Published in Applied and Environmental Microbiology 79, issue 17, 5302-5312, 2013
which should be used for any reference to this work

Quantification of Endospore-Forming *Firmicutes* by Quantitative PCR with the Functional Gene *spo0A*

Matthieu Bueche, Tina Wunderlin, Ludovic Roussel-Delif, Thomas Junier, Loic Sauvain, Nicole Jeanneret, Pilar Junier

Laboratory of Microbiology, Institute of Biology, University of Neuchatel, Neuchatel, Switzerland

Bacterial endospores are highly specialized cellular forms that allow endospore-forming *Firmicutes* (EFF) to tolerate harsh environmental conditions. EFF are considered ubiquitous in natural environments, in particular, those subjected to stress conditions. In addition to natural habitats, EFF are often the cause of contamination problems in anthropogenic environments, such as industrial production plants or hospitals. It is therefore desirable to assess their prevalence in environmental and industrial fields. To this end, a high-sensitivity detection method is still needed. The aim of this study was to develop and evaluate an approach based on quantitative PCR (qPCR). For this, the suitability of functional genes specific for and common to all EFF were evaluated. Seven genes were considered, but only *spo0A* was retained to identify conserved regions for qPCR primer design. An approach based on multivariate analysis was developed for primer design. Two primer sets were obtained and evaluated with 16 pure cultures, including representatives of the genera *Bacillus*, *Paenibacillus*, *Brevibacillus*, *Geobacillus*, *Alicyclobacillus*, *Sulfobacillus*, *Clostridium*, and *Desulfotomaculum*, as well as with environmental samples. The primer sets developed gave a reliable quantification when tested on laboratory strains, with the exception of *Sulfobacillus* and *Desulfotomaculum*. A test using sediment samples with a diverse EFF community also gave a reliable quantification compared to 16S rRNA gene pyrosequencing. A detection limit of about 10^4 cells (or spores) per gram of initial material was calculated, indicating this method has a promising potential for the detection of EFF over a wide range of applications.

Endospore formers are a paraphyletic group within the *Firmicutes* (1). This phylum comprises three classes: *Bacilli*, *Clostridia*, and *Erysipelotrichi* (2), but only the first two classes contain endospore-forming species. *Bacilli* principally include aerobic bacteria, whereas *Clostridia* encompass mainly anaerobic strains. Since the discovery of endospore-forming *Firmicutes* (EFF) in the late 19th century, most of the knowledge on their biology has been gained from laboratory studies on cultivable strains (3). In laboratory experiments, endosporulation is commonly triggered by starvation (4) or is coupled to other physiological changes, such as solventogenesis (5), as the ultimate mechanism when all other survival strategies have failed (6).

The survival advantage conferred by sporulation is considered one of the factors that has led to the ubiquitous distribution of EFF (7, 8). As spores remain viable for long periods of time, EFF can survive long distance transportation by climatic or biological agents, leading to a higher dispersion rate than non-EFF organisms (9, 10). In addition, the broad metabolic diversity found within this group is probably another factor that explains their ubiquity (11).

The presence of EFF is a positive factor in many domains related to human life, such as plant growth promotion, biocontrol (12), or the production of fermented food and probiotics by the health care industry (13). However, EFF are often the cause of contamination problems in the food (14) and medical (15) industries. Since endospores resist many disinfection treatments (chemical, heat, and UV sterilization) (7), it is difficult to remove them without a reduction in the quality of the end product. Also, the formation of biofilms is a major challenge encountered in production plants (16, 17). In the case of the food industry, in addition to the health concerns related to contamination by pathogenic species (e.g., *Bacillus cereus* or *Clostridium botulinum*), there is also the aspect of spoilage and short shelf life caused by nonpathogenic species. All these issues represent potentially significant economic losses (18).

Until now, the methods available to detect EFF were either laborious and time-consuming or not sensitive enough. For example, due to the heterogeneity under culture conditions, multiple assays have to be carried out to retrieve as many different EFF species as possible. Culture-independent approaches are mainly based on the quantification of dipicolinic acid (DPA), a unique biomarker for endospores (19). Different methods have been developed to extract and quantify DPA from endospores (10, 19–22), with detection limits ranging from 10^8 endospores/g of sediment (via high-performance liquid chromatography) (23), 10^5 endospores/g of soil (via complexation to terbium and time-resolved fluorometry) (20, 23), or 10^3 endospores/ml (via complexation to terbium and direct fluorescence microscopy) (24). Despite the relatively high sensitivity of DPA-based methods, a disadvantage of those assays is that they cannot take into account vegetative cells of EFF. Furthermore, terbium fluorescence is easily quenched with substances commonly found in environmental samples (e.g., humic acids [23] or organophosphates [19]), leading to a dramatic decrease in the detection limit.

