain where the absence of  $\sigma^F$  leads to the formation of two forespore compartments, the level of  $\beta$ -galactosidase accumulation is roughly double of what is found in a wild-type strain. Altogether, these results fit perfectly with the presence of a promoter recognized by both  $\sigma^F$  and  $\sigma^G$ , and activated exclusively during sporulation, in the forespore. This was further confirmed by inducing expression of the *lacZ* fusion in exponentially growing cells that were engineered to produce  $\sigma^F$  or  $\sigma^G$  in response to the addition of IPTG. The results shown in Figure 3B indicate that both sigma factors are equally efficient in activating transcription of the fusion. From now on, the gene encoding this sRNA will be called *csfG* (controlled by sigma-F and sigma-G).

Most of the *csfG* gene was replaced by a kanamycin-resistance cassette, eliminating 9 bp from the promoter region and 123 bp from the RNA encoding region. A *B. subtilis* strain harboring this mutation produces as much heat-resistant spores as the wild-type strain



**Figure 3.** Forespore-specific transcription of the *csfG* gene. (A) Time course expression of a transcriptional *csfG-lacZ* fusion integrated into the *B. subtilis* chromosome in various mutant backgrounds. Cells were grown in DSM. (B) Induction of the same fusion in cells growing exponentially in 2x YT medium and harboring a plasmid allowing synthesis of  $\sigma^F$  or  $\sigma^G$  in the presence of IPTG. A control experiment was carried out in the absence of IPTG.

and does not show any obvious growth defect (data not shown). To investigate more subtle phenotypes, we subjected the mutant strain to successive cycles of growth, sporulation and germination, in competition with a wild-type strain. After 10 such cycles the ratio of mutant vs. wild-type bacteria shifted from 0.5–0.5 to 0.02–0.98, indicating a loss of about 29% of the mutant strain at each cycle, presumably as a consequence of a slight germination delay of the mutant spores (data not shown).

**A highly conserved motif.** As expected from its belonging to a cluster of genes co-occurring in several endospore formers and strongly enriched in sporulation genes *csfG* is present in a wide range of Bacilli and Clostridia. It is always found immediately upstream of and diverging from a *ylbH* ortholog. We identified a *csfG* ortholog in the 21 *Bacillaceae* that we analyzed, in *Alicyclobacillaceae* and in *Paenibacillaceae*, but not in *Planococcaceae* (e.g., *Lysinibacillus sphaericus*). We also found a *csfG* ortholog in various Clostridia, all of them (but *Dethiobacter alkaliphilus*) being thermophiles. Conversely, *csfG* is absent from bacteria that do not form endospores.

A multiple sequence alignment of 24 *CsfG* orthologs is shown in **Figure 4** where potential base pairing is indicated. Whereas *CsfG* orthologs from closely related Bacilli are conserved throughout their whole length (**Fig. 1B**) this extended

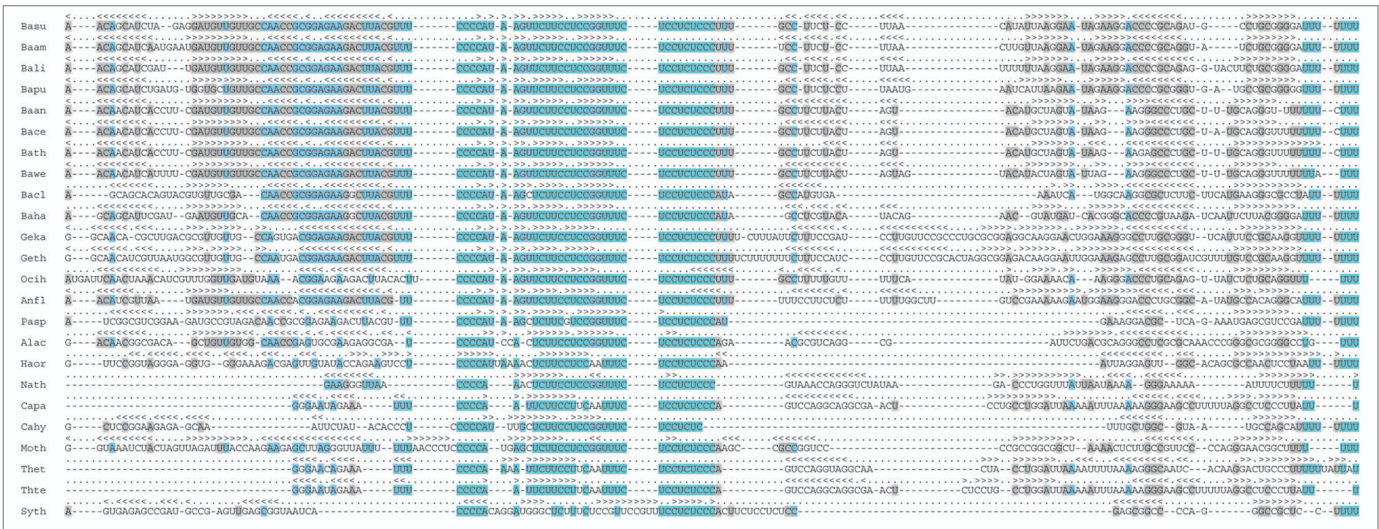
alignment pinpoints the striking conservation of a central sequence that constitutes the second half of a stem-loop structure and its adjacent unpaired region. All *CsfG* sequences are predicted to also contain a rho-independent transcription terminator. The remaining predicted secondary structures are variable, with fewer stem loops in sequences originating from Clostridia in accordance with their shorter size. Examples of the predicted secondary structures of a few *CsfG* RNAs are shown in **Figure 5** where the conserved central motif is highlighted.

All *CsfG* sequences from Bacilli share an extended conserved motif consisting of UCCCCAU, followed at 4 nt by UCUUC, and one nt further by a strictly conserved 18-residue, pyrimidine-rich sequence. A shorter subset of this motif is present in the *CsfG* sequences from Clostridia. The conserved CCCCCA block is predicted to be in a loop, the UCUUC block to be included in a stem, and the long UC-rich sequence to be at least partially unpaired. The remarkable conservation of this sequence and its specific arrangement suggest that this part of *CsfG* could be interacting with an RNA target, following a strand displacement mechanism similar to what has been described for DsrA in *E. coli*.<sup>17</sup> It could also be the recognition motif for an RNA binding protein as previously described for some other highly conserved RNAs.<sup>8</sup>

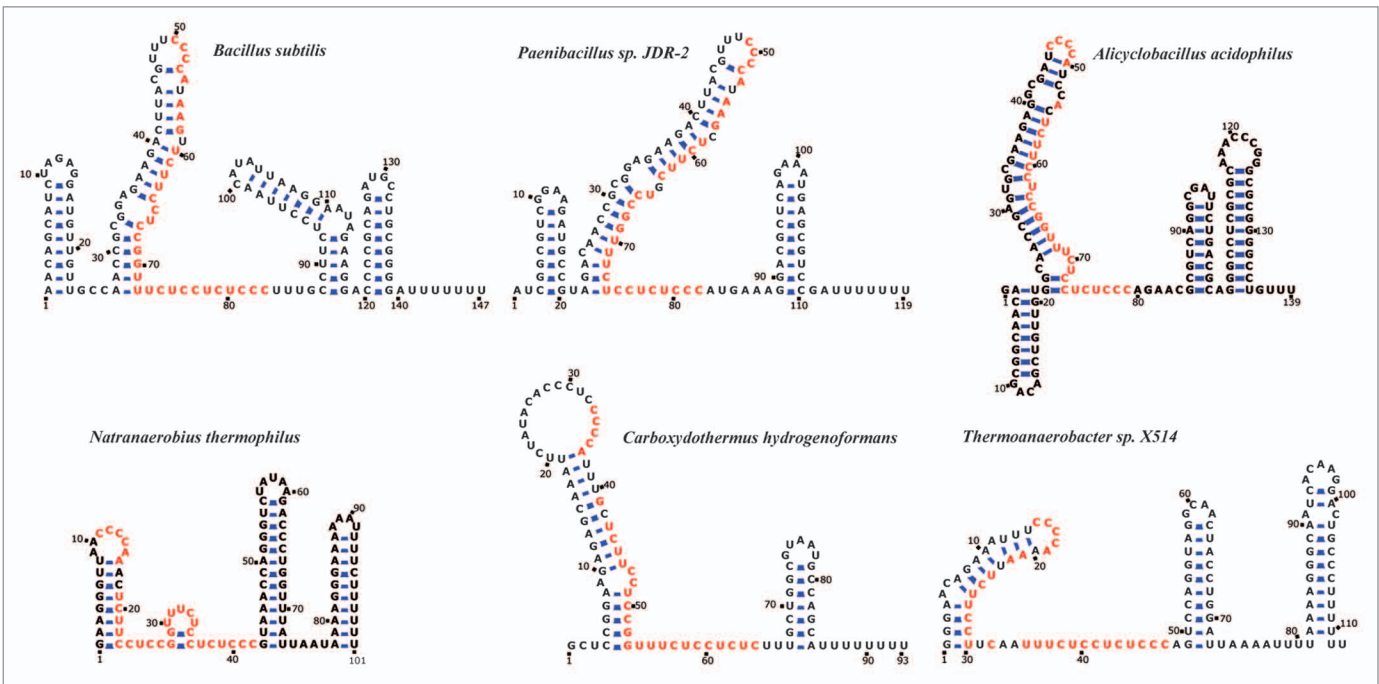
## Discussion

*CsfG* is a small, non-coding RNA synthesized exclusively in the forespore all along the sporulation process, first by  $\sigma^F$  then by  $\sigma^G$ . Its presence in all *Bacillaceae* and in several other endospore formers, including bacteria that diverged from *B. subtilis* billions of years ago, indicates that *CsfG* was present in the earliest endospore formers and that its preservation is somehow beneficial for the organism. Our competition experiments point to a potential germination defect of spores lacking *CsfG*. Interestingly, in several Clostridia where *csfG* is missing, it is replaced upstream of *ylbH* by *gpr*, a gene that is also transcribed exclusively in the forespore by  $\sigma^F$ - and  $\sigma^G$ -associated RNA polymerase. In some Clostridia containing *csfG* the *gpr* gene is located only a few kb downstream of *csfG*, next to a small cluster of  $\sigma^G$ -controlled *ssp* genes. These genes encode small proteins stored in the mature spore where they become the substrate of the Gpr protease during germination. These genomic relationships might reflect an ancient genetic linkage and they fit with a possible functional involvement of *CsfG* in germination.

If *CsfG* plays a negative role on the expression of some genes, as it is often the case with trans-acting sRNAs, potential targets could be genes controlling vegetative functions not required in the forespore, such as replication initiation or septation. However, synthesizing *CsfG* in exponentially growing bacteria, from an inducible promoter on a replicative plasmid, did not lead to a reduced growth rate nor to filamentation (data not shown). We used several programs to predict potential targets for *B. subtilis* *CsfG*. However, its conserved pyrimidine-rich region has the ability to pair with purine-rich ribosome binding sites, leading to a long list of targets with no clear winners (data not shown). We noticed that the highly conserved *CsfG* motif UCUUCCUCCGGUUU is perfectly complementary to a segment of stem-loop 40 in 16S rRNA whereas the *ylbH* gene, next to which *csfG* is always located, encodes a protein belonging to the methyltransferase superfamily, with highest similarity to RsmD, a 16S rRNA methylase. The significance of these observations remains to be investigated.



**Figure 4.** Multiple structural alignment of 24 CsfG orthologs. Bacterial names are indicated by their first letters and correspond to *B. subtilis*, *B. amyloxyfaciens*, *B. licheniformis*, *B. pumilus*, *Bacillus anthracis*, *Bacillus cereus*, *Bacillus thuringiensis*, *Bacillus weihenstephanensis*, *Bacillus clausii*, *Bacillus halodurans*, *Geobacillus kaustophilus*, *Geobacillus thermodentrificans*, *Oceanobacillus iheyensis*, *Anoxybacillus flavithermus*, *Paenibacillus sp. JDR-2*, *Alicyclobacillus acidocaldarius*, *Halothermotrix orenii*, *Natranaerobius thermophilus*, *Carboxydibrachium pacificum*, *Carboxydotherrmus hydrogeniformans*, *Moorella thermoacetica*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter tengcongensis*, *Symbiobacterium thermophilum*. Text is highlighted in a turquoise/blue/grey gradient to indicate regions of high/medium/low conservation. Arrowheads indicate potential base pairing.



