JB Accepts, published online ahead of print on 2 November 2012
J. Bacteriol. doi:10.1128/JB.01778-12

Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*

Bjorn A. Traag[1], Antonia Pugliese[1], Jonathan A. Eisen[2] and Richard Losick[1, #]

[1]Department of Molecular and Cellular Biology, Harvard University, [2]Department of Evolution and Ecology, University of California Davis Genome Center

Running title: *Discovery of sporulation genes using gene conservation*

Corresponding author: losick@mcb.harvard.edu

Abstract

The capacity to form endospores is unique to certain members of the low-G+C group of Gram-positive bacteria (*Firmicutes*) and requires signature sporulation genes that are highly conserved across members of distantly related genera, such as *Clostridium* and *Bacillus*. Using gene conservation among endospore-forming bacteria, we identified eight previously uncharacterized genes that are enriched among endospore-forming species. Expression of five of these genes was dependent on sporulation-specific transcription factors. Mutants of none of the genes exhibited a conspicuous defect in sporulation, but mutants of two, *ylxY* and *ylyA*, were outcompeted by a wild type strain under sporulation-inducing conditions but not during growth. In contrast, a *ylmC* mutant displayed a slight competitive advantage over the wild type specific to sporulation-inducing conditions. The phenotype of a *ylyA* mutant was ascribed to a defect in spore germination efficiency. This work demonstrates the power of combining phylogenetic profiling with reverse genetics and gene regulatory studies to identify unrecognized genes that contribute to a conserved developmental process.

Introduction

The formation of endospores is a distinctive developmental process wherein a dormant cell type (the endospore) is formed inside of another cell (the mother cell) and ultimately released into the environment by lysis of the mother cell (29, 38). Endospores are metabolically inactive and highly resistant to environmental stresses, such as heat, radiation, chemicals and desiccation (34). At the same time, these spores monitor the environment and are capable of rapidly resuming growth when conditions are favorable (33). Endospore formation is unique to the low-G+C group of Gram-positive bacteria (*Firmicutes*). For the most part, it is restricted to the family *Bacillaceae* and the class *Clostridia*, but members of the less well-studied family *Veillonellaceae* (*e.g. Acetonema longum* (21, 39)) also produce endospores. The last common ancestor of *Clostridium* and *Bacillus* predates the initial rise of oxygen in the atmosphere (3), approximately 2.3 billion years ago. And yet, remarkably, orthologs of signature sporulation genes are shared between the genomes of these distantly related bacteria (17, 38). We wondered whether gene conservation among endospore-formers could be exploited to discover previously unrecognized genes involved in sporulation.

Sporulation has been most extensively studied in the model organism *B. subtilis*. Entry into sporulation is governed by the master regulator Spo0A, which is activated by phosphorylation through a multicomponent signal transduction pathway (19). Spo0A~P directly regulates (activates or represses) the expression of 121 genes (26) and significantly influences the expression of over 500 genes (12). Sporulation is initiated when Spo0A~P levels reach a threshold (14). Sporulating cells undergo several, successive morphological changes, a hallmark of which is the formation of a two-compartment sporangium consisting of forespore and mother-cell compartments. As development proceeds, the forespore is wholly engulfed by the mother

cell to create a cell-within-a-cell. The inner cell becomes the dormant spore and is released from the mother cell by lysis (38). Upon release, the mature spore can remain dormant for long periods of time or in response to germinants give rise to a vegetative cell. The developmental program of sporulation is governed in part by the successive action of four compartment-specific sigma factors (appearing in the order σ^F, σ^E, σ^G and σ^K), whose activities are confined to the forespore (σ^F and σ^G) or the mother cell (σ^E and σ^K) (24).

Traditional approaches of forward genetics have identified many, if not all, genes that are essential for sporulation (*spo*). These approaches rely on conspicuous phenotypes for the identification of target genes. Complementary approaches, such as the identification of sporulation-specific proteins (e.g. SASP and coat proteins) and transcriptome analysis (11, 40), revealed additional genes under sporulation control, including genes that contribute to efficient sporulation and spore resistance properties. Here we report the application of phylogenetic profiling in an effort to discover additional genes involved in sporulation that might have gone undetected in previous approaches. We therefore sought to identify previously unrecognized genes involved in sporulation in *B. subtilis* on the basis of their conservation among endospore-forming bacteria. Using phylogenetic profile analysis for the initial identification, and transcriptional and mutational analyses, we discovered previously overlooked genes under sporulation control, including two genes, mutants of which caused a small but detectable developmental defect.

Materials and Methods

Bacterial strains, culturing conditions

Escherichia coli strain DH5α was used for propagating plasmids and grown and transformed using standard procedures (31). The *B. subtilis* strains used in this work are listed in Table 1. All strains were derived from the prototrophic laboratory strain PY79 (45). Transformation of *Bacillus* strains with double-stranded PCR fragments, plasmid or genomic DNA was done as previously described (42). Sporulation was induced by exhaustion in Difco sporulation (DS) medium, or, in the case of β-galactosidase activity assays (see below), by resuspension in Sterlini-Mandelstam (SM) medium (27).

Plasmid construction

Oligonucleotides used for polymerase chain reaction (PCR) in this study are listed in Table 2. Fragments of the upstream regions of candidate genes, including the respective ribosome binding sites (RBS) and start codons, were amplified by PCR and cloned into pAH124 (6) using the appropriate restriction sites (Table 2). In this way, *lacZ* reporter constructs were obtained with the start codon of the gene fused directly to the *lacZ* gene. Constructs were introduced into PY79 by transformation.

Complementation constructs for the *ylmC, ylxY* and *ylyA* mutants were made by PCR amplification of fragments carrying the gene promoters and entire the open reading frames (ORF). These PCR fragments were cloned using the appropriate restriction enzymes into pDG1662 (18), in the case of *ylmC* and *ylxY*, or pSac-Cam (25), in the case of *ylyA*. Constructs were introduced into the respective mutant strains by transformation.