To overcome these limitations, a molecular assay could provide a valid alternative. Although an assay for the detection of specific thermophilic *Bacilli* has been reported (25), a universal

tool is still needed. The aim of this study was to develop and evaluate a molecular approach. Genes involved in the endosporulation process were considered for the identification of a functional molecular marker. Endosporulation is a complex mechanism that involves many regulatory genes (6). Several of them could be used to develop universal primers for endosporulation (T. Wunderlin, T. Junier, L. Roussel-Delif, N. Jeanneret, and P. Junier, unpublished data) and they were tested here. *In silico* tests led to the selection of *spo0A* as a promising marker gene. Quantitative PCR (qPCR) is an accurate method to quantify the frequency of a gene from a DNA extract (26), and therefore in this study qPCR primers targeting the gene *spo0A* were designed, tested, and validated in pure cultures and environmental samples.

MATERIALS AND METHODS

Evaluation of target genes involved in endosporulation. Candidate genes for the development of universal primers to target EFF by PCR assays were identified in a previous study of 27 EFF genomes (Wunderlin et al., unpublished) and included the following: *spo0A*, *gpr*, *spoIVB*, *spoVT*, *spoVAC*, and *spoVAD*. The gene coding for sigma-H factor (*sigH*) was added to this list because it has been shown to be a regulator of sporulation, as *spo0A* is (6). In order to estimate the prevalence of these seven genes among endospore-forming species, an initial list of 90 endospore-forming genera was established and confirmed manually by checking for sporulation as a feature in the *Bergey's Manual of Systematic Bacteriology* (2). This list was key to verify the prevalence of the selected sporulation genes retrieved from the automatic annotation of bacterial genomes that can contain annotation errors (27).

To check for the presence of the seven selected genes in EFF, TIGRFAM collections of proteins (28) generated by the TIGR annotation engine of the J. Craig Venter Institute (29) were downloaded from the Integrated Microbial Genome (IMG) system (30). The seven genes selected in this study corresponded to TIGR01441 (*gpr*), TIGR02875 (*spo0A*), TIGR02851 (*spoVT*), TIGR02860 (*spoIVB*), TIGR02845 (*spoVAD*), TIGR02838 (*spoVAC*), and TIGR02859 (*sigH*). The species in the TIGRFAM list were compared to the list of 90 endospore-forming genera described above. Only species corresponding to endospore formers were retained.

Refined database of EFF sequences. For the design of universal primers for EFF, only the *spo0A* gene was considered. A functional search was used to retrieve *spo0A* sequences in all microbial genomes published in the IMG-JGI database. A query with the keywords *spo0A* or "stage 0 sporulation" retrieved sequences from genes that were not linked to endospore formation. A manual selection was carried out to find the annotated functions that corresponded to the *spo0A* gene. Based on sequence size (250 to 290 amino acids), as well as name, the following keyword annotations seemed to correspond to *spo0A* and were retained for this study: spo0a, spo0a protein, sporulation initiation factor spo0a, sporulation transcription factor spo0a, sporulation transcriptional activator spo0a, and stage 0 sporulation protein A.

A database of 216 gene sequences corresponding to these annotations was constructed. The "seqret" command from the EMBOSS package (31) was used to check for the integrity of the generated fasta file. A multiple alignment was performed using the default settings in MAFFT (32). In order to obtain a database containing a minimum of homologous sequences, a Jukes-Cantor distance matrix was calculated with the command "distmat" from EMBOSS. Sorting of the distances and removal of the closest sequences was carried out with a customized R script (see the supplemental material for further details) (33). As the aim of this sorting step was to avoid redundancy and overrepresentation of specific groups of organisms in the database, an experimentally calibrated threshold value of 2 was set to accomplish this. Gaps were removed, and the 99 remaining sequences were realigned with MAFFT. The "consambig" command (EMBOSS) was used to produce a consensus sequence. To determine the most conserved regions in the alignment, the number of conserved bases for each position was calculated using an R script that ran a 20-bp sliding window with a skip of 1 base in the alignment. A nonmetric multidimensional scaling (NMDS) analysis was performed (vegan library [34]) to determine the homogeneity of the sequences in four conserved regions identified with the R script. This first analysis allowed the detection of atypical sequences from non-endospore-forming species (*Ethanoligenens* spp., *Acetivibrio* spp., and *Blautia* spp., among others) that strongly decreased the conservation signal. These atypical sequences were removed. Finally, a refined database containing 80 sequences was obtained and realigned using MAFFT.

Primer development. The 80 aligned sequences were used for a similar procedure to identify conserved regions. Taking into account the suitability of the amplicon for qPCR (50 to 250 bp) (35, 36), only a few sites were retained. Furthermore, in order to have amplicons of a constant length for all endospore-forming species, the alignment between the annealing sites for the forward and reverse primers could not contain gaps. Based on these parameters, only 11 sites were kept and evaluated by hierarchical clustering (Jaccard distance and average linkage) to identify the least variable sites (vegan library in R). A threshold of 0.4 was selected to retain only the most conserved regions. A Jaccard distance of 0.5 corresponds to an average change of 1 in 3 bases, which matches the variability of the genetic code for the majority of amino acids. Thus, for a 0.4 threshold, only five out of six codons contained in a 20-bp primer present one degeneracy. The 20-bp sequences from the most promising sites were analyzed manually to find degenerate primers that matched the majority of the endospore-forming species. The exact position of each primer in the alignment was then fine-tuned to include a perfectly conserved region for the 3′ end (the five last base pairs) to increase specificity and efficiency during PCR amplification (37).

Finally, an overview of the variability of the sequences inside the retained sites was obtained by a principal components analysis [PCA; rda() function of vegan library in R]. The heterogeneity for the three major sets of EFF sequences (from the genera *Bacillus*, *Clostridium*, and *Geobacillus*) was plotted as an ellipse corresponding to the standard deviation of the PCA coordinates for the members of these groups. All the combinations corresponding to the degenerate positions in the primers were also projected on the two axes of the ordination plot, and an ellipse was drawn around these points to estimate the specificity coverage of the primers. The script used for primer design is provided in the supplemental material.

***In silico* specificity test using a BLAST search.** The Primer-BLAST online tool (38) provided by the NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to evaluate the specificity of the two different primer pairs that were designed. As this tool cannot take into account the degenerate positions, all the different combinations for the forward and reverse primers were inserted manually in the Web interface. The "primer pair specificity checking" parameter was set to use the nucleotide reference database and to restrict the query to only bacterial taxa (excluding uncultured sequences, which are not informative). All the other parameters were set to default values. For each primer set, a text file containing each individual query per degenerate combination was created. The redundancies were removed using low-level Unix commands (grep, awk, sort, and uniq) based on the gene identification number. Sequences were then grouped together by genus, and mismatches were quantified for both forward and reverse primers. This work was done for the entire annealing site (20 or 23 bp). The same analysis was carried out for only the last 6 bp on the 3′ end of the primers, to confirm the conservation and to more accurately evaluate the annealing specificity (data not shown).

Strains and culture conditions. In order to test the efficiencies of the newly developed primers, 16 strains of EFF and 2 non-EFF strains were cultivated for DNA extraction (Table 1). DNA extractions were performed with the innuPREP bacteria DNA kit (Analytik Jena AG, Jena, Germany) and following the manufacturer's instructions. The media

TABLE 1 Strains used for *in vitro* PCR and qPCR tests and culture conditions

Strain	Species	Medium	Temp (°C)	EFF
Neu 1062	*Bacillus licheniformis*	NB	37	Yes
Neu 1070	*Bacillus thuringiensis*	NB	30	Yes
Neu 1121	*Bacillus subtilis*	NB	30	Yes
Neu 1261	*Bacillus amyloliquefaciens*	NB	30	Yes
Neu 91	*Paenibacillus alvei*	NB	30	Yes
Neu1005	*Paenibacillus alvei*	NB	30	Yes
Neu 1003	*Lysinibacillus sphaericus*	NB	30	Yes
Neu 1014	*Brevibacillus brevis*	NB	30	Yes
Neu 1040	*Brevibacillus thermoruber*	DSM 572	45	Yes
Neu 1149	*Geobacillus thermoglucosidasius*	DSM 305	65	Yes
HFF 3	*Geobacillus thermoleovorans*	PTYG, pH 5.0	50	Yes
B 18	*Alicyclobacillus acidocaldarius*	PTYG, pH 2.9	50	Yes
JWO 13	*Sulfobacillus acidophilus*	FeSO	50	Yes
Neu 28	*Clostridium beijerinckii*	ST1 anaerobic	30	Yes
Neu 1155	*Clostridium pasteurianum*	DSM 54 anaerobic	37	Yes
MI-1	*Desulfotomaculum reducens*	Difco 2216*	37	Yes
Neu 1021	*Escherichia coli*	NB	37	No
Neu 1037	*Lactococcus lactis* subsp. *lactis*	DSM 92	37	No

used were standard nutrient broth (NB); DSM medium (http://www.dsmz.de/); standard I nutrient broth (ST1; Merck Millipore, Darmstadt, Germany); PTYG medium composed of 0.25 g/liter peptone, 0.25 g/liter tryptone, 0.5 g/liter yeast extract, 0.5 g/liter glucose, 0.6 g/liter $MgSO_4 \cdot 7H_2O$, 0.07 g/liter $CaCl_2 \cdot 2\ H_2O$ (pH was adjusted with NaOH and H_2SO_4); FeSO medium (described elsewhere [39]); Difco2216*, corresponding to a modified version of marine broth (also described elsewhere [40]) supplemented with 20 mM sodium pyruvate and 1 mg/liter of resazurin to verify anoxia. All cultures were in liquid medium under agitation, except for Neu 1155 (*Clostridium pasteurianum*) and Neu 28 (*Clostridium beijerinckii*), which were cultivated in petri dishes by spreading them on solid medium (15 g/liter agar). The incubation was at 37°C under anaerobic conditions attained by placing the plates in sealed plastic bags in an Anaerocult A mini apparatus and one strip of Anaerotest (Merck Millipore).

Experimental accuracy test. To assess the accuracy of the qPCR quantification, eight strains of EFF (*Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus thuringiensis*, *Brevibacillus brevis*, *Paenibacillus alvei*, *Bacillus amyloliquefaciens*, *Lysinibacillus sphaericus*, and *Clostridium beijerinckii*) and two strains of a non-EFF (*Escherichia coli* and *Herbaspirillum autotrophicum*) were cultivated and quantified by microscopic counting using a Neubauer improved chamber with a special depth of 10 μm (0642010; Paul Marienfeld, Lauda-Königshofen, Germany) prior to DNA extraction. *B. subtilis* was grown in 2× SG medium (41) and incubated at 30°C under agitation at 150 rpm. Vegetative cells were harvested from a fresh overnight culture, whereas endospores were obtained from an older culture in which phase-contrast microscopy revealed over 99% endospores. Cells from other strains were obtained by cultivation in NB and incubation at 37°C (*B. licheniformis* and *E. coli*), 24°C (*H. autotrophicum*), or 30°C (other strains) with agitation at 150 rpm. Cells were fixed in a solution with 4% Formalin and 1× phosphate-buffered saline until counting. Before counting, cells were dispersed in a solution of 1% sodium hexametaphosphate and vortexed for 30 s. Samples were then diluted in physiological solution and loaded by capillarity into the Neubauer improved counting chamber. Counting was carried out with a bright-field microscope with a magnification of ×400. Cells in several squares containing 2.5×10^{-5} μl of cell suspension were quantified, and the average number of cells per μl of fresh culture was calculated.

Finally, DNA extraction from cells and endospores was carried out with the FastDNA spin kit for soil (MP Biomedicals, Santa Ana, CA) following a modified procedure to extract DNA from resistant structures, including endospores (Wunderlin et al., unpublished). This kit was selected because it results in a better extraction yield than others when working with endospores of EFF (42).

PCR conditions. All PCR amplifications were done using *Taq* DNA polymerase with ThermoPol buffer from New-England Biolabs (Ipswich, MA). Reaction mixtures with a final volume of 20 μl contained 1× ThermoPol Buffer, either 0.75 μM (primer set 1, spo0A655f-spo0A923r [Table 2]) or 1 μM (primer set 2, spo0A655f-spo0A834r [Table 2]) of the forward primer, 0.45 μM reverse primer (both primer sets), 200 μM deoxynucleoside triphosphate mix (Promega Corporation, Madison, WI), 1.25 U *Taq* DNA polymerase, and 1 μl of diluted sample (1 to 2 ng of DNA). The volume was completed with PCR-grade water. PCRs were carried out with the Arktik thermal cycler from Thermo Scientific (VWR International, Radnor, PA). The PCR program consisted of an initial denaturation at 95°C for 5 min, followed by 35 cycles composed of a denaturation step at 95°C for 30 s, an annealing step at 52°C (primer set 1) or 54°C (primer set 2) for 30 s, and an elongation step at 68°C for 30 s. A final elongation step at 68°C for 10 min was performed. The PCR results were verified by electrophoresis in 2% agarose gels, loading 2 μl of the PCR product diluted in 10 μl of loading buffer (0.05% bromophenol blue, 8% sucrose, 1.6 mM Tris-HCl at pH 8, 0.16 mM EDTA at pH 8). DNA was revealed by staining for 30 min in a 3× GelRed bath (Biotium Inc., Hayward, CA). Gel imaging was carried out with a GenoPlex chamber (VWR).

qPCR conditions for the *spo0A* gene. qPCRs were carried out in a final reaction volume of 10 μl with 5 μl Rotor-Gene SYBR green PCR master mix (Qiagen GmbH, Hilden, Germany), 0.75 μM and 0.45 μM each primer (forward and reverse, respectively) for set 1 (spo0A655f-spo0A923r); and 1 μM and 0.45 μM each primer for set 2 (spo0A655f-spo0A834r). Two-microliter aliquots of diluted samples (approximately 5 ng) were added to the reaction tubes. The Corbett Rotor-Gene RG-3000A thermocycler (Qiagen) was used for the amplifications. An initial denaturation/polymerase activation step at 95°C for 10 min was first performed. For primer set 1, the 45 following cycles consisted of denaturation at 95°C for 30 s, annealing at 52°C for 30 s, and elongation at 72°C for 30 s. For primer set 2, the 45 following cycles consisted of denaturation 95°C for 15 s, annealing 54°C for 15 s, and elongation 68°C for 10 s. For both primers sets, a melting analysis was performed at the end of the reaction by

TABLE 2 Summary information on the three designed primers[a]

Site no.	Primer	Sense	Sequence (5′–3′)	Length (bp)	% GC content	Annealing temp (°C)	Hairpin formation	3′ complementarity
14	spo0a655f	Forward	5′-GGH GTD CCN GCN CAT ATH AA	20	40–65	47.7–57.9	None	None
15	spo0a923r	Reverse	5′-GCD ATG AAY TCD GAG TTN GTN GG	23	39–61	51.7–60.6	None	None
9	spo0a834r	Reverse	5′-CCA HGC NAC TTC WAT NGC RT	20	40–60	47.7–55.9	None	None

[a] GC content, annealing temperature, hairpin formation, and 3′ complementarity were obtained by using the Oligo Calc Web tool (60).

gradually increasing the temperature by steps of 1°C from the elongation temperature up to 95°C.

The standard curve for quantification was prepared from 10-fold dilutions (10^8 to 10^2 copies/μl) of a plasmid in which the *spo0A* gene of *B. subtilis* was inserted. The TOPO TA cloning kit (Invitrogen, Carlsbad, CA) was used to produce this plasmid in One Shot TOP10F′ chemically competent *E. coli* cells (Invitrogen), following the manufacturer's guidelines. Plasmid DNA was extracted with the Wizard Plus SV Miniprep DNA purification system (Promega) following the manufacturer's instructions. Finally, the number of gene copies was deduced from the DNA quantification carried out with a Qubit 2.0 fluorometer (Invitrogen).

qPCR conditions for the 16S rRNA gene. An additional qPCR quantification of the 16S rRNA gene was performed with DNA extracted from the pure cultures to verify the accuracy of the *spo0A* gene quantification. The qPCR was done on the hypervariable V3 region of the 16S rRNA gene. The primers used were 338f (5′-ACTCCTACGGGAGGCAGCAG-3′) and 520r (5′-ATTACCGCGGCTGCTGG-3′) (43, 44) at a concentration of 0.3 μM each. qPCR was carried out as described for the *spo0A* gene with a modified program consisting of an initial denaturation/activation step at 95°C for 15 min, followed by 40 cycles composed of denaturation at 95°C for 10 s, annealing at 55°C for 15 s, and elongation at 72°C for 20 s. The standard curve consisted of 10-fold dilutions (10^8 to 10^2 copies/μl) of a plasmid in which the V3 region of an environmental clone was inserted. The Zero Blunt TOPO PCR cloning kit for sequencing (Invitrogen) was used to produce the plasmid. Transformation, plasmid extraction, and quantification were performed as mentioned above for the *spo0A* gene.

The ratio between the qPCR quantification of both genes was calculated (qPCR 16S rRNA/qPCR *spo0A*) and divided by the number of copies of the 16S rRNA gene found in each individual species that was tested. The number of 16S rRNA gene copies was obtained from the rrnDB published by Michigan State University (45). A ratio of 1 would be expected in case of no bias, but to take into account different factors of variability such as, for example, differences of annealing specificity, pipetting errors, differences in amplification efficiency (for the *spo0A* sequences), and differences in the number of copies of the 16S rRNA gene between the tested strains and the rrnDB, an arbitrary value between 0.1 and 10 was deemed accurate.

Environmental sequences. Sediments from the Bay of Vidy (Lake Geneva, Switzerland) were used to test the efficiency of the qPCR EFF quantification method in environmental samples. In a first sampling campaign, six cores were retrieved from different areas (Table 3) and subsampled in two layers: depths of 0 to 3 cm and 3 to 9 cm. During a second sampling, four additional cores were retrieved and subsampled in three layers: 0 to 1.5 cm, 1.5 to 3 cm, and 3 to 9 cm. A modified protocol, including cell separation from the environmental matrix and sequential beadbeating steps, was used for DNA extraction (Wunderlin et al., unpublished). Inhibitor-free DNA was then retrieved by an extra ethanol purification step at the end of the extraction procedure.

Samples were sent to Eurofins MWG Operon (Ebersberg, Germany) for 454 pyrosequencing of a fragment of the 16S rRNA gene. Fragments of approximately 500 bp were retrieved using primers Eub8f (5′-AGAGTTTGATCCTGGCTCAG-3′) and Eub519r (5′-GTATTACCGCGGCTGCTGG-3′) as described elsewhere (46). Bioinformatic processing was performed with QIIME (47), following the default pipeline recommended in the online tutorial for 454 pyrosequences. The BIOM format table (48) was retrieved from the QIIME analyses and converted into a text file (convert_biom.py) suitable for importing data in R and to compute the ratio of *Firmicutes* sequences in each sediment sample. Because the reference database to assign taxonomy in QIIME does not allow in every case the identification of all the operational taxonomix units (OTUs) at the genus level, the ratio of *Firmicutes* was evaluated in place of the ratio of EFF. Nonetheless, a similar ratio has been obtained when considering only the *Bacilliales* and *Clostridiales* orders.

Bacterial abundance in environmental samples. In order to assess the abundance of total bacteria in each sample, a qPCR for the hypervariable V3 region of the 16S rRNA gene was carried out as described above for pure cultures. The ratio between the 16S rRNA gene and the *spo0A* gene was calculated.

TABLE 3 Bacterial communities in environmental samples based on pyrosequencing data[a]

Core	Sampling depth (m)	Swiss coordinate		Depth of layer (cm)	Result from bioinformatics analysis of pyrosequencing data		
		East	North		No. of sequences	No. of OTUs	*Firmicutes*
C2.21	47	534509	151342	0–3	22,805	4,778	12,822
				3–9	22,573	3,794	13,398
C8.22	60	534368	151306	0–3	19,440	4,954	2,224
				3–9	20,393	4,157	11,898
C1.03	36	533735	151351	0–3	20,640	6,618	985
				3–9	21,977	6,745	2,424
C2.03	35	534063	151491	0–3	19,357	6,048	931
				3–9	17,830	5,624	2,351
C3.03	32	534676	151543	0–3	20,521	4,200	2,383
				3–9	11,995	4,346	2,385
C4.03	32	534676	151543	0–3	20,873	4,674	2,821
				3–9	31,024	5,415	17,973
C1b.24	32	534709	151526	0–1.5	31,533	5,729	7,169
				1.5–3	9,463	3,064	2,355
				3–9	9,409	3,409	2,767
C2b.24	31.5	534648	151567	0–1.5	10,812	3,221	1,337
				1.5–3	13,349	3,683	1,222
				3–9	5,848	2,217	1,370
C3b.24	31	534803	151513	0–1.5	10,199	2,885	1,147
				1.5–3	5980	1,852	567
				3–9	8,191	2,400	3,189
C4b.24	31	534961	151369	0–1.5	26,725	5,779	708
				1.5–3	7,203	2,570	509
				3–9	6,531	2,380	1,725

[a] All the environmental samples were retrieved from Lake Geneva, in the Bay of Vidy (Lausanne, Switzerland). Sampling locations are given in Swiss national coordinates (CH1903). Information on bacterial communities for layer subsamples were obtained by bioinformatics analysis on pyrosequencing data.

RESULTS

Selection of molecular markers for EFF. A search for EFF clearly demonstrated that the trait of endospore formation is not distributed uniformly in all bacteria from the phylum *Firmicutes* and can be found scattered in various genera of the classes *Clostridia* and *Bacilli*. A first set of genes annotated as part of the endosporulation process and common to EFF with different phylogenetic affiliations was previously identified in our laboratory (Wunderlin et al., unpublished). The comparison between TIGRFAM annotations and a list of 90 endospore-forming genera showed that six out of the seven chosen genes (*gpr*, *spo0A*, *spoVT*, *spoIVB*, *spoVAD*, and *spoVAC*) appeared to be very conserved and widespread in endospore formers (present in 277 to 282 out of 300 EFF species) (Fig. 1). For each of these genes, a large number of sequences from species belonging to the genera *Bacillus*, *Clostridium*, *Paenibacillus*, and *Geobacillus* were found. However, in the case of *Bacillus* and *Clostridium*, this was partially due to the fact that these genera were overrepresented in the data set relative to other EFF. Only the gene coding for the sigma-H factor (*sigH*) was clearly not as well represented as the other six (present in 256 out of 300 EFF species). Based on the results, all the genes but *sigH* could be used as molecular markers for endosporulation. Since the phylogeny of *spo0A* is similar to the phylogeny obtained with the 16S rRNA gene. As a consequence, the gene selected for the design of the qPCR primers was *spo0A*.

Reported sporulation gene in TIGR database

gpr	***spo0A***	*spoVT*	*spoIVB*	*spoVAD*	*spoVAC*	*sigH*	
TIGR01441	**TIGR02875**	TIGR02851	TIGR02860	TIGR02845	TIGR02838	TIGR02859	
123	**124**	124	123	126	126	117	
11	**12**	12	12	12	12	11	
9	**9**	9	9	9	9	9	
9	**9**	9	7	9	9	9	*Alicyclobacillus, Lysinibacillus, Brevibacillus, Others Bacillaceae (28 genera)*
2	**2**	2	2	2	2	2	*Candidatus / Tuberibacillus(0) /Desmospora, 5 other Thermoactinomycetaceae(0)*
85	**86**	83	85	85	86	72	
2	**2**	2	2	2	2	2	*Alkaliphilus, Caloramator(0), Caminicella(0), Metabacterium(0), Oxobacter(0), Tepidibacter(0)*
4	**4**	4	4	4	4	2	*Coprococcus, Sporobacterium(0)*
2	**2**	2	2	2	2	2	*Syntrophothermus, Syntrophomonas, Pelospora(0), Syntrophospora(0)*
15	**15**	15	15	15	15	15	*Peptococcaceae / Acidaminococcaceae / Heliobacteriaceae / Ruminococcaceae(0) / Peptostreptococcaceae(0) / Incertae sedis*
15	**15**	14	15	15	14	14	*Thermoanaerobacter and 11 other genera*
1	**1**	1	1	1	1	1	*Halonatronum(0), Natroniella(0), Orenia(0), Sporohalobacter(0) / Acetohalobium*
278	**281**	277	277	282	282	256	**On 300 Endospore-forming species**

FIG 1 Phylogenetic overview of the major genera of endospore-forming *Firmicutes* (left) and frequency of sporulation related genes in each clade (table in the middle). *spo0A* (in bold) is the most prevalent gene in endospore-forming bacteria from different clades. For the branches "Other" the genera are indicated on the right. The number 0 indicates the genera that have no match for any of the queried genes in the IMG/JGI database.

Primer design. To overcome the variability of the sequences of a functional gene, an approach based on multivariate analysis was developed for the design of universal primers, a process that could possibly be automated in a primer design pipeline. Six highly conserved regions suitable for primer design were identified in a refined database of 80 *spo0A* sequences. In total, 15 starting positions for 20-bp primers containing at least 9 conserved bases (maximum 11 bases) could be defined in the alignment (see Fig. S1 in the supplemental material). However, due to the existence of several gaps in the alignment between the second and the third regions, only the last four (between positions 649 and 923 of the alignment, primer sites 5 to 15 [see Fig. S1]) were effectively suitable for designing qPCR primers, yielding amplicons of a constant length of approximately 150 bp.

From the 11 sites evaluated by clustering, 8 had sequences sufficiently conserved for grouping them together within a Jaccard distance below 0.4 (see Materials and Methods). These eight sites were grouped into three regions corresponding to positions 652 to 686, 815 to 834, and 901 to 923 in the alignment (see Fig. S1 in the supplemental material). To develop primers of 20 bp with this threshold, still an average of 5 degenerate bases were present within the region, and therefore the exact position and the length of the primers was fine-tuned manually to reduce degeneracies.

Initially, two sites were selected (site 14, positions 655 to 674, and site 15, 901 to 923) to develop a primer pair giving an amplicon of 268 bp. In the first site, the forward primer spo0A655f was defined (20 bp), with three degenerate bases and two inosines (Table 2). In the second site, a reverse primer (spo0A923r) of 23 bp, also with three degenerate bases and two inosines, was designed. A third site was selected to improve matching to *Clostridia* (see results of the specificity test below). In this third site, a new reverse primer (spo0A834r) of 20 bp with three degenerate bases and two inosines was designed. The combination of spo0A655f and spo0A834r resulted in a shorter amplicon of 180 bp, which was better suited for qPCR.

A PCA performed on the portions of the sequences corresponding to the primer sites (Fig. 2) showed clear differences in the specificities of the two reverse primers. The reverse primer spo0A923r did not have the same level of coverage as the other two primers. Indeed, in the PCA, the annealing sites for spo0A923r were clearly separated into two distinct groups that were not considered by the overall coverage of the degenerate primer that matched mainly *Bacillus*, and not *Clostridium* or *Geobacillus*. Concerning the two other sites, Fig. 2 illustrates that the sequence variations within the annealing sites for the primers spo0A655f and spo0A834r were much lower and that the degenerate primers covered almost all of the variation displayed in the two main principal components of the ordination plot.

Specificity test *in silico*. In the evaluation of the first primer set (spo0A655f -spo0A923r) with the NCBI Primer-BLAST online tool (38), 145 different species belonging to 26 genera were detected. For the second set (spo0A655f-spo0A834r), 150 species were distributed in 24 genera. Even though the major genera detected by Primer-BLAST were the same for the two different sets, there was some variation in the minor groups found.

The annealing specificity of both sets was evaluated by quantifying the mismatches reported between the primers used and the sequences detected in the NCBI database. These values were averaged by genera and are represented as a bar plot in Fig. 3. In agreement with the results of the PCA, the average number of mismatches for the reverse primer in the second set was much lower than that for the first set, and therefore the second set appears to be more universal.

Regarding specificity, only a few non-endospore-forming genera (*Eubacterium*, *Ruminococcus*, and *Faecalibacterium*) were detected with the two primer sets. These genera are members of the class of *Clostridia*. Furthermore, a closer look at the results in the TIGRfam *spo0A* collection showed that these genera possess a protein with a structure that is closely related to Spo0A. In addition, a recent genomic profiling for sporulation-related genes (49) showed that *Eubacterium* and *Ruminococcus* can be predicted as endospore-forming *Clostridia*, based on their genetic signature. In the case of *Faecalibacterium*, a search for endosporulation-related genes in the genome of *Faecalibacterium prausnitzii* A2-165, recently available from the Genome Institute at Washington University (50), using BioCyc revealed the presence of more than 30 additional genes related to endospore formation. Genera belong-

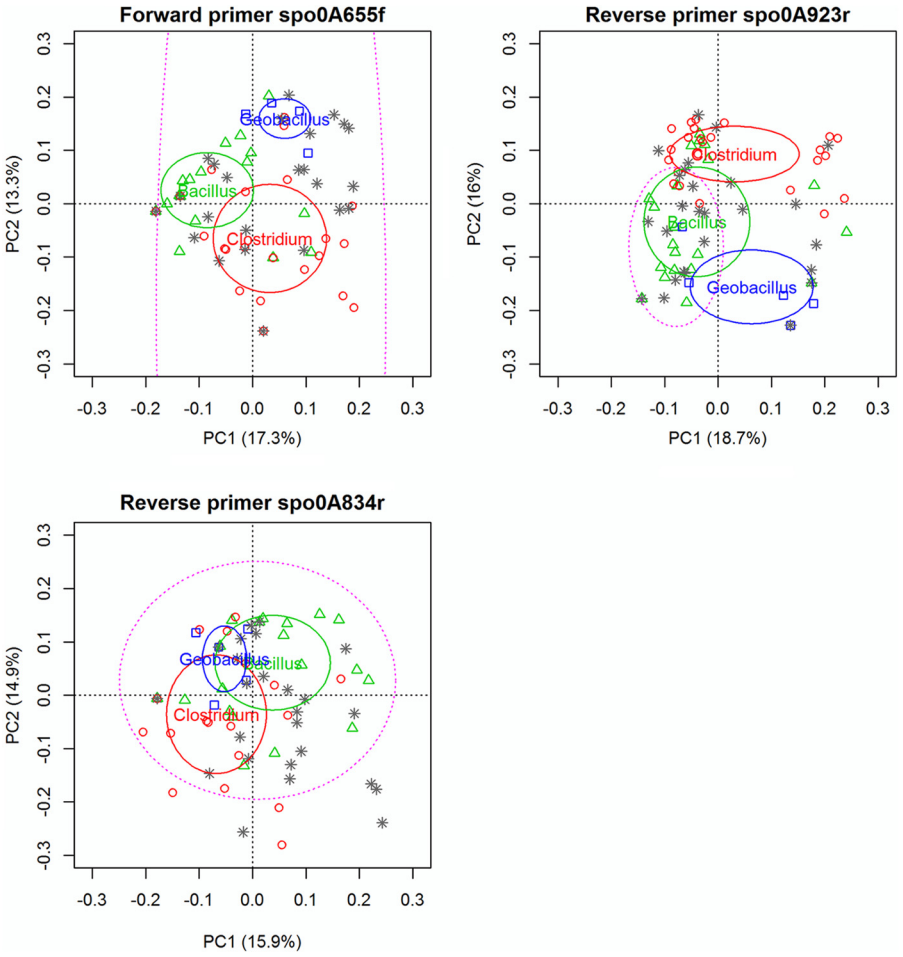


FIG 2 Principal component analysis of the three selected regions used for primer design. This representation gives an overview of the homogeneity of the sequences inside the annealing regions. The dimension of the ellipses (solid lines) gives an estimation of the sequence variability for the main genera found in endospore-forming *Firmicutes* (EFF). The individual points correspond to species used for the alignment. Δ, *Bacillus* spp.; □, *Geobacillus* spp.; O, *Clostridium* spp.; ∗, other EFF species. The ellipse in dashed lines gives an estimate of the coverage of the degenerate primers designed for each particular site.

ing to non-*Firmicutes* phyla were not reported from the Primer-BLAST analysis.

Laboratory experiments. The validation of both primer sets in laboratory experiments was first done by regular PCR with a set of seven species. The two non-endospore-forming strains, *Escherichia coli* and *Lactococcus lactis* subsp. *lactis*, gave no PCR amplification, whereas four out of five endospore formers gave positive results (*Bacillus subtilis*, *Clostridium pasteurianum*, *Alicyclