Small Genes under Sporulation Control in the *Bacillus subtilis* genome^{∇†}

Matthias Schmalisch,¹ Elisa Maiques,¹ Lachezar Nikolov,¹ Amy H. Camp,¹ Bastien Chevreux,²
Andrea Muffler,² Sabrina Rodriguez,³ John Perkins,³ and Richard Losick^{1*}

The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138¹; Research and Development, DSM Nutritional Products, Ltd., Kaiseraugst, Switzerland²; and DBC, DFS Food Specialties B.V., Delft, Netherlands³

Received 10 May 2010/Accepted 28 July 2010

Using an oligonucleotide microarray, we searched for previously unrecognized transcription units in intergenic regions in the genome of *Bacillus subtilis*, with an emphasis on identifying small genes activated during spore formation. Nineteen transcription units were identified, 11 of which were shown to depend on one or more sporulation-regulatory proteins for their expression. A high proportion of the transcription units contained small, functional open reading frames (ORFs). One such newly identified ORF is a member of a family of six structurally similar genes that are transcribed under the control of sporulation transcription factor σ^E or σ^K . A multiple mutant lacking all six genes was found to sporulate with slightly higher efficiency than the wild type, suggesting that under standard laboratory conditions the expression of these genes imposes a small cost on the production of heat-resistant spores. Finally, three of the transcription units specified small, noncoding RNAs; one of these was under the control of the sporulation transcription factor σ^E , and another was under the control of the motility sigma factor σ^D .

Bacillus subtilis, also known as the hay bacillus and commonly found in soil, is the best-characterized member of the Gram-positive group of bacteria. It is highly accessible for genetic manipulation and has therefore served as a model organism for laboratory studies, especially for cell differentiation processes (for reviews see references 1 and 29). With the publication of its genome sequence by Kunst et al. in 1997 (18), it also became the first Gram-positive bacterium with a complete genome sequence. At the time of publication, 4,100 protein-coding genes were annotated in a total of 4,214,810 bp. These protein-coding genes constitute 87% of the genome, a number which has been found to be very similar to the composition of other bacterial genomes such as the genome of *Escherichia coli* (4). Although the ratio of noncoding DNA to coding DNA is very consistent in bacterial genomes, ranging from 6 to 14% of noncoding DNA for the majority of sequenced genomes (28), it has become apparent in recent years that the percentage of coding RNAs is actually higher than previously anticipated.

In the last decade a wealth of additional genes has been discovered in previously annotated genomes that play important roles in a range of cellular processes. Most of these genes encode small regulatory RNAs (sRNAs or ncRNAs, for noncoding RNAs) that were overlooked in the annotation due to their lack of an open reading frame (ORF). Since screening for small RNAs in *E. coli* began in 2001, more than 120 sRNAs have been discovered in *E. coli* alone by several methods, including high-density oligoarrays, cDNA cloning of small RNAs, or computational analyses (for reviews on small RNAs,

see references 27, 13, and 32). While the majority of the discovered sRNAs are still of unknown function, some of them have been extensively characterized. As early as 1997, it was discovered that OxyS, a small noncoding RNA that is produced in *E. coli* in response to oxidative stress (2), regulates several target mRNAs by influencing mRNA stability and translation (for a microreview on OxyS and its target *rpoS*, see reference 26).

Small RNAs for a number of other bacteria have been discovered, including both Gram-positive as well as Gram-negative species. Examples are *Staphylococcus aureus* (16), *Pseudomonas aeruginosa* (45) and *Clostridium acetobutylicum* (5). In addition to experimentally verified small RNAs, several hundred RNAs are predicted by computational methods. Using a novel prediction algorithm, Tran et al. (41) reported the finding of 601 potential small RNAs in *E. coli*.

In addition to small RNAs, the focus of recent microbial genome studies has also been on small ORFs in intergenic regions. The identification or discovery of small ORFs (for example, those of less than 50 codons) can be challenging. They often cannot be reliably annotated, and their products are difficult to detect with standard biochemical methods, such as two-dimensional gel electrophoresis. Applying new computational models for sequence comparison and ribosome binding to the genome of *E. coli*, Hemm et al. (15) discovered 18 additional small ORFs and confirmed the synthesis of 20 previously annotated proteins smaller than 50 amino acids. Previously nonannotated ORFs have also been reported in *Pseudomonas fluorescens*, where 16 new ORFs were found, with 9 of them being smaller than 100 amino acids (17).

In *B. subtilis*, several new small RNAs have been discovered in recent years, including RatA, an antisense RNA for a peptide toxin (35), FsrA, a small RNA regulated by the iron-response regulator Fur (12), SR1, a small RNA controlled by the gluconeogenesis regulator CcpN (20), and several novel small RNAs under sporulation control (36).

In addition to these noncoding RNAs, additional small

* Corresponding author. Mailing address: The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138. Phone: (617) 495-4905. Fax: (617) 496-4642. E-mail: Losick@mcb.harvard.edu.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

[∇] Published ahead of print on 13 August 2010.

ORFs in previously “empty” intergenic regions have been discovered. Perhaps the first to be identified and still one of the most striking examples is the 26-codon-long *spoVM* gene, which plays an essential role in spore formation (8). Gaballa et al. (12) characterized three small ORFs, of which two were previously not annotated: *fbpB* (intergenic region between *fbpA* and *ydbM*) and *fbpC* (between *ypbQ* and *ypbR*). Another example is *mciZ*, a small 40-codon ORF that was discovered by Handler et al. (14) in the intergenic region of *nudF* and *yqkF* and that has been shown to interact with the cell division protein FtsZ. In addition to these findings, genome-wide transcriptional studies have led to the discovery of previously unknown genes. Lee et al. (19) used an antisense microarray that covered not only known ORFs but also some intergenic regions. RNA expression analysis suggested 35 independent transcripts in intergenic regions, with 20 of them containing putative ORFs. Finally, the most recent findings come from a genome-wide transcriptional study done by Rasmussen et al. (25). This group used a microarray to investigate all transcriptionally active regions in *B. subtilis*. Observed transcriptional activity in intergenic regions suggests 84 putative small RNAs. To date, none of these has been confirmed by other methods.

The focus of this investigation was a transcriptional analysis of the intergenic regions of *B. subtilis*, with a special emphasis on unknown small genes that are turned on during sporulation. By carrying out analysis on a genome-wide basis, we sought to extend earlier work in which several genes for small RNAs that are transcribed in response to sporulation (36) were discovered. For this purpose, we used a densely tiled microarray that covered all intergenic regions larger than 50 nucleotides. As we report, 19 transcript units were identified, 11 of which proved to be under sporulation control. Utilizing optimized bacterial luciferase reporters and other assays to characterize the newly identified transcription units, we found that a high proportion (16) contained one or more functional mRNAs for novel small proteins. Three of the transcription units are small, noncoding RNAs, one of which is under sporulation control and one of which is under the control of the motility transcription factor σ^P .

MATERIALS AND METHODS

Growth conditions. Strains were grown at 37°C in Luria-Bertani (LB) or Difco sporulation (DS) medium (23, 30). For the fluorescence microscopy experiments, sporulation was induced by resuspension as described earlier (38). Ampicillin (100 µg/ml), chloramphenicol (5 µg/ml), tetracycline (10 µg/ml), spectinomycin (100 µg/ml), and macrolide-lincosamide-streptogramin (MLS) antibiotics (1 µg/ml erythromycin and 25 µg/ml lincomycin) were added to the growth medium or agar plates where appropriate.

Plasmid and strain construction. The *B. subtilis* strains used in this study were PY79 (SP β^+ prototroph) and isogenic derivatives thereof. All strain constructions were done using standard laboratory techniques. For cloning purposes, *E. coli* DH5 α strains were used, and plasmids were transferred to *E. coli* using standard laboratory techniques. PCR products were amplified from chromosomal DNA derived from the *B. subtilis* strain PY79 unless otherwise noted. Primer sequences used for the PCRs are given in Table S1 in the supplemental material. Plasmids, chromosomal DNA, or products from PCRs of long flanking homologous regions were transferred into *B. subtilis* by a simple one-step transformation protocol as described by Wilson and Bott (46).

To generate a plasmid harboring an optimized bacterial luciferase reporter, a 5.7-kb PvuII/EcoRI DNA fragment containing a derivative of the bacterial luciferase operon (*luxABCDE*) from *Photobacterium luminescens* was subcloned from pSB2025 (24) into pSac-Cm (directs integration at the *sacA* locus [21]), which had been digested with BamHI, blunted with Klenow, and subsequently

digested with EcoRI. Importantly, the ribosome binding sites (RBSs) upstream of *luxA*, *luxC*, and *luxE* in pSB2025 were modified by Qazi et al. (24) to more closely resemble consensus *B. subtilis* RBSs. Next, the resulting plasmid was digested with EcoRI/Sall and ligated with an EcoRI/Sall-digested PCR product containing the sporulation-dependent *P_{spoB}* promoter (amplified using primers AH60 and AH61). Finally, the RBSs upstream of *luxB* and *luxD* were modified to be more similar to optimal *B. subtilis* RBSs (43). The original *luxB* RBS was switched to TAAGGAGGTAAAGAAATG (core RBS in bold, inserted nucleotides in italics, and start codon of *luxB* underlined) using primers AH272 and AH273, and the original *luxD* RBS was switched to AATGGAGGTAAAGTATG (core RBS in bold, inserted nucleotides in italics, and start codon of *luxD* underlined) using primers AH275 and AH276. The resulting luciferase reporter plasmid (*sacA::P_{spoB}-luxABCDE cat*) harboring optimized RBSs upstream of all five *lux* cistrons was designated pAH321. We found that sporulation-specific expression of *luxABCDE* from this fully optimized construct in *B. subtilis* was significantly improved (~8-fold) compared to the original, partially optimized *luxABCDE* present in pSB2025 (data not shown). Of note, we found that the fully optimized *luxABCDE* operon was more robustly expressed when induced during vegetative growth under the control of the *P_{hyperspank}* promoter than the original partially optimized operon (up to a 30-fold improvement) (9; data not shown).

To construct transcriptional fusions of the intergenic regions in *B. subtilis* to the optimized luciferase reporter, PCR products (see Table S2 in supplemental materials) were digested with EcoRI and Sall and ligated to the EcoRI/Sall-digested pAH321 backbone (lacking the *P_{spoB}* promoter). For translational luciferase fusions, digested PCR products were cloned into a modified version of pAH321, designated pMS25, which lacks the *luxA* RBS and start codon. To generate pMS25, the *luxA* gene (starting from the second codon) was amplified from pAH321 using primers MS@L67 and MS@L68, which added in-frame EcoRI and Sall restriction sites to the 5' end of *luxA*. The resulting PCR product was digested with EcoRI and AgeI and ligated into the EcoRI/AgeI-digested pAH321 backbone. Importantly, transcription of the *luxABCDE* operon and translation of *luxA* from pMS25 require cloning a DNA fragment bearing a promoter, an RBS, and a start codon in frame with the *luxA* gene.

Deletion/insertion mutants were created by the technique of long flanking homology PCR as described previously (44) with the following sets of four primers: MS@L103 to MS@L106 for intergenic region (IGR) *yocL-yocM*, MS@L107 to MS@L110 for IGR *yocG-yocH*, MS@L111 to MS@L114 for IGR *ydcl-ydcK*, MS@L126 to MS@L129 for IGR *yocK-yocN*, MS@L130/MS@L131 and MS@L128/MS@L129 for *yozN-yocN*, MS@L134 to MS@L135 for IGR *ywqC-ywqB*, MS@L138 to MS@L141 for IGR *yzdH-ydJ*, MS@L142 to MS@L145 for IGR *yznH-thyA*, MS@L146 to MS@L149 for IGR *ypcP-ypbS*, MS@L180 to MS@L183 for IGR *ydiM-ydiN*, MS@L205 to MS@L208 for IGR *ygeN-comEC*, MS@L213 to MS@L216 for IGR *yrhK-yrhJ*, MS@L243 to MS@L246 for *ypzD*, MS@L259 to MS@L262 for *yzdH*, MS@L267 to MS@L270 for *ydgAB*. For the deletion of the region covering IGR *yocLM* to *yozN-yocN*, primers MS@L107/MS@L108 were used for the upstream region and MS@L128/MS@L129 for the downstream region. Multiple mutants were generated accordingly using the appropriate single mutant strain as a source. The sources of antibiotic resistance genes used in creating the deletion/insertion mutations were pDG780 (*kan*), pAH52 (*erm*), and pDG1515 (*tet*) (9, 10). Chromosomal DNA was isolated from the mutant strains, and insertions were verified by PCR analysis.

The strains expressing green fluorescent protein (GFP) fused in frame to the coding sequence of known and newly identified ORFs were obtained by Campbell-like recombination of plasmids based on pCVO119 (7). PCR products with an average length of 300 to 350 bp for the different genes were obtained by amplification with the following primers: MS@L209/MS@L210 for IGR *ygeN-comEC*, MS@L219/MS@L220 for IGR *yjbS-cofF*, MS@L221/MS@L222 for IGR *yocL-yocM*, MS@L223/MS@L224 for IGR *ywqC-ywqB*, MS@L225/MS@L226 for IGR *ydcl-ydcK*, MS@L227/MS@L228 for IGR *yznH-thyA*, MS@L229/MS@L230 for IGR *yzdH-ydJ*, MS@L233/MS@L234 for IGR *ypcP-ypbS*, MS@L235/MS@L236 for IGR *yozI-rapK*, MS@L237/MS@L238 for IGR *yrkD-yrkC*, MS@L241/MS@L242 for IGR *yrhI-ytqI*, MS@L247/MS@L248 for *ypzD*, MS@L263/MS@L264 for *yzdH*, and MS@L301/MS@L302 for *ygdA*. The PCR products and the plasmid pCVO119 were digested with BamHI and Sall, and the fragments were ligated. The resulting plasmids were verified by PCR analysis and transferred to the *B. subtilis* wild-type strain PY79.

Luciferase assay. To monitor expression of luciferase during cell growth and sporulation, *B. subtilis* cells harboring *luxABCDE* reporter fusions were grown in LB medium to stationary phase, back-diluted 100-fold in water, and spotted in 10-µl volumes in triplicate onto 200 µl of LB or DS agar pads in white, clear-bottom 96-well plates. Plates were incubated at 37°C in a Synergy 2 plate reader (BioTek) and bioluminescence from each well was measured in real time every 15 min. Cell growth was simultaneously monitored by recording absorbance at

600 nm. Bioluminescence is reported in arbitrary units. Of note, we found that cell growth/sporulation on agar pads, as opposed to in liquid medium, resulted in more robust and reproducible luciferase measurements.

Heat kill assay. To quantify spore formation, cells were induced to sporulate by nutrient exhaustion in Difco sporulation medium (DSM) (30). A single colony was picked from a fresh agar plate and used to inoculate 5 ml of DSM and allowed to grow at 37°C for 24 h. The frequency of sporulation was calculated from the number of CFU before and after heat treatment (20 min at 80°C), which was determined by plating various dilutions on agar plates. For each experiment, wild-type *B. subtilis* PY79 was used as a control, and for comparison the number of spores obtained with the wild type (1×10^8 to 5×10^8 per ml) was set as 100%.

Fluorescence microscopy. For the localization of the sporulation-dependent GFP fusion proteins, cultures of the different strains were grown overnight in LB medium and diluted in medium used for resuspension. At mid-exponential phase, sporulation was induced by resuspension in SM medium (38). Cells were harvested at different times after induction of sporulation by centrifugation and resuspended in phosphate-buffered saline (PBS) buffer. Where necessary, membranes were visualized by adding the membrane dye FM4-64 [*N*-(3-triethylammoniumpropyl)-4-(*p*-diethylaminophenyl)-hexatrienyl]pyridinium dibromide; green fluorescent (Invitrogen) or TMA-DPH (1-[4-(trimethylamino)phenyl]-6-phenyl-1,3,5-hexatriene; blue fluorescent) (Molecular Probes) to the PBS buffer (final concentration, 1 μ g/ml). For the fluorescence microscopy, cells were immobilized on a 1% agarose pad covered with a polylysine-treated coverslip. The same equipment and methods were used as described in previous work by this lab (see reference 11 for details).

β -Galactosidase assay. For measurements of gene activity with *lacZ* reporter genes, cells were grown in LB medium or induced to sporulate by the resuspension method (23, 38). Aliquots of cells were collected at intervals, stored at -20°C, and processed for β -galactosidase activity essentially as previously described (6).