**Figure 5.** Predicted secondary structure of representative CsfG RNAs. A varna view (21) of the predicted secondary structure of CsfG RNAs from six distant endospore formers is shown. Highly conserved nucleotides are indicated in red.

Whatever its role it is noteworthy that the CsfG sRNA was initially identified by a computational screen for elements showing phylogenetic co-occurrence in endospore formers. Making a spore involves a large number of genes, but only a subset

of them lead to a clearcut sporulation defect when disrupted.<sup>4</sup> Moreover, despite the conservation among endospore formers of a core of sporulation genes, mostly the main regulators and their principal targets, the presence or absence of a large

number of sporulation-expressed genes depend on the specific habitat of the bacterium.<sup>7</sup> Phylogenetic profiles used in the NAPP algorithm are particularly fit to capture such complex patterns as shown by the identification of CsfG.

## Materials and Methods

**Computational analyses.** The complete NAPP procedure is described in Marchais et al. In short, coding sequences (CDS) and conserved non-coding elements (CNE) from a species of interest (here *B. subtilis*) are sought systematically in 420 bacterial genomes using Blastn,<sup>18</sup> producing a profile of normalized Blast scores. We then measure pairwise Pearson's distances between profiles and we cluster profiles using either K-means or hierarchical clustering.<sup>15</sup> The "sporulation" cluster was identified by visual inspection of the hierarchical clustering representation, as part of a larger cluster enriched in known non-coding RNAs identified by Marchais et al. Searches for CsfG homologs were performed against the NCBI nr databases using Blastn. The multiple structural alignment of CsfG homologs was performed using the LocARNA web server<sup>19</sup> and further manually edited using the Ralee Emacs plugin<sup>20</sup> to correct local errors. Drawing and editing of RNA secondary structure was performed with the varna software.<sup>21</sup> The CsfG entry was submitted to the RFAM database.<sup>22</sup> A pre-release is available at: [en.wikipedia.org/wiki/User:Amarchais/CsfG\\_RNA](http://en.wikipedia.org/wiki/User:Amarchais/CsfG_RNA).

**Growing bacteria.** Bacterial growth conditions and genetic methods were as described in reference 23. All *B. subtilis* strains are derivatives of JH642. The *sig* mutants are from our laboratory collection. Sporulation was induced by exhaustion in Difco sporulation medium (DSM) in the absence of antibiotics. Induction of  $\sigma^F$  or  $\sigma^G$  activity during exponential growth was carried out in 2x YT medium, in strains harboring a plasmid with the *sigF* or the *sigG* sequence under the control of the *spac* promoter, by addition of 1-mM isopropyl  $\beta$ -D-thiogalactoside (IPTG) at  $OD_{570} = 0.25$ .  $\beta$ -galactosidase activity is expressed as nanomoles of 2-nitrophenyl- $\beta$ -D-galactopyranoside hydrolyzed per minute per milligram of protein.

**Cloning *csfG*.** The *csfG* region was cloned by PCR using oligonucleotides (sequences available on request) that created an *EcoRI* site 130 bp upstream of the predicted *csfG* transcription start. The natural *SacII* site located at position 32 in

CsfG was used to create a transcriptional fusion with *lacZ* after cloning into the pDG1661 vector that allows integration by a double recombination event at the *amyE* locus.<sup>24</sup> Similarly, a *NcoI* site and a *PstI* site were introduced by PCR at position -9 and +123 respectively, according to the predicted *csfG* transcription start. A kanamycin resistance cassette was cloned between these two restriction sites and transferred onto the chromosome by a double recombination event that led to inactivation of *csfG*.

**Northern blot analysis.** RNA was prepared according to a published protocol.<sup>25</sup> Total RNA (10  $\mu$ g) was separated on a denaturing 8% polyacrylamide-8 M urea gel and transferred to a Hybond-N<sup>+</sup> Membrane (GE Healthcare) for 4 h at 60 V in 0.5x TBE. The membrane was pre-hybridized overnight in UltraHyb buffer (Ambion) and probed with a <sup>32</sup>P-labelled oligonucleotide (GGA GAA GGC AAA GGG AGA GGA GAA ACC GGA GG). Hybridization and wash steps were as described in reference 26. Briefly, the membrane was hybridized 4 h at 42°C in UltraHyb buffer, washed once in 2x SSC, 0.01% SDS for 10 minutes at 42°C and 5 times at room temperature in 0.2x SSC, 0.01% SDS. The signal was revealed using a Typhoon Trio (GE Healthcare).

## Acknowledgements

We are grateful to C. Condon for his advice and encouragement. This work was supported by CNRS (IFR 115 and UPR9073).

## References

1. Errington J. Regulation of endospore formation in *Bacillus subtilis*. *Nat Rev Microbiol* 2003; 1:117-26.
2. Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, et al. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol* 2004; 2:1664-83.
3. Steil L, Serrano M, Henriques AO, Völker U. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. *Microbiology* 2005; 151:399-420.
4. Wang ST, Setlow B, Conlon EM, Lyon JL, Imamura D, Sato T, et al. The forespore line of gene expression in *Bacillus subtilis*. *J Mol Biol* 2006; 358:16-37.
5. Hilbert DW, Piggot PJ. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol Mol Biol Rev* 2004; 68:234-62.
6. Paredes CJ, Alsaker KV, Papoutsakis ET. A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol* 2005; 3:969-78.

7. Stragier P. A gene odyssey: exploring the genomes of endospore-forming bacteria. In: Sonenshein AL, Hoch JA, Losick R, Eds. *Bacillus subtilis* and its Closest Relatives: From Genes to Cells. Washington, DC: ASM Press 2002; 519-26.
8. Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* 2007; 10:156-63.
9. Yu YTN, Yuan X, Velicer GJ. Adaptive evolution of an sRNA that controls *Myxococcus* development. *Science* 2010; 328:993.
10. Toledo-Arana A, Repoila F, Cossart P. Small non-coding RNAs controlling pathogenesis. *Curr Opin Microbiol* 2007; 10:182-8.
11. Silvaggi JM, Perkins JB, Losick R. Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. *J Bacteriol* 2006; 188:532-41.
12. Schmalisch M, Maiques E, Nikolov L, Camp AH, Chevreux B, Muffler A, et al. Small genes under sporulation control in the *Bacillus subtilis* genome. *J Bacteriol* 2010; 192:5402-12.
13. Saito S, Kakeshita H, Nakamura K. Novel small RNA-encoding genes in the intergenic regions of *Bacillus subtilis*. *Gene* 2009; 428:2-8.
14. Rasmussen S, Nielsen HB, Jarmer H. The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol Microbiol* 2009; 73:1043-57.
15. Marchais A, Naville M, Bohn C, Bouloc P, Gautheret D. Single-pass classification of all noncoding sequences in a bacterial genome using phylogenetic profiles. *Genome Res* 2009; 19:1084-92.
16. Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, et al. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci USA* 2004; 101:6421-6.
17. Lease RA, Cusick ME, Belfort M. Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc Natl Acad Sci USA* 1998; 95:12456-61.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; 215:403-10.
19. Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol* 2007; 3:65.
20. Griffiths-Jones S. RALEE—RNA Alignment Editor in Emacs. *Bioinformatics* 2005; 21:257-9.
21. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 2009; 25:1974-5.
22. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, et al. Rfam: updates to the RNA families database. *Nucleic Acids Res* 2009; 37:136-40.
23. Karmazyn-Campelli C, Rhayat L, Carballido-López R, Duperrier S, Frandsen N, Stragier P. How the early sporulation sigma factor  $\sigma^E$  delays the switch to late development in *Bacillus subtilis*. *Mol Microbiol* 2008; 67:1169-80.
24. Guéroult-Fleury AM, Frandsen N, Stragier P. Plasmids for ectopic integration in *Bacillus subtilis*. *Gene* 1996; 180:57-61.
25. Bechhofer DH, Oussenko I, Yao S, Mathy N, Condon C. Analysis of mRNA decay in *Bacillus subtilis*. *Methods Enzymol* 2008; 447:259-76.
26. Opdyke JA, Kang JG, Storz G, GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol* 2004; 186:6698-705.