

The σ^E Regulon and the Identification of Additional Sporulation Genes in *Bacillus subtilis*

Patrick Eichenberger¹, Shane T. Jensen², Erin M. Conlon²
Christiaan van Ooij¹, Jessica Silvaggi¹
José-Eduardo González-Pastor¹, Masaya Fujita¹, Sigal Ben-Yehuda¹
Patrick Stragier³, Jun S. Liu² and Richard Losick^{1*}

¹Department of Molecular and Cellular Biology, Harvard University Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138 USA

²Department of Statistics Harvard University, 1 Oxford Street, Cambridge, MA 02138 USA

³Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France

We report the identification and characterization on a genome-wide basis of genes under the control of the developmental transcription factor σ^E in *Bacillus subtilis*. The σ^E factor governs gene expression in the larger of the two cellular compartments (the mother cell) created by polar division during the developmental process of sporulation. Using transcriptional profiling and bioinformatics we show that 253 genes (organized in 157 operons) appear to be controlled by σ^E . Among these, 181 genes (organized in 121 operons) had not been previously described as members of this regulon. Promoters for many of the newly identified genes were located by transcription start site mapping. To assess the role of these genes in sporulation, we created null mutations in 98 of the newly identified genes and operons. Of the resulting mutants, 12 (in *prkA*, *ybaN*, *yhbH*, *ykoV*, *ylbJ*, *ypjB*, *yqfC*, *yqfD*, *ytrH*, *ytrI*, *ytvI* and *yunB*) exhibited defects in spore formation. In addition, subcellular localization studies were carried out using in-frame fusions of several of the genes to the coding sequence for GFP. A majority of the fusion proteins localized either to the membrane surrounding the developing spore or to specific layers of the spore coat, although some fusions showed a uniform distribution in the mother cell cytoplasm. Finally, we used comparative genomics to determine that 46 of the σ^E -controlled genes in *B. subtilis* were present in all of the Gram-positive endospore-forming bacteria whose genome has been sequenced, but absent from the genome of the closely related but not endospore-forming bacterium *Listeria monocytogenes*, thereby defining a core of conserved sporulation genes of probable common ancestral origin. Our findings set the stage for a comprehensive understanding of the contribution of a cell-specific transcription factor to development and morphogenesis.

© 2003 Published by Elsevier Science Ltd

Keywords: *Bacillus subtilis*; sporulation; sigma factors; bacterial genomics; protein localization

*Corresponding author

Introduction

A crucial issue in developmental biology is the generation of distinct cell types by the establishment of differential gene expression programs in cells of common origin. The process of endospore formation (sporulation) in *Bacillus subtilis* and other related Gram-positive bacteria, is an ideal

system in which to address this question experimentally.^{1,2} Early in the process of sporulation, *B. subtilis* cells divide asymmetrically to generate two compartments of unequal size and dissimilar developmental fates. The smaller compartment, the forespore, develops into the spore, whereas the larger compartment, the mother cell, which nurtures the developing spore, lyses when morphogenesis is complete to liberate the mature spore. Differentiation involves the action of four cell-specific sigma factors: σ^F , σ^E , σ^G , and σ^K . The σ^F and σ^E factors are activated shortly after asymmetric division when σ^F directs gene expression in

Abbreviations used: Cy, cyanine; GFP, green fluorescent protein.

E-mail address of the corresponding author: losick@mcb.harvard.edu

the forespore³ and σ^E directs gene expression in the mother cell.⁴ The σ^F and σ^E factors direct the transcription of largely non-overlapping sets of genes that have distinct functions in the sporulation process. Later in sporulation, σ^F is replaced in the forespore by σ^G and σ^E is replaced in the mother cell by σ^K .^{5,6}

Here we are concerned with the identification of genes under the control of the mother cell transcription factor σ^E . The σ^E factor is initially synthesized as an inactive proprotein called pro- σ^E .⁷ Synthesis of pro- σ^E commences prior to septation,⁸ but the conversion of pro- σ^E to mature σ^E does not take place until after asymmetric division when the active form of the transcription factor is found in the mother cell. Proteolytic processing of pro- σ^E is mediated by the proprotein processing enzyme SpoIIIGA,⁹⁻¹¹ which is activated by a secreted signaling protein (SpoIIR) that is produced in the forespore under the control of σ^F .¹²⁻¹⁴

The development of methods for carrying out transcriptional profiling on a genome-wide basis (for a review see Rhodius *et al.*¹⁵) has made it possible to identify genes that are activated during the various stages of sporulation in a comprehensive manner. Recently, we have used *B. subtilis* DNA arrays to characterize the regulons for the DNA-binding protein Spo0A,¹⁶ the principal regulator for entry into sporulation, and σ^H ,¹⁷ an alternative sigma factor governing the transcription of genes involved in stationary phase and in the early stages of sporulation. Here we present the results obtained for the σ^E regulon. Previous reports have indicated that more than 70 genes are directly controlled by σ^E ,¹⁸ already constituting the largest regulon for any of the four cell-specific sigma factors active during sporulation. Genes controlled by σ^E are involved in a variety of morphological and regulatory functions, as illustrated by the phenotype of strains in which the gene (*sigE*) encoding σ^E has been inactivated. A *sigE* mutant is blocked in sporulation after polar septation (stage II) and is impaired in the morphological process of engulfment of the forespore by the mother cell.¹⁹⁻²¹ Instead, in the *sigE* mutant, a second polar septum is formed at the opposite pole of the cell generating a second forespore-like compartment, a phenotype that has been termed "disporic".^{22,23} It has been shown recently that the same group of σ^E -controlled genes is required for engulfment and inhibition of division at the second cell pole.^{24,25} Furthermore, σ^E controls the expression of several genes involved in the pathways leading to the synthesis and activation of σ^G and σ^K , and accordingly, cells with a mutation in *sigE* exhibit neither σ^G nor σ^K activity.² Finally, a large group of genes controlled by σ^E is necessary for the synthesis and assembly of two elaborate structures that protect the spore against adverse environmental conditions: the cortex and the coat. The cortex is a modified form of peptidoglycan,²⁶ which is synthesized between the internal and

external membranes of the forespore, whereas the coat is a shell of proteins surrounding the mature spore.²⁷

Recently, several genome sequences of other Gram-positive endospore-forming bacteria related to *B. subtilis* have been released. These bacteria have attracted the attention of researchers for a variety of reasons: *Bacillus anthracis*,²⁸ the causative agent of anthrax, and *Clostridium perfringens*,²⁹ the agent of gas gangrene, are important because of public health concerns; *Clostridium acetobutylicum*,³⁰ a solvent-producing bacterium is of considerable industrial importance, whereas *Bacillus halodurans*³¹ and *Oceanobacillus iheyensis*³² are of interest because they are adapted to live in extreme environments. Additionally, the analysis of the complete genome sequence of *Thermoanaerobacter tengcongensis* indicates that, even though this thermophilic bacterium has not yet been shown to sporulate, it possesses an apparently complete set of genes required for the process of sporulation.³³ Interestingly, genes involved in the pathways leading to the activation of the sporulation sigma factors appear to be strictly conserved among all of these bacteria, even though sporulation takes place in largely different environments.³⁴ However, it is likely that the set of genes controlled by each of the sporulation sigma factors in these organisms varies to a certain extent. By identifying the complete set of genes regulated by each sigma factor in *B. subtilis*, and studying the conservation of these genes in the genomes of related endospore formers, one can hope to distinguish between genes that are generally required for sporulation and those that are less widely conserved and confer species-specific characteristics.

Here we report that σ^E directs the transcription of more than 250 genes during sporulation in *B. subtilis*. In addition, by systematic inactivation of most of these σ^E -controlled genes, we have identified several genes that were not previously known to be required for efficient sporulation in *B. subtilis*. We show that these genes are usually also present in the genomes of other endospore formers, but absent from the genome of the highly related but non-endospore-forming bacterium *Listeria monocytogenes*. Finally, we present the results of cytological studies, which led to the identification of new coat-associated proteins.

Results

Genome-wide identification of genes under the control of σ^E

To identify genes transcribed under the control of σ^E , two complementary transcriptional profiling strategies were used. In one, RNA was extracted from *sigE*⁺ and *sigE* mutant cells at a time during sporulation when σ^E activity was at its highest, i.e. 2.5 hours after suspension in sporulation medium. The σ^E factor is known to direct the synthesis in

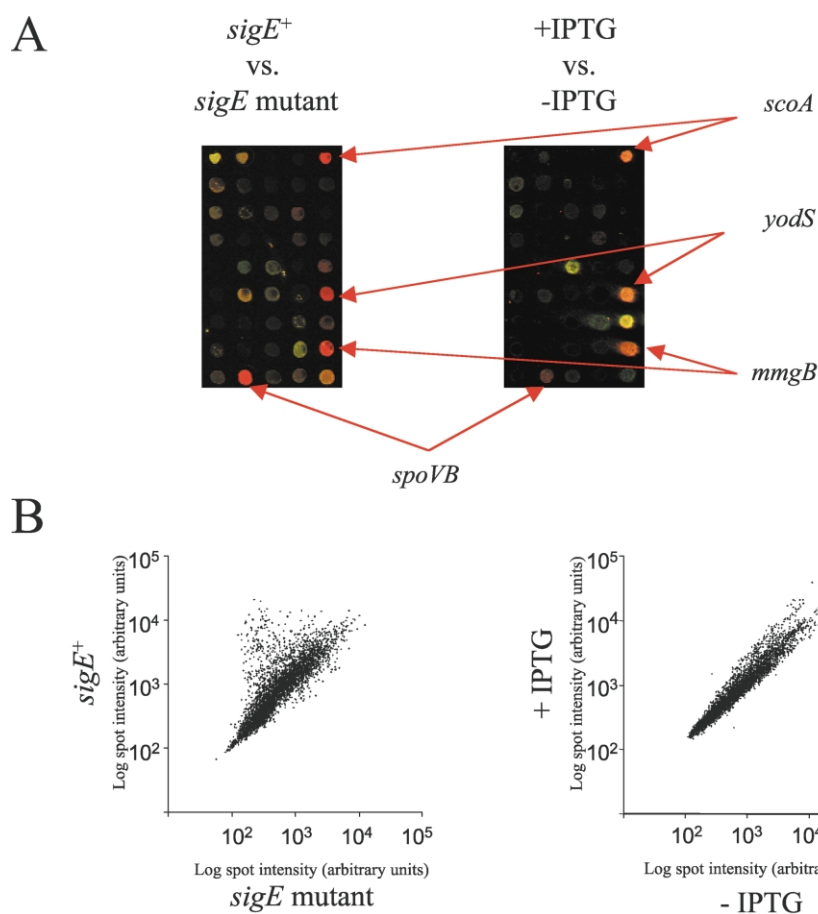


Figure 1. Identification of σ^E -regulated genes by transcriptional profiling. (A) The Figure shows a small region of the microarray chosen to give examples of the identification of σ^E -controlled genes. The left-hand array was hybridized with Cy5-labeled cDNA (red) generated with RNA from sporulating cells of the σ^E strain PE436 and with Cy3-labeled cDNA (green) from sporulating cells of the σ^E mutant strain PE437. The right-hand array was hybridized with Cy5-labeled cDNA generated with RNA from growing cells of strain MF453 that had been treated with IPTG and with Cy3-labeled cDNA generated with RNA from cells of MF453 that had not been treated with inducer. The arrows point to the known σ^E -controlled genes *spoVB* and *mmgB* and to *scoA* and *yodS*, two genes that had not been previously known to be under the control of σ^E . Genes whose expression was up-regulated by σ^E appear as red or orange spots on the microarray. (B) Logarithmic scale plots of spot intensities (arbitrary units). The left-hand panel shows the signal intensity for each spot on the entire array of *B. subtilis* ORFs obtained with cDNA generated from RNA from

sporulating cells of the σ^E strain PE436 plotted against the corresponding signal obtained with probe generated from sporulating cells of the σ^E mutant PE437. The right-hand panel shows the signal intensity for each spot in the microarray obtained with probe generated from growing cells of strain MF453 that had been treated with IPTG plotted against the corresponding signal obtained with probe against growing cells of MF453 that had not been treated with inducer. Data presented are from a representative experiment and are not normalized.

the mother cell of the DNA-binding protein SpoIIID,^{35–37} which, in turn, leads to the appearance of the later-acting regulatory proteins σ^{K38-41} and GerE.⁴² The σ^E factor also triggers the appearance in the forespore of the transcription factor σ^G .^{43,44} Therefore, to restrict our analysis to genes likely to be under the direct control of σ^E , we used σ^E and σ^E mutant strains (PE436 and PE437, respectively) that also contained null mutations in the genes for SpoIIID and σ^G .

The other strategy was to artificially produce σ^E in vegetatively growing cells. Efforts to produce σ^E directly using an inducible copy of *sigE* from which the pro-amino acid coding sequence had been removed were unsuccessful (M.F., unpublished results). Therefore, we used strain MF453, which was engineered to produce pro- σ^E and its processing enzyme SpoIIIGA during growth and to produce the SpoIIR signaling protein in response to the inducer IPTG.⁴⁵ In experiments with MF453, RNA was isolated from growing cells that had been collected one hour after the addition of IPTG and, in parallel, from cells that had not been treated with inducer.

Thus, *in toto*, RNA was prepared from four cell cultures: two from cells of the σ^E (PE436) and the σ^E mutant (PE437) cells undergoing sporulation and two from vegetative cells of strain MF453 grown with and without IPTG. Next, the four RNAs were used to generate fluorescently labeled cDNAs that were differentially tagged with either Cyanine5 (red) or Cyanine3 (green). For example in the experiment of Figure 1(A), Cy5 was used to label cDNAs from σ^E cells and cells grown in the presence of inducer and Cy3 was used to label cDNAs from σ^E mutant cells and cells grown without inducer (in other experiments the labels were switched). Finally, mixtures of each of the two pairs of differentially labeled cDNAs were incubated under hybridization conditions with DNA microarrays. The microarrays contained PCR-amplified DNAs corresponding to 4074 of the 4106 open reading frames (ORFs) in the *B. subtilis* genome as described.¹⁷

An example of the results is shown in Figure 1(A), which displays a region of the microarrays containing two genes (*spoVB* and *mmgB*) that were known to be under the control of σ^E ^{46,47} and two

Table 1. Known σ^E -controlled genes

Operon organization	Gene	Ratio ^a sporulation	Ratio ^b overexpression	Function ^c
<i>asnO</i> (<i>visO</i>)	<i>asnO</i>	3.9	1.9 ^d	Asparagine synthetase
<i>bofA</i>	<i>bofA</i>	9.6	15.1	Inhibition of the pro- σ^K processing machinery
<i>cotE</i>	<i>cotE</i>	49.4	2.1	Spore coat protein (outer), morphogenetic protein
<i>cotJABC-yesJ</i>	<i>cotJA</i>	50.8	3.8	Spore coat protein (inner)
	<i>cotJB</i>	48.9	9	Spore coat protein (inner)
	<i>cotJC</i>	46.2	3.2	Spore coat protein (inner)
	<i>yesJ</i>	29.1	4.2	Similar to acetyltransferase (GCN5 family)
<i>cwlD</i>	<i>cwlD</i>	16.7	9.7	<i>N</i> -acetylmuramoyl-L-alanine amidase (sporulation), germination
				Cell wall hydrolase, germination
<i>cwlJ</i>	<i>cwlJ</i>	2.8	1.2 ^d	Cell wall hydrolase, germination
<i>dacB-spmAB</i>	<i>dacB</i>	9.9	3.3	Penicillin-binding protein 5* (D-alanyl-D-alanine carboxypeptidase)
	<i>spmA</i>	- ^e	5.7	Spore maturation protein (spore core dehydration)
	<i>spmB</i>	18.1	1.9	Spore maturation protein (spore core dehydration)
				1,4- α -glucan branching enzyme
<i>glgBCDAP</i>	<i>glgB</i>	2.1	2.9	Glucose (Glc)-1-phosphate adenylyltransferase
	<i>glgC</i>	2.1 ^f	2.4	Required for glycogen biosynthesis
	<i>glgD</i>	2.1	2.9	Glycogen synthase
	<i>glgA</i>	2.3	2.5	Glycogen phosphorylase
	<i>glgP</i>	2.2	2.9	Acetyl-coenzyme A (CoA) acetyltransferase
<i>mmgABCDE-yqiQ</i>	<i>mmgA</i>	53.3	2.4	3-Hydroxybutyryl-CoA dehydrogenase
	<i>mmgB</i>	27.5	2.6	Acyl-CoA dehydrogenase, similar to <i>AcdA</i> and <i>YngJ</i>
	<i>mmgC</i>	67.1	2.2	Citrate synthase III
	<i>mmgD</i>	108.4	2.5	Unknown, possibly involved in propionate metabolism
	<i>mmgE</i>	60.7	2.2	Similar to carboxyvinyl-carboxyphosphonate phosphorylmutase
	<i>yqiQ</i>	9.3	1.9	Sporulation-specific extracellular nuclease
<i>nucB</i>	<i>nucB</i>	1.9	- ^g	Alkaline phosphatase III
<i>phoB-ydhF</i>	<i>phoB</i>	29.4	2.2	Induced by phosphate starvation
	<i>ydhF</i>	17.8	3.2	Spore coat protein (inner), morphogenetic protein, associated with <i>SpoVID</i>
<i>safA</i> (<i>yrbA</i>)- <i>coxA</i> (<i>yrbB</i>)	<i>safA</i>	31.8	1.8 ^d	Spore cortex protein
<i>spoIID</i>	<i>coxA</i>	3.5	3.3	Dissolution of the septal cell wall (engulfment and inhibition of septation)
	<i>spoIID</i>	55.2	2.4	Dissolution of the septal cell wall (engulfment and inhibition of septation)
<i>spoIIM</i>	<i>spoIIM</i>	4.1	4.6	Dissolution of the septal cell wall (engulfment and inhibition of septation)
<i>spoIIP</i>	<i>spoIIP</i>	1.6 ^h	1.9 ^h	Dissolution of the septal cell wall (engulfment and inhibition of septation)
<i>spoIIIAABCDEFGH</i>	<i>spoIIIAA</i>	24.2	2.7	σ^G regulation
	<i>spoIIIAB</i>	57.5	4.4	σ^G regulation
	<i>spoIIIAC</i>	9.6	4.1	σ^G regulation
	<i>spoIIIID</i>	27.1	2.9	σ^G regulation
	<i>spoIIIAE</i>	32.3	2.5	σ^G regulation
	<i>spoIIIAF</i>	48	2.6	σ^G regulation
	<i>spoIIIAG</i>	21.1	2.7	σ^G regulation
	<i>spoIIIAH</i>	9	2.8	σ^G regulation
	<i>spoIIIAI</i>	47.4	2.4	Proper spore cortex formation and coat assembly, morphogenetic protein
<i>spoIVCA</i>	<i>spoIVCA</i>	1.2 ^f	1.9	DNA recombinase required for creating the <i>sigK</i> gene, activated by <i>SpoIIID</i>
<i>spoIVCB</i>	<i>spoIVCB</i>	2.6	1.1 ^d	N-terminal half of σ^K , activated by <i>SpoIIID</i>
<i>spoIVFAB</i>	<i>spoIVFA</i>	21	3.2	Inhibition of <i>SpoIVFB</i>
	<i>spoIVFB</i>	36.1	5.6	Membrane metalloprotease (processing of Pro- σ^K to active σ^K)
<i>spoVB</i>	<i>spoVB</i>	15.7	2.7	Spore cortex synthesis
<i>spoVD</i>	<i>spoVD</i>	43.5	3.1	Penicillin-binding protein (spore cortex)
<i>spoVE-murGB</i>	<i>spoVE</i>	3.2	5.9	Spore cortex peptidoglycan synthesis
	<i>murG</i>	2.7	1.9	UDP-GlcNAc-MurNAc-(pentapeptide) PPi-undecaprenol GlcNAc transferase
<i>spoVK</i>	<i>murB</i>	3.1	3.3	UDP-N-acetylenolpyruvoylglucosamine reductase
	<i>spoVK</i>	6 ^h	1.5 ^h	Sporulation (AAA family of ATPases), activated by <i>SpoIIID</i>
<i>spoVM</i>	<i>spoVM</i>	- ⁱ	- ⁱ	Cortex and coat

(continued)

Table 1 Continued

Operon organization	Gene	Ratio ^a sporulation	Ratio ^b overexpression	Function ^c
<i>spoVR</i>	<u>spoVR</u>	148.3	2.5	Spore cortex synthesis
<i>spoVID-ysxE</i>	<u>spoVID</u>	112.2	2.1	Spore coat protein, morphogenetic protein
	<i>ysxE</i>	24.9	2.2	Unknown
<i>usd-spoIIID-mbl</i>	<u>usd</u>	4.1	– ⁱ	Translation of <i>spoIIID</i>
	<u>spoIIID</u>	4.9 ^f	2.2	Transcriptional regulator (DeoR family)
	<i>mbl</i>	65.9	2.4	MreB-like protein
<i>yaaH</i>	<u>yaaH</u>	205.6	3	Germination (L-Ala-stimulated pathway), similar to YdhD
<i>yabPQ-divIC</i>	<u>yabP</u>	14.2	4	Unknown
	<u>yabQ</u>	14.3	4	Spore cortex formation
	<u>divIC</u>	9.4	2.9	Cell-division initiation protein
<i>ydhD</i>	<u>ydhD</u>	51.4	2.2	Similar to YaaH
<i>yfhS</i>	<u>yfhS</u>	– ⁱ	– ⁱ	Unknown
<i>yjmCD-uxuA(yjmE)-yjmF-exuTR</i> (<i>yjmGH</i>)	<u>yjmC</u>	55.5	5.4	Hexuronate utilization operon, malate dehydrogenase
	<u>yjmD</u>	64.7	3.5	Hexuronate utilization operon, alcohol dehydrogenase
	<i>uxuA</i>	84.1	8.8	Hexuronate utilization operon, D-mannonate hydrolase
	<i>yjmF</i>	18.7	9.3	Hexuronate utilization operon, D-mannonate oxidoreductase
	<i>exuT</i>	71	8.9	Hexuronate utilization operon, hexuronate transporter
	<i>exuR</i>	15.4	2.9	Transcriptional repressor of the hexuronate utilization operon
<i>yknT</i>	<u>yknT</u>	53.2	1.7	Unknown
<i>yoaW</i>	<u>yoaW</u>	1.6 ^h	1.4 ^h	Unknown
<i>yteV</i>	<u>yteV</u>	1.3 ^f	2.6	Unknown

Inactivation of genes indicated in bold led to a sporulation defect. Genes underlined are conserved in *B. anthracis*, *B. halodurans*, *O. iheyensis*, *C. acetobutylicum*, *C. perfringens* and absent in *Listeria*.

^a Ratios of relative RNA levels in *sigE*⁺ versus *sigE* mutant.

^b Ratios of relative RNA levels from *sigE* overexpression (+IPTG) one hour after induction compared to the same time point without overexpression (–IPTG).

^c Function according to SubtiList (<http://genolist.pasteur.fr/SubtiList>)

^d Missed cut-off in the overexpression experiment.

^e Spot excluded from the data analysis in the *sigE*⁺ versus *sigE* mutant experiment.

^f Missed cut-off in the *sigE*⁺ versus *sigE* mutant experiment.

^g Spot excluded from the data analysis in the overexpression experiment (signal intensity not significantly above background level).

^h Missed cut-off in both experiments.

ⁱ Spot excluded from the data analysis in both experiments (signal intensity not significantly above background level).

^j Missed cut-off in the *sigE*⁺ versus *sigE* mutant experiment and spot excluded from the data analysis in the overexpression experiment (signal intensity not significantly above background level).

genes (*scoA* and *yodS*) that were not previously recognized as being targets of the sporulation transcription factor. Importantly, all four genes were transcribed in a σ^E -dependent manner (as indicated by the red or orange color) both in sporulating cells and in cells engineered to produce σ^E during growth. However, in all cases a noticeably stronger signal was obtained during sporulation than during growth for which we offer two explanations. First, the level of σ^E was higher in sporulating cells than was achieved in the engineered cells. Second, the IPTG-inducible regulatory system used in triggering pro- σ^E processing during growth was leaky, and a substantial level of σ^E activity was observed even in the absence of inducer (probably because a small amount of the SpoIIR signaling protein was sufficient to trigger a significant level of pro- σ^E processing). For example, when σ^E activity was monitored in a liquid culture using a *lacZ* gene fused to a known σ^E -controlled gene (*spoIIID*) the level of β -galacto-

sidase observed in the presence of inducer was twofold higher than in the absence of inducer (data not shown). This value is similar to the corresponding ratio of 2.4 obtained for *spoIIID* in the microarray experiment (Table 1). Nonetheless, and as we shall see, the IPTG-inducible system was highly effective as judged by the fact that it revealed a high proportion of the previously known σ^E -controlled genes as well as many previously uncharacterized genes.

A comprehensive representation of the results obtained with the microarrays is presented in the scatter plot of Figure 1(B) and in Tables S1 and S2 of the supplementary material. The results are based on three independent preparations of RNA for each of two cultures of sporulating cells and each of two cultures of vegetative cells. Moreover, in some cases, the same RNA preparation was used more than once (see Materials and Methods for a description of the data analysis). The scatter plots of Figure 1(B) compare the intensities of the

Table 2. Newly identified σ^E -controlled genes

Operon organization	Gene	Ratio ^a sporulation	Ratio ^b overexpression	Function ^c
σ^E promoter confirmed by 5'-RACE-PCR				
<i>germ-rph-ysnAB</i>	<i>gerM</i>	40.8	3.1	Sporulation ^d and germination (cortex hydrolysis)
	<i>rph</i>	2.3	2	Ribonuclease PH
	<i>ysnA</i>	2.3	2.8	Similar to xanthosine triphosphate pyrophosphatase
	<i>ysnB</i>	2.1	2.4	Similar to calcineurin-like phosphoesterase
<i>prkA</i>	<i>prkA</i>	8.2	2.2	Sporulation ^d , serine protein kinase, inner coat
<i>ybaN</i>	<i>ybaN</i>	14.9	2.2	Sporulation ^d , similar to polysaccharide deacetylase, similar to YlxY
<i>ycgFG</i>	<i>ycgF</i>	7.4	2.7	Unknown (putative amino acid transporter, lysine exporter family)
	<i>ycgG</i>	11.4	4.7	Unknown
<i>ydcA</i>	<i>ydcA</i>	4.6	2.2	Similar to intramembrane serine protease (Rhomboid family), growth ^e
<i>ydcC⁺-alr (dal)</i>	<i>ydcC</i>	48.6	3.2	Unknown
	<i>alr</i>	4.5	2.8	D-Alanine racemase, similar to YncD
<i>yhaX⁺-hemZ (yhaVW)</i>	<i>yhaX</i>	22.6	2.8	Similar to hydrolase (HAD family), inner coat, PhoP-regulated ¹⁰⁵
	<i>hemZ</i>	6.5	2.3	Coproporphyrinogen III oxidase (heme metabolism)
<i>yhbH</i>	<i>yhbH</i>	42.1	2.5	Sporulation ^d and germination ^e , PhoP-regulated ¹⁰⁵
<i>yjbX</i>	<i>yjbX</i>	34.4	2.1	Glutamic acid rich protein, outer coat
<i>ykvI</i>	<i>ykvI</i>	40.1	4.6	Similar to YyaD
<i>ykvUV</i>	<i>ykvU</i>	61.7	4	Similar to spore cortex membrane protein SpoVB
	<i>ykvV</i>	7	- ^h	Sporulation ^d , similar to thioredoxin, secretion sequence, mutant produces phase-dark spores
<i>ylbJ</i>	<i>ylbJ</i>	31.6	2.4	Sporulation ^d , mutant produces phase-dark spores
<i>yngJIHGFE</i>	<i>yngJ</i>	37.7	2.2	Similar to butyryl-CoA dehydrogenase, similar to MmgC and AcdA
	<i>yngI</i>	61.9	2.2	Similar to long-chain acyl-CoA synthetase
	<i>yngH</i>	57.9	2.2	Similar to biotin carboxylase
	<i>yngG</i>	72.6	2.2	Similar to hydroxymethylglutaryl-CoA lyase
	<i>yngF</i>	35.7	2.5	Similar to 3-hydroxybutyryl-CoA dehydratase
	<i>yngE</i>	38.6	2.4	Similar to propionyl-CoA carboxylase
	<i>yngJ</i>	<i>yngJ</i>	8.2	3.8
<i>yqfCD</i>	<i>yqfC</i>	12.2	3.8	Sporulation ^d , mutant produces phase-dark spores
	<i>yqfD</i>	50.7	3.1	Sporulation ^d , mutant produces phase-dark spores
<i>yqfZY</i>	<i>yqfZ</i>	38	3.2	Unknown
	<i>yqfY⁺</i>	9.7	1.9	Similar to GcpE (IspG), non-mevalonate pathway to terpenoids synthesis
<i>ytrHI</i>	<i>ytrH⁺</i>	-	-	Sporulation ^d
	<i>ytrI</i>	22.7	5.1	Sporulation ^d
<i>ytvI</i>	<i>ytvI</i>	70.3	2	Sporulation ^d (putative permease), similar to YbdI
<i>ytxC</i>	<i>ytxC</i>	28.2	6.6	Unknown
<i>yunB</i>	<i>yunB</i>	16.3	3.3	Sporulation ^{d,k}
<i>yuzC</i>	<i>yuzC</i>	25	5.1	Inner coat
<i>yxjC-scoAB-yxjF</i>	<i>yxjC</i>	31.6	3	Unknown (putative permease)
	<i>scoA</i>	63.4	2.9	Succinyl CoA:3-oxoacid CoA-transferase (subunit A), similar to YodS
	<i>scoB</i>	4.8	2.7	Succinyl CoA:3-oxoacid CoA-transferase (subunit B), similar to YodR
	<i>yxjF</i>	17.5	2.7	Similar to gluconate 5-dehydrogenase, short chain alcohol dehydrogenase (adh)
<i>yyaD</i>	<i>yyaD</i>	63.2	3.4	Similar to YkvI
<i>yybI</i>	<i>yybI</i>	10.7	2.3	Inner coat
σ^E promoter predicted by bioinformatics				
<i>alkA</i>	<i>alkA</i>	6.2	2.2	DNA-3-methyladenine glycosylase (hydrolysis of alkylated DNA)
<i>citH (yxiQ)</i>	<i>citH</i>	62.8	3.3	Transporter of divalent metal ions/citrate complexes, similar to SpmA
<i>comER</i>	<i>comER</i>	9.4	2.1	Similar to pyrroline-5'-carboxylate reductases, similar to ProH
<i>cotB</i>	<i>cotB</i>	6.2	2.5	Outer coat, also controlled by σ^K
<i>cotG</i>	<i>cotG</i>	5.1	3.6	Coat (incorporation of CotB into the coat), also controlled by σ^K
<i>cotH</i>	<i>cotH</i>	9.8	3.1	Inner coat, also controlled by σ^K
<i>cotYZ</i>	<i>cotY</i>	6.3	2.1	Coat, also controlled by σ^K

(continued)

Table 2 Continued

Operon organization	Gene	Ratio ^a sporulation	Ratio ^b overexpression	Function ^c
<i>cypA</i> <i>glnQHMP</i>	<i>cotZ</i>	7.6	9.8	Coat, also controlled by σ^K
	<i>cypA</i>	24	4	Cytochrome P450-like enzyme
	<i>glnQ</i>	37.7	2.5	Glutamine ABC transporter (ATP-binding)
	<i>glnH</i>	73.1	1.7	Glutamine ABC transporter (glutamine-binding)
	<i>glnM</i>	38.2	1.2 ¹	Glutamine ABC transporter (membrane protein)
<i>ispE (yabH)-purR-yabJ</i>	<i>glnP</i>	36.2	6.4	Glutamine ABC transporter (membrane protein)
	<i>ispE¹</i>	6	1.6	4-Diphosphocytidyl-2-C-methyl-D-erythritol kinase, non-mevalonate pathway
	<i>purR</i>	4.9	1 ¹	Transcriptional repressor of the purine operons
	<i>yabJ</i>	6	1.4	Similar to translation initiation inhibitor, adenine-mediated repression of <i>purA</i> ¹⁰⁶
<i>kapD</i>	<i>kapD</i>	11.3	3	Inhibitor of the KinA pathway to sporulation
<i>kbaA</i>	<i>kbaA^{mn}</i>	11.1	1.4	Activation of KinB
<i>lplD</i>	<i>lplDⁿ</i>	4	1.6	Similar to glycosyl hydrolase
<i>nadA</i>	<i>nadAⁿ</i>	2.4	1.7	Quinolinate synthetase (pyridine biosynthesis)
<i>patA</i>	<i>patA</i>	14.5	2.2	Aminotransferase, similar to SpsC and YhjL
<i>pepF (yjbG)</i>	<i>pepF</i>	3.7	1.3	Similar to oligopeptidase (M3 family), over-expression inhibits sporulation ⁷³
<i>phoPR</i>	<i>phoP</i>	4.4	1.3	Two-component response regulator involved in phosphate regulation
	<i>phoR</i>	4.1	1.4	Sensor histidine kinase involved in phosphate regulation
<i>proHJ</i>	<i>proH</i>	5.9	7.7	Pyrroline-5-carboxylate reductase, similar to ComER
	<i>proJ</i>	7.1	1.8	Glutamate 5-kinase
<i>pucH (yunH)</i>	<i>pucH</i>	7.5	2.5	Allantoinase (purine degradation)
	<i>sipS</i>	3	2.3	Signal peptidase I (protein secretion)
	<i>spsC</i>	<i>spsCⁿ</i>	2.8	2.6
<i>spsGIJK</i>	<i>spsGⁿ</i>	6.8	7.7	Spore coat polysaccharide synthesis, also controlled by σ^K
	<i>spsIⁿ</i>	3.7	1.7	Spore coat polysaccharide synthesis, also controlled by σ^K
	<i>spsJⁿ</i>	19.1	3.9	Spore coat polysaccharide synthesis, UDP-GlcNAc epimerase, also controlled by σ^K
	<i>spsKⁿ</i>	7.7	3.4	Spore coat polysaccharide synthesis, DTDP-L-rhamnose synthetase, also controlled by σ^K
<i>sqhC</i>	<i>sqhC</i>	21	1.4	Squalene-hopene cyclase
<i>tmrB</i>	<i>tmrBⁿ</i>	4.1	2	Tunicamycin resistance
<i>ybaS</i> <i>ybbFEDC</i>	<i>ybaS</i>	2.8	1.7	Similar to sodium-dependent transporter
	<i>ybbF</i>	3.4	1.6	Similar to sucrose phosphotransferase enzyme II
	<i>ybbE</i>	4.6	1.6	Similar to glycosyl hydrolase (β -lactamase)
	<i>ybbD</i>	3	1.4	Similar to glycosyl hydrolase (β -hexosaminidase)
	<i>ybbC</i>	3	1.4	Unknown
<i>ycgL</i> <i>yclF</i>	<i>ycgL</i>	3.7	2	Unknown
	<i>yclF</i>	18.2	2.1	Similar to di-tripeptide ABC transporter (membrane protein)
<i>ydbI</i> <i>yeaA-ydjP</i>	<i>ydbI</i>	12.5	1.8	Similar to YtvI (putative permease)
	<i>yeaA</i>	5.3	2	Unknown
	<i>ydjP</i>	6.4	2	Similar to chloride peroxidase (α/β hydrolase family), similar to YisY
<i>yflB</i> <i>yfnED</i>	<i>yflBⁱ</i>	5.1	3	Unknown
	<i>yfnEⁿ</i>	2.9	2.2	Similar to glycosyl transferase, similar to SpsA
	<i>yfnD</i>	2.1	1.7	Unknown
<i>ygaE</i> <i>yhaST</i>	<i>ygaEⁿ</i>	2	1.6	Unknown
	<i>yhaS</i>	5.9	1 ¹	Unknown
	<i>yhaT</i>	5	1.8	Similar to TrkA (potassium uptake)
<i>yhbB</i> <i>yheCD</i>	<i>yhbB</i>	8.4	2.1	Inner coat
	<i>yheC</i>	4 ^o	1.5 ^o	Similar to YheD, inner coat ^P
	<i>yheD</i>	32	1.7	Similar to YheC, inner coat ^P
<i>yhfN</i> <i>yhfO</i>	<i>yhfN</i>	14	2.7	Similar to metalloendopeptidase (M48 family)
	<i>yhfOⁿ</i>	2.5	2.6	Similar to acetyltransferase (GCN5 family), similar to SpsD
<i>yhjL</i>	<i>yhjL</i>	20.8	3.2	Similar to DegT/DnrJ/EryC1/StrS aminotransferase, similar to SpsC and PatA
<i>yhxC</i>	<i>yhxC</i>	41.7	3.4	Similar to alcohol dehydrogenase (short chain adh family), similar to YkvO
<i>yisY</i>	<i>yisY</i>	7	2.1	Similar to chloride peroxidase (α/β hydrolase family), similar to YdjP
<i>yitE</i> <i>yjaV</i>	<i>yitE</i>	33.7	4.8	Unknown
	<i>yjaV</i>	13.4	3.7	Similar to UDP-Glc 4-epimerase

(continued)

Table 2 Continued

Operon organization	Gene	Ratio ^a sporulation	Ratio ^b overexpression	Function ^c
<i>yjaZ</i>	<i>yjaZ</i>	6.4	1.9	Unknown
<i>yjbE</i>	<i>yjbE</i>	19.6	2.9	Similar to toxic anion resistance protein (tellurium efflux)
<i>yjfA</i>	<i>yjfA</i>	17.2	4.3	Unknown
<i>yknUV</i>	<i>yknU</i>	9.7	1.4 ^d	Similar to ABC transporter (ATP-binding protein)
	<i>yknV</i>	10.7	1.6	Similar to ABC transporter (ATP-binding protein)
<i>ykoP</i>	<i>ykoP</i>	3.1	4.3	Unknown
<i>ykrL</i>	<i>ykrLⁿ</i>	2.1	2.2	Similar to heat-shock protein (putative protease)
<i>ykvO</i>	<i>ykvO^a</i>	2.2	2.4	Similar to Glc 1-dehydrogenase (short chain adh family), similar to YhxC
<i>ylaK</i>	<i>ylaK</i>	50.5	2.7	Similar to PhoH ATPase
<i>ylaM</i>	<i>ylaM</i>	31.6	2.7	Similar to glutaminase
<i>ylbK</i>	<i>ylbK</i>	4.2	5.3	Similar to papatin-like phospholipase, similar to YqhO
<i>ylbO</i>	<i>ylbO</i>	23.4	2.8	Similar to DNA-binding protein (Myb family), similar to RsfA (YwfN)
<i>yloB</i>	<i>yloB</i>	45.1	4.5	Similar to calcium-transporting ATPase, sporulation ^a and germination, ¹⁰⁷
<i>ylxY-mlpA</i>	<i>ylxY</i>	6.5	3.3	Similar to polysaccharide deacetylase, sporulation ^{a,t} , similar to YbaN
	<i>mlpA</i>	11.5	3.2	Similar to metalloendopeptidase (M16 family)
<i>ymaF</i>	<i>ymaF</i>	30.3	52.4	Sporulation ^{a,t}
<i>yncD</i>	<i>yncD</i>	14.1	2.3	Similar to Alr ^s
<i>yndA</i>	<i>yndA</i>	9.4	2.7	Unknown
<i>ynG</i>	<i>ynGD^{n,t}</i>	2.4	2.4	Similar to phosphoesterase (DHH family)
<i>ynGK</i>	<i>ynGK</i>	4.3	2.1	Unknown
<i>yobN</i>	<i>yobN</i>	10.9	4.7	Similar to L-amino acid oxidase
<i>yocMLK</i>	<i>yocM</i>	8.1	1.7	Similar to CotP and CotM (α -crystallin domain of small heat shock proteins)
	<i>yocL</i>	3	1 ⁱ	Unknown
	<i>yocK</i>	5.5	3.3	Similar to general stress protein (DksA/TraR C4-type Zn finger family)
<i>yodN</i>	<i>yodN</i>	35.7	3.3	Unknown
<i>yodTSRQ (argE)-yodP-kamA (yodO)</i>	<i>yodT</i>	50.4	2.2	Similar to adenosylmethionine-8-amino-7-oxononanoate aminotransferase
	<i>yodS</i>	20.8	2.3	Similar to 3-oxoadipate CoA-transferase, similar to ScoA
	<i>yodR</i>	52.8	2.4	Similar to butyrate-acetoacetate CoA-transferase, similar to ScoB
	<i>yodQ</i>	7.8	2.2	Similar to succinyl-diaminopimelate desuccinylase
	<i>yodP</i>	43	2.2	Unknown (putative β -lysine-acetyltransferase)
	<i>kamA</i>	27.7	2	Lysine 2,3-aminomutase
<i>yozN-yocN</i>	<i>yozN</i>	8.3	- ^h	Similar to YocN
	<i>yocN</i>	3	2.3	Similar to YozN
<i>ypbG</i>	<i>ypbG</i>	17.1	1.7	Similar to calcineurin-like phosphoesterase
<i>ypiA</i>	<i>ypiA</i>	20.9	1.8	Unknown (putative GlcNAc transferase)
<i>ypjD-dapB-mgsA(ypjF)-ypjGH-cca (papS)-birA</i>	<i>ypjD</i>	6.7	2.2	Unknown
	<i>dapB</i>	3.4	1.6	Dihydrodipicolinate-reductase (diaminopimelate synthesis)
	<i>mgsA</i>	1.6	2.6	Probable methylglyoxal synthase
	<i>ypjG</i>	3.3	6.6	Unknown
	<i>ypjH</i>	3.3	0.4 ^d	Similar to glycosyltransferase (lipopolysaccharide biosynthesis)
	<i>cca</i>	4.8	1.5	tRNA nucleotidyltransferase
	<i>birA</i>	3.2	1.9	Biotin acetyl-CoA-carboxylase synthetase and transcriptional regulator
<i>yqeW</i>	<i>yqeW</i>	19.6	2.6	Similar to sodium/phosphate cotransporter
<i>yqgT</i>	<i>yqgT</i>	2.7	2	Similar to DL-endopeptidase I, sporulation specific in <i>B. sphaericus</i> ¹⁰⁸
<i>yqhO</i>	<i>yqhO</i>	4.3	2.1	Similar to papatin-like phospholipase, similar to YlbK
<i>yqhV</i>	<i>yqhV</i>	8.4	3.4	Unknown
<i>yqkF</i>	<i>yqkFⁿ</i>	2.4	1.5	Similar to oxidoreductase (aldo-keto reductase family)
<i>yrbD</i>	<i>yrbD</i>	7	2.1	Similar to Na ⁺ /H ⁺ -dependent alanine carrier protein
<i>yrbG</i>	<i>yrbG^a</i>	14.6	11.3	Unknown
<i>yrrRS</i>	<i>yrrR</i>	3.2	2.7	Similar to penicillin-binding protein SpoVD
	<i>yrrS</i>	4.7	2.1	Unknown

(continued)

Table 2 Continued

Operon organization	Gene	Ratio ^a sporulation	Ratio ^b overexpression	Function ^c
<i>yrzE</i>	<i>yrzE</i>	61	2.7	Unknown
<i>ysnE</i>	<i>ysnE</i>	2.8	1.7	Similar to acetyltransferase (GCN5 family)
<i>ytbD</i>	<i>ytbD</i>	4.3	2.3	Similar to antibiotic resistance protein (sugar transporter family)
<i>ytpAB</i>	<i>ytpA</i>	3	1.5	Similar to triglyceride lipase (lysophospholipase)
	<i>ytpB</i>	4	1.8	Unknown
<i>ytvB</i>	<i>ytvB</i>	9.4	2.4	Unknown
<i>yugT</i>	<i>yugT</i>	13.7	2.9	Similar to glycosyl hydrolase (exo- α -1,4-glucosidase), similar to GlgB
<i>yuiE</i>	<i>yuiE</i> ⁿ	3.2	1.3	Similar to leucyl aminopeptidase
<i>yumB</i>	<i>yumB</i>	2.5	2.6	Similar to NADH dehydrogenase
<i>ywcBA</i>	<i>ywcB</i>	20.7	4.3	Unknown
	<i>ywcA</i>	18.8	2.9	Similar to sodium-dependent symport
<i>ywdL</i>	<i>ywdL</i>	56.6	1.7	Germination ^{g, h}
<i>ywjE</i>	<i>ywjE</i>	11.8	2.6	Similar to cardiolipin synthetase (phospholipase D family), similar to YwnE
<i>ywjF-acdA</i>	<i>ywjF</i>	2.6 ^o	1.7 ^o	Similar to iron-sulphur-binding reductase
	<u><i>acdA</i></u>	3.7	2.7	Acyl-CoA dehydrogenase, similar to MmgC and YngJ
<i>ywlB^f-ywlC</i>	<i>ywlB</i>	3.5	3.5	Unknown
	<i>ywlC</i>	3.4	1.9	Similar to Sua5 (RNA-binding protein required for normal growth in yeast), growth ^h
<i>ywnE</i>	<i>ywnE</i> ⁿ	2.1	3.6	Similar to cardiolipin synthase (phospholipase D family), similar to YwjE
<i>ywqF</i>	<i>ywqF</i>	3.8	1.9	Similar to UDP-Glc/GDP mannose dehydrogenase
<i>yxiLM-deaD</i>	<i>yxiL</i>	- ^v	- ^v	Unknown
	<i>yxiM</i>	3.8	2	Similar to rhamnogalacturonan acetyltransferase
	<i>deaD</i>	2.5	2.5	ATP-dependent RNA helicase
<i>yxiO</i>	<i>yxiO</i>	3.9	1.5	Unknown (putative permease)
<i>yyaE</i>	<i>yyaE</i> ^{n,w}	2.2	1.6	Similar to formate dehydrogenase (molybdopterin oxidoreductase family)
<i>yycN</i>	<i>yycN</i> ⁿ	4.5	2	Similar to acetyltransferase (GCN5 family)
<i>yycOP</i>	<i>yycO</i>	5.1	3.8	Unknown
	<i>yycP</i>	4.8	2	Unknown

Inactivation of genes indicated in bold led to a sporulation defect. Genes underlined are conserved in *B. anthracis*, *B. halodurans*, *O. iheyensis*, *C. acetobutylicum*, *C. perfringens* and absent in *Listeria*.

^a Ratios of relative RNA levels in *sigE*⁺ versus *sigE* mutant.

^b Ratios of relative RNA levels from *sigE* overexpression (+IPTG) one hour after induction compared to the same time point without overexpression (-IPTG).

^c Possible function according to SubtiList (<http://genolist.pasteur.fr/SubtiList>) with additional comments based on BLAST searches results (<http://www.ncbi.nlm.nih.gov/BLAST/>), SwissProt annotations (<http://us.expasy.org/sprot/>), Pfam annotations (<http://www.sanger.ac.uk/Software/Pfam/>) and contributions from this study.

^d Mutant impaired in sporulation as judged by decreased heat-resistance of spores (see Table 4).

^e Reduced growth rate in LB medium (unpublished observation).

^f A predicted terminator sequence is present between the two genes.

^g Germination defect according to MICADO database, (<http://locus.jouy.inra.fr/cgi-bin/genmic/madbase/progs/madbase.operl>)

^h Spot excluded from the data analysis in the overexpression experiment (signal intensity not significantly above background level), this gene was included in the Table because it seems to be part of an operon with another gene that was successfully identified.

ⁱ This gene could not be inactivated by long-flanking homology PCR.

^j Gene not annotated in *B. subtilis* genome sequence and not present on the microarrays.

^k Activation of σ^k is delayed in the mutant (J.S., P.E. & R.L., unpublished observation).

^l Missed cut-off in the overexpression experiment, this gene was included in the Table because it seems to be part of an operon with gene(s) that was (were) successfully identified.

^m Signal could also be due to readthrough from the upstream σ^E promoter for *ybaN* (transcribed in the opposite direction); *kbaA* stop is 1 bp downstream of *ybaN* stop.

ⁿ Inactivation of this gene has not been attempted.

^o Missed cut-off in both experiments, this gene was included in the Table because it could be part of an operon with another gene that was successfully identified.

^p C. van O. & R.L. (unpublished observation).

^q Previously reported sporulation defect not observed under our conditions.

^r Sporulation defect according to MICADO database, (<http://locus.jouy.inra.fr/cgi-bin/genmic/madbase/progs/madbase.operl>)

^s Mutant requires exogenous D-Ala for growth (P.E. & R.L., unpublished observation).

^t Signal could also result from readthrough from the upstream σ^E promoter for *yngJ* (transcribed in the same direction).

^u Signal could also be due to readthrough from the upstream σ^E promoter for *spoVB* (transcribed in the opposite direction), *yrbG* stop is 3 bp downstream of *spoVB* stop.

^v Spot excluded from the data analysis in both experiments (signal intensity not significantly above background level), this gene was included in the Table because it could be part of an operon with two other genes that were successfully identified. In addition, it is a very small gene and small genes frequently give signals of low intensity in microarray experiments.

^w Signal could also be due to readthrough from the upstream σ^E promoter for *yyaD* (transcribed in the same direction); a predicted terminator sequence is present between the two genes.

fluorescent signals for each gene between the *sigE*⁺ strain and the *sigE* mutant strain (left panel) and between cells of MF453 grown in the presence and in the absence of inducer (right panel). The majority of the points fall on or near the line with a slope of 1, indicating that the RNA levels for the large majority of genes was not significantly influenced by σ^E . In contrast, points that were significantly off the line corresponded to the genes whose expression was directly or indirectly altered by σ^E . In total, 372 genes were identified whose expression was significantly higher in the *sigE*⁺ cells than in the *sigE* mutant cells, whereas 279 genes were expressed at lower levels. Although the number of down-regulated genes was comparable to the number of up-regulated genes, the fold changes in expression were much higher for up-regulated genes. Indeed, 237 genes were up-regulated more than fourfold (and 132 more than tenfold) in the *sigE*⁺ strain. Conversely, the changes in expression ratios for the majority of down-regulated genes were modest (234 genes with less than fourfold changes in expression). Interestingly, a high proportion of the down-regulated genes were previously characterized sporulation genes whose expression is known to be under the control of regulatory proteins that are active prior to σ^E activation. Examples are *kinA*, *citG* and *yisK*, which are transcribed under the control of the early-acting sporulation regulatory protein $\sigma^{H17,18}$ and *spoIIR*, *bofC* and *ywhE*, that are transcribed in the forespore under the control of σ^F .^{18,48} Hereafter we are principally concerned with the identification of genes that were up-regulated by σ^E and likely to be under the direct control of σ^E .

Of the 372 genes that were up-regulated in sporulating cells of the *sigE*⁺ strain, 225 were also up-regulated in the vegetative cells that were induced to produce σ^E during growth. These 225 genes, which are listed in Tables 1 and 2, are our strongest candidates for genes transcribed by σ^E -containing RNA polymerase. A total of 27 additional genes that did not meet our most stringent criteria in the transcriptional profiling experiments (plus one that was missed during the

annotation of the *B. subtilis* genome, see below) were nonetheless considered likely to be under the direct control of σ^E for reasons explained below and in the footnotes to the Tables, resulting in a grand total of 253 putative members of the σ^E regulon (Table 3(A)). Of these 253 genes, 72 had previously been shown to be under the direct control of σ^E and are listed in Table 1. Of these 72 genes, 56 were identified in both kinds of transcriptional profiling experiments, whereas ten genes missed the cut-off under one of the two conditions. Of the six genes that were not identified by either transcriptional profiling strategy, three (*spoIIP*, *spoVK* and *yoaW*) would have been included under less restrictive criteria and three are very short genes (*spoVM*, *usd* and *yfhS*). It has been reported that short transcripts often give signals of low intensity that are difficult to detect in arrays based on PCR products of full-length ORFs.^{17,49}

Grouping genes under the control of σ^E into operons

Of the 253 genes described above, 181 had not been previously reported to be regulated by σ^E and are listed in Table 2. As a further criterion for assigning genes to the σ^E regulon, we sought to identify sequences conforming to the consensus for σ^E -recognized promoters in the regions upstream of these genes. In doing so, however, it was necessary to ascertain which genes were likely to be part of an operon and hence under the control of a common promoter. Based on previously reported distances between genes that are known to be contained within operons, we assumed for simplicity that two adjacent σ^E -controlled genes in the same orientation that were separated by less than 80 bp were likely to lie in the same operon. When the distance between the two adjacent genes was greater than 80 bp, we searched the intergenic region for the presence of a possible σ^E -recognized promoter (see below). If no such candidate promoter sequence was present (and if no apparent transcriptional terminator could be detected), we attributed transcription of the downstream gene to transcriptional read-through from

Table 3. Genes and groupings

	Previously known	Newly identified	Total
A. σ^E-controlled genes			
Identified in both profiling experiments	56	169	225
Identified in <i>sigE</i> ⁺ versus <i>sigE</i> only	5	8	13
Identified in IPTG ⁺ versus IPTG ⁻ only	5	–	5
Previously missed in genome annotation	–	1	1
Not identified in profiling experiments	6	3 ^a	9
Total genes regulated by σ^E	72	181	253
B. Grouping genes under the control of σ^E into operons			
Operons (number of genes)	13 (49 genes)	34 (94 genes)	47 (143 genes)
Single-gene transcription units	23 genes	87 genes	110 genes
Total transcription units (number of genes)	36 (72 genes)	121 (181 genes)	157 (253 genes)

^a In operon and up-regulated (but below cut-off value).

the upstream gene. In other words, the genes were assigned to a common operon.

Two previously described examples of transcriptional read-through across a larger than 80 bp intergenic region were confirmed in our analysis. Transcription of *coxA*, which is located 228 bp downstream of *safA*, and *mbl*, which is 164 bp downstream of *spoIIID*, is known to originate in part from σ^E -controlled promoters located upstream of *safA* and *spoIIID*, respectively.^{50–52} Our transcriptional profiling confirmed that *coxA* and *mbl* are under the control of σ^E and our promoter search algorithm (see below) failed to detect candidates for σ^E -controlled promoters in the intergenic regions. However, neither gene is regulated solely by σ^E and in both cases a second promoter (not recognized by σ^E) has been identified within the large intergenic region. Two additional examples of read-through across large intergenic regions were revealed by transcription start site mapping experiments (see below). Transcription of *yesJ*, which is located 123 bp downstream of *cotJ*C, and *yqfY*, which is located 121 bp downstream of *yqfZ*, was found to originate from σ^E -controlled promoters located upstream of *cotJ*A and *yqfZ*, respectively.

Finally, informatics led us to conclude that *yjbH*, *ylaL* and *ypjC*, which were identified in our transcriptional profiling experiments, were each likely to be transcribed by read-through from a downstream gene of opposite orientation. Hence, σ^E -directed transcription across these genes would have generated anti-sense RNAs rather than protein-coding sequences. We do not consider these as members of the σ^E regulon and they are not included in Tables 2 and 3.

In total, we provisionally assign 94 of the newly identified σ^E -controlled genes to 34 operons. This left us with a total of 87 genes that we infer are transcribed as single transcription units from a promoter that is close to, or immediately upstream of, the protein-coding sequence. The entire list of individual transcription units and of predicted operons is presented in Table 2 and is summarized in Table 3(B), whereas the distances between adjacent genes in the operons are shown in Table S1 of supplementary material.

Searching for σ^E -controlled promoters using computational methods

Based on the above analysis we infer that the newly identified σ^E -controlled genes listed in Table 2 are transcribed from approximately 121 σ^E -recognized promoters (34 operons plus 87 genes). We used two complementary computational methods in an effort to identify these promoters. In the first method, 38 previously known σ^E -recognized promoters (including cases where two promoters under σ^E control govern the same gene) were compiled to design a consensus weight matrix.⁵³ Each element in the matrix represented the reported frequency of the corresponding

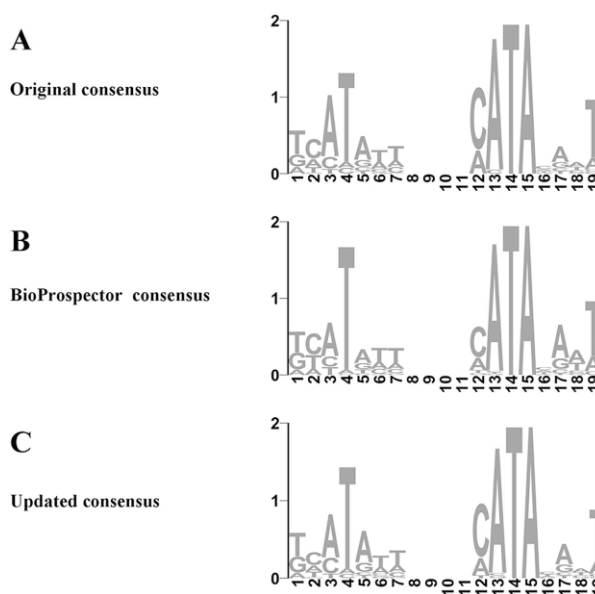


Figure 2. Consensus promoter sequences recognized by σ^E -containing RNA polymerase. (A) Consensus sequence logo obtained from the compilation of 38 previously identified σ^E -binding sites. The height of the letters indicates the frequency that a given base is represented at each position. Positions 1–7 on the horizontal axis correspond to the seven base-pairs of the -35 region and positions 12–19 to the eight base-pairs of the -10 region. The spacing between the two regions is $14(\pm 1)$ bp but was shortened to 4 bp (positions 8–11) in the Figure for simplicity. (B) Consensus sequence logo obtained from the compilation of sequences identified as common motifs in regions upstream of σ^E -regulated genes. The genes were selected according to the results of the transcriptional profiling experiments and their upstream regions analyzed with BioProspector. The noise in the consensus sequence was reduced by the use of an optimization algorithm. (C) Consensus sequence logo obtained from the compilation of 38 previously identified σ^E -binding sites and 24 newly identified σ^E promoters confirmed by transcription start site mapping.

nucleotide at every position of the -35 or -10 elements of the promoter. The resulting consensus sequence logo is presented in Figure 2(A) in which the height of the letters indicates the frequency that a given base is represented at that position.⁵⁴ Next, we analyzed 200 bp-long regions extending upstream from every gene and the first gene of every operon identified in both transcriptional profiling experiments or that was previously reported to be under the control of σ^E . These regions were searched for sequences conforming to the consensus weight matrix and the results were sorted according to the score associated with the sequences identified as possible σ^E -controlled promoters. As an additional criterion, only a fixed distance of $14(\pm 1)$ bp was allowed between the -35 and -10 elements. The predicted sites are listed along with their associated scores in Table S1 of the supplementary material. To assess the accuracy of the consensus weight matrix, we assigned scores to the previously known

<i>gerM</i>	ttcg GTCTATT tcctaagaggctcg TATACATA atagt AC aaacatcct taggaggg aaaagtATG
<i>prkA</i>	caaa GCATGTA tcacgag-atggcgg CATAGATT gtatc AAAG tac-25 bp-gatag aggaggt ccttATG
<i>ybaN</i>	tcgg TTATATT caattgt-ccatgct CATAAGAT gt AAAACA aga-10 bp-gata aggaggg catcatcaGTG
<i>ycgF</i>	ttgt GCATAGC ttggccc-gttccc GAATAAAT Tgtac AAG ttacataagaga aggaggt acgggcccGTG
<i>ydcA</i>	tacg TACTATT taaattgg-tttgtct CATAAAGC gttact AT agata-10 bp-aagaaa agggg gctcggATG
<i>ydcC</i>	gtct GCATATT agggaaa-ccccact CATATATT tgatagt G catt aggagag acaagttgtTTG
<i>yhaX</i>	gttg TTCTAAA cacagag-acagggt CATATCCT atacac AAG tacagtcacaa ggagggg gacatctctGTG
<i>yhbH</i>	gttt GTCAATT aaacgtgcatatgtg CATATGAT gaatat AAAT ct-65 bp-tt aggaggg gaattcATG
<i>yjbX</i>	tttc TTCTGAT tttcagc-tttctgt CATATAGA tagaat AT gacacaat cgagggg ATG
<i>ykvI</i>	ttcg TCTTTT Caaaggaa--cttgtct CATAGGTT at AAAAAG gca-20 bp-tcgt cgaggg gaagaaATG
<i>ykvU</i>	aata AAATAAT ttttgaa-cttgtct CATATGAT gttgg AG tacaagagaca aggagag acatcATG
<i>ylbJ</i>	tatg GTCTAAA actgaaccctatgct CGTATATT agtac AAAG aatca-75 bp-gcat ggggagg taagaATG
<i> yngJ</i>	cagg GAATGAT tatagaa-ctgcct AATAGGAT gttac AAAG atgtga aggagg aaacgatGTG
<i>yjpJ</i>	gttg GTCTATC cctgtcc-ctttct CATATTC Tgtaagg AG agagagt gggagg gacctgcATG
<i>yqfC</i>	aaag TGTTAGA acctcct-ttcaaat CATACACA tgagatg Aaaggggg ttcttttATG
<i>yqfZ</i>	tttg TCATGAA tcgaaat-ggacaag CATACTAT ggtat AAAAAT ttcgtatg ggggagg aaaagcgATG
<i>ytrH</i>	atca TCATACT tgctcctg-aaagctc AATATGAT at AAAA ggtaggtgataaagATG
<i>ytvI</i>	gtat TCATATT cgccgc--agcgtg AATACATA - AAAAAT aggacatgctgacgag aggagg accgtttttGTG
<i>ytxC</i>	ttac GTCTATT tttaaaaa-catcccc CATATACT tgtaac AG atgccgta agggggg acaacATG
<i>yunB</i>	ctat TACTATT gtcccctc-ttacaag CATACATT gtgat AT gta agggggg attttcTTG
<i>yuzC</i>	tttg TCATATT cggaattagggatc TATACATA tag AAAC atccttttt ggaggg tgctcccATG
<i>yxjC</i>	atcg GAATAGT ttgccc-a-cgcttct CGTACATA tg AAAG ggg-50 bp-gaa aggggg gatgtgtGTG
<i>yyaD</i>	tatg GCATGTT tgcttct-ctttatt TATATAGT aacaat AAC ggg-15 bp-gtctat gaggt gataGTG
<i>yybI</i>	tatg TCTTTGA gtgcca--ttctcg CATATAAT gattc AT tcagctgattag ggagg tgacacaaTTG
Consensus	DYMRWW CATAHAWT

D is A or G or T, H is A or C or T, M is A or C, R is A or G, W is A or T, Y is C or T

Figure 3. Mapping of transcription start sites by 5'-RACE-PCR. The underlined uppercase bold letters identify the 5' ends of mRNAs from σ^E -controlled genes as determined by RACE-PCR. Also indicated are the corresponding -35 and -10 regions (uppercase letters in bold), the ribosome-binding site (double underlining) and the translation start site (uppercase letters).

σ^E -controlled promoters. The matrix identified the correct promoter in 32 out of 38 cases.

The second computational method, which used a program known as BioProspector†, did not depend on prior knowledge of recognition sequences for σ^E -containing RNA polymerase. BioProspector uses a Gibbs sampling algorithm designed to identify common motifs in selected sequences.^{55,56} The program was asked to find a dimer motif, with one block of 7 bp corresponding to the -35 element and one block of 8 bp corresponding to the -10 element separated by a gap of length 11–15 bp (see Materials and Methods). Because the algorithm used by BioProspector is stochastic, each BioProspector run does not give the same results for the same input. Therefore, 100 independent BioProspector runs were performed on the same set of upstream sequences that had been analyzed with the consensus weight matrix and the top five most conserved motifs were selected. Subsequently, an optimization algorithm was used to reduce the noise in the motif signal and to limit the predictions to binding sites that have a spacing of 14(±)1 bp between the -35 and the -10 elements. The algorithm goes through each position of each upstream region and adds a new motif starting at that position only if the score of the resulting consensus sequence is increased (see

Materials and Methods for details). The consensus sequence logo obtained from BioProspector following the optimization procedure is given in Figure 2(B). Site predictions for each upstream sequence for the genes in the dataset were obtained directly from the optimization algorithm and presented in Table S1 of the supplementary material. BioProspector appeared slightly less successful than the weight matrix-based method in identifying known σ^E binding sites in that the program recognized 30 of the 38 known promoters. However, given that BioProspector was uninformed by previously known promoter sequences (except for the range of spacing between the -10 and -35 elements), the success rate was impressively high. In fact, known σ^E -controlled promoters for six genes (*asnO*, *glgB*, *phoB*, *spoVM*, *yfhS* and *yoaW*) failed to be identified by either computational method. Most of these genes also fell below the cut-off value in at least one of the profiling experiments, indicating that they are likely to have weakly recognized promoters. Out of the 121 possible new σ^E -controlled promoters predicted by the consensus weight matrix method, 71 were also selected by BioProspector. In addition, in five cases, BioProspector identified a possible promoter different from the promoter that had been predicted by the consensus weight matrix. Finally, in 14 cases BioProspector selected more than one possible σ^E -binding site per upstream sequence, suggesting that some

† <http://bioprospector.stanford.edu>

Table 4. Functional analysis

Gene	Sporulation efficiency (%)	Orthologs					
		<i>B. anthracis</i>	<i>B. halodurans</i>	<i>O. iheyensis</i>	<i>C. acetobutylicum</i>	<i>C. perfringens</i>	<i>L. monocytogenes L. innocua</i>
<i>prkA</i>	27	<i>ba1123</i>	<i>prkA(bh1029)</i>	<i>ob2654</i>	<i>prkA(cac0579)</i>	<i>prkA(cpe1334)</i>	–
<i>ybaN</i>	12	<i>ba0732</i>	<i>bh0243</i>	<i>ob0199</i>	–	–	–
<i>yhbH</i>	25	<i>ba1124</i>	<i>bh1031</i>	<i>ob2647</i>	<i>cac0580</i>	<i>cpe1333</i>	–
<i>ykvoV</i>	0.001	<i>ba4160</i>	–	–	–	–	–
<i>ylbJ</i>	0.0001	<i>ba4605</i>	<i>bh2588</i>	<i>ob1452</i>	<i>cac1740</i>	<i>cpe1727</i>	–
<i>ypjB</i>	27	<i>ba2068</i>	<i>bh1676</i>	–	–	–	–
<i>yqfC</i>	0.0001	<i>ba4979</i>	<i>bh1359</i>	<i>ob1957</i>	<i>cac1290</i>	<i>cpe2021</i>	–
<i>yqfD</i>	0.0001	<i>ba4978</i>	<i>bh1360</i>	<i>ob1956</i>	<i>cac1291</i>	<i>spoIV(cpe2020)</i>	–
<i>ytrH</i>	11	<i>ba5273</i>	<i>bh3170</i>	<i>ob2178</i>	–	–	–
<i>ytrI</i>	19	<i>ba5274</i>	<i>bh3171</i>	<i>ob2179</i>	–	–	–
<i>ytrI</i>	29	<i>ba5264</i>	<i>bh0463</i>	<i>ob2169</i>	<i>cac1794</i>	<i>cpe2648</i>	–
<i>yunB</i>	14	<i>ba0080</i>	<i>bh1036</i>	<i>ob2370</i>	<i>cac2300</i>	<i>cpe1961</i>	–

upstream regions may contain more than one σ^E promoter.

Transcriptional mapping of σ^E -controlled promoters

To evaluate the accuracy of the computer predictions, we mapped the promoters for 24 of the newly identified σ^E -controlled genes (Figure 3) using 5'-RACE PCR^{57,58} (see Materials and Methods). These genes were selected because our systematic functional analysis (see below) indicated that they play important roles in sporulation. Of the 24, 18 had been predicted by both BioProspector and the consensus weight matrix. Of the remaining six, four promoters had been identified by the weight matrix only, one by BioProspector only and one (for *ylbJ*) by neither. Nonetheless, the promoter for *ylbJ* exhibited sufficient similarity to the σ^E consensus (at least in hindsight) to be considered a *bona fide* σ^E -controlled promoter. This promoter was probably missed in our informatics approaches because of its unusual –10 region sequence, in the sense that it starts with the sequence CGTA instead of the canonical CATA.

An updated consensus was obtained by combining the previously known promoter sequences with the newly determined ones, corresponding to a total of 62 mapped σ^E -binding sites (Figure 2(C)). The updated consensus was then used to re-examine all of the promoter predictions. When the predictions differed between the two computational methods or when more than one site was predicted by BioProspector, the motif that matched best with the updated consensus was selected as the most likely promoter and indicated as such in Table S1 of the supplemental material. Interestingly, out of the 121 genes and operons identified as being under the control of σ^E , five were previously reported as being under the control of σ^K (*cotB*, *cotG*, *cotH*, *cotYZ* and *spsGIJK*).¹⁹ Promoters controlled by σ^K and σ^E have almost identical –10 consensus sequences with the principal difference

between the two classes of promoters lying in their –35 regions. Therefore, it is possible that some promoters under the control of σ^K might also be recognized to some extent by σ^E .

Functional analysis of σ^E regulon members by gene inactivation

As a result of transcriptional profiling, 5'-RACE-PCR, and the use of computational methods, a total of 181 genes transcribed from 121 σ^E -controlled promoters were identified as new members of the σ^E regulon. To analyze the role of these genes in sporulation, we used the technique of the long-flanking homology PCR⁵⁹ to create null mutations in many of the newly identified genes and operons. In the case of operons, we generally disrupted only the first gene in the transcription unit, although in 20 cases more than one gene in the operon was mutated. We failed to obtain null mutants of *ispE* and *yflB* (probably for technical reasons in the case of *yflB*, whereas we suspect that *ispE* is an essential gene because it is presumed to be part of the non-mevalonate pathway to isoprenoid synthesis⁶⁰), and we did not attempt to create mutations of *kbaA*, *lplD*, *nadA*, *spsC*, *spsGIJK*, *tmrB*, *yfnED*, *ygaE*, *yhfO*, *ykrL*, *ykvoO*, *yingD*, *yqkF*, *yuiE*, *ywnE*, *yyaE* and *yycN* because these genes are up-regulated by σ^E only modestly. Thus, a total of 98 mutants were created (see Table S1 of supplementary material for the entire list of mutants). Of these, 11 (as well as a twelfth introduced below) were found to be impaired in sporulation as judged by measurements of the production of heat-resistant spores (Table 4 and Table S1 of supplementary material). Mutants of *ykvoV* and *ylbJ* and of the *yqfCD* operon were severely impaired in sporulation (which was reduced by more than 1000-fold), whereas the other mutants sporulated at a frequency that was three to eightfold lower than the wild-type strain. The *ykvoV*, *ylbJ* and *yqfCD* mutants produced spores that appeared dark under phase

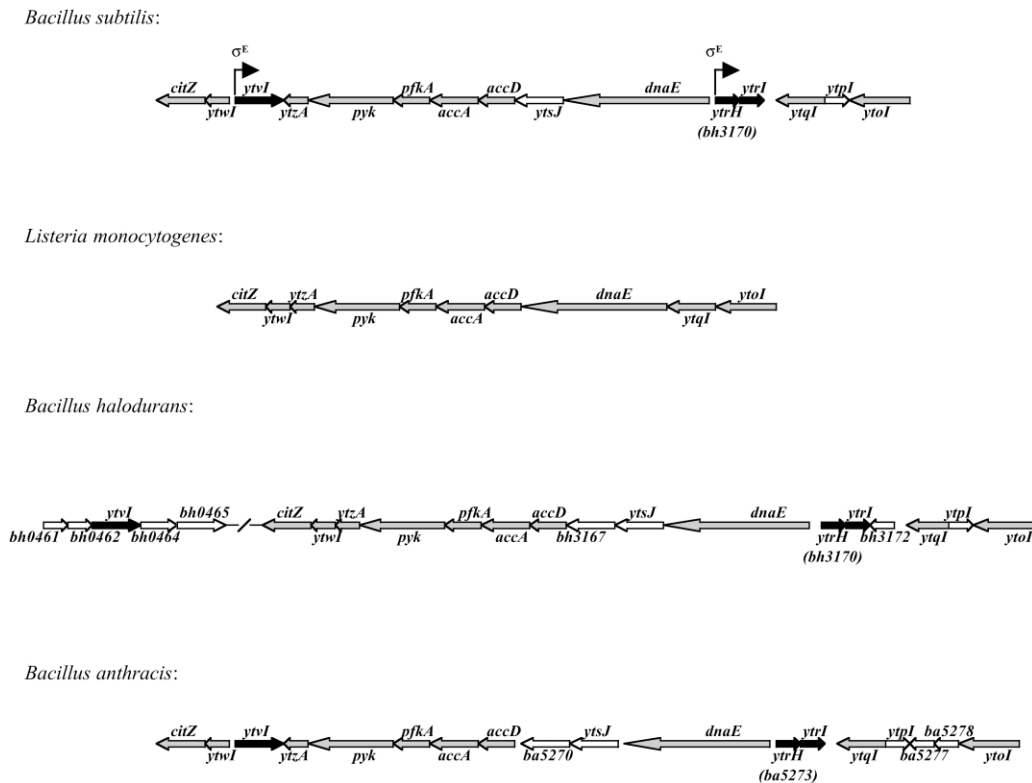


Figure 4. Conservation of gene organization in related Gram-positive bacteria. The chromosomal region in the vicinity of *dnaE* is displayed for four related species: *B. subtilis*, *L. monocytogenes*, *B. halodurans* and *B. anthracis*. Genes are shown as arrows indicating the direction of transcription. Filled black arrows are σ^E -controlled genes, filled gray arrows are genes conserved in all four organisms, open arrows are genes that are not conserved. The location of σ^E -controlled promoters is also indicated. Genes are named according to that of the corresponding *B. subtilis* ortholog, with the exception of genes having no *B. subtilis* ortholog, which are designated by their original annotation. In the case of *B. halodurans*, a second chromosomal region is shown in order to include the ortholog to *ytvI*, the separation between the two regions being indicated by the broken line. The *ytrH* gene was identified in the present work as an ortholog of the *B. halodurans* gene *bh3170* and had not been annotated as a gene in the *B. subtilis* genome sequence.

contrast microscopy (as opposed to the bright appearance of wild-type spores), suggesting a defect in maturation of the spore core or the spore cortex. That mutants of *yqfCD* and *yIbJ* were defective in sporulation has been independently observed by the Japanese functional analysis network† and by S. Trewhitt, H. Chirakkal & A. Moir (personal communication).

Some of the newly identified sporulation genes show significant similarity to genes of known function‡: *ybaN* is homologous to a polysaccharide deacetylase, *ytvI* encodes a putative permease and *ykV* is homologous to thioredoxin. The *prkA* gene encodes a previously identified protein kinase but was not previously known to be involved in sporulation.⁶¹

Sporulation genes are generally conserved among endospore-forming members of the genus *Bacillus* and to some extent among members of the distantly related endospore-forming genus

*Clostridium*³⁴ but not among members of the closely related but non-endospore-forming genus *Listeria*.⁶² At this time, in addition to *B. subtilis*, the genomes of five endospore-forming species (*B. anthracis*,²⁸ *B. halodurans*,³¹ *O. iheyensis*,³² *C. acetobutylicum*,³⁰ and *C. perfringens*²⁹) and two non-endospore-forming *Listeria* species (*L. monocytogenes* and *Listeria innocua*)⁶² have been entirely sequenced and annotated. We analyzed the conservation of the σ^E -controlled genes whose disruption impairs sporulation, and observed the following (Table 4). Orthologs were found in *B. anthracis* and with one exception in *B. halodurans* and two exceptions in *O. iheyensis*. In the case of *Clostridium*, orthologs for seven of the 12 genes were detected. In the case of *Listeria*, however, orthologs for the all 12 genes were absent, even though the regions surrounding the missing sporulation genes were usually highly conserved. As an example, Figure 4 compares the regions of the *B. subtilis* and *Listeria* chromosomes containing the newly discovered sporulation genes *ytvI* and *ytrI*. Orthologs of almost every *B. subtilis* gene in the vicinity of *ytvI* and *ytrI*, but not the two sporulation genes

† <http://bacillus.genome.ad.jp>

‡ <http://genolist.pasteur.fr/SubtiList>

Table 5. Protein localization

Gene	GFP fusion localization in wild type cells	GFP fusion localization in <i>spoIVA</i> mutant	GFP fusion localization in <i>cotE</i> mutant	Number of membrane domains ^a
<i>prkA</i>	Inner coat	–	+	–
<i>yhaX</i>	Inner coat and discrete dot in mother cell	–	+	–
<i>yhbB</i>	Inner coat	–	+	–
<i>yuzC</i>	Inner coat	–	+	–
<i>yybI</i>	Inner coat	–	+	1 (Cter inside)
<i>yjbX</i>	Outer coat	–	–	–
<i>ybaN</i>	Outer forespore membrane	ND	ND	1 (Cter outside)
<i>ykvI</i>	Outer forespore membrane	ND	ND	10 (Cter inside)
<i>ykvU</i>	Outer forespore membrane	ND	ND	9 (Cter outside)
<i>ypjB</i>	Outer forespore membrane	ND	ND	2 (Cter inside)
<i>ytrI</i>	Outer forespore membrane	ND	ND	8 (Cter outside)
<i>ytrH</i>	Outer forespore membrane	ND	ND	3 (Cter outside)
<i>yhaT</i>	Mother cell cytoplasm	ND	ND	–
<i>yhbH</i>	Mother cell cytoplasm	ND	ND	–
<i>ymaF</i>	Mother cell cytoplasm	ND	ND	–
<i>yocN</i>	Mother cell cytoplasm	ND	ND	–
<i>ytxC</i>	Mother cell cytoplasm	ND	ND	–
<i>ywlB</i>	Mother cell cytoplasm	ND	ND	–

^a Predictions of membrane domains from http://www.ch.embnet.org/software/TMPRED_form.html

themselves, are present and in the same order in the corresponding region of the *L. monocytogenes* genome. Thus, as noted previously,⁶² *B. subtilis* and *L. monocytogenes* are closely related except for the absence of many sporulation genes in the latter.

The same region was examined in the related endospore-forming bacterium *B. halodurans* and here again a high degree of conservation was observed. However, two major differences were noted. First, the ortholog of *ytrI* is found in a different region of the *B. halodurans* chromosome than in *B. subtilis*. Second, *B. halodurans* seems to contain an additional gene (*bh3170*) in the gap between *dnaE* and *ytrI*. Upon close inspection we discovered that *B. subtilis* does contain an ortholog of *bh3170*, which was missed in the annotation of the *B. subtilis* genome and which is located in the 350-bp gap between the σ^E promoter that controls *ytrI* transcription and the beginning of the *ytrI* ORF. We designate this newly identified gene *ytrH* and conclude that *ytrH* and *ytrI* constitute a two-gene operon, which is transcribed from a σ^E -controlled promoter located just upstream of *ytrH*. We inactivated the *ytrH* gene, resulting in a mutant that was slightly more impaired in spore formation (tenfold) than was the corresponding *ytrI* mutant (fivefold). A similar two-gene operon containing orthologs of *ytrH* and *ytrI* is also found in *B. anthracis* (Figure 4) and *O. iheyensis* (not shown).

Subcellular localization of proteins expressed from σ^E -controlled genes

To gain further insight into the function of some of the newly identified members of the σ^E regulon, we carried out subcellular localization studies by creating in-frame fusions to the coding sequence

for the green fluorescent protein (GFP). Priority was given to genes whose inactivation resulted in reduced sporulation. The results of the subcellular localization studies are presented in Table 5 and Figure 5, in which the green color corresponds to fluorescence from GFP and the red color to fluorescence from FM4-64, which was used to stain the membranes of the sporangia.

First, we present the results for GFP fusions to six proteins (Table 5), YkvU (Figure 5(A)), YpjB (Figure 5(B)), YbaN, YkvI, YtrH and YtrI (data not shown), which were inferred to be integral membrane proteins based on the prediction of the TMpred server[†]. In all six cases, the fusion proteins were found to localize to the sporulation septum shortly after the initiation of engulfment (as indicated by the slight curvature of the septum) and were found to remain associated with the membrane that surrounds the forespore (the outer forespore membrane) during later stages of sporulation. In contrast, little or no localization was detected in all cases at the cytoplasmic membrane that surrounds the mother cell. This pattern of selective localization to the outer forespore membrane is essentially equivalent to that observed previously for the integral membrane proteins BofA, SpoIVFA, and SpoIVFB, which are known to be produced under the control of σ^E .^{63,64} Thus, selective localization to the outer forespore membrane is emerging as a common feature of integral membrane proteins produced under the control of the mother-cell transcription factor.

Next, we examined gene products that were predicted to be non-membrane proteins on the

[†] http://www.ch.embnet.org/software/TMPRED_form.html

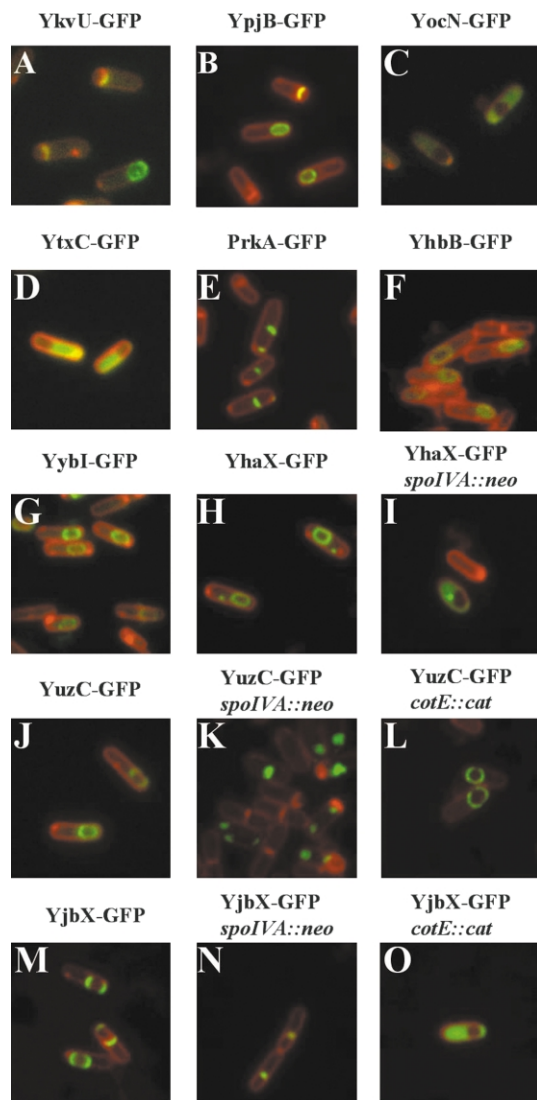


Figure 5. Subcellular localization of fusions to the GFP. Fluorescent micrographs of representative cells producing GFP fused in-frame to the C terminus of the protein products of several σ^E -controlled genes. Cells were collected after growth in DS medium for 15–18 hours at 25 or 37 °C and stained with the membrane dye FM4-64. In the indicated cases, results are presented for the localization of the fusion protein in cells mutant for coat morphogenesis genes *spoIVA* and *cotE* as well as in cells of the wild-type. The Figure shows merged images of GFP (green) and FM4-64 (red). The strains were as follows: (A) PE392 (YkvU-GFP), (B) PE339 (YpjB-GFP), (C) PE404 (YocN-GFP), (D) PE394 (YtxC-GFP), (E) PE370 (PrkA-GFP), (F) PE389 (YhbB-GFP), (G) PE405 (YybI-GFP), (H) PE388 (YhaX-GFP), (I) PE397 (YhaX-GFP in *spoIVA* mutant), (J) PE391 (YuzC-GFP), (K) PE399 (YuzC-GFP in *spoIVA* mutant), (L) PE400 (YuzC-GFP in *cotE* mutant), (M) PE371 (YjbX-GFP), (N) PE386 (YjbX-GFP in *spoIVA* mutant), (O) PE387 (YjbX-GFP in *cotE* mutant).

grounds that such proteins would be candidates for components of the spore coat. The coat is a proteinaceous shell that is produced in the mother cell and surrounds the developing spore and ultimately the mature spore.^{27,65} We created GFP

fusions to 12 such proteins with the following results: YocN-GFP (Figure 5(C)) and YtxC-GFP (Figure 5(D)), YhaT-GFP, YhbH-GFP, YmaF-GFP, YwIB-GFP (Table 5 and data not shown) were found to be uniformly distributed in the mother cell cytoplasm. The failure to observe subcellular localization could indicate that these six proteins do not localize in a specific manner. Alternatively, in some or all of the cases, the fusion protein may have undergone cleavage, liberating the GFP moiety, or may not have localized properly due to interference by the GFP moiety. In contrast, PrkA-GFP (Figure 5(E)), YhbB-GFP (Figure 5(F)), YybI-GFP (Figure 5(G)), YhaX-GFP (Figure 5(H)), YuzC-GFP (Figure 5(J)) and YjbX-GFP (Figure 5(M)) were found to exhibit specific patterns of subcellular localization, forming rings or cap-like structures around the forespore. This pattern of localization is characteristic of proteins that are associated with the spore coat. In support of the idea that these are indeed coat proteins, in each case localization around the forespore was found to be dependent upon SpoIVA, a σ^E -produced protein that is known to anchor the coat to the surface of the developing spore.^{66–68} Examples of the effect of a *spoIVA* mutation in disrupting the normal pattern of localization are shown in Figure 5(I) for YhaX-GFP, Figure 5(K) for YuzC-GFP and in Figure 5(N) for YjbX-GFP. We comment further on the case of YhaX-GFP because of a novel feature of its localization. Specifically, in addition to appearing to associate with the spore coat, the fusion protein localized as a discrete dot in the mother cell cytoplasm (Figure 5(H)). Because the dot was observed both in the presence and absence of a *spoIVA* mutation (Figure 5(I)), we infer that it is unrelated in its formation to the spore coat.

The coat consists of an outer layer whose assembly depends on the morphogenetic protein CotE (itself an outer coat protein) and an inner layer that does not.^{68,69} To distinguish inner coat proteins from outer coat proteins, we investigated the localization pattern of the GFP fusions to the six putative coat proteins (PrkA, YhaX, YhbB, YjbX, YuzC, and YybI) on CotE (Figure 5(L) and (O), Table 5 and data not shown). In only one case, that of YjbX-GFP, was localization disrupted by the presence of a *cotE* mutation (Figure 5(O)), on which basis we provisionally conclude that YjbX is an outer coat protein and that PrkA, YhaX, YhbB, YuzC, and YybI are inner coat proteins. A recent characterization of the protein composition of the spore by a proteomics-based approach by Kuwana *et al.*⁷⁰ also identified YjbX, YuzC and YybI as protein components of the mature spore.

Discussion

Here, we have taken a genome-wide approach combining transcriptional profiling and bioinformatics to characterize the regulon for the mother cell-specific transcription factor σ^E . We have

shown that as many as 253 genes (~6% of the annotated ORFs in the *B. subtilis* genome) appear to be controlled by σ^E and among these, 181 had not been previously described as members of this regulon. Recently, similar experiments have been carried out in *B. subtilis* to characterize the regulons for the general stress sigma factor, σ^B (127 genes)⁵⁷ and one of the ECF (extracytoplasmic function) sigma factors, σ^W (60 genes).⁷¹ Unlike σ^B and σ^W , which are involved in transient adaptive responses, σ^E is active for an extended period of time as an essential component of an elaborate differentiation program during which striking morphological changes are observed. Therefore, it is perhaps not surprising that a larger number of genes appear to be controlled by σ^E than by σ^B or σ^W . In addition, transcriptional profiling experiments were carried out for the stationary phase and early sporulation sigma factor, σ^H (87 genes)¹⁷ and the DNA-binding protein Spo0A, which together with σ^H governs entry into sporulation.¹⁶ About 586 genes were identified whose transcription was significantly influenced by Spo0A, but chromatin immunoprecipitation and biochemical experiments indicate that only about 50 of these genes are under the direct control of Spo0A (V. Molle & R.L., unpublished results). Spo0A and σ^H and σ^E act at successive stages of sporulation. Indeed, the gene for σ^E is under the direct control of Spo0A⁷² and the indirect control of σ^H . We have also shown that many of the genes that are expressed at high levels early in sporulation are down-regulated after σ^E activation. Therefore, the appearance of σ^E represents a major turning point in the sporulation process in that the down-regulation of many genes expressed at the onset of sporulation is replaced by the activation of a new and very large set of genes in the mother cell. Interestingly, *pepF* (*yjbG*), which has recently been reported to inhibit sporulation when overexpressed,⁷³ is controlled by σ^E and may play a role in the observed down-regulation of early sporulation genes upon σ^E activation.

Functional roles of the genes controlled by σ^E

It still comes as a surprise that the σ^E regulon is so large. Many of the previously known σ^E -controlled genes had been identified on the basis of their role in the sporulation process, and, after many years of intensive genetic research, it was assumed that the repertoire of yet undiscovered sporulation-essential genes was limited. We have inactivated most of the newly identified σ^E -controlled genes to determine if they are required for efficient sporulation. Of the 98 mutants, 12 are impaired in sporulation, but only four (*ykvV*, *yIbJ*, *yqfC* and *yqfD*) are down by more than a 1000-fold. This brings the total of σ^E -controlled genes required for sporulation under our laboratory conditions to 45 (i.e. ~20% of the genes in the regulon; these genes are emphasized in bold characters in Tables 1 and 2). It was already apparent from pre-

vious work that the inactivation of a gene expressed during sporulation does not necessarily result in a conspicuous sporulation defect. For instance, only three σ^E -controlled genes (*spoIIR*, *spoIIQ*, *spoIIIG*) are essential for sporulation from a total of about 15 known σ^E -controlled operons, and similarly, only two (*spoIVB* and *spoVA*) of the 32 known σ^G -controlled operons are essential for sporulation.²

Comparisons between the genomes of endospore-forming bacteria define a conserved core of sporulation genes of probable common ancestral origin. Orthologs of 64 of the 253 genes that appear to be controlled by σ^E in *B. subtilis* are present in the genomes of the three *Bacillus* (*B. anthracis*, *B. halodurans* and *O. iheyensis*) and the two *Clostridium* (*C. acetobutylicum* and *C. perfringens*) species that we considered (see Table S1 of supplementary material). Among these 64 genes, 18 are also found in the genomes of the related non-endospore formers *L. monocytogenes* and *L. innocua*, implying that their functions are not restricted to the process of sporulation, and incidentally that in *Listeria* these genes need to be expressed from a different promoter. Thus, we end up with 46 σ^E -controlled genes (underlined in Tables 1 and 2) that form the set of genes that are likely to be expressed in the mother cell of many or all endospore formers, assuming that a σ^E recognition sequence is present in their regulatory regions. Among these 46 conserved genes, only about 70% are essential for efficient sporulation in *B. subtilis*, whereas some of the 45 σ^E -controlled genes known to be required for sporulation in *B. subtilis* are not conserved in all endospore formers.

More general sequence comparisons allowed us to make predictions about the functions performed by many of the newly identified σ^E -controlled genes (Table 2). We were able to sort most of these genes in four functional groups that also include most of the previously identified σ^E -controlled genes. The first major function accomplished by genes under the control of σ^E is to promote the engulfment of the forespore by the mother cell. Three genes have been shown to be essential for this process (*spoIID*, *spoIIM* and *spoIIP*),^{20,74–76} and they are absolutely conserved (sometimes even duplicated) in all endospore formers. Their products are involved in degradation of the septal peptidoglycan and mutations in *spoIID*, *spoIIM* or *spoIIP* block sporulation at morphological stage II, prior to the stage of engulfment. Furthermore, it has been shown recently that these genes are also necessary for preventing septation at the other pole of the sporangium. When the three mutations are combined in the same strain, the stage II block is even more severe and two septa are observed, one at each pole of the cell.^{24,25} This is the closest one could get by inactivation of selected σ^E -controlled genes to reconstituting the characteristic phenotype of a *sigE* mutant. However, this triple mutant is still significantly different from a *sigE* mutant.²⁵ In a *sigE* null mutant, the second

chromosome is completely translocated into the second forepore-like compartment at the distal end of the sporangium, leaving the central compartment of the sporangium free of any DNA, whereas in the triple mutant the second chromosome is not translocated. None of the newly identified σ^E -controlled genes appear to play a major role in engulfment, since all the new sporulation mutants complete this step normally. However, it will be interesting to introduce these new mutations in a strain already lacking *spoIID*, *spoIIM* and *spoIIP* to check if chromosome translocation into the second forespore-like compartment can be restored.

The second major function of σ^E -controlled genes is the synthesis of the protective envelopes around the spore, the cortex and the different spore-coat layers. The cortex is a modified form of peptidoglycan synthesized between the inner and outer membranes of the forespore. Many σ^E -controlled genes required for its formation have already been identified, such as *spoVB*,⁴⁶ *spoVD*⁷⁷ and *spoVE*.⁷⁸ Usually, defects in cortex architecture reduce spore heat-resistance dramatically and prevent spores from brightening, a phenotype easily recognized by phase-contrast microscopy. We have identified four new σ^E -controlled genes that fulfill these criteria: *ylbJ*, *yqfC*, *yqfD* and *ykvV*. Orthologs of *ylbJ*, *yqfC* and *yqfD* were identified in the genomes of all endospore-formers, whereas *ykvV* is not always conserved, although more distant homologs can be found in a variety of other microbial genomes. YkvV is related to thioredoxin, a protein involved in disulfide bond formation in *E. coli* and bears a putative secretion sequence. Because proteins related to thioredoxin (but not thioredoxin itself) are found in the *E. coli* periplasm where the peptidoglycan is synthesized,⁷⁹ YkvV might be acting in the space located between the inner and outer membranes of the forespore where the cortex is synthesized. Additional genome comparisons indicate that many genes required for cortex synthesis are strictly conserved among endospore formers, suggesting that some aspects of the cortex structure are universal. However, certain characteristics of the spore cortex are expected to vary since some genes are less widely conserved. For instance, the *gerM*,^{80,81} *cwlJ*^{82,83} and *ywdL*⁸⁴ genes, which are required for efficient germination presumably through their involvement in cortex hydrolysis, are conserved among *Bacilli* but are absent from *Clostridia*. Another example is YbaN, a putative polysaccharide deacetylase, which we found to be required for efficient sporulation in *B. subtilis*, but has no ortholog in *Clostridia*.

Even more variability is observed in the conservation of the components of the second protective envelope of the spore, the spore coat. Striking structural differences are indeed observed by electron microscopy between the spore coats of *B. subtilis* and *B. anthracis*,⁶⁵ even though the set of sporulation genes between these two organisms is usually very similar. It is possible that the coat

structure has evolved influenced by the type of environment from which the spores have to be protected. We identified eight new σ^E -controlled genes encoding components of the spore coat and we suspect that more may be present. Since defects in spore-coat assembly usually do not affect spore heat-resistance, it is difficult to screen a large quantity of mutants for coat assembly defects. We used cytology techniques to demonstrate that seven proteins are localized in the inner layer of the spore coat: PrkA, YhaX, YhbB, YuzC, YybI, as well as YheC and YheD (C. van O. & R.L., unpublished results), whereas one protein, YjbX, is localized in the outer layer. Thus, most of the coat proteins synthesized under σ^E control are targeted to the inner coat (CotE and YjbX being the only exceptions), whereas most of the known proteins of the outer coat are dependent on σ^K .²⁷ This temporal regulation probably helps the ordered assembly of the different layers of the spore coat. Among the newly identified coat-associated proteins, PrkA and YhbB seem to be the only ones conserved in all endospore-formers, whereas in contrast YuzC is unique to *B. subtilis*.

The third major function, carried out by a large group of genes under the control of σ^E , is to maintain a sufficient level of metabolic activity to enable the progression of the sporulation process under conditions of limiting nutrient availability. The different pathways used to achieve this goal are presumably largely influenced by the environment in which the sporulation process takes place and have probably evolved accordingly. Considering that some spore formers are strict anaerobes, whereas others sporulate in aerobic environments, a high degree of variability is expected in the conservation of the σ^E -controlled genes involved in these metabolic functions. In addition, very few genes belonging to this group appear to be essential for sporulation, which is expected since it seems unlikely that the sporulating cell would rely on a unique pathway to provide the energy required for progression of the sporulation process. Many genes in this group (*yngJIHGFE*, *yodTSRPQ*, *ytpAB*, *ywjF-acdA*, *yxjC-scoAB-yxjF*, *ylbK*, *yqhO*, *ywjE* and *ywnE*) appear to be involved in lipid metabolism. This suggests a possible way of generating energy in the absence of nutrients by oxidation of fatty acids from the cytoplasmic membrane. An alternative explanation would be that these genes encode proteins required for the catabolism of polyhydroxyalkanoates.⁸⁵ Protein degradation is another way of generating nutrients and some genes (*mlpA*, *pepF*, *ycdA*, *yhfN*, *yqgT* and *yuiE*) encode putative peptidases and proteases. Finally, the presence of more than 20 σ^E -controlled genes with similarity to transporters suggests that nutrients can be scavenged from the external environment, in particular, amino acids or nucleotides that have been generated by the action of nucleases and proteases secreted in an earlier phase of sporulation under the control of σ^H .¹⁷ Also, genes belonging to this group may have

functions that are not restricted to the process of sporulation and may be transcribed under non-sporulation conditions from additional, non- σ^E -controlled promoters.

The fourth major function of σ^E is to set the stage for the next and final steps of the sporulation process. Three major transcriptional regulators are controlled by σ^E : SpoIID, σ^G , and σ^K .^{35,38,39,40,41,43,44,86–88} All three, as well as the pathways leading to their activation are strictly conserved among all endospore formers, with the sole exception of the σ^K -activation pathway in *Clostridium difficile*.³⁴ None of the new σ^E -controlled genes appears to be essential for σ^G activation, whereas only one of them, *yunB*, seems to code for a protein involved, at least indirectly, in the pathway leading to the activation of σ^K . Inactivation of *yunB* delays σ^K activation (J.S., P.E. & R.L., unpublished results) and results in reduced sporulation efficiency.

Finally, about 40 genes could not be related to any of these four major functions and further investigation will be required to determine their role in the sporulation process. Inactivation of three genes from this group, *ypjB* and the two genes of the *ytrHI* operon, resulted in decreased sporulation efficiency. However, this decrease was modest and could not be associated to a specific stage of sporulation.

In summary, one of the intriguing conclusions of this study is that inactivation of the new σ^E -controlled genes rarely impairs sporulation. It might be that some of the functions performed by proteins encoded by the σ^E -controlled genes are redundant. In that case, combinations of mutations would have to be generated to observe substantial sporulation deficiencies. We also believe that backup mechanisms can be switched on to compensate for the inactivation of a specific pathway, especially in the case of genes involved in metabolic functions. Finally, given that sporulation efficiency has only been tested by measurements of the production of heat-resistant spores in very specific laboratory conditions, it is possible that some sporulation defects have been overlooked. Some of the genes of unknown function might then turn out to be coding for components of the spore coat or to be required for spore germination.

Distribution of σ^E -controlled genes on the *B. subtilis* chromosome

The genome-wide identification of genes belonging to a specific regulon provides a way to study the distribution of these genes on the bacterial chromosome. For this purpose, a genetic map of the σ^E regulon is presented on Figure 6. An issue to consider is whether the observed distribution is random or influenced by specific constraints. It has been reported that a change in the localization of certain genes on the *B. subtilis* chromosome has considerable functional consequences during sporulation.^{89–92} In all known examples, this effect

is explained by the concept of transient genetic asymmetry.⁸⁹ In the course of asymmetric division, the polar septum is formed over what will become the forespore chromosome. As a consequence, about 70% of this chromosome is still present in the mother cell after polar septation has been completed, and needs to be pumped into the forespore by the action of the SpoIIIE DNA translocase.⁹³ Therefore, during a short period of time, which is simultaneous to the time when σ^F and σ^E are activated, some genes are excluded from the forespore and present in two copies in the mother cell. Furthermore, since the chromosome is oriented with the region corresponding to the origin of replication anchored at the extreme pole of the cell, it is always the same region of the chromosome, i.e. the origin-proximal region, that is located in the forespore following polar septation.^{93,94} For instance, it has been shown that the location of *spoIIR*, a σ^F -controlled gene encoding a signaling protein essential for σ^E activation, is critical for its function.^{91,92} This gene is found close to the origin region and thus is present in the forespore upon polar septation, ensuring that SpoIIR is synthesized immediately after σ^F activation. Accordingly, σ^E activation quickly follows σ^F activation. In contrast, when the gene is moved to the origin-distal region of the chromosome and is therefore excluded from the forespore by transient genetic asymmetry, σ^E activation is delayed, presumably until the region containing the *spoIIR* gene is translocated into the forespore. In the case of σ^E -controlled genes, gene exclusion by transient genetic asymmetry cannot occur since a complete copy of the chromosome (the mother cell chromosome) is always present in the mother cell. However, transient genetic asymmetry could, at least in principle, result in higher levels of expression for σ^E -controlled genes located in the origin-distal region and therefore present in the mother cell in two copies, as opposed to the origin-proximal σ^E -controlled genes present in only one copy in the mother cell. A total of 75% of the σ^E -controlled genes are found in the region of the forespore chromosome that is located in the mother cell after polar septation (Figure 6). This is not significantly different from what is expected for a random distribution, which is in the 70% range. Therefore, the global distribution of σ^E -controlled genes on the *B. subtilis* chromosome does not seem to contribute drastically to transient genetic asymmetry, although we cannot rule out that the chromosomal localization of some σ^E -controlled genes is important.

Another conclusion that can be drawn from Figure 6 is that the density of σ^E -controlled genes is not uniform along the *B. subtilis* chromosome. Strikingly, one region centered around position 300° on the genetic map (from *pucH* to *cotB*) is completely free of σ^E -controlled genes. This represents a considerable gap of 387 kb, almost 10% of the *B. subtilis* chromosome. The similar region on the other side of the chromosome (centered around

actually represent clusters of sporulation genes rather than clusters of σ^E -controlled genes. Furthermore, these clusters are for the most part well conserved in the genomes of *B. anthracis*, *B. halodurans* and *O. iheyensis* (data not shown), although clusters IV, V and VI are inverted in *B. anthracis* and *B. halodurans*. Conversely, these clusters are not found in the more distantly related genomes of *C. acetobutylicum* and *C. perfringens*, even though gene order on a more local scale is often conserved between *Bacilli* and *Clostridia* where other clusters of sporulation genes are sometimes observed (data not shown). In any case, it is not yet clear if the observed clustering of sporulation genes in *Bacilli* reflects a functional requirement or if it is just the remnant of an ancient more compact organization.

Conclusion

We have shown here that identification on a genome-wide basis of a complete regulon, in this case the regulon of the mother-cell early sigma factor σ^E , can generate a considerable amount of novel information, even for processes that have already been the focus of many detailed studies, such as sporulation in *B. subtilis*. We have also explored ways to take advantage of the increasing availability of complete genome sequences of related bacteria, and we were able to extend some of our observations to a whole class of bacteria, in this case the Gram-positive endospore formers. Additionally, comparative genomics have emphasized some differences that will warrant further investigation, for instance the indication that the protein composition of the spore coat varies considerably among endospore formers. Finally, our characterization of the σ^E regulon constitutes the first step towards a comprehensive identification of the genes expressed in each cell compartment during sporulation of *B. subtilis* and related bacteria.

Materials and Methods

Strains, plasmids and growth conditions

All strains used here were derivatives of the wild-type strain PY79.⁹⁵ The strains used for transcriptional profiling under conditions of sporulation were PE436 (*sigE*⁺) and PE437 (*sigE* mutant). PE436 was generated by transformation of PE239 (*spoIIID* :: *tet*)²⁵ with chromosomal DNA from strain RL560 (*sigG* :: *cat*)⁴³ and selection for chloramphenicol resistance. The mutations in *sigG* and *spoIIID* were included to eliminate σ^G and σ^K -dependent expression from the experiments. PE437 was built from PE239 and RL1061 (*sigE* :: *erm*).⁹⁶ PE436 and PE437 were grown in parallel in hydrolyzed casein medium at 37 °C to an $A_{600\text{ nm}}$ of 0.6. Sporulation was induced by resuspension in Sterlini–Mandelstam medium at 37 °C.^{97,98} At two hours and 30 minutes after resuspension, 25 ml of each culture were collected for RNA isolation (see below).

Overexpression of σ^E during growth was the result of the treatment of strain MF453 with IPTG. MF453

(*spoIIA-sigE* :: *tet*, *amyE* :: P_{spac}^C -*spoIIA-sigE* *spc*, *thrC* :: $P_{\text{spac}}^{\text{IN}}$ -*spoIIR* *erm*, *spoIID* :: *spoIID-lacZ* *cat*) was obtained by transformation of MF444⁴⁵ with chromosomal DNA from MF372 (*spoIID* Ω *spoIID-lacZ* *cat*), which is a derivative of MO692 (*spoIID* Ω *spoIID-lacZ* *cat*, *trpC2*, *pheA1*). For transcriptional profiling, a culture of strain MF453 was grown in LB medium at 37 °C to an $A_{600\text{ nm}}$ of 0.5, at which time the culture was split in two and IPTG (1 mM final concentration) was added to one culture. At one hour after induction, a 25 ml sample was taken from the induced culture and compared to the same time point of the parallel culture without IPTG.

All deletion strains were generated using the technique of long-flanking homology PCR.⁵⁹ Chromosomal DNA obtained from these strains was analyzed by PCR to confirm the integration of the resistance cassette at the expected locus. A complete list of the strains is presented in Table S1 of the supplementary material. The sequence of the primers used for the constructions is available upon request.

The strains expressing GFP fused in-frame to the coding sequence of σ^E -controlled genes were obtained by Campbell-like recombination of plasmids generated by cloning of *Xho*I/*Bam*HI-digested PCR fragments of the gene of interest (the sequence of the primers used for the amplification is available upon request) into pCVO119. Plasmid pCVO119 was synthesized by inserting the multiple cloning site from pBluescript SKII (Stratagene) into pKL147.⁹⁹ Plasmid pBluescript SKII was digested with *Sac*I and the overhang removed with T4 DNA polymerase (New England Biolabs) following the manufacturer's instructions. It was then digested with *Xho*I and the released 87 bp fragment was isolated. pKL147 was digested with *Eco*RI, and the overhang was filled in with Klenow fragment DNA polymerase I (New England Biolabs) following the manufacturer's instructions, and subsequently digested with *Xho*I. The 87 bp fragment from pBluescript SKII was then ligated to the digested pKL147 vector to form pCVO119. The following plasmids were obtained [(pCVO287 (*yhbH*-GFP), pCVO288 (*ymaF*-GFP), pPE34 (*ypjB*-GFP), pPE35 (*ytvI*-GFP), pPE37 (*yjbX*-GFP), pPE39 (*prkA*-GFP), pPE40 (*ykvI*-GFP), pPE41 (*yhbB*-GFP), pPE42 (*yhaX*-GFP), pPE43 (*yuzC*-GFP), pPE44 (*ykvU*-GFP), pPE46 (*ytxC*-GFP), pPE51 (*yocN*-GFP), pPE53 (*yybI*-GFP), pPE55 (*ywlB*-GFP), pPE56 (*ytrH*-GFP) and pPE57 (*yhaT*-GFP)] and used to transform PY79 to spectinomycin resistance and generate strains CVO1722, CVO1723, PE339, PE340, PE371, PE370, PE390, PE389, PE388, PE391, PE392, PE393, PE394, PE407, PE404, PE405, PE450, PE447 and PE449, respectively. To analyze localization of putative coat proteins in null mutants of *cotE* or *spoIVA*, the GFP fusions were introduced into strains RL364 (*cotE* :: *cat*)⁶⁸ and RL1397 (*spoIVA* :: *neo*).⁶⁸ The following strains were obtained: PE383 (*prkA*-GFP *spc*, *spoIVA* :: *neo*), PE384 (*prkA*-GFP *spc*, *cotE* :: *cat*), PE386 (*yjbX*-GFP *spc*, *spoIVA* :: *neo*), PE396 (*yjbX*-GFP *spc*, *cotE* :: *cat*), PE395 (*yhbB*-GFP *spc*, *spoIVA* :: *neo*), PE387 (*yhbB*-GFP *spc*, *cotE* :: *cat*), PE397 (*yhaX*-GFP *spc*, *spoIVA* :: *neo*), PE398 (*yhaX*-GFP *spc*, *cotE* :: *cat*), PE399 (*yuzC*-GFP *spc*, *spoIVA* :: *neo*), PE400 (*yuzC*-GFP *spc*, *cotE* :: *cat*), PE415 (*yybI*-GFP *spc*, *spoIVA* :: *neo*), PE416 (*yybI*-GFP *spc*, *cotE* :: *cat*).

Transcriptional profiling

DNA microarrays

The construction of our microarrays has been described in detail by Britton *et al.*¹⁷

Sample preparation

Samples collected for RNA isolation were immediately mixed with an equal volume of methanol at -20°C . After centrifugation for five minutes at 4000g, the cell pellets were stored at -80°C . RNA was extracted by a hot acid/phenol isolation protocol.¹⁶

Labeling and hybridization

This procedure has been described in detail by Britton *et al.*¹⁷ Briefly, 50 μg of RNA were reverse transcribed using the superscript first strand synthesis system for RT-PCR (Invitrogen) in the presence of [Cy3]dUTP or [Cy5]dUTP (Perkin Elmer Life Sciences) to generate labeled cDNA. Samples from two different conditions were labeled in parallel and mixed prior to removal of unincorporated nucleotides using QiaQuick purification spin columns (Qiagen). Hybridizations were performed overnight in CMT-hybridization chambers (Corning) plunged in a water bath at 60°C . Finally, slides were scanned on a GenePix 4000B scanner (Axon Instruments). Each profiling experiment was performed three times independently (see below).

Statistical data analysis

Statistical analysis was carried out using R statistical software.¹⁰⁰ The transcriptional profiling strategy, which compared *sigE*⁺ cells to *sigE* mutant cells undergoing sporulation, had five replicate slides and two dye swaps and was based on RNAs isolated from six independently grown cultures. The strategy, which compared vegetative cells grown with and without IPTG, had four replicate slides from three independent experiments, with one dye swap. We performed analyses on these two complementary sets of experiments separately.

Intensity values for the Cy3 and Cy5 channels were calculated as mean foreground minus mean background intensities (i.e. the local backgrounds) and were directly obtained from the Genepix Pro 4.0 software (Axon Instruments). Spots with negative intensities for both Cy3 and Cy5 were removed from analysis. Spots with negative intensity in one channel but positive in the other were given a minimum positive intensity value for the negative channel. Each slide was normalized using an iterative "rank invariant" method introduced by Tseng *et al.*¹⁰¹ Here, M is the log-ratio of expression for each gene, and A is the averaged log intensity.¹⁰² A locally weighted scatterplot smoother "Lowess", a built-in function of R, was used to fit a normalization curve \hat{M} through the rank invariant genes, where $\hat{M} = \hat{f}(A)$. The normalized log-ratios are defined as $\hat{M} = M - \hat{M}$. The ranks of the Cy3 and Cy5 channels were calculated for each gene. Genes with ranks within a percentage p (p chosen to be 2%) of the total number of genes and with average Cy3 and Cy5 ranks not among the 25 highest or lowest ranks were included in the initial set of rank invariant genes. This set was reduced iteratively, by retaining genes with Cy3 and Cy5 ranks within $p\%$ of the current set of rank invariant genes. The iteration was stopped when the set of rank invariant genes did not change after selection. The Lowess procedure was used to normalize the entire data set for each slide. It was not necessary to normalize across slides to account for differential spread of the distributions, since the replicate slides had similar distributions.

The normalized replicate slides were combined using the Bayesian hierarchical model introduced by Tseng *et al.*¹⁰¹ which incorporates variation across slides and experiments. We used a Markov chain Monte Carlo implementation of the model and sampled 2000 iterations of the posterior distributions, which was sufficient for convergence. We used posterior medians of the log-ratios of expression for each gene as estimates of the true log-ratio, and the corresponding 95% posterior interval (Bayesian confidence interval).

The score of a gene is defined as the posterior probability of its log-expression ratio greater than zero. A complete list of scores for each gene on the microarrays can be found in Table S2 of the supplementary material. Scores above or equal to 0.95 were considered significant for up-regulated genes in the experiment comparing *sigE*⁺ to *sigE* mutant cells. In the same experiment, scores below or equal to 0.05 were considered significant for down-regulated genes. The cut-off values were selected by comparison to the scores associated to known σ^E -regulated genes. In the σ^E overexpression experiment, a cut-off value of 0.9 was considered significant for up-regulated genes. A slightly less restrictive cut-off value was selected in this case because the observed ratios in the σ^E overexpression experiment were usually much smaller than the ratios in the experiment comparing *sigE*⁺ to *sigE* mutant cells.

The data are also available in minimum information about a microarray experiment (MIAME)¹⁰³ compliant format†.

Promoter search

Upstream scoring procedure

An estimate of the consensus matrix for the σ^E motif was created by counting the nucleotides in the 38 confirmed sites from the 36 known σ^E -controlled genes. For each gene in the set of 157 possible σ^E -controlled operons, we detected the upstream position that best corresponded to the σ^E -binding motif by "scoring" each position j in the 200 bp upstream region of gene i with the following score function:

$$\text{Score}_{ij} = \log \left(\frac{P(\text{motif starting in pos}_{ij})}{P(\text{background continuing in pos}_{ij})} \right)$$

Let each parameter in the motif consensus matrix be denoted as θ_{jk} where j indexes the 15 positions in the motif (7 bp - 35 site + 8 bp - 10 site) and k indexes the nucleotide ($A = 1, \dots, T = 4$) and let the four background frequencies be denoted as θ_{0k} , which were estimated by simply counting the nucleotide frequencies in all the upstream regions. Assuming an independence model between neighboring nucleotides, we can easily calculate the score for a pair of segments, (r_1, r_2, \dots, r_7) and $(r_8, r_9, \dots, r_{15})$ with a gap of length 13-15 as:

$$\begin{aligned} \text{Score}_{ij} &= \log \left[\left(\frac{P(\text{nuc}_j \text{ from motif})}{P(\text{nuc}_j \text{ from bg})} \right) \right. \\ &\quad \left. \times \left(\frac{P(\text{nuc}_{j+1} \text{ from motif})}{P(\text{nuc}_{j+1} \text{ from bg})} \right) \dots \right] \\ &= \log \left[\left(\frac{\theta_{1,r1}}{\theta_{0,r1}} \right) \left(\frac{\theta_{2,r2}}{\theta_{0,r2}} \right) \dots \left(\frac{\theta_{15,r15}}{\theta_{0,r15}} \right) \right] \end{aligned}$$

† <http://mcb.harvard.edu/losick>

The position in each upstream sequence with the maximum score over all possible positions and all allowable gap lengths (13–15) was considered the best candidate to be a σ^E -binding site.

Scores were calculated for 10,000 random sequences generated from using the background frequencies. The distribution of these 10,000 scores approximates the score distribution under the null hypothesis that a given sequence contains no σ^E site. We used this distribution to calculate a p -value for our observed scores. We used a p -value cutoff of 0.05, which corresponded to a threshold score that was the 9500th highest out of the 10,000 random scores. Any observed scores above this threshold were considered statistically significant.

BioProspector and optimization algorithm

BioProspector⁵⁶ is an algorithm designed to find short strings of nucleotides (motifs) that are highly conserved across multiple sequences of DNA[†]. The algorithm is based on the probability model for conserved motifs introduced by Lawrence *et al.*⁵⁵ with an extension to allow for dimeric motifs with variable gaps. This model is implemented by a Gibbs sampling strategy, which is a simulation-based approximation technique often used to fit models that are too complicated to have simple analytic estimation procedures (see Liu *et al.*¹⁰⁴ for more statistical detail).

The program is capable of finding multiple sites per upstream region, as well as allowing for the possibility that certain upstream regions may have no sites at all. BioProspector can be specified to handle dimers with variable gap length, such as is the case with the σ^E promoter.

A natural consequence of this simulation-based algorithm is that the results are approximate and vary somewhat between multiple uses of BioProspector on the same dataset. With this in mind, we performed 100 separate runs of BioProspector for the same data set under identical settings. Each run of BioProspector returned the top five most conserved motifs found, for a total of 500 discovered motifs. For each discovered motif, the output that we get from BioProspector consists of site predictions in each upstream region, an estimate of the consensus frequency matrix and background frequencies, and a “motif score” that measures the conservation of the motif’s frequency matrix (for details refer to Liu *et al.*⁵⁶). We used this motif score to rank all the discovered motifs and kept the five highest scoring motifs:

Consensus		Degree		Score
Block 1	Block 2	Block 1	Block 2	
TCATATT	CATATAAT	DYMTRWW	CATAHAWT	2.528
TTATATT	CATATAAT	DYMTRWW	CATAHRWT	2.516
CATATTT	CATACATT	YATRWWY	CATAYAWT	2.506
GTATATT	CATATAAT	KYATRWH	MATAHAAT	2.506
TTTATAT	CATATAAT	KKYMTAW	MATAYAWT	2.494

In the above list, we show the top five motifs summarized in terms of the most-frequent-nucleotide consensus sequence as well as the IUPAC-coded consensus sequence and the motif score. We see that these discovered motifs are similar in terms of the consensus sequence, though there are noticeable differences.

Ideally, we would like to find an estimate of the consensus that dominates each of these top scoring motifs in terms of some measure of high conservation, i.e. further reducing the noise within the motif signal. To achieve this end, we refined our top scoring motif estimates *via* a simple optimization algorithm to maximize function:

$$S = A[I(\theta||\theta_0) - \log(1/p_0)]$$

where A is the number of motifs, $\log(1/p_0)$ is a penalty term, and $I(\theta||\theta_0)$ is a measure of the entropy between the current motif parameter estimates and the background parameter estimates, i.e.:

$$I(\theta||\theta_0) = \sum_{j=1}^{16} \sum_{k=1}^4 \theta_{jk} \log\left(\frac{\theta_{jk}}{\theta_{0k}}\right)$$

This scoring function is an approximation to the log posterior distribution of the motif model by Liu *et al.*¹⁰⁴ so this optimization algorithm attempts to find the posterior mode of that model. High motif conservation will lead to motif parameter estimates that are far away from the background parameter estimates, which will result in a large entropy measure $I(\theta||\theta_0)$. Thus, by maximizing S , we favor motif estimates that have high conservation. We also prefer to find as many sites as possible, so we multiply the entropy measure by the number of motifs. The parameter $p_0 < 1$ can be interpreted as a prior representation of motif abundance. For example, if $p_0 = 1/500$, then we expect *a priori* to get one site per every 500 bp. A small prior probability of sites means that the algorithm will be more restrictive in terms of the addition of new sites.

Our algorithm goes through each position of each upstream region and adds a new motif starting at that position if the score S is increased. If a motif is already present at that position, then this motif is removed if the score would be increased by its absence. The algorithm iterates through all the upstream regions multiple times until there is no longer a change in the score S . Only gaps of the desired range (13–15 bp) were accepted in the optimization algorithm. We also force the algorithm to pick only non-overlapping motif sites. We used $p_0 = 1/300$ for the dataset.

The practical difference is minimal between the near-optimal estimates achieved by our optimization algorithm starting from each of the five top scoring BioProspector motifs. The only difference worth noting is that the –35 site is shifted over by one nucleotide between some of the different optimal results, indicating that the width of the –35 element should perhaps be 8 (see the list with consensus sequences above).

5'-RACE-PCR

The 5'-end of several σ^E -controlled mRNAs was determined by the rapid amplification of cDNAs ends (RACE)-PCR procedure.^{57,58} Total RNA was extracted as described for the transcriptional profiling experiments (see above) from strains PE436 (*sigE*⁺) and PE437 (*sigE* mutant). An aliquot of 20–50 μ g of total RNA were reverse transcribed in the presence of 70 pmol of a gene-specific primer (located at about 400 bp downstream from the predicted promoter) using the superscript first strand synthesis system for RT-PCR (Invitrogen). In each series of experiments, up to 16 gene-specific primers (sequence available upon request) were used in the same reaction to generate cDNAs from mRNAs

[†] <http://bioprospector.stanford.edu>

corresponding to up to 16 putative σ^E -controlled genes simultaneously. The sample was subsequently purified using a QiaQuick purification spin column (Qiagen). A homopolymeric T-tail was added to the 3'-end of the cDNAs with terminal transferase (Roche) following the instructions of the manufacturer and the sample was again purified using a QiaQuick column. The tailed cDNAs served as templates for PCR amplification in the presence of a poly(dA) primer and a second gene-specific primer (located at about 150 bp downstream from the predicted promoter). Only one gene-specific primer was used per PCR amplification and up to 16 PCR reactions were run in parallel. The PCR products were sequenced using either the second gene-specific primer or a third gene-specific primer located closer to the predicted promoter. In a limited number of cases, the PCR products had to be purified on a 1% (w/v) agarose gel and recovered with the QiaQuick Gel Extraction Kit (Qiagen) prior to sequencing. The transcription start site was defined as the nucleotide immediately preceding the stretch of A residues complementary to the T-tail in the sequence of the PCR product. When a stretch of A residues was found in the promoter sequence overlapping with the stretch of A residues in the PCR product, the start site could not be determined with a 1 bp precision. However, the level of accuracy was more than sufficient to validate the promoters predicted by bioinformatics.

Comparative genomics

B. subtilis genome sequence was obtained from the SubtiList website[†]; *L. monocytogenes* and *L. innocua* genome sequences were obtained from ListiList[‡]; (*B. anthracis* (strain A2012), *B. halodurans*, *C. acetobutylicum*, *C. perfringens*, *O. iheyensis* and *T. tencongensis* genome sequences were obtained from the NCBI website[§]. Orthologs of *B. subtilis* genes in related genomes were identified in two steps. First, the corresponding protein sequence of the gene of interest was blasted against the selected bacterial genome using either the blastp or the tblastn program in the microbial genome section of the BLAST webpage[¶]. Secondly, the homolog with the best score was selected and blasted against the *B. subtilis* genome[¶]. Only when the top-scoring gene in the second BLAST was identical to the gene of interest selected at the beginning of the procedure, was the homolog in the other bacterial genome considered to be a true ortholog.

Microscopy

Fluorescence microscopy was performed exactly as described by Eichenberger *et al.*²⁵ Prior to microscopy, 1 ml of Difco sporulation medium (DSM) was inoculated from a single colony of the strain of interest and grown for 16–18 hours at 25 °C or 37 °C. Then, the culture was concentrated in 0.1 ml of PBS supplemented with FM4-64 (Molecular Probes) at 1.5 $\mu\text{g ml}^{-1}$ final concentration. Next, 2 μl of the concentrated culture were placed on

the microscope slide and covered by a poly-L-lysine-treated coverslip.

Acknowledgements

We thank Katerina Ragkousi and Peter Setlow, and Sarah Trewhitt, Haridasan Chirakkal and Anne Moir for communicating results prior to publication. We are grateful to Virginie Molle for sharing results prior to publication and assistance in the production of the microarrays. We thank Paul Fawcett for helpful comments on the manuscript, Tyler Aldredge, Keith Morneau and Rachel Erlich from the Bauer Center for Genomics Research at Harvard University for advice, Robert Britton at the Massachusetts Institute of Technology for help at an early stage of the project and members of the laboratory for helpful discussions. P.E. was a post-doctoral fellow of the Swiss National Science Foundation and of the Human Frontier Science Program. S.T.J., E.M.C. and J.S.L. were supported by grant DMS 0094613 from the National Science Foundation and grant R01 HG02518-01 from the National Institutes of Health. C.v.O. was supported by NIH NRSA fellowship GM20165. J.E.G.-P. acknowledges the support of the Ministerio de Educacion y Ciencia Post-doctoral Fellowship (Spain). M.F. was supported by a grant-in-aid for research abroad from the Ministry of Education, Science and Culture (Japan). S.B.-Y. was a post-doctoral fellow of the Human Frontier Science Program. This work was supported by National Institutes of Health grant GM18568 to R.L.

References

1. Stragier, P. & Losick, R. (1996). Molecular genetics of sporulation in *Bacillus subtilis*. *Annu. Rev. Genet.* **30**, 297–341.
2. Piggot, P. J. & Losick, R. (2002). Sporulation genes and intercompartmental regulation. In *Bacillus subtilis and its Closest Relatives: From Genes to Cells* (Sonenshein, A. L., Hoch, J. A. & Losick, R., eds), pp. 483–518, American Society for Microbiology, Washington, DC.
3. Margolis, P., Driks, A. & Losick, R. (1991). Establishment of cell type by compartmentalized activation of a transcription factor. *Science*, **254**, 562–565.
4. Driks, A. & Losick, R. (1991). Compartmentalized expression of a gene under the control of sporulation transcription factor sigma E in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA*, **88**, 9934–9938.
5. Losick, R. & Stragier, P. (1992). Crisscross regulation of cell-type-specific gene expression during development in *B. subtilis*. *Nature*, **355**, 601–604.
6. Li, Z. & Piggot, P. J. (2001). Development of a two-part transcription probe to determine the completeness of temporal and spatial compartmentalization of gene expression during bacterial development. *Proc. Natl Acad. Sci. USA*, **98**, 12538–12543.
7. LaBell, T. L., Trempey, J. E. & Haldenwang, W. G. (1987). Sporulation-specific sigma factor sigma 29

[†] <http://genolist.pasteur.fr/SubtiList/>

[‡] <http://genolist.pasteur.fr/ListiList/>

[§] http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/eub_g.html

[¶] http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/genom_table.cgi

- of *Bacillus subtilis* is synthesized from a precursor protein, p31. *Proc. Natl Acad. Sci. USA*, **84**, 1784–1788.
8. Satola, S. W., Baldus, J. M. & Moran, C. P., Jr (1992). Binding of Spo0A stimulates *spoIIIG* promoter activity in *Bacillus subtilis*. *J. Bacteriol.* **174**, 1448–1453.
 9. Stragier, P., Bonamy, C. & Karmazyn-Campelli, C. (1988). Processing of a sporulation sigma factor in *Bacillus subtilis*: how morphological structure could control gene expression. *Cell*, **52**, 697–704.
 10. Jonas, R. M., Weaver, E. A., Kenney, T. J., Moran, C. P., Jr & Haldenwang, W. G. (1988). The *Bacillus subtilis* *spoIIIG* operon encodes both sigma E and a gene necessary for sigma E activation. *J. Bacteriol.* **170**, 507–511.
 11. Peters, H. K., III & Haldenwang, W. G. (1994). Isolation of a *Bacillus subtilis* *spoIIIGA* allele that suppresses processing-negative mutations in the pro-sigma E gene (*sigE*). *J. Bacteriol.* **176**, 7763–7766.
 12. Londono-Vallejo, J. A. & Stragier, P. (1995). Cell–cell signaling pathway activating a developmental transcription factor in *Bacillus subtilis*. *Genes Dev.* **9**, 503–508.
 13. Hofmeister, A. E., Londono-Vallejo, A., Harry, E., Stragier, P. & Losick, R. (1995). Extracellular signal protein triggering the proteolytic activation of a developmental transcription factor in *B. subtilis*. *Cell*, **83**, 219–226.
 14. Karow, M. L., Glaser, P. & Piggot, P. J. (1995). Identification of a gene, *spoIIIR*, that links the activation of sigma E to the transcriptional activity of sigma F during sporulation in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA*, **92**, 2012–2016.
 15. Rhodius, V., Van Dyk, T. K., Gross, C. & LaRossa, R. A. (2002). Impact of genomic technologies on studies of bacterial gene expression. *Annu. Rev. Microbiol.* **56**, 599–624.
 16. Fawcett, P., Eichenberger, P., Losick, R. & Youngman, P. (2000). The transcriptional profile of early to middle sporulation in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA*, **97**, 8063–8068.
 17. Britton, R. A., Eichenberger, P., Gonzalez-Pastor, J. E., Fawcett, P., Monson, R., Losick, R. & Grossman, A. D. (2002). Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis*. *J. Bacteriol.* **184**, 4881–4890.
 18. Helmann, J. D. & Moran, C. P. Jr (2002). RNA polymerase and sigma factors. In *Bacillus subtilis and its Closest Relatives: From Genes to Cells* (Sonenshein, A. L., Hoch, J. A. & Losick, R., eds), pp. 289–312, American Society for Microbiology, Washington, DC.
 19. Ryter, A., Schaeffer, P. & Ionesco, H. (1966). Classification cytologique, par leur stade de blocage, des mutants de sporulation de *Bacillus subtilis*, Marburg [cytologic classification, by their blockage stage, of sporulation mutants of *Bacillus subtilis* Marburg]. *Ann. Inst. Pasteur (Paris)*, **110**, 305–315.
 20. Piggot, P. J. & Coote, J. G. (1976). Genetic aspects of bacterial endospore formation. *Bacteriol. Rev.* **40**, 908–962.
 21. Illing, N. & Errington, J. (1991). Genetic regulation of morphogenesis in *Bacillus subtilis*: roles of sigma E and sigma F in prespore engulfment. *J. Bacteriol.* **173**, 3159–3169.
 22. Setlow, B., Magill, N., Febroriello, P., Nakhimovsky, L., Koppel, D. E. & Setlow, P. (1991). Condensation of the forespore nucleoid early in sporulation of *Bacillus* species. *J. Bacteriol.* **173**, 6270–6278.
 23. Lewis, P. J., Partridge, S. R. & Errington, J. (1994). Sigma factors, asymmetry, and the determination of cell fate in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA*, **91**, 3849–3853.
 24. Pogliano, J., Osborne, N., Sharp, M. D., Abanes-De Mello, A., Perez, A., Sun, Y. L. & Pogliano, K. (1999). A vital stain for studying membrane dynamics in bacteria: a novel mechanism controlling septation during *Bacillus subtilis* sporulation. *Mol. Microbiol.* **31**, 1149–1159.
 25. Eichenberger, P., Fawcett, P. & Losick, R. (2001). A three-protein inhibitor of polar septation during sporulation in *Bacillus subtilis*. *Mol. Microbiol.* **42**, 1147–1162.
 26. Foster, S. J. & Popham, D. L. (2002). Structure and synthesis of cell wall, spore cortex, teichoic acids, S-layers, and capsules. In *Bacillus subtilis and its Closest Relatives: From Genes to Cells* (Sonenshein, A. L., Hoch, J. A. & Losick, R., eds), pp. 21–42, American Society for Microbiology, Washington, DC.
 27. Driks, A. (2002). Proteins of the spore coat and coat. In *Bacillus subtilis and its Closest Relatives: From Genes to Cells* (Sonenshein, A. L., Hoch, J. A. & Losick, R., eds), pp. 527–536, American Society for Microbiology, Washington, DC.
 28. Read, T. D., Salzberg, S. L., Pop, M., Shumway, M., Umayam, L., Jiang, L. *et al.* (2002). Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science*, **296**, 2028–2033.
 29. Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T. *et al.* (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc. Natl Acad. Sci. USA*, **99**, 996–1001.
 30. Nolling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q., Gibson, R. *et al.* (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**, 4823–4838.
 31. Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N. *et al.* (2000). Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucl. Acids Res.* **28**, 4317–4331.
 32. Takami, H., Takaki, Y. & Uchiyama, I. (2002). Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya ridge and its unexpected adaptive capabilities to extreme environments. *Nucl. Acids Res.* **30**, 3927–3935.
 33. Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S. *et al.* (2002). A complete sequence of the *T. tengcongensis* genome. *Genome Res.* **12**, 689–700.
 34. Stragier, P. (2002). A gene odyssey: exploring the genomes of endospore-forming bacteria. In *Bacillus subtilis and its Closest Relatives: From Genes to Cells* (Sonenshein, A. L., Hoch, J. A. & Losick, R., eds), pp. 519–526, American Society for Microbiology, Washington, DC.
 35. Kunkel, B., Kroos, L., Poth, H., Youngman, P. & Losick, R. (1989). Temporal and spatial control of the mother-cell regulatory gene *spoIIID* of *Bacillus subtilis*. *Genes Dev.* **3**, 1735–1744.
 36. Stevens, C. M. & Errington, J. (1990). Differential gene expression during sporulation in *Bacillus*

- subtilis*: structure and regulation of the *spoIIID* gene. *Mol. Microbiol.* **4**, 543–551.
37. Tatti, K. M., Jones, C. H. & Moran, C. P., Jr (1991). Genetic evidence for interaction of sigma E with the *spoIIID* promoter in *Bacillus subtilis*. *J. Bacteriol.* **173**, 7828–7833.
 38. Stragier, P., Kunkel, B., Kroos, L. & Losick, R. (1989). Chromosomal rearrangement generating a composite gene for a developmental transcription factor. *Science*, **243**, 507–512.
 39. Kroos, L., Kunkel, B. & Losick, R. (1989). Switch protein alters specificity of RNA polymerase containing a compartment-specific sigma factor. *Science*, **243**, 526–529.
 40. Halberg, R. & Kroos, L. (1994). Sporulation regulatory protein SpoIIID from *Bacillus subtilis* activates and represses transcription by both mother-cell-specific forms of RNA polymerase. *J. Mol. Biol.* **243**, 425–436.
 41. Sato, T., Harada, K., Ohta, Y. & Kobayashi, Y. (1994). Expression of the *Bacillus subtilis* *spoIVCA* gene, which encodes a site-specific recombinase, depends on the *spoIIGB* product. *J. Bacteriol.* **176**, 935–937.
 42. Cutting, S., Panzer, S. & Losick, R. (1989). Regulatory studies on the promoter for a gene governing synthesis and assembly of the spore coat in *Bacillus subtilis*. *J. Mol. Biol.* **207**, 393–404.
 43. Karmazyn-Campelli, C., Bonamy, C., Savelli, B. & Stragier, P. (1989). Tandem genes encoding sigma-factors for consecutive steps of development in *Bacillus subtilis*. *Genes Dev.* **3**, 150–157.
 44. Illing, N. & Errington, J. (1991). The *spoIIIA* operon of *Bacillus subtilis* defines a new temporal class of mother-cell-specific sporulation genes under the control of the sigma E form of RNA polymerase. *Mol. Microbiol.* **5**, 1927–1940.
 45. Fujita, M. & Losick, R. (2002). An investigation into the compartmentalization of the sporulation transcription factor sigmaE in *Bacillus subtilis*. *Mol. Microbiol.* **43**, 27–38.
 46. Popham, D. L. & Stragier, P. (1991). Cloning, characterization, and expression of the *spoVB* gene of *Bacillus subtilis*. *J. Bacteriol.* **173**, 7942–7949.
 47. Bryan, E. M., Beall, B. W. & Moran, C. P., Jr (1996). A sigma E dependent operon subject to catabolite repression during sporulation in *Bacillus subtilis*. *J. Bacteriol.* **178**, 4778–4786.
 48. Amaya, E., Khvorova, A. & Piggot, P. J. (2001). Analysis of promoter recognition *in vivo* directed by sigma(F) of *Bacillus subtilis* by using random-sequence oligonucleotides. *J. Bacteriol.* **183**, 3623–3630.
 49. Richmond, C. S., Glasner, J. D., Mau, R., Jin, H. & Blattner, F. R. (1999). Genome-wide expression profiling in *Escherichia coli* K-12. *Nucl. Acids Res.* **27**, 3821–3835.
 50. Takamatsu, H., Hiraoka, T., Kodama, T., Koide, H., Kozuka, S., Tochikubo, K. & Watabe, K. (1998). Cloning of a novel gene *yrbB*, encoding a protein located in the spore integument of *Bacillus subtilis*. *FEMS Microbiol. Letters*, **166**, 361–367.
 51. Halberg, R. & Kroos, L. (1992). Fate of the SpoIIID switch protein during *Bacillus subtilis* sporulation depends on the mother-cell sigma factor, sigma K. *J. Mol. Biol.* **228**, 840–849.
 52. Abhayawardhane, Y. & Stewart, G. C. (1995). *Bacillus subtilis* possesses a second determinant with extensive sequence similarity to the *Escherichia coli mreB* morphogene. *J. Bacteriol.* **177**, 765–773.
 53. Stormo, G. D. & Tan, K. (2002). Mining genome databases to identify and understand new gene regulatory systems. *Curr. Opin. Microbiol.* **5**, 149–153.
 54. Schneider, T. D. & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucl. Acids Res.* **18**, 6097–6100.
 55. Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F. & Wootton, J. C. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, **262**, 208–214.
 56. Liu, X., Brutlag, D. L. & Liu, J. S. (2001). Bioprospector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput.*, 127–138.
 57. Price, C. W., Fawcett, P., Ceremonie, H., Su, N., Murphy, C. K. & Youngman, P. (2001). Genome-wide analysis of the general stress response in *Bacillus subtilis*. *Mol. Microbiol.* **41**, 757–774.
 58. Frohman, M. A. (1994). On beyond classic RACE (rapid amplification of cDNA ends). *PCR Methods Appl.* **4**, S40–S58.
 59. Wach, A. (1996). PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast*, **12**, 259–265.
 60. Hecht, S., Eisenreich, W., Adam, P., Amslinger, S., Kis, K., Bacher, A. *et al.* (2001). Studies on the non-mevalonate pathway to terpenes: the role of the GcpE (IspG) protein. *Proc. Natl Acad. Sci. USA*, **98**, 14837–14842.
 61. Fischer, C., Geourjon, C., Bourson, C. & Deutscher, J. (1996). Cloning and characterization of the *Bacillus subtilis* *prkA* gene encoding a novel serine protein kinase. *Gene*, **168**, 55–60.
 62. Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F. *et al.* (2001). Comparative genomics of *Listeria* species. *Science*, **294**, 849–852.
 63. Resnekov, O., Alper, S. & Losick, R. (1996). Subcellular localization of proteins governing the proteolytic activation of a developmental transcription factor in *Bacillus subtilis*. *Genes Cells*, **1**, 529–542.
 64. Rudner, D. Z., Pan, Q. & Losick, R. M. (2002). Evidence that subcellular localization of a bacterial membrane protein is achieved by diffusion and capture. *Proc. Natl Acad. Sci. USA*, **99**, 8701–8706.
 65. Driks, A. (2002). Maximum shields: the assembly and function of the bacterial spore coat. *Trends Microbiol.* **10**, 251–254.
 66. Roels, S., Driks, A. & Losick, R. (1992). Characterization of *spoIVA*, a sporulation gene involved in coat morphogenesis in *Bacillus subtilis*. *J. Bacteriol.* **174**, 575–585.
 67. Stevens, C. M., Daniel, R., Illing, N. & Errington, J. (1992). Characterization of a sporulation gene, *spoIVA*, involved in spore coat morphogenesis in *Bacillus subtilis*. *J. Bacteriol.* **174**, 586–594.
 68. Driks, A., Roels, S., Beall, B., Moran, C. P., Jr & Losick, R. (1994). Subcellular localization of proteins involved in the assembly of the spore coat of *Bacillus subtilis*. *Genes Dev.* **8**, 234–244.
 69. Zheng, L. B., Donovan, W. P., Fitz-James, P. C. & Losick, R. (1988). Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. *Genes Dev.* **2**, 1047–1054.
 70. Kuwana, R., Kasahara, Y., Fujibayashi, M., Takamatsu, H., Ogasawara, N. & Watabe, K. (2002). Proteomics characterization of novel spore proteins of *Bacillus subtilis*. *Microbiology*, **148**, 3971–3982.

71. Cao, M., Kobel, P. A., Morshedi, M. M., Wu, M. F., Paddon, C. & Helmann, J. D. (2002). Defining the *Bacillus subtilis* sigma(W) regulon: a comparative analysis of promoter consensus search, run-off transcription/microarray analysis (ROMA), and transcriptional profiling approaches. *J. Mol. Biol.* **316**, 443–457.
72. Baldus, J. M., Green, B. D., Youngman, P. & Moran, C. P., Jr (1994). Phosphorylation of *Bacillus subtilis* transcription factor Spo0A stimulates transcription from the *spolIG* promoter by enhancing binding to weak 0A boxes. *J. Bacteriol.* **176**, 296–306.
73. Kanamaru, K., Stephenson, S. & Perego, M. (2002). Overexpression of the *pepF* oligopeptidase inhibits sporulation initiation in *Bacillus subtilis*. *J. Bacteriol.* **184**, 43–50.
74. Lopez-Diaz, I., Clarke, S. & Mandelstam, J. (1986). *SpoIID* operon of *Bacillus subtilis*: cloning and sequence. *J. Gen. Microbiol.* **132**, 341–354.
75. Smith, K., Bayer, M. E. & Youngman, P. (1993). Physical and functional characterization of the *Bacillus subtilis* *spoIIM* gene. *J. Bacteriol.* **175**, 3607–3617.
76. Frandsen, N. & Stragier, P. (1995). Identification and characterization of the *Bacillus subtilis* *spoIIP* locus. *J. Bacteriol.* **177**, 716–722.
77. Daniel, R. A., Drake, S., Buchanan, C. E., Scholle, R. & Errington, J. (1994). The *Bacillus subtilis* *spoVD* gene encodes a mother-cell-specific penicillin-binding protein required for spore morphogenesis. *J. Mol. Biol.* **235**, 209–220.
78. Piggot, P. J., Chak, K. F. & Bugaichuk, U. D. (1986). Isolation and characterization of a clone of the *spoVE* locus of *Bacillus subtilis*. *J. Gen. Microbiol.* **132**, 1875–1881.
79. Debarbieux, L. & Beckwith, J. (1999). Electron avenue: pathways of disulfide bond formation and isomerization. *Cell*, **99**, 117–119.
80. Sammons, R. L., Slynn, G. M. & Smith, D. A. (1987). Genetical and molecular studies on *gerM*, a new developmental locus of *Bacillus subtilis*. *J. Gen. Microbiol.* **133**, 3299–3312.
81. Slynn, G. M., Sammons, R. L., Smith, D. A., Moir, A. & Corfe, B. M. (1994). Molecular genetical and phenotypical analysis of the *gerM* spore germination gene of *Bacillus subtilis* 168. *FEMS Microbiol. Letters*, **121**, 315–320.
82. Ishikawa, S., Yamane, K. & Sekiguchi, J. (1998). Regulation and characterization of a newly deduced cell wall hydrolase gene (*cwlI*) which affects germination of *Bacillus subtilis* spores. *J. Bacteriol.* **180**, 1375–1380.
83. Paidhungat, M., Ragkousi, K. & Setlow, P. (2001). Genetic requirements for induction of germination of spores of *Bacillus subtilis* by Ca(2+)-dipicolinate. *J. Bacteriol.* **183**, 4886–4893.
84. Ragkousi, K., Eichenberger, P., van Ooij, C. & Setlow, P. (2003). Identification of a new gene essential for germination of *Bacillus subtilis* spores with Ca²⁺-dipicolinate. *J. Bacteriol.* In press.
85. Jendrossek, D. & Handrick, R. (2002). Microbial degradation of polyhydroxyalkanoates. *Annu. Rev. Microbiol.* **56**, 403–432.
86. Kunkel, B., Sandman, K., Panzer, S., Youngman, P. & Losick, R. (1988). The promoter for a sporulation gene in the *spoIVC* locus of *Bacillus subtilis* and its use in studies of temporal and spatial control of gene expression. *J. Bacteriol.* **170**, 3513–3522.
87. Cutting, S., Roels, S. & Losick, R. (1991). Sporulation operon *spoIVF* and the characterization of mutations that uncouple mother-cell from forespore gene expression in *Bacillus subtilis*. *J. Mol. Biol.* **221**, 1237–1256.
88. Ireton, K. & Grossman, A. D. (1992). Interactions among mutations that cause altered timing of gene expression during sporulation in *Bacillus subtilis*. *J. Bacteriol.* **174**, 3185–3195.
89. Frandsen, N., Barak, I., Karmazyn-Campelli, C. & Stragier, P. (1999). Transient gene asymmetry during sporulation and establishment of cell specificity in *Bacillus subtilis*. *Genes Dev.* **13**, 394–399.
90. Dworkin, J. & Losick, R. (2001). Differential gene expression governed by chromosomal spatial asymmetry. *Cell*, **107**, 339–346.
91. Khvorova, A., Chary, V. K., Hilbert, D. W. & Piggot, P. J. (2000). The chromosomal location of the *Bacillus subtilis* sporulation gene *spoIIR* is important for its function. *J. Bacteriol.* **182**, 4425–4429.
92. Zupancic, M. L., Tran, H. & Hofmeister, A. E. (2001). Chromosomal organization governs the timing of cell type-specific gene expression required for spore formation in *Bacillus subtilis*. *Mol. Microbiol.* **39**, 1471–1481.
93. Wu, L. J. & Errington, J. (1994). *Bacillus subtilis* SpoIIIE protein required for DNA segregation during asymmetric cell division. *Science*, **264**, 572–575.
94. Webb, C. D., Teleman, A., Gordon, S., Straight, A., Belmont, A., Lin, D. C. *et al.* (1997). Bipolar localization of the replication origin regions of chromosomes in vegetative and sporulating cells of *B. subtilis*. *Cell*, **88**, 667–674.
95. Youngman, P., Perkins, J. B. & Losick, R. (1984). Construction of a cloning site near one end of *Tn917* into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne *erm* gene. *Plasmid*, **12**, 1–9.
96. Kenney, T. J. & Moran, C. P., Jr (1987). Organization and regulation of an operon that encodes a sporulation-essential sigma factor in *Bacillus subtilis*. *J. Bacteriol.* **169**, 3329–3339.
97. Sterlini, J. M. & Mandelstam, J. (1969). Commitment to sporulation in *Bacillus subtilis* and its relationship to development of actinomycin resistance. *Biochem. J.* **113**, 29–37.
98. Harwood, C. R. & Cutting, S. S. (1990). *Molecular Biological Methods for Bacillus*, Wiley, Chichester.
99. Lemon, K. P. & Grossman, A. D. (1998). Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science*, **282**, 1516–1519.
100. Ihaka, R. & Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314.
101. Tseng, G. C., Oh, M. K., Rohlin, L., Liao, J. C. & Wong, W. H. (2001). Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucl. Acids Res.* **29**, 2549–2557.
102. Dudoit, S., Yang, Y. H., Callow, M. J. & Speed, T. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statist. Sin.* **12**, 111–140.
103. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C. *et al.* (2001). Minimum information about a microarray

- experiment (MIAME)-toward standards for microarray data. *Nature Genet.* **29**, 365–371.
104. Liu, J. S., Neuwald, A. F. & Lawrence, C. E. (1995). Bayesian models for multiple local sequence alignment and Gibbs sampling strategies. *J. Am. Stat. Assoc.* **90**, 1156–1170.
105. Pragai, Z. & Harwood, C. R. (2002). Regulatory interactions between the Pho and sigma(B)-dependent general stress regulons of *Bacillus subtilis*. *Microbiology*, **148**, 1593–1602.
106. Rappu, P., Shin, B. S., Zalkin, H. & Mantsala, P. (1999). A role for a highly conserved protein of unknown function in regulation of *Bacillus subtilis* *purA* by the purine repressor. *J. Bacteriol.* **181**, 3810–3815.
107. Raeymaekers, L., Wuytack, E., Willems, I., Michiels, C. & Wuytack, F. (2002). Expression of a P-type Ca(2+) -transport ATPase in *Bacillus subtilis* during sporulation. *Cell Calcium*, **32**, 93.
108. Hourdou, M. L., Guinand, M., Vacheron, M. J., Michel, G., Denoroy, L., Duez, C. *et al.* (1993). Characterization of the sporulation-related gamma-D-glutamyl-(L)-meso-diaminopimelic-acid-hydrolysing peptidase I of *Bacillus sphaericus* NCTC 9602 as a

member of the metallo(zinc) carboxypeptidase A family. Modular design of the protein. *Biochem. J.* **292**, 563–570.

Edited by M. Gottesman

(Received 29 October 2002; received in revised form 22 January 2003; accepted 27 January 2003)

SCIENCE @ DIRECT®
www.sciencedirect.com

Supplementary Material for this paper comprising two Tables is available on Science Direct