Competition experiments. To identify a sporulation phenotype of mutant strains lacking one or more proteins of the *yozN* homologs, including the newly identified gene in the *yocL-yocM* intergenic regions, we challenged these strains against the parental wild-type strain in a competition experiment. To allow easy identification of the strains on an agar plate, an isopropyl- β -D-thiogalactopyranoside (IPTG)-inducible β -galactosidase gene was utilized. For each experiment, this gene was present in either the wild-type or the mutant strain. At the beginning of the experiment, optical densities (OD) of cell cultures of the wild type and of the mutant strain grown overnight in LB medium were measured. Equal amounts (1 OD unit) of both cell cultures were mixed and diluted 1:1,000 in 10 ml of DSM. After 24 h of growth at 37°C, the cell culture was heat treated (for details, see "Heat kill assay" above) and diluted 1:1,000 in 10 ml of fresh DSM. These steps were repeated at least five times, and dilutions of the cell cultures were plated on LB agar plates containing 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal; 40 μ g/ml). Numbers of blue and white colonies were counted, reflecting the ratio of the wild type to the mutant strain. A similar protocol was followed for growth in LB medium with the exception that no heat treatment was applied. As a control to show that no strain had a general growth disadvantage in DSM, an experiment was set up in which the cell culture was continuously transferred into fresh DSM before cells initiated sporulation. In this experiment, the cell culture was diluted 1:100,000 in fresh medium, and the optical cell density was measured to ensure that the cells were still in logarithmic growth phase (OD of <0.8) prior to inoculation of new DSM.

Spore purification and germination. Spores were prepared essentially as described in *Molecular Biological Methods for Bacillus* (section 9.8.1) (23), with a few modifications. In brief, sporulation was induced by growth in DSM for 72 h at 37°C. Spores and other cells/cellular debris were collected by centrifugation and washed twice with ice-cold distilled water. The pellet was then resuspended in 10 ml of ice-cold distilled water and stored overnight at 4°C. The next day, the solution was pelleted by centrifugation and again washed and resuspended in ice-cold distilled water. Samples were checked under the microscope for phase-bright spores. These steps were repeated (usually after 5 to 6 days) until no vegetative cells were seen under the microscope and 90% of the spores were phase bright.

Germination of the pure spore preparations was initiated by the addition of L-alanine at a final concentration of 10 mM. Germination was monitored by the loss of optical density.

Computational target identification. To identify potential targets for the putative small noncoding RNA in the intergenic region of *yrhJ-yrhK*, we used the online available program TargetRNA in the advanced setting (<http://snowwhite.wellesley.edu/targetRNA/advanced.html>) (40). Since no determination of the 5' or 3' end of the *yrhJ-yrhK* RNA was done in the previous study, we used the

observed hybridization region as obtained by the microarray analysis. The sequence used for the target search was 5'-GCGCGAUACUCCCUAAUAC AUCCUUUCAGUAGAUUUCAUUAUCGUGCCGGUUUUCUCAUA UGCAAAGCAAUCCCGCCAAUCAGAUUGGCGGGGAUU. The key parameters for the search were as follows: genome, *B. subtilis*; search around, start codon; region, -30 to +20; *P* value threshold, 0.01.

Microarray design. Using the *B. subtilis* sequence information from the NC_000964.2 GenBank file (10 November 2004 release), intergenic regions with a length at least equal to 50 nucleotides between the stop and the start codon of the neighboring annotated genes were selected. The regions selected did not include coding sequence (CDS), tRNA, rRNA, or misc-RNA (RNA of unknown function) features. They did include small cytoplasmic RNAs (scRNAs). Parsing of the GenBank file and sequence generation were achieved by using Perl scripts. Both strands of the selected intergenic regions (two strands of 470 kb each, for a total of 940 kb) were represented on an Affymetrix array, DSMbsrnab530219 (NimbleExpress format, 282,000 features, and 17- μ m feature size), allowing a tiling step of six bases. In addition, spike-in controls of *Arabidopsis thaliana* genes and *B. subtilis* genes whose expression patterns were known were displayed on the array (see supplemental material for details).

RNA isolation. Cell cultures for RNA isolation were harvested by a brief centrifugation step (5 min at 8,500 rpm at room temperature), and the pellet was immediately frozen in liquid nitrogen and kept at -80°C until further processing. The pellet was resuspended in 1 ml of lysis buffer containing 500 μ l of phenol-chloroform and 100 μ l of 10% sodium dodecyl sulfate. The solution was transferred into lysis tubes (MP Biomedicals) containing glass beads (lysis matrix B) and processed in a bead beater (Qbiogene) two times for 45 s at 6,000 rpm. The lysis tubes were centrifuged at 13,000 rpm at 4°C, and supernatant without cell debris or the matrix was transferred into a new microcentrifuge tube. After two phenol-chloroform steps with 1 ml, RNA was precipitated by addition of an equal volume of ethanol and incubation at -20°C overnight. RNA was pelleted by centrifugation (30 min at 13,000 rpm at 4°C) and washed twice with 1 ml of 70% ethanol. The pellet was dried and resuspended in 100 μ l of diethyl pyrocarbonate (DEPC)-water. RNA was treated with DNase I to remove residual chromosomal DNA and purified with either a Qiagen RNeasy minikit or Ambion mirVana microRNA ([miRNA] for small-RNA enrichment) according to the manufacturer's recommendations. Quality of the RNA preparation was verified by using an Agilent Bioanalyzer.

cDNA labeling, hybridization, and microarray scanning. Preparation of cDNA targets and hybridization were based on methods described previously (19). Washing and staining with GeneChip were performed according to Affymetrix's standard protocol. Microarrays were processed with an Affymetrix scanner and analyzed by using the Affymetrix gene expression analysis suite. Data were normalized according to the mean of the sum of all of the comparable experiments (robust multichip average [RMA] method). Expression data were visualized using the Affymetrix Integrated Genome Browser.

Homolog search and alignment. The ORFs in the intergenic regions were queried by using the protein sequence against the translated nucleotide database (tBlastN [<http://www.ncbi.nlm.nih.gov/BLAST/>]). The homologs for the ORF in the *yocL-yocM* intergenic regions were found by using protein-protein BLAST against the *B. subtilis* genome only. The protein alignment for *ypzD* and IGR *yzdH-ydfR* and the multiple alignments for the *yozN* homologs were done using the ClustalW2 algorithm and the standard parameters (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>).

RESULTS AND DISCUSSION

Identifying transcripts in intergenic regions. We developed an oligonucleotide microarray using Affymetrix GeneChip technology to search for transcripts in intergenic regions of the *B. subtilis* genome. The microarray was designed to include all intergenic regions with a length of at least 50 nucleotides between the stop and the start codons of the neighboring genes. Regions containing tRNA or rRNA genes were excluded from the microarray.

Hybridization experiments were carried out with RNA isolated from the prototrophic wild-type strain PY79. Cells were grown in sporulation medium and harvested at the mid-exponential phase of growth or at 2 or 5 h after the onset of sporulation. We enriched for small RNAs, as described in