Phylogenetic profile analysis

For each predicted gene product in *B. subtilis subsp. subtilis* 168, its presence or absence in 626 complete archaeal and bacterial genomes that were available at the time of the initial phylogenetic analysis was determined by asking whether a putative ortholog was present (1) or absent (0) in that species, similar to previously described (44). This analysis included 46 genomes of endospore-forming bacteria belonging to the family *Bacillaceae* and the class *Clostridia*, namely: *Bacillus amyloliquefaciens* FZB42, *Bacillus anthracis* str. Ames, *Bacillus anthracis* str. 'Ames Ancestor', *Bacillus anthracis* str. Sterne, *Bacillus cereus* ATCC 10987, *Bacillus cereus* ATCC 14579, *Bacillus cereus* E33L, *Bacillus cereus subsp. cytotoxis* NVH 391-98, *Bacillus clausii* KSM-K16, *Bacillus halodurans* C-125, *Bacillus licheniformis* ATCC 14580, *Bacillus pumilus* SAFR-032, *Bacillus thuringiensis serovar konkukian* str. 97-27, *Bacillus thuringiensis* str. Al Hakam, *Bacillus weihenstephanensis* KBAB4, *Geobacillus kaustophilus* HTA426, *Geobacillus thermodenitrificans* NG80-2, *Oceanobacillus iheyensis* HTE831, *Alkaliphilus metalliredigens* QYMF, *Alkaliphilus oremlandii* OhILAs, *Caldicellulosiruptor saccharolyticus* DSM 8903, *Carboxydothermus hydrogenoformans* Z-2901, *Clostridium acetobutylicum* ATCC 824, *Clostridium beijerinckii* NCIMB 8052, *Clostridium botulinum A str.* ATCC 3502, *Clostridium botulinum A str.* ATCC 19397, *Clostridium botulinum A str.* Hall, *Clostridium botulinum F str.* Langeland, *Clostridium difficile* 630, *Clostridium kluyveri* DSM 555, *Clostridium perfringens* ATCC 13124, *Clostridium perfringens* SM101, *Clostridium perfringens* str. 13, *Clostridium phytofermentans* ISDg, *Clostridium tetani* E88, *Clostridium thermocellum* ATCC 27405, *Desulfitobacterium hafniense* Y51, *Desulfotomaculum reducens* MI-1, *Moorella thermoacetica* ATCC 39073, *Symbiobacterium thermophilum* IAM 14863, *Syntrophomonas wolfei subsp. wolfei* str. Goettingen, *Thermoanaerobacter pseudethanolicus* ATCC 33223, *Thermoanaerobacter tengcongensis* MB4, *Pelotomaculum thermopropionicum* SI,

116 *Thermoanaerobacter sp.* X514. Proteins were then grouped by their distribution patterns across
117 species. In this way, a set of 58 genes, many of which are signature sporulation genes (see
118 below) and all of which are highly enriched among endospore-forming species, was obtained.

119

120 *β-galactosidase activity assays*

121 Samples were collected from shaking cultures in duplicate at various time points after
122 sporulation was induced by resuspension into SM medium. β-galactosidase activity was
123 measured in a Synergy 2 plate reader (BioTek) as previously described (6). The experiment was
124 repeated to ensure reproducibility. β-galactosidase activity is reported in arbitrary (AU) units as
125 the rate of ONPG conversion (*i.e.* V_{max}, with units of OD_{420} per minute) divided by the OD_{600} of
126 the sample at the time of collection, as previously described (6).

127

128 *Competition experiments*

129 A wild type reference strain was competed against strains mutant for candidate genes as
130 previously described (32), with a few differences. Typically, starting from overnight LB cultures
131 the wild type reference strain (RL5360), which carried an isopropyl β-D-1-thiogalactopyranoside
132 (IPTG)-inducible *lacZ*, was mixed with an excess mutant strain in 5 ml DS medium. In
133 competition assays using *ylmC* mutants, which displayed an advantage under sporulation-
134 inducing conditions, we started with an excess of the wild type strain. Cultures were grown and
135 allowed to sporulate at 37˚C in DS medium for 24 hours. Cultures were then heat-treated at 80˚C
136 for 20 minutes, briefly cooled at room temperature, and diluted in 5ml fresh DS medium. At

appropriate intervals, dilutions were plated on agar plates containing 0.008% w/v 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and 1 mM IPTG. Numbers of blue and white colonies were counted, reflecting the ratio of wild type reference to mutant strain, respectively. Similar competition experiments were done in LB to establish if observed phenotypes were specific to sporulation-inducing conditions. In these experiments, we grew the co-cultures at 25˚C to minimize the time the culture was in stationary phase. In short, wild type reference and mutant strains were grown at 25˚C in LB for 24 hours, and subsequently diluted in 5ml fresh LB for another round of growth. The heat-treatment step was omitted in these experiments.

Germination of purified spores

Spores were essentially purified as previously described (27). In short, strains were induced to sporulate by growing in DS medium for 48-72 hours. Cells were pelleted by centrifugation, and the pellet was resuspended in ice-cold water. The cells were washed twice with ice-cold water and stored overnight at 4˚C. The next day, cells were pelleted by centrifugation and washed another 6-8 times with ice-cold water. In this way, preparation were obtained that were more than 95% phase-bright spores, as judged by phase-contrast microscopy.

For germination assays, spores at an OD_{600} of 10 were activated by heat-treatment at 80˚C for 20 minutes and cooled on ice for 2 minutes. Activated spores were germinated at an OD_{600} of 0.5 in LB medium. Germination was recorded as a loss of optical density in a Synergy 2 plate reader (BioTekin).

Results

Candidates for uncharacterized sporulation genes

We used phylogenetic profile analysis to search for genes in the genome of *B. subtilis subsp. subtilis* 168 (hereafter simply referred to as *B. subtilis*) that are specifically conserved among endospore-forming, low-G+C Gram-positive bacteria. A similar method was previously used to identify signature sporulation genes in the thermophilic firmicute *Carboxydothermus hydrogenoformans* (44). Phylogenetic profiling works by grouping genes according to their distribution patterns in different species. For each predicted gene product in *B. subtilis,* its presence or absence in all complete archaeal and bacterial genomes available at the time of this analysis (626 genomes in total) was determined by asking whether an ortholog was present in that species. This analysis included 46 genomes of endospore-forming bacteria belonging to the family *Bacillaceae* and the class *Clostridia*. Orthologs were then grouped by their distribution patterns across species. In this way, 58 genes were identified that were highly and specifically conserved among endospore-forming bacteria (Fig S1). Many of these are signature sporulation genes with well-studied roles in spore formation, such as *spoIIR* and *spoIIGA*, which mediate the activation of the mother-cell-specific transcription factor σ^E (20), *spoIID*, *spoIIM* and *spoIIP*, required for forespore engulfment by the mother cell (8, 13, 36), and *spoIVA*, which encodes a morphogenetic protein required for coat assembly (30). Five genes (*i.e. bkdR*, *ylmC*, *ymxH*, *ylzA* (formerly designated *remA*) and *ymfB* (formerly designated *tepA*)), however, had no previously documented role in spore formation, and an additional three genes (*i.e. ylxY*, *yteA* and *ylyA*) had previously been shown to be under sporulation control, but no roles in sporulation had been described (10, 40). We hypothesized that these genes might play previously unrecognized or overlooked roles in sporulation.

Six of the eight genes were widely conserved among the 46 endospore-forming *Bacillaceae* and *Clostridia* genomes examined. In the cases of *ymxH* and *ylyA*, however, only 25 and 11 orthologs, respectively, were identified. *ymxH* orthologs are present in almost all endospore-forming *Bacillaceae* species but missing from the majority of endospore-forming *Clostridia* species. Nonetheless, *ymxH* exhibits significant sequence similarity to one of the other genes on our list *ylmC*, which is abundant among endospore-forming bacteria. Meanwhile, a close look at *ylyA*, which is homologous to *yteA*, revealed that several genes identified as *yteA* orthologs in endospore-forming bacteria share similar gene synteny with *B. subtilis ylyA*. *B. subtilis ylyA* is flanked by several well-characterized genes, including, but not restricted to: *divIVA*, encoding a cell division protein, *ileS*, encoding a isoleucyl-tRNA synthetase, *lspA*, encoding a type II signal peptidase, and *rluD*, encoding a pseudouridylate synthase, many of which are found in the vicinity of *ylyA* orthologs in *Bacillaceae* and *Clostridia*. This suggests that these genes are actually *ylyA* orthologs, raising the number of orthologs to 29 (Table 3). Thus, all eight genes are widely conserved among endospore-forming bacteria or are homologous to genes that are.

Finally, two additional genes were identified, *yvjA* and *buk*, that had orthologs both in endospore-forming bacteria and in all four species of *Listeria* included in this analysis (Fig S1). *Listeria* species are closely related to *B. subtilis* but do not form spores and lack almost all signature sporulation genes (37). Because of their presence in non-endospore-forming bacteria, we considered *yvjA* and *buk* as unlikely candidates for unrecognized sporulation genes. Nonetheless, and as a control, we retained one of these genes, *buk*, in our analysis. Thus, a total of nine genes were carried forward for further investigation.

Transcription under sporulation-inducing conditions

We next asked if the nine candidate genes are transcribed under conditions that induce sporulation. For this purpose we built transcriptional reporter constructs, typically cloning a 300-400 base pair fragment directly upstream of the gene and fusing its start codon to that of the *lacZ* gene. In the case of *buk*, which is the third gene in a seven-gene (*bkd*) operon that is involved in branched-chain amino acid utilization (9), we instead cloned an approximately 300 base pair fragment upstream of *ptb*, the first gene of the operon. *B. subtilis* strains carrying these constructs integrated at the *amyE* locus were induced to sporulate by resuspension in SM medium (27), and samples taken at various times were analyzed for β-galactosidase activity. Seven of the nine reporters were expressed during sporulation, six of which were induced at various times after the induction of sporulation (Fig 1). We did not observe activity for the *bkdR* and *ylzA* reporters under the tested conditions (data not shown).

The *bkd* reporter was active from the time of the induction of sporulation, decreasing only slightly during the time it was monitored (Fig 1A). As previously reported, expression of the *bkd* operon depends on the alternative sigma factor σ^L and BkdR (9). Indeed, in a *bkdR* mutant expression was abolished (Fig 1A). We next measured expression in a strain mutant for *spo0A*. The *spo0A* gene encodes the master regulator for entry into sporulation Spo0A, which is active in its phosphorylated form Spo0A~P (19, 38). Activity was approximately two-fold higher compared to wild type at the times tested. Interestingly, in a strain doubly mutant for *spo0A* and *bkdR* expression levels were similar to that of wild type (Fig 1A), showing that in the absence of Spo0A BkdR is not required for expression. That is, a *spo0A* mutation is epistatic to a *bkdR* mutation. This suggests that BkdR antagonizes Spo0A~P to activate transcription from the *bkd* operon promoter. We found a potential Spo0A binding site (GTCGAAA; Fig S2), with high

similarity to the consensus binding sequence (TTTGTCGAAA (26)), located immediately downstream to the σ^L-dependent transcriptional start site (9). Just upstream of the promoter are tandem sequences previously shown to be important for BkdR-mediated activation (9). *In toto*, these observations suggest that the binding of BkdR upstream of the promoter overcomes the repressive effect of the binding of Spo0A~P just downstream of the start site.

The *ymxH* reporter was expressed from an early time, increasing slightly but constantly during the first five hours (Fig 1B). We tested the activity in strains mutant for *spo0A* and *sigF*, which encodes the first forespore-specific sigma factor σ^F (38). Expression in a *spo0A* mutant was up-regulated, steadily increasing from one hour after sporulation was induced. In contrast, activity in a *sigF* mutant was unchanged compared to wild type (Fig 1B). These findings suggest that *ymxH* is directly or indirectly under the negative control of Spo0A~P but is not otherwise under sporulation control.

Expression from the *ylxY* and *ylmC* reporters was induced between hour one and two of sporulation (Fig 1C, D). Previous work indicated that *ylxY* expression is σ^E-dependent and under the negative control of the mother cell-specific regulator SpoIIID (10-11). We tested the activity of the *ylxY* and *ylmC* reporters in strains mutant for *sigF*, *sigE*, which encodes the mother cell-specific sigma factor σ^E, or *sigG*, which encodes the late-appearing, forespore-specific sigma factor σ^G. Activity for both was abolished in the *sigF* and *sigE* mutants, but reached wild type levels in the *sigG* mutant (Fig 1C, D). Thus, these genes are transcribed in a σ^E-dependent manner.

Finally, three reporters were induced between two and three hours after sporulation was induced, namely those for *ymfB*, *yteA* and *ylyA* (Fig 1E-G). Previous transcriptome analyses

indicated that *yteA* and *ylyA* are indeed part of the σ^G regulon (40), whereas *ymfB* was not known to be under sporulation control. The activity of all three reporters was abolished in a *sigG* mutant (Fig 1E-G). In addition, *ylyA* was previously shown to be under the control of SpoVT, a modulator of σ^G-dependent transcription (2, 40). We tested the activity of all three reporters in a strain mutant for *spoVT*. The activity of the *yteA* and *ylyA* reporters was markedly reduced in a *spoVT* mutant, whereas the activity of *ymfB* reporter increased in a *spoVT* mutant (Fig 1E-G).

In summary, we conclude that five of the nine genes in our investigation are under sporulation control with two, namely *ylxY* and *ylmC*, being under the control of σ^E and three, *ymfB*, *yteA* and *ylyA*, under the control of σ^G.

Competition-based analysis of candidate gene mutants

We constructed mutant strains for eight of the nine genes by deleting and replacing their ORF with antibiotic-resistance cassettes. We were unable to obtain a mutant for the ninth gene *ylzA*. Previously, others obtained transposon insertions directly upstream of *ylzA*; however, no report was made of transposon insertions internal to the *ylzA* coding sequence (43).

None of the eight mutants had a conspicuous phenotype as judged by colony morphology or spore formation (data not shown). Thus, if any of these genes represent previously uncharacterized sporulation genes, their contribution to spore formation must be subtle. To test for such a subtle role, we carried out competition experiments in which mutant strains were competed for several rounds of sporulation against a wild type reference strain marked by an IPTG-inducible *lacZ* gene. Typically, mutant cells were several fold in excess to the wild type at

the start of the experiment. Co-cultures of the mutant and wild type were grown and allowed to sporulate at 37˚C in DS medium for 24 hours. The culture was then heat-treated at 80˚C and diluted in fresh DS medium for another round of sporulation. At appropriate intervals, dilutions were plated on agar plates containing X-gal and IPTG and blue (wild type) and white (mutant) colonies were counted. Mutant strains that were out competed (or in one case slightly under competed) by the wild type were next subjected to competition experiments in growth medium to determine if the observed competition phenotype was an indirect consequence of a growth defect rather than a defect in sporulation. In these experiments, co-cultures were grown at 25˚C in LB medium for 24 hours and subsequently diluted in fresh LB medium for another round of growth. The heat-treatment step was omitted in these experiments.

The mutant strains fell into three categories: those that did not have a competition phenotype under sporulation-inducing conditions; those that had a phenotype under sporulation-inducing but not under growth conditions; and lastly, those that had a phenotype under both conditions. The first category comprised the *bkdR*, *ymxH*, *ymfB* and *yteA* mutants. In a competition experiment in which the wild type reference strain was competed against an unmarked wild type strain the percentage of the wild type reference strain remained constant during several rounds of competition (Fig S3), indicating that the inducible *lacZ* construct did not affect the fitness of the reference strain. Similarly, in competition experiments with strains mutant for *bkdR*, *ymxH*, *ymfB* or *yteA* the percentage of the wild type reference strain remained constant (Fig S3).

The second category comprised the *ylxY*, *ylyA* and *ylmC* mutants. Strains mutant for *ylxY* and *ylyA* had clear competition deficits under sporulation-inducing conditions. Over the course of nine and five rounds of competition, respectively, the percentage of wild type reference strain

had increased from approximately 20% to 90% of the population (Fig 2, 3). Genetic complementation by reintroducing a copy of the respective gene at an ectopic locus restored competitiveness to the mutant strains (Fig 2, 3), confirming that the observed phenotypes resulted from deletion of the genes. In contrast to the apparent deficit of *ylxY* and *ylyA* mutants, the *ylmC* mutant exhibited a slight competitive advantage under sporulation-inducing conditions (Fig 4). As before, genetic complementation by reintroducing a copy of *ylmC* at the ectopic *amyE* locus reversed this phenotype (Fig 4). All three competition phenotypes were found to be specific to competition experiments under sporulation-inducing conditions, because no changes from the starting ratio were observed in competition experiments during growth (Fig 5).

Finally, the third category contained only one strain, the *buk* mutant, which had a clear competition deficit under sporulation-inducing and growth conditions (Fig S4).

A ylyA mutant is defective in germination

We do not know the precise step in sporulation at which the *ylxY* mutation impedes spore formation or the step at which the *ylmC* mutation confers a competitive advantage. However, in the case of *ylyA*, we can, at least in part, attribute the competitive disadvantage to impaired spore germination. As shown in Fig. 6, spores from a *ylyA* mutant are slower to germinate than either the wild type parent or a *ylyA* mutant strain harboring a wild type copy of the gene at the ectopic *sacA* locus.

Discussion

Using phylogenetic profiling, we identified eight genes (*i.e. bkdR*, *ylmC*, *ymxH*, *ylxY*, *ylzA*, *ymfB*, *yteA* and *ylyA*) that are widely conserved among endospore-forming species of *Bacillaceae* and *Clostridia*, but were not previously reported to be involved in sporulation or, in the case of *ylxY*, *yteA* and *ylyA*, not well characterized. *ymfB*, which is homologous to ClpP-like proteases (Table 3), was previously suggested to be involved in translocation and processing of the α-amylase AmyQ ((4); *ymfB* was therein named *tepA*). Researchers from the same laboratory, however, later reported that they were unable to replicate the initial results with a clean knock-out of *ymfB* (41). *ylzA*, which encodes a hypothetical protein with no clear homology to known proteins (Table 3), was previously shown to be involved in the regulation of extracellular matrix components during biofilm formation in *B.subtilis* ((43); *ylzA* is therein designated *remA*). Five of the eight genes were found to be under the control of sporulation-specific transcription factors, with *ylmC* and *ylxY* being under the control of the mother cell-specific sigma factor σ^E, and *ymfB*, *yteA* and *ylyA* under the control of the forespore-specific factors σ^G and SpoVT. Inactivation of *ylxY*, *ylyA* and *ylmC*, resulted in measurable changes in competitiveness under sporulation-inducing conditions, but not under growth conditions.

We included *buk* as a control in our investigation. Like the eight other genes, orthologs of *buk* are enriched among endospore-forming bacteria but are also found in some non-endospore-forming species, most notably *Listeria*, a close relative of *B. subtilis* that is asporogenic. The *buk* gene, which codes for a butyrate kinase, is part of a seven-gene operon in *B. subtilis* that is involved in the utilization of branched-chain amino acids as a nitrogen source (9). Orthologs of other members of the operon are widespread among bacteria, much more so than *buk* itself. In *B. subtilis* the operon is under the control of the alternative sigma factor σ^L and the transcription activator BkdR (9), which as we have shown is itself highly conserved among endospore-

337 forming bacteria. Inactivation of *buk* resulted in defects that were not specific to sporulation, and
338 the *bkd* operon promoter was constitutively active under the tested conditions. Interestingly, we
339 found a putative Spo0A binding site adjacent to the predicted transcriptional start site, and our
340 results indicate that BkdR antagonizes Spo0A~P to activate expression. The presence of BkdR
341 orthologs in almost all endospore-forming species analyzed (Table 3), but not *Listeria* species,
342 supports the idea that BkdR is conserved among endospore-forming bacteria to counteract the
343 effect of Spo0A~P on transcription.

344 *ylmC* and *ymxH* are paralogs that code for small hypothetical proteins that resemble a
345 motif known as photosynthetic reaction center (PRC) beta-barrel domain (Table 3). The PRC
346 domain, which is itself widespread among photosynthetic and non-photosynthetic bacteria,
347 archaea and plants, is thought to mediate protein-protein interactions (1). Deletion of *ylmC*
348 resulted in a strain that had a slight competitive advantage over the wild type under sporulation-
349 inducing conditions. It is unclear what causes this unexpected phenotype. A double mutant of
350 *ylmC* and *ymxH* essentially had the same phenotype as the *ylmC* single mutant (data not shown).
351 We infer that YlmC must confer some fitness advantage during spore formation under unknown
352 environmental conditions (*e.g.* a spore resistance property) but that production of the protein
353 evidently imposes a slight cost on spore formation that impedes development.

354 Deletion of *ylxY* and *ylyA* resulted in clear competitive deficits specific to sporulation-
355 inducing conditions. *ylxY* encodes a probable polysaccharide deacetylase (Table 3), and exhibits
356 some similarity to two *B. subtilis* genes under sporulation control, namely *pdaA* and *pdaB*,
357 mutants of which display defects in spore cortex maturation (15-16, 35). The nature of the
358 competition deficit of the *ylxY* mutant is currently unknown. In contrast, we have determined that
359 the competition defect of the *ylyA* mutant stems not from a defect in spore formation *per se* but

rather from impaired germination of the mutant spores. Gourse and co-workers previously reported that YlyA and YteA share some sequence similarity to the transcription factor DksA (23). DksA inhibits the transcription of ribosomal RNA (rRNA) genes by direct interaction with RNA polymerase (28). Mutation of *ylyA* or *yteA*, however, seemingly did not affect the activity of the P1 promoter of the rRNA gene *rrnB* under any of the tested growth conditions (23). It will be interesting to see whether YlyA similarly modifies RNA polymerase activity and if so whether this modification influences the expression of genes involved in spore germination.

Our results reinforce the view that phylogenetic profiling in combination with reverse genetics and gene regulatory studies can be a powerful tool for the discovery of genes that play a subtle role in a complex developmental process and whose contributions might otherwise be overlooked by traditional approaches of forward genetics.

Acknowledgments

We thank Martin Wu for help with phylogenetic profiling, and Rick Gourse for bringing to our attention the similarity of *yteA* and *ylyA* to *dksA*. We thank members of the Losick laboratory for helpful discussion. This work was supported by a Netherlands Organisation for Scientific Research (NWO) Rubicon grant to B.A.T., Gordon and Betty Moore Foundation grant 1660 to J.A.E., and National Institute of Health grant GM18568 to R.L.

1. **Anantharaman, V., and L. Aravind.** 2002. The PRC-barrel: a widespread, conserved domain shared by photosynthetic reaction center subunits and proteins of RNA metabolism. Genome Biol **3:**RESEARCH0061.

2. **Bagyan, I., J. Hobot, and S. Cutting.** 1996. A compartmentalized regulator of developmental gene expression in Bacillus subtilis. J Bacteriol **178:**4500-4507.

3. **Battistuzzi, F. U., A. Feijao, and S. B. Hedges.** 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol Biol **4:**44.

4. **Bolhuis, A., A. Matzen, H. L. Hyyrylainen, V. P. Kontinen, R. Meima, J. Chapuis, G. Venema, S. Bron, R. Freudl, and J. M. van Dijl.** 1999. Signal peptide peptidase- and ClpP-like proteins of Bacillus subtilis required for efficient translocation and processing of secretory proteins. J Biol Chem **274:**24585-24592.

5. **Branda, S. S., J. E. Gonzalez-Pastor, S. Ben-Yehuda, R. Losick, and R. Kolter.** 2001. Fruiting body formation by Bacillus subtilis. Proc Natl Acad Sci U S A **98:**11621-11626.

6. **Camp, A. H., and R. Losick.** 2009. A feeding tube model for activation of a cell-specific transcription factor during sporulation in Bacillus subtilis. Genes Dev **23:**1014-1024.

7. **Camp, A. H., and R. Losick.** 2008. A novel pathway of intercellular signalling in Bacillus subtilis involves a protein with similarity to a component of type III secretion channels. Mol Microbiol **69:**402-417.

8. **Coote, J. G.** 1972. Sporulation in Bacillus subtilis. Characterization of oligosporogenous mutants and comparison of their phenotypes with those of asporogenous mutants. J Gen Microbiol **71:**1-15.

9. **Debarbouille, M., R. Gardan, M. Arnaud, and G. Rapoport.** 1999. Role of bkdR, a transcriptional activator of the sigL-dependent isoleucine and valine degradation pathway in Bacillus subtilis. J Bacteriol **181:**2059-2066.

10. **Eichenberger, P., M. Fujita, S. T. Jensen, E. M. Conlon, D. Z. Rudner, S. T. Wang, C. Ferguson, K. Haga, T. Sato, J. S. Liu, and R. Losick.** 2004. The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis. PLoS Biol **2:**e328.

11. **Eichenberger, P., S. T. Jensen, E. M. Conlon, C. van Ooij, J. Silvaggi, J. E. Gonzalez-Pastor, M. Fujita, S. Ben-Yehuda, P. Stragier, J. S. Liu, and R. Losick.** 2003. The sigmaE regulon and the identification of additional sporulation genes in Bacillus subtilis. J Mol Biol **327:**945-972.

12. **Fawcett, P., P. Eichenberger, R. Losick, and P. Youngman.** 2000. The transcriptional profile of early to middle sporulation in Bacillus subtilis. Proc Natl Acad Sci U S A **97:**8063-8068.

13. **Frandsen, N., and P. Stragier.** 1995. Identification and characterization of the Bacillus subtilis spoIIP locus. J Bacteriol **177:**716-722.

14. **Fujita, M., J. E. Gonzalez-Pastor, and R. Losick.** 2005. High- and low-threshold genes in the Spo0A regulon of Bacillus subtilis. J Bacteriol **187:**1357-1368.

15. **Fukushima, T., T. Tanabe, H. Yamamoto, S. Hosoya, T. Sato, H. Yoshikawa, and J. Sekiguchi.** 2004. Characterization of a polysaccharide deacetylase gene homologue (pdaB) on sporulation of Bacillus subtilis. J Biochem **136:**283-291.

16. **Fukushima, T., H. Yamamoto, A. Atrih, S. J. Foster, and J. Sekiguchi.** 2002. A polysaccharide deacetylase gene (pdaA) is required for germination and for production of muramic delta-lactam residues in the spore cortex of Bacillus subtilis. J Bacteriol **184:**6007-6015.

17. **Galperin, M. Y., S. L. Mekhedov, P. Puigbo, S. Smirnov, Y. I. Wolf, and D. J. Rigden.** 2012. Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of

sporulation-specific genes. Environ Microbiol **published online 13 Aug 2012 doi:**10.1111/j.1462-2920.2012.02841.x.

18. **Guerout-Fleury, A. M., N. Frandsen, and P. Stragier.** 1996. Plasmids for ectopic integration in Bacillus subtilis. Gene **180:**57-61.

19. **Hoch, J. A.** 1993. Regulation of the phosphorelay and the initiation of sporulation in Bacillus subtilis. Annu Rev Microbiol **47:**441-465.

20. **Hofmeister, A. E., A. Londono-Vallejo, E. Harry, P. Stragier, and R. Losick.** 1995. Extracellular signal protein triggering the proteolytic activation of a developmental transcription factor in B. subtilis. Cell **83:**219-226.

21. **Kane, M. D., and J. A. Breznak.** 1991. Acetonema longum gen. nov. sp. nov., an H2/CO2 acetogenic bacterium from the termite, Pterotermes occidentis. Arch Microbiol **156:**91-98.

22. **Kenney, T. J., and C. P. Moran, Jr.** 1987. Organization and regulation of an operon that encodes a sporulation-essential sigma factor in Bacillus subtilis. J Bacteriol **169:**3329-3339.

23. **Krasny, L., and R. L. Gourse.** 2004. An alternative strategy for bacterial ribosome synthesis: Bacillus subtilis rRNA transcription regulation. EMBO J **23:**4473-4483.

24. **Losick, R., and P. Stragier.** 1992. Crisscross regulation of cell-type-specific gene expression during development in B. subtilis. Nature **355:**601-604.

25. **Middleton, R., and A. Hofmeister.** 2004. New shuttle vectors for ectopic insertion of genes into Bacillus subtilis. Plasmid **51:**238-245.

26. **Molle, V., M. Fujita, S. T. Jensen, P. Eichenberger, J. E. Gonzalez-Pastor, J. S. Liu, and R. Losick.** 2003. The Spo0A regulon of Bacillus subtilis. Mol Microbiol **50:**1683-1701.

27. **Nicholson, W. L., and P. Setlow.** 1990. Sporulation, Germination and Outgrowth. C.R. Harwood and S.M. Cutting (Eds.), Molecular Biological Methods for Bacillus. Wiley-Interscience Publication**:**391-450.

28. **Paul, B. J., M. M. Barker, W. Ross, D. A. Schneider, C. Webb, J. W. Foster, and R. L. Gourse.** 2004. DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. Cell **118:**311-322.

29. **Piggot, P. J., and J. G. Coote.** 1976. Genetic aspects of bacterial endospore formation. Bacteriol Rev **40:**908-962.

30. **Roels, S., A. Driks, and R. Losick.** 1992. Characterization of spoIVA, a sporulation gene involved in coat morphogenesis in Bacillus subtilis. J Bacteriol **174:**575-585.

31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning : a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

32. **Schmalisch, M., E. Maiques, L. Nikolov, A. H. Camp, B. Chevreux, A. Muffler, S. Rodriguez, J. Perkins, and R. Losick.** 2010. Small genes under sporulation control in the Bacillus subtilis genome. J Bacteriol **192:**5402-5412.

33. **Setlow, P.** 2003. Spore germination. Curr Opin Microbiol **6:**550-556.

34. **Setlow, P.** 2006. Spores of Bacillus subtilis: their resistance to and killing by radiation, heat and chemicals. J Appl Microbiol **101:**514-525.

35. **Silvaggi, J. M., D. L. Popham, A. Driks, P. Eichenberger, and R. Losick.** 2004. Unmasking novel sporulation genes in Bacillus subtilis. J Bacteriol **186:**8089-8095.

36. **Smith, K., M. E. Bayer, and P. Youngman.** 1993. Physical and functional characterization of the Bacillus subtilis spoIIM gene. J Bacteriol **175:**3607-3617.

37. **Stragier, P.** 2002. A gene odyssey: exploring the genomes of endospore-forming bacteria. A.L. Sonenshein, J.A. Hoch, R. Losick (Eds.), Bacillus subtilis and its Closest Relatives: From Genes to Cells, American Society for Microbiology, Press 519–526.

38. **Stragier, P., and R. Losick.** 1996. Molecular genetics of sporulation in Bacillus subtilis. Annu Rev Genet **30:**297-241.

39. **Tocheva, E. I., E. G. Matson, D. M. Morris, F. Moussavi, J. R. Leadbetter, and G. J. Jensen.** 2011. Peptidoglycan remodeling and conversion of an inner membrane into an outer membrane during sporulation. Cell **146:**799-812.

40. **Wang, S. T., B. Setlow, E. M. Conlon, J. L. Lyon, D. Imamura, T. Sato, P. Setlow, R. Losick, and P. Eichenberger.** 2006. The forespore line of gene expression in Bacillus subtilis. J Mol Biol **358:**16-37.

41. **Westers, H., E. Darmon, G. Zanen, J. W. Veening, O. P. Kuipers, S. Bron, W. J. Quax, and J. M. van Dijl.** 2004. The Bacillus secretion stress response is an indicator for alpha-amylase production levels. Lett Appl Microbiol **39:**65-73.

42. **Wilson, G. A., and K. F. Bott.** 1968. Nutritional factors influencing the development of competence in the Bacillus subtilis transformation system. J Bacteriol **95:**1439-1449.

43. **Winkelman, J. T., K. M. Blair, and D. B. Kearns.** 2009. RemA (YlzA) and RemB (YaaB) regulate extracellular matrix operon expression and biofilm formation in Bacillus subtilis. J Bacteriol **191:**3981-3991.

44. **Wu, M., Q. Ren, A. S. Durkin, S. C. Daugherty, L. M. Brinkac, R. J. Dodson, R. Madupu, S. A. Sullivan, J. F. Kolonay, D. H. Haft, W. C. Nelson, L. J. Tallon, K. M. Jones, L. E. Ulrich, J. M. Gonzalez, I. B. Zhulin, F. T. Robb, and J. A. Eisen.** 2005. Life in hot carbon monoxide: the complete genome sequence of Carboxydothermus hydrogenoformans Z-2901. PLoS Genet **1:**e65.

45. **Youngman, P., J. B. Perkins, and R. Losick.** 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.

Figure legends

Figure 1. Transcription and regulatory gene-dependence of candidate sporulation genes. Strains carrying *lacZ* transcriptional reporter constructs for *bkd* (operon including *buk*; A), *ymxH* (B), *ylxY* (C), *ylmC* (D), *ymfB* (E), *yteA* (F) and *ylyA* (G) were induced to sporulate and β-galactosidase activity was monitored for samples taken at indicated time points after sporulation-induction. Activity (AU) was assayed in wild type (□) or strains mutant for *spo0A* (×), *sigF* (◊), *sigE* (○), *sigG* (Δ), *spoVT* (●), *bkdR* (+) or *spo0A* and *bkdR* (∗).

Figure 2. A *ylxY* mutant exhibits a competition deficit under sporulation-inducing conditions. A wild type reference strain carrying an IPTG-inducible *lacZ* gene (RL5360) was competed in DS medium against a *ylxY* mutant, starting with approximately 20% wild type strain. Cultures were incubated at 37˚C for 24 hours, heat-treated at 80˚C, and diluted in fresh DS medium. After indicated rounds of competition, dilutions of the culture were plated on agar plates containing IPTG and X-gal, and blue (wild type) and white (mutant) colonies were counted. Bar graphs indicate the percentage of the wild type reference strain. The wild type strain was competed against a *ylxY* mutant (dark grey), and a complemented strain carrying a copy of *ylxY* at the ectopic *amyE* locus (light grey).

Figure 3. A *ylyA* mutant exhibits a competition deficit under sporulation-inducing conditions. A wild type reference strain carrying an IPTG-inducible *lacZ* gene (RL5360) was

competed in DS medium against a *ylyA* mutant, starting with approximately 20% wild type strain. Cultures were incubated at 37˚C for 24 hours, heat-treated at 80˚C, and diluted in fresh DS medium. After indicated rounds of competition, dilutions of the culture were plated on agar plates containing IPTG and X-gal, and blue (wild type) and white (mutant) colonies were counted. Bar graphs indicate the percentage of the wild type reference strain. The wild type strain was competed against a *ylyA* mutant (dark grey), and a complemented strain carrying a copy of *ylyA* at the ectopic *sacA* locus (light grey).

Figure 4. A *ylmC* mutant exhibits a slight competitive advantage under sporulation-inducing conditions. A wild type reference strain carrying an IPTG-inducible *lacZ* gene (RL5360) was competed in DS medium against a *ylmC* mutant, starting with approximately 75% wild type strain. Cultures were incubated at 37˚C for 24 hours, heat-treated at 80˚C, and diluted in fresh DS medium. After indicated rounds of competition, dilutions of the culture were plated on agar plates containing IPTG and X-gal, and blue (wild type) and white (mutant) colonies were counted. Bar graphs indicate the percentage of the wild type reference strain. The wild type strain was competed against a *ylmC* mutant (dark grey), and a complemented strain carrying a copy of *ylmC* at the ectopic *amyE* locus (light grey).

Figure 5. *ylxY*, *ylyA* and *ylmC* mutants do not exhibit a significant competitive deficit during growth. A wild type reference strain carrying an IPTG-inducible *lacZ* gene (RL5360) was competed in LB medium against an unmarked wild type (dark grey), and strains mutant for *ylxY* (light grey), *ylyA* (black), and *ylmC* (white), starting with approximately 20% wild type

strain. Cultures were incubated at 25˚C for 24 hours, and diluted in fresh LB. After indicated rounds of competition, dilutions of the culture were plated on agar plates containing IPTG and X-gal, and blue and white colonies were counted. Bar graphs indicate the percentage of the wild type reference strain.

Figure 6. A *ylyA* mutant is delayed in germination. Purified spores were heat-treated and cooled on ice. Activated spores were germinated at an OD_{600} of 0.5 by dilution in LB medium. Every 4 minutes after germination-induction the OD_{600} was measured, and germination is reported as the percentage in the drop in optical density.

548 Table 1. Strains used in this study

Strain[a]	Genotype	Source or reference
PY79	Prototrophic derivative of *B. subtilis subsp. subtilis* 168	(45)
RL5360	*amyE:*$P_{hyperspank}$*-lacZ spec*	This study
RL5361	*bkdR::spec*	This study
RL5362	*bkdR::erm*	This study
RL5363	*buk::erm*	This study
RL5364	*ylmC::erm*	This study
RL5365	*ylmC::erm amyE::ylmC spec*	This study
RL5366	*ymxH::spec*	This study
RL5367	*ylxY::spec*	This study
RL5368	*ylxY::spec amyE::ylxY cat*	This study
RL5369	*ymfB::spec*	This study
RL5370	*yteA::erm*	This study
RL5371	*ylyA::erm*	This study
RL5372	*ylyA::erm sacA::ylyA kan*	This study
RL5373	*yocK::tet*	This study
RL5374	*amyE::*P_{bkdR}*-lacZ cam*	This study
RL5375	*amyE::*P_{bkd}*-lacZ cam*	This study
RL5376	*spo0A::spec amyE::*P_{bkd}*-lacZ cam*	This study
RL5377	*bkdR::erm amyE::*P_{bkd}*-lacZ cam*	This study
RL5378	*spo0A::spec bkdR::erm amyE::*P_{bkd}*-lacZ cam*	This study
RL5379	*amyE::*P_{ylmC}*-lacZ cam*	This study
RL5380	*sigF::kan amyE::*P_{ylmC}*-lacZ cam*	This study
RL5381	*sigE::erm amyE::*P_{ylmC}*-lacZ cam*	This study
RL5382	*sigG::kan amyE::*P_{ylmC}*-lacZ cam*	This study
RL5383	*spo0A::spec amyE::*P_{ymxH}*-lacZ cam*	This study
RL5384	*sigF::kan amyE::*P_{ymxH}*-lacZ cam*	This study
RL5385	*amyE::*P_{ylxY}*-lacZ cam*	This study

RL5386	*sigF::kan amyE::*P_{ylxY}*-lacZ cam*	This study
RL5387	*sigE::erm amyE::*P_{ylxY}*-lacZ cam*	This study
RL5388	*sigG::kan amyE::*P_{ylxY}*-lacZ cam*	This study
RL5389	*amyE::*P_{ylzJ}*-lacZ cam*	This study
RL5390	*amyE::*P_{ymfB}*-lacZ cam*	This study
RL5391	*sigF::kan amyE::*P_{ymfB}*-lacZ cam*	This study
RL5392	*sigG::kan amyE::*P_{ymfB}*-lacZ cam*	This study
RL5393	*spoVT::spec amyE::*P_{ymfB}*-lacZ cam*	This study
RL5394	*amyE::*P_{yteA}*-lacZ cam*	This study
RL5395	*sigG::kan amyE::*P_{yteA}*-lacZ cam*	This study
RL5396	*spoVT::spec amyE::*P_{yteA}*-lacZ cam*	This study
RL5397	*amyE::*P_{ylyA}*-lacZ cam*	This study
RL5398	*sigG::kan amyE::*P_{ylyA}*-lacZ cam*	This study
RL5399	*spoVT::spec amyE::*P_{ylyA}*-lacZ cam*	This study
RL2242	*spo0A::spec*	(12)
RL1265	*sigF::kan*	(5)
RL1061	*sigE::erm*	(22)
RL4962	*sigG::kan*	(7)
RL3873	*spoVT::spec*	(40)

549 [a] All strains are isogenic with PY79 unless otherwise marked.

550 Table 2. Oligonucleotides used to construct reporter and complementation constructs

Primer[a]	Sequence (5'-3')
bkdR-296E	ctggaattcgatgaatcctgacaacccttg
bkdR+3H	ctgaagcttcatcccgatacccctttgtat
bkd-298E	ctggaattcgaaggcgaaaagctgtctgt
bkd+3H	ctgaagcttcatctgttaccacctttcttg
ylmC-300E	ctggaattcaagtgaaacgggagtgtcca
ylmC+3H	ctgaagcttcattccatcacgtccttttc
ylmC+362B	ctgaggatcccttattttaccacatcttactg
ymxH-282E	ctgagaattcaatgctgagcttagaaagcac
ymxH+3H	ctgaaagcttcatgtctgtcacccccttg
ylxY-330E	ctggaattcttcggggctttcgttgaaatt
ylxY+3E	ctggaattccatgttctgtcccccctcac
ylxY+1076B	ctgaggatccatcgcaacagaacggactgtc
ylzA-380E	ctggaattcaatcaaagaatggactgaagacg
ylzA+3E	ctggaattccatcttctacgttccccctgt
ymfB-234E	ctggaattcaaacatcaaatgtcgaatggtc
ymfB+3H	ctgaagcttcataatgctgtccttcgcatc
yteA-351E	ctggaattcttggctttatgtaatgcatgtag
yteA+3E	ctggaattccattgtgatcgcctcgtttct
ylyA-613E	ctggaattcgtggtctcatttaacatttgttg
ylyA+3E	ctggaattccattcttcacaactcctgctc
ylyA+514B	ctgggatccctgcaataataagtagtgcaatc

551 [a]numbers refer to the 5' nucleotide position relative to the first nucleotide of the start codon (+1)

552 of the respective gene. B=*Bam*HI; E=*Eco*RI; H=*Hin*dIII.

553 Table 3. Candidates for uncharacterized sporulation genes among genes conserved in endospore-
554 forming bacteria.

gene	Number of orthologs in		predicted product
	all bacteria/ archaea[a]	*Bacillaceae*/ *Clostridia*[b]	
bkdR	87	43	DNA-binding transcriptional regulator
buk	62	34	butyrate kinase
ylmC	45	45	hypothetical protein; PRC-barrel domain
ymxH	25	25	hypothetical protein; PRC-barrel domain
ylxY	42	38	polysaccharide deacetylase
ylzA	73	43	hypothetical protein
ymfB	44	44	ClpP-like protease
yteA	51 (33)[c]	41 (23)[c]	DksA-like regulator
ylyA	11 (29)[c]	11 (29)[c]	DksA-like regulator

555 [a]626 bacterial and archaeal genomes were considered

556 [b]46 *Bacillus* and *Clostridia* genomes were considered

557 [c]Numbers between brackets are number of orthologs for *yteA* and *ylyA* corrected for similarity in
558 gene synteny with *B. subtilis ylyA*