TABLE 1. Novel genes and homologs

Gene orientation ^a	Left gene	Right gene	Size ^b	Homolog(s) ^c	Previous report(s)
←→←	<i>yxbC</i>	<i>yxbB</i>	28 aa	<i>B. licheniformis</i>	
←←→	<i>yybS</i>	<i>cotF</i>	49 aa	<i>B. licheniformis</i> , <i>B. amyloliquefaciens</i>	3, 25
→→←	<i>yrpD</i>	<i>yrpE</i>	25 aa	<i>G. thermodenitrificans</i> , <i>B. amyloliquefaciens</i> , <i>B. pumilis</i>	
←←←	<i>yrkD</i>	<i>yrkC</i>	79 aa	<i>G. kaustophilus</i> , <i>G. uraniumreducens</i> , <i>D. psychrophila</i>	3, 25
←→→→	<i>yoZJ</i>	<i>rapK</i>	90 aa, 56 aa	<i>B. amyloliquefaciens</i>	3, 25
→→←	<i>ytrI</i>	<i>ytqI</i>	63 aa	<i>B. pumilis</i> , <i>B. licheniformis</i> , <i>B. halodurans</i> , <i>O. iheyensis</i> , <i>E. sibiricum</i> , <i>H. modesticaldum</i>	3, 25
→→←	<i>purD</i>	<i>yezC</i>	44 aa	<i>B. anthracis</i> , <i>B. cereus</i> , <i>B. amyloliquefaciens</i> , <i>L. sphaericus</i>	
←←→	<i>ywqC</i>	<i>ywqB</i>	55 aa	<i>B. licheniformis</i> , <i>B. pumilis</i>	3, 25
←←←	<i>ypcP</i>	<i>ypbS</i>	48 aa	<i>B. amyloliquefaciens</i> , <i>B. licheniformis</i> , <i>G. kaustophilus</i>	3, 25
←→←	<i>yocL</i>	<i>yocM</i>	75 aa	<i>B. subtilis</i> , <i>B. amyloliquefaciens</i> , <i>B. pumilis</i>	3
→→→	<i>ydcI</i>	<i>yrcK</i>	37 aa	<i>B. licheniformis</i> , <i>B. amyloliquefaciens</i> , <i>B. anthracis</i> , <i>B. clausii</i> , <i>B. coagulans</i>	
←←←←	<i>ydZH</i>	<i>ydfR</i>	76 aa, 101 aa	<i>B. subtilis</i> , <i>B. amyloliquefaciens</i> , <i>B. licheniformis</i> , <i>G. kaustophilus</i> , <i>G. thermodenitrificans</i> , <i>B. pumilis</i>	3
←→→	<i>ynzH</i>	<i>thyA</i>	58 aa	—	3, 25
←→→	<i>yoaH</i>	<i>yoaI</i>	39 aa	<i>B. halodurans</i> , <i>G. kaustophilus</i> , <i>B. clausii</i> , <i>B. cereus</i>	
→→→	<i>yqeN</i>	<i>comEC</i>	44 aa	<i>B. amyloliquefaciens</i> , <i>B. pumilis</i> , <i>B. anthracis</i> , <i>B. weihenstephanensis</i> , <i>B. cereus</i> , <i>B. thuringiensis</i> , <i>B. megaterium</i> , <i>B. licheniformis</i> , <i>G. thermodenitrificans</i> , <i>G. kaustophilus</i> , <i>E. sibiricum</i> , <i>L. sphaericus</i> , <i>B. halodurans</i>	3, 25
→←←	<i>yjbG</i>	<i>yjbH</i>	55 aa	<i>B. amyloliquefaciens</i> , <i>B. licheniformis</i> , <i>G. kaustophilus</i> , <i>B. pumilis</i> , <i>G. thermodenitrificans</i> , <i>B. cereus</i> , <i>B. weihenstephanensis</i> , <i>B. anthracis</i>	3
→→←	<i>yrhK</i>	<i>yrhJ</i>	~100 nt	<i>B. amyloliquefaciens</i>	
→→→	<i>ykuI</i>	<i>ykuJ</i>	~120 nt	<i>B. pumilis</i> , <i>B. licheniformis</i>	12
→←←	<i>yocG</i>	<i>yocH</i>	~220 nt	—	

^a Outside arrows represent left and right known genes. Internal arrows indicate newly identified genes based on luciferase or GFP fusion studies; in two cases two new genes were identified.

^b Size of proteins predicted from ORFs in the IGR. In two cases more than one ORF was detected. In the last three examples no coding region was identified, but the size of the putative transcript is provided. aa, amino acid; nt, nucleotide.

^c Similarity determined at the DNA sequence level. In two cases paralogs to *B. subtilis* genes (bold) were detected. In two other cases, no similarities were found (—). The following are *Bacillus* species: *B. licheniformis*, *B. amyloliquefaciens*, *B. pumilis*, *B. halodurans*, *B. anthracis*, *B. cereus*, *B. clausii*, *B. coagulans*, *B. weihenstephanensis*, *B. thuringiensis*, *B. megaterium*. Other organisms are as follows: *D. psychrophila*, *Desulfotalea psychrophila*; *E. sibiricum*, *Exiguobacterium sibiricum*; *G. kaustophilus*, *Geobacillus kaustophilus*; *G. thermodenitrificans*, *Geobacillus thermodenitrificans*; *G. uraniumreducens*, *Geobacter uraniumreducens*; *H. modesticaldum*, *Heliobacterium modesticaldum*; *L. sphaericus*, *Lysinibacillus sphaericus*; *O. iheyensis*, *Oceanobacillus iheyensis*.

Materials and Methods, and then created fluorescently labeled cDNAs by reverse transcription. Finally, the fluorescent probes were annealed with the microarray (for details, see Materials and Methods). Analysis of the microarray data showed hybridization signals above background in most intergenic regions (>75%). We used three criteria in an initial effort to distinguish transcripts that arose from initiation within an intergenic region from the 3' or 5' untranslated regions of adjacent genes: (i) the minimum length of the transcript was 30 bp as judged by having five contiguous hybridization signals above background, (ii) the distance of the transcript from flanking genes was greater than 30 bp if the adjacent gene was transcribed from the same DNA strand, and (iii) no known or predicted transcription termination sequence was present in the intergenic region.

Thirty-four transcripts that met these criteria and that in addition appeared to be present after the onset of sporulation but absent in control experiments with RNA from cells isolated late in stationary phase in LB medium (data not shown) were chosen for further study. As a next step, we asked whether the chromosomal regions specifying these transcripts contained a promoter. To facilitate rapid screening of the candidate intergenic regions, we developed a luciferase-based reporter assay that allowed real-time, automated measurement of promoter activity in a 96-well format (see Materials and Methods for details). We fused 29 of the 34 regions to a promoterless copy

of the bacterial luciferase operon (*luxABCDE*) that was modified to harbor optimal translational signals for expression in *B. subtilis* (24; also the present study). Of the 29 fusions, 10 failed to show promoter activity under all conditions tested (for a list of the fusions with no activity, see Table S3 in the supplemental material). In contrast, the remaining 19 fusions exhibited luciferase activity in either sporulation medium (DS) and/or in complex medium (LB), indicating the apparent presence of promoters in these intergenic regions (see Fig. S1, gray solid lines, in the supplemental material). Table 1 lists the intergenic regions for which we identified an apparent promoter.

Recently, Rasmussen et al. (25) globally identified regions of the *B. subtilis* genome that are transcriptionally active during growth in minimal medium or LB medium. Our experiments were designed to detect genes in intergenic regions that are induced in DS sporulation medium. Nonetheless, 8 out of the 19 intergenic transcription units identified in our investigation were also detected in the study of Rasmussen et al. The newly identified transcripts were in these intergenic regions: *yybS-cotF*, *yrkD-yrkC*, *yoZJ-rapK*, *ytrI-ytqI*, *ywqC-ywqB*, *ypcP-ypbS*, *ynzH-thyA*, and *yqeN-comEC*.

Small ORFs. We looked for possible ORFs in the intergenic regions containing transcripts. If a putative ORF was found, we investigated whether the predicted amino acid sequence was homologous to amino acid sequences in the databases using tBlastN (NCBI). ORFs were found in 16 of the intergenic

yybS-cotF (<i>yyzH</i>)	TTGAGAACATGCACAAAGAGGTGAAACACA ATG --- TAA TTGTGATGGG	147 bp
yrpD-yrpE	GTACTATCCTCTTTT AAGGAAGGTGATGTTATG --- TAA GAACACACCT	75 bp
yrkD-yrkC (<i>yrzN</i>)	CTTGATATCCAAAATAGATAAATTAGAGGCT ATG --- TGAAAGAAAGGAT	237 bp
yoZJ-rapK (<i>yoZY</i>)	TTTAACCTCATAATTAATATATAAAATGTAA ATG --- TAA ACTTATAGGG	270 bp
ytrI-ytqI (<i>ytzJ</i>)	GCTTAATGGCGAAAGGAAGGAGGAAGGAA ATG --- TAA GAGAGAGGGT	189 bp
purD-yezC	AACCACCGCATCACGGGGATGTACAGCTAT GTG --- TGA TAAACAAAAA	132 bp
ywqC-ywqB (<i>ywzD</i>)	GCATCATATAATAATCGTGTCTTCAACTTA ATG --- TAA AAATCCCGCA	165 bp
ypcP-ypbS (<i>ypzF</i>)	ACTATATGCATCCGACGTAAGGGGGGATTC ATG --- TGA TAAATATTCA	144 bp
yocL-yocM (<i>yoyB</i>)	AGGTATATCCCATAGAAAGGAGAGGTTTTC ATG --- TAA CAATGGTTTT	225 bp
ydcI-ydcK	TTATGATAAAAGCTGAGGAGGAATGAGCCT ATG --- TAA GCAGTGCTTT	111 bp
ydzH-ydfR (<i>ydzR</i>)	ATAAACAGTCATAACAGAGAGGAGAGATGA ATG --- TAA AAATTAAGAA	228 bp
ydzH-ydfR	TTAAGAAATAAACACTTCATGAGGTGCCGT ATG --- TAG ATGACATATA	291 bp
ynzH-thyA (<i>ynzJ</i>)	AAATCAGAAAAAGCTATTGGAGGATTTAAT GTG --- TAA AAGAGGGGGA	174 bp
yoaH-yoaI	CGCAGCATAATATGAAGGAGCGAAACT TAAATG --- TAA TCAACAAAAA	117 bp
yqeN-comEC (<i>yqzM</i>)	TTATACTTAAGTCAGTCAGGAGGGATAAAG TTG --- TAA CAGCGACTGC	132 bp
yjbg-yjbH (<i>yizD</i>)	TTATAAGGAAAAAGGACGAGGAGGACCGCC ATG --- TAA AAAAAGCCTGT	165 bp

FIG. 1. Newly identified small ORFs. Predicted ORFs (identified using ORF finder at <http://www.ncbi.nlm.nih.gov>) in intergenic regions in which transcriptional activity was observed by microarray analysis and confirmed by luciferase and GFP reporter fusions. The start codon and stop codon of each ORF are underlined and in bold. Potential RBSs upstream of each ORF are underlined. In three cases (*ywqC-ywqB*, *yrkD-yrkC*, and *yoZJ-rapK*) regions of significant complementarity to the 3' end of 16S rRNA were not detected. Where an ORF was predicted by the latest reannotation of the *B. subtilis* genome sequence (3), the corresponding new gene name is given in parentheses.

regions, with 15 of the 16 ORFs having sequence similarity to ORFs in other species (Table 1). Two of these intergenic regions, *yoZJ-rapK* and *ydzH-ydfR*, each contained two putative ORFs. While both ORFs in the *ydzH-ydfR* intergenic region showed high similarity to known or predicted proteins in other species, only one of the two putative ORFs in the *yoZJ-rapK* region has a known homolog.

Thus, we identified a total of 18 putative protein-coding sequences. Figure 1 shows the beginning and end of the ORFs in their genetic context. For the two cases of intergenic regions containing tandem pairs of ORFs (*yoZJ-rapK* and *ydzH-ydfR*), no promoter activity was detected for the region between the first and second ORFs (data not shown). Therefore, in both cases the pairs of ORFs appear to constitute single transcriptional units (operons).

During the course of these experiments, Barbe et al. (3) reported a reannotation of the *B. subtilis* genome sequence, which included ORFs in intergenic regions where none had been previously annotated. As a result, 11 of the 18 putative ORFs identified by us are also present in the most current *B. subtilis* genome annotation (<http://genodb.pasteur.fr/cgi-bin/WebObjects/GenoList>). All of these ORFs are either homologs to genes of unknown function or do not have homologs to previously reported genes (3). Because the newly annotated ORFs were not experimentally confirmed, we simply refer to them according to the intergenic region in which they are located. However, to facilitate a cross-comparison, we have included in Fig. 1 the new annotation names for the 11 cases as they appear in the annotation of Barbe et al. (3).

To investigate whether the newly discovered ORFs were indeed functional and translated into proteins, we fused the

promoter regions including the first 8 to 10 codons of each ORF (only the first ORF in the case of the two tandem pairs of ORFs) in frame to a variant of the luciferase operon in which the RBS and start codon of the first gene (*luxA*) were absent. In this way, we created translational fusions for all but three (*yxbC-yxbB*, *yoZJ-rapK*, and *yrkD-yrkC*) of the 16 ORFs. Luciferase activity was observed for all 13 of these fusions (see Fig. S1, dotted gray lines, in the supplemental material), results that confirm that these ORFs are indeed translated. However, some of the translational luciferase fusions showed low activity relative to their corresponding transcriptional fusion (*ytrI-ytqI*, *purD-yezC*, and *ypcP-ypbS*) (see Fig. S1), whereas other translation fusions showed activity comparable to their corresponding transcriptional fusion (see Fig. S1). Because the *luxA* gene in the transcriptional fusions had an optimal RBS for *B. subtilis* and an ATG start, these findings seem to reflect variations in the strength of the translational signals for the ORFs. Putative RBSs and start and stop codons for each ORF are shown in Fig. 1.

In general, we observed weaker translational activity for ORFs beginning with GTG and TTG start codons than for those beginning with ATG. These findings are in agreement with previous results in *B. subtilis*, showing decreased translation initiation and mRNA stability associated with GTG and TTG start codons (33). An apparent exception was the *ynzH-thyA* translational fusion, which, despite its GTG start codon, showed almost the same level of activities for both the transcriptional and translational fusions (Fig. S1). However, in this case, the putative RBS (the sequence TTGGAGG located 7 bp upstream of the start codon) (Fig. 1) is close to the optimal for *B. subtilis* (22).

FIG. 2. Cell-specific expression of newly identified genes under sporulation control. Fluorescent micrographs of cells expressing *gfp* fused in frame to the 3' terminus of the indicated genes. Sporulation was induced by resuspension in SM medium. Cells were stained with the membrane dye TMA-DPH (Molecular Probes). Shown is a merged image of the GFP (green) and the membrane dye (blue). The *ywqC-ywqB* gene (IGR CB-GFP), which is under Spo0A control, was expressed in the predivisional sporangium (visualized at hour 1); the *ytrI-ytqI* gene (IGR II-GFP), which is under σ^E control, and *ynzH-thyA* (IGR HA-GFP), which is under σ^K control, were expressed in the mother cell (visualized at hours 3 and 5, respectively); and *ypcP-ypbS* (IGR PS-GFP), which is under σ^G control, was expressed in the forespore (visualized at hour 3).

Finally, and of note, there is one exceptional case. The ORF in the *ywqC-ywqB* intergenic region exhibited no apparent RBS upstream of the start codon (Fig. 1); however, the translational fusion showed high expression levels (see Fig. S1 in the supplemental material). No other potential start codons or RBS could be identified. Expression of *ywqC-ywqB* in the absence of an apparent RBS was confirmed with an in-frame fusion to the gene for the green fluorescent protein (Fig. 2; IGR CB-GFP). Also setting a boundary on the region in which translation could commence was the cloned segment of DNA that was used for the promoter fusion. This DNA segment contained a consensus match for a σ^A -dependent promoter located 18 bp upstream of the start codon, reinforcing the conclusion that the putative ATG start codon is, indeed, the translation start point.

Sporulation-dependent expression. As indicated above, the intergenic regions that were chosen for further study specified transcripts that were preferentially produced during sporulation and, hence, candidates for genes under sporulation control. To further investigate whether production of these transcripts was indeed under sporulation control, the transcriptional fusions were tested for luciferase activity in a series of mutants for various sporulation-regulatory proteins. Entry into sporulation is governed by the response regulator Spo0A and the alternative sigma factor σ^H (37). Spo0A is active in the predivisional sporangium but later, after the stage of asymmetric division, becomes active selectively in the mother cell. Subsequent gene expression is governed by the sequential appearance of the sigma factors σ^F , σ^E , σ^G , and σ^K , with σ^F and σ^G being specific to the forespore and σ^E and σ^K being specific to the mother cell (39).

Using the above described reporter strains, we found that 11 of the newly identified promoters were under sporulation control; i.e., expression was at least dependent on Spo0A. These 11 included one putative, non-protein-coding gene, *yocG-yocH*, and 10 protein-coding genes (Table 2). Transcription from promoters for three of the genes (those contained in the *ywqC-ywqB*, *yjbG-yjbH*, and *yqeN-comEC* regions) was dependent on Spo0A but not on any of the sporulation sigma factors. Of the remaining eight, promoter activity for three protein-coding genes (*ydzH-ydfR*, *yocL-yocM*, and *ytrI-ytqI*) as well for the noncoding gene (*yocG-yocH*) was dependent on both Spo0A and σ^E . (We consider the *yocG-yocH* gene further below.) Finally, promoter activity for two genes, those in the intergenic regions *ypcP-ypbS* and *ynzH-thyA*, were dependent

on only σ^G and σ^K , respectively. For the remaining two cases, those of the promoters for *yybS-cotF* and *yoaH-yoaI*, expression depended on Spo0A, but additional dependencies on sporulation sigma factors were not investigated.

Reinforcing the results from the dependency experiments were images from fluorescent microscopy experiments with in-frame fusions of newly discovered protein-coding genes to the gene for GFP. For example, Fig. 2 shows that the Spo0A-dependent gene in the *ywqC-ywqB* intergenic region was expressed in the predivisional sporangium. Likewise, and as expected, expression of the σ^E -dependent *ytrI-ytqI* gene was limited to the mother cell. In contrast, expression of the σ^G -dependent gene contained in the *ypcP-ypbS* region was restricted to the forespore. Finally, and in contrast to the whole compartment fluorescence patterns seen for the above fusion proteins, GFP fused to the σ^K -dependent gene contained in the *ynzH-thyA* region localized in a ring around the forespore (Fig. 2, the forespore at the far right). This is a localization pattern characteristic of coat proteins. It seems likely, therefore, that the product of the newly discovered ORF in the *ynzH-thyA* region is a previously unrecognized component of the coat.

Finally, we return to the eight genes whose expression did not exhibit a dependence on Spo0A. Thus, despite the results from the microarray data (stationary-phase expression in DS but not in LB medium), these genes appeared not to be under sporulation control. Instead, we conclude that these are genes

TABLE 2. Sigma factor dependency

IGR	Sigma factor
<i>ywqC-ywqB</i>	<i>spo0A</i> , <i>sigH</i>
<i>yjbG-yjbH</i>	<i>spo0A</i> , <i>sigH</i>
<i>yqeN-comEC</i>	<i>spo0A</i> , <i>sigH</i>
<i>ydzH-ydfR</i>	<i>sigE</i>
<i>yocL-yocM</i>	<i>sigE</i>
<i>ytrI-ytqI</i>	<i>sigE</i>
<i>yocG-yocH</i>	<i>sigE</i>
<i>ypcP-ypbS</i>	<i>sigG</i>
<i>ynzH-thyA</i>	<i>sigK</i>
<i>yybS-cotF</i>	ND ^a
<i>yoaH-yoaI</i>	ND ^a

^a ND, not determined. Expression was sporulation dependent, but dependency was not further investigated.

FIG. 3. The *ydzH-ydfR* intergenic gene and its paralog are expressed in the mother cell. (A) Fluorescent micrographs of cells expressing *gfp* fused in frame to the 3' terminus of the small ORF in the *ydzH-ydfR* intergenic region (IGR HR1-GFP) and to the 3' terminus of its paralog *ypzD* (YpzD-GRP). The cells were collected 2 h after induction of sporulation by resuspension and stained with a fluorescent membrane dye. The merged images (GFP, green; membrane, blue) show representative cells expressing GFP in the mother cell. (B) Alignment of inferred amino acid sequences for the newly identified ORF in the *ydzH-ydfR* region (IGR HR1) and YpzD. The proteins have 47% identical and 67% similar amino acids, as indicated by the consensus amino acid and a plus sign, respectively.

that are simply more actively expressed in the postexponential phase of growth in DS medium.

Paralogs to newly identified genes. As shown in Table 1 and discussed earlier, two of the newly identified protein-coding genes have apparent paralogs in the *B. subtilis* genome. The first ORF in the *ydzH-ydfR* intergenic region shares significant sequence similarity with *ypzD*. An alignment of the predicted products of the two genes reveals 47% identical and 67% similar amino acids over the entire length of the two proteins (Fig. 3B). Reinforcing the striking similarity in protein-coding sequences, the two genes are regulated in a similar manner. This was demonstrated by constructing in-frame fusions of *ypzD* to the luciferase operon and *gfp* gene. The results with fusions showed that, as with the *ydzH-ydfR* ORF (above), expression of *ypzD* was dependent on σ^E (data not shown) and expressed specifically in the mother cell (Fig. 3A). It seems likely that the *ydzH-ydfR* and *ypzD* genes play similar roles in sporulation.

Even more striking is the case of the newly identified gene in the *yocL-yocM* intergenic region. Its predicted product has significant similarity to the proteins encoded by the *B. subtilis* genes: *yocN*, *yozN*, *ydzH*, *ydgA*, and *ydgB*. Although all six predicted protein products share a high number of conserved amino acids (Fig. 4B), neighbor-joining shows that the six proteins subdivide into two groups. The predicted product of the newly identified gene in the intergenic region of *yocL-yocM* shows the highest similarity to the *yocN* and *yozN* gene products, with 59% and 49% similar amino acids, respectively,

while similarity to the other three inferred gene products was below 40%. These three, *ydzH*, *ydgA*, and *ydgB*, appear more phylogenetically related to each other, with a sequence similarity greater than 60%.

Because *yocL-yocM* was identified as being under sporulation control, we asked if any of the apparent paralogs are also involved in sporulation. Two of the genes, *yocN* and *yozN*, have been previously identified as being under the control of the sporulation-dependent sigma factor σ^E and being expressed in a mother cell-specific fashion (7). To investigate whether the remaining three genes are also involved in sporulation, we constructed in-frame fusions to both the luciferase operon and the *gfp* gene. Because *ydgB* and *ydgA* seem to be in an operon (their ORFs are separated only by 13 bp), we examined the expression of only the upstream member of the operon, *ydgB*. The luciferase operon was fused to the upstream region of *ydzH* and *ydgB*, and *gfp* was fused to the 3' ends of *yozN*, the *yocL-yocM* IGR, *ydgA*, and *ydzH* genes. Interestingly, luciferase fusions to *ydzH* and *ydgB* were switched on later in sporulation than the other three paralogs. Furthermore, dependence studies showed that *ydgB* and *ydzH* were under the control of the late-appearing, mother cell-specific sigma factor σ^K (data not shown). Finally, fluorescent microscopy experiments with the *gfp* fusions confirmed, as expected, that expression was confined to the mother cell (Fig. 4A). In addition, in the cases of *ydgA* and *ydzH*, striking punctate patterns of localization were observed in which one or more foci were seen around the forespore (Fig. 4A). Such a punctate pattern of localization has

FIG. 4. The YozN protein family. (A) Shown are fluorescence micrographs of sporangia containing *gfp* fused in frame to the 3' of the ORF found in the *yocL-yocM* intergenic region (IGR LM-GFP) and to the 3' ends of *yozN* (YozN-GFP), *ydgA* (YdgA-GFP), and *ydzh* (YdzH-GFP). *yozN* (7) and the *yocL-yocM* intergenic gene are under σ^E control and were expressed in the mother cell (visualized at hour 3). *ydgA-gfp* and *ydzh-gfp* were expressed later in sporulation (visualized at hour 4). Whereas the YozN-GFP and IGR LM-GFP proteins were distributed throughout the mother cell, the YdgA-GFP and YdzH-GFP proteins formed foci around the outside of the forespore. (B) BLAST analysis revealed the indicated five paralogs of the inferred protein product of the *yocL-yocM* intergenic region (IGF-LM ORF). The IGR-LM ORF exhibited 59% and 49% similarity to YozN and YocN, respectively. Similarity to YdgA, YdgB, and YdzH was lower. The multiple alignment was created using ClustalW (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Identical residues are indicated by the program with asterisks, and similar residues are marked with colons (highly similar, like valine and leucine) or periods (less similar).

been seen previously for GFP fusions to other proteins produced in the mother cell and suggests that YdgB and YdzH are likely to be components of the spore coat (42).

A multiple mutant exhibits enhanced sporulation. We built deletion mutants for the 11 newly identified genes under sporulation control (the 10 protein-coding genes and the putative noncoding gene) as well as for the four additional paralogous genes (*ypzD*, *ydgA*, *ydgB*, and *ydzh*) herein found to be under sporulation control. No conspicuous defect in the production of heat-resistant spores (see Table S4 in the supplemental material) or in the timing of sporulation, as judged by microscopy (data not shown), was detected for any of the mutants. Because the newly discovered gene in the *yocL-yocM* intergenic region has five paralogs, we constructed mutants for all six genes individually and together. Neither mutants of individual genes alone nor a multiple mutant lacking all six genes exhibited a significant defect in sporulation (see Table S5 in the supplemental material).

As a more sensitive test for a subtle defect in growth or sporulation, we carried out competition experiments between the wild type and various mutants through cycles of growth and sporulation. To distinguish the wild type from the mutants, we inserted an IPTG-inducible *lacZ* gene in either the wild type or the mutants. The experiment was started by mixing exponentially growing wild-type and mutant cells and inoculating Difco sporulation medium. After 24 h, the nonsporulated cells were

inactivated (at 80°C for 20 min), and fresh Difco sporulation medium was inoculated (1:1,000). The ratio of the two strains was determined by plating dilutions of the culture on plates with X-Gal and IPTG, with either the wild type or mutant showing up as blue colonies. Experiments with five different mutants (*ywqC-ywqB*, *ydcl-ydcK*, *yocG-yocH*, *ypcP-ypbS*, and *ynzH-thyA*) exhibited no differences in growth and sporulation between the wild type and the mutants (data not shown).

However, the multiple mutant lacking all six genes, similar to *yocL-yocM*, exhibited a phenotype. Figure 5A shows the ratios of wild-type to mutant cells through six cycles of sporulation, with the mutant harboring the *lacZ* marker. Surprisingly, the multiple mutant out-competed the wild-type cells. The presence of the marker gene had no effect on the experiment: inserting *lacZ* into the parent and repeating the competition experiment again showed that the multiple mutant out-competed the wild type (data not shown).

The competition advantage was specific for sporulation because growth in nonsporulation medium, LB medium, showed no advantage of the mutant over the wild type (Fig. 5B). To further investigate whether the ability of the multiple mutant to out-compete the wild type occurred during the growth phase of the cycles, we carried out an additional competition experiment, this time keeping the cells continually growing in Difco sporulation medium by continuous transfer while in the exponential phase of growth. Under these continuous growth con-

FIG. 5. Competition between wild-type and mutant strains during growth and sporulation. Wild-type cells and cells of the mutant for Δ IGR *yocL-yocM*, Δ *yocN-yozN*, Δ *ydzH*, and Δ *ydgA-ydgB* and also carrying an IPTG-inducible *lacZ* gene were mixed in equal parts. At time zero (start), 100 ml of Difco sporulation medium (A) and 100 ml of Luria broth (B) were inoculated with 1 ml of the mixed cells. After incubation for 24 h, 1 ml of the cultures was transferred to fresh sporulation (A) and LB (B) media. In the case of the sporulation culture, vegetative cells were inactivated by incubation at 80°C for 20 min prior to the transfer into new Difco sporulation medium. This growing and reinoculating cycle was repeated six times (time points 1 to 6). At time zero as well as after each transfer, samples of cells were diluted and plated onto LB agar containing IPTG and X-Gal, and CFU counts were determined for both blue and white colonies. The columns represent the total CFU counts as 100%, with the percentage of wild-type colonies shown in yellow and the percentage of the mutant strain shown in blue. While the initial ratio of the wild type to the mutant strain does not change during growth in LB medium (B), cells of the mutant strains do accumulate in the sporulation medium (A).

ditions, the multiple mutant showed no growth advantage over the wild type (data not shown).

Based on these observations, we conclude that the mutant has an advantage over the wild type during either spore formation or spore germination. In an effort to distinguish between these possibilities, spores were purified from both the wild-type and the multiple mutant strains, and germination was initiated by the addition of alanine. No differences were observed between the two strains (data not shown). We therefore favor the view that expression of *yocL-yocM* and its structurally similar genes together imposes a small cost to the production of heat-resistant spores under the conditions tested and that spore formation is more efficient in the absence of the paralogs than in their presence. Presumably, the newly discovered family imparts a fitness advantage to spore formation in the environment but evidently not under standard laboratory conditions.

Genes for noncoding RNAs. As we have seen, most (16) of the newly identified genes in intergenic regions proved to be protein-coding genes. However, we discovered three transcription units in regions in which no apparent ORFs were present. Thus, these regions contain candidates for genes for small, non-protein-coding RNAs. During the course of this work, one of the intergenic regions, *ykuI-ykuJ*, was found to harbor a gene for a small noncoding RNA involved in the response to iron (12). The size of this small RNA, termed FsrA, correlates with the boundaries of the signal detected in the microarray. Evidence that the remaining two intergenic regions, those between *yocG* and *yocH* and between *yrhJ* and *yrhK*, contain previously uncharacterized genes for noncoding RNA is as follows.

One of these non-protein-coding genes lies in the region

between the convergently oriented *yocG* and *yocH* genes in the same orientation as *yocH* (Table 1). The gene is flanked by terminators downstream from each of the neighboring genes. These observations, as well as the detection of promoter activity downstream of *yocH* as shown by the use of a reporter fusion (see Fig. S1 in the supplemental material), suggest that this gene is not part of the *yocH* transcriptional unit but, rather, is an independent transcriptional unit. As mentioned above, the promoter was shown to be dependent on σ^E (Table 2), and a good match to the consensus for promoters recognized by σ^E was found in the region downstream of *yocH*. The entire intergenic region was analyzed for potential ORFs, revealing only two and neither with an apparent RBS. We conclude that the newly identified gene is a noncoding gene. (Conceivably, it is translated but, if so, without a recognizable RBS.) A deletion mutant of the gene did not exhibit a measurable defect in sporulation, as judged both by measurements of sporulation efficiency (see Table S4 in the supplemental material) and by a competition experiment similar to that described above (data not shown).

The gene for the second noncoding RNA lies between *yrhJ* and *yrhK* in the same orientation as *yrhK* and opposite to *yrhJ* (Table 1 and Fig. 6A). No ORF is present between *yrhJ* and *yrhK*, and the gene must therefore specify a noncoding RNA. Strong promoter activity was detected downstream of *yrhJ*. This promoter was active during the exponential phase of growth, shutting down at the beginning of stationary phase (see Fig. S1, *yrhJ-yrhK*, in the supplemental material). In preliminary work, we carried out a rapid amplification of cDNA ends (RACE) experiment to attempt to identify the promoter for the *yrhJ-yrhK* RNA (data not shown). We noticed that the region so identified contained a perfect match to the consensus

FIG. 6. The expression of IGR *yrhK-yrhJ* is under the control of the alternative sigma factor σ^D . (A) Shown is the nucleotide sequence of the IGR *yrhK-yrhJ*. The boxed nucleotides conform to the consensus for promoter recognized by σ^D -containing RNA polymerase [TAAA (-35)-N₁₅-GCCGATAT(-10)]. The sequence of the putative small RNA is in bold, and the terminator is underlined. (B) β -Galactosidase activity (open symbols) was monitored at various times during the growth curve of either the wild type ($\square$; EMF287) or a mutant strain for σ^D ($\circ$; EMF327), both harboring *amyE::P_{IGR yrhK-yrhJ}-lacZ*. The OD at 600 nm (OD₆₀₀) was monitored at the same times ($\blacksquare$). Cells were grown in liquid LB medium at 37°C.

-35 (TAAA) and -10 (GCCGATAT) sequences for promoters recognized by σ^D (31, 34). To test whether this region contained a functional σ^D promoter, a *lacZ* fusion to the putative promoter was created. The results shown in Fig. 6B demonstrate that it was expressed in a manner that was completely dependent on σ^D .

Next, we carried out a computational search for potential targets using the program TargetRNA (40) with the sequence extending from the promoter region to the downstream terminator. The search yielded 32 targets with substantial levels of complementarity (see Table S6 in the supplemental material). Among those with the highest levels of complementarity were *citZ*, *aroB*, *abrB*, *hisB*, and *kinC*, with the most impressive complementarity being seen for *citZ*. It will be interesting to see in future work if the *yrhJ-yrhK* RNA is a σ^D -dependent antisense RNA for one or more of these putative target genes.

Summary. We have identified, and demonstrated promoter activity for, 19 transcription units in intergenic regions of the genome. Two of these transcription units contained two ORFs, bringing the total number of genes to 21. Three of the transcription units did not contain ORFs and represent small, noncoding RNAs.

Of the 19 newly identified transcription units, 11 were under the control of Spo0A or a later-acting sporulation transcription factor. Among the genes under sporulation control, one was a small, noncoding RNA, and two were genes that were apparent paralogs of previously recognized ORFs. Thus, an ORF in the *yzdH-ydzR* intergenic region is apparently paralogous to *ypzD*. Additional analysis showed that both paralogs were under the

control of σ^E . Likewise, and even more striking, the ORF in the *yocL-yocM* intergenic region is apparently paralogous to five other genes; two of these were known to be under sporulation control, and the remaining three were here shown (two inferred to be members of the same operon) to be under sporulation control. Thus, our analysis brings to 15 the number of genes newly recognized as being under sporulation control.

A multiple mutant lacking the ORF in the *yocL-yocM* intergenic region and its five paralogs exhibited a novel phenotype. Instead of being impaired in spore formation, as might have been anticipated, it produced spores more efficiently than the wild-type parent in competition experiments involving cycles of spore formation. We conclude that under laboratory conditions, expression of the six paralogs imposes a small cost to spore formation. Presumably, under natural conditions this cost is offset by a fitness benefit that was not replicated in the laboratory.

A final noteworthy feature of this investigation was the identification of three small, non-protein-coding RNAs, two of which are newly identified. One of the newly identified sRNAs (the *yocG-yocH* RNA) was under the control of the sporulation sigma factor σ^E , whereas the other (the *yrhJ-yrhK* RNA) was under the control of the motility sigma factor σ^D . A striking feature of the *yrhJ-yrhK* RNA is its conspicuous complementarity to multiple protein-coding genes, raising the possibility that it influences the stability or translation of the corresponding mRNAs.

ACKNOWLEDGMENTS

We thank Thomas Albert for assistance in the preparation of the *Bacillus* sRNA microarray and Markus Wyss for helpful comments on the paper.

This work was supported by DSM and by NIH grant GMA18568 to R.L. M.S. was supported by a postdoctoral fellowship from the DFG (German Research Foundation). A.H.C. was supported by a Helen Hay Whitney postdoctoral fellowship. E.M. is the recipient of a MEC postdoctoral fellowship from the Secretaría General de Estado de Universidades e Investigación del Ministerio de Educación y Ciencia (Spain).

REFERENCES

1. Aguilar, C., H. Vlamakis, R. Losick, and R. Kolter. 2007. Thinking about *Bacillus subtilis* as a multicellular organism. *Curr. Opin. Microbiol.* 10:638-643.
2. Altuvia, S., D. Weinstein-Fischer, A. Zhang, L. Postow, and G. Storz. 1997. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* 90:43-53.
3. Barbe, V., S. Cruveiller, F. Kunst, P. Lenoble, G. Meurice, et al. 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155:1758-1775.
4. Blattner, F., G. Plunkett III, C. A. Bloch, N. T. Perna, M. Riley, et al. 1997. The complete genome sequence of *Escherichia coli*. *Science* 277:1453-1462.
5. Borden, J. R., S. W. Jones, D. Indurthi, Y. Chen, and T. E. Papoutsakis. 2010. A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in *Clostridium acetobutylicum* involving non-coding RNAs and ribosomal RNA processing. *Metab. Eng.* 12:268-281.
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. Eichenberger, P., S. T. Jensen, E. M. Colon, C. van Ooij, J. Silvaggi, et al. 2003. The σ^E regulon and the identification of additional sporulation genes in *Bacillus subtilis*. *J. Mol. Biol.* 327:945-972.
8. Fan, N., S. Cutting, and R. Losick. 1992. Characterization of the *Bacillus subtilis* sporulation gene, *spoVK*. *J. Bacteriol.* 174:1053-1054.
9. Ferguson, C. C., A. H. Camp, and R. Losick. 2007. *gerT*, a newly discovered germination gene under the control of the sporulation transcription factor σ^K in *Bacillus subtilis*. *J. Bacteriol.* 189:7681-7689.
10. Guerout-Fleury, A. M., K. Shazand, N. Frandsen, and P. Stragier. 1995. Antibiotic-resistance cassettes for *Bacillus subtilis*. *Gene* 167:335-336.

11. Fujita, M., and R. Losick. 2002. An investigation into the compartmentalization of the sporulation transcription factor sigmaE in *Bacillus subtilis*. *Mol. Microbiol.* **43**:27–38.
12. Gaballa, A., H. Antelmann, C. Aguilar, S. K. Khakh, S. Kyung-Bok, et al. 2008. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small basic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105**:11927–11932.
13. Gottesman, S. 2004. The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu. Rev. Microbiol.* **58**:308–328.
14. Handler, A. A., J. E. Lim, and R. Losick. 2008. Peptide inhibitor of cytokinesis during sporulation in *Bacillus subtilis*. *Mol. Microbiol.* **68**:588–599.
15. Hemm, M. R., B. J. Paul, T. D. Schneider, G. Storz, and K. E. Rudd. 2008. Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol. Microbiol.* **70**:1487–1501.
16. Johansson, J., and P. Cossart. 2003. RNA-mediated control of virulence gene expression in bacterial pathogens. *Trends Microbiol.* **11**:280–285.
17. Kim, W., M. W. Silby, S. O. Purvine, J. S. Nicoll, and K. K. Hixson. 2009. Proteomic detection of non-annotated protein-coding genes in *Pseudomonas fluorescens* Pf0-1. *PLoS One* **4**:e8455.
18. Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **393**:249–256.
19. Lee, J. M., S. Zhang, S. Saha, S. Santa Anna, C. Jiang, and J. Perkins. 2001. RNA expression analysis using an antisense *Bacillus subtilis* genome array. *J. Bacteriol.* **183**:7371–7380.
20. Licht, A., S. Preis, and S. Brantl. 2005. Implication of CcpN in the regulation of a novel untranslated RNA (SR1) in *Bacillus subtilis*. *Mol. Microbiol.* **58**:189–206.
21. Middleton, R., and A. Hofmeister. 2004. New shuttle vectors for ectopic insertion of genes into *Bacillus subtilis*. *Plasmid* **51**:238–245.
22. Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, et al. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol. Gen. Genet.* **186**:339–346.
23. Nicholson, W. L., and P. Setlow. 1990. Sporulation, germination, and outgrowth, p. 391–450. In C. R. Harwood and S. M. Cutting (ed.), *Molecular biological methods for Bacillus*. John Wiley and Sons, New York, NY.
24. Qazi, S. N., E. Counil, J. Morrissey, C. E. Rees, A. Cockayne, K. Winzer, W. C. Chan, P. Williams, and P. J. Hill. 2001. *agr* expression precedes escape of internalized *Staphylococcus aureus* from the host endosome. *Infect. Immun.* **69**:7074–7082.
25. Rasmussen, S., H. B. Nielsen, and H. Jarmer. 2009. The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol. Microbiol.* **73**:1043–1057.
26. Repoila, F., N. Majdalani, and S. Gottesman. 2003. Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. *Mol. Microbiol.* **48**:855–861.
27. Repoila, F., and F. Darfeuille. 2009. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biol. Cell* **101**:117–131.
28. Rogozin, I. B., K. S. Makarova, D. A. Natale, A. N. Spiridonov, R. L. Tatusov, et al. 2002. Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. *Nucleic Acid Res.* **30**:4264–4271.
29. Rudner, D. Z., and R. Losick. 2001. Morphological coupling in development: Lessons from prokaryotes. *Dev. Cell* **1**:733–742.
30. Schaeffer, P., J. Millet, and J. P. Aubert. 1965. Catabolic repression of bacterial sporulation. *Proc. Natl. Acad. Sci. U. S. A.* **54**:704–711.
31. Serizawa, M., H. Yamamoto, H. Yamaguchi, Y. Fujita, K. Kobayashi, et al. 2004. Systematic analysis of SigD-regulated genes in *Bacillus subtilis* by DNA microarray and Northern blotting analyses. *Gene* **329**:125–136.
32. Sharma, C. M., and J. Vogel. 2009. Experimental approaches for the discovery and characterization of regulatory small RNAs. *Curr. Opin. Microbiol.* **12**:536–546.
33. Sharp, J. S., and D. H. Bechhofer. 2003. Effect of translational signals on mRNA decay in *Bacillus subtilis*. *J. Bacteriol.* **185**:5372–5379.
34. Sierro, N., Y. Makita, M. J. L. de Hoon, and K. Nakai. 2008. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res.* **36**:93–96.
35. Silvaggi, J. M., J. B. Perkins, and R. Losick. 2005. Small untranslated RNA antitoxin in *Bacillus subtilis*. *J. Bacteriol.* **187**:6641–6650.
36. Silvaggi, J. M., J. B. Perkins, and R. Losick. 2006. Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. *J. Bacteriol.* **188**:532–541.
37. Sonenshein, A. L. 2000. Control of sporulation initiation in *Bacillus subtilis*. *Curr. Opin. Microbiol.* **3**:561–566.
38. Sterlini, J. M., and J. Mandelstam. 1969. Commitment to sporulation in *Bacillus subtilis* and its relationship to development of actinomycin resistance. *Biochem. J.* **113**:29–37.
39. Stragier, P., and R. Losick. 1990. Cascades of sigma factors revisited. *Mol. Microbiol.* **4**:1801–1806.
40. Tjaden, B., S. S. Goodwin, J. A. Opdyke, M. Guillier, D. X. Fu, S. Gottesman, and G. Storz. 2006. Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res.* **34**:2791–2802.
41. Tran, T. T., F. Zhou, S. Marshburn, M. Stead, S. R. Kushner, and Y. Xu. 2009. De novo computational prediction of non-coding RNA genes in prokaryotic genomes. *Bioinformatics* **25**:2897–2905.
42. van Ooij, C., P. Eichenberger, and R. Losick. 2004. Dynamic patterns of subcellular protein localization during spore coat morphogenesis in *Bacillus subtilis*. *J. Bacteriol.* **186**:4441–4448.
43. Vellanoweth, R. L., and J. C. Rabinowitz. 1992. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. *Mol. Microbiol.* **6**:1105–1114.
44. Wach, A. 1996. PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast* **12**:259–265.
45. Wilderman, P. J., N. A. Sowa, D. J. FitzGerald, P. C. FitzGerald, S. Gottesman, et al. 2004. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* **101**:9792–9807.
46. 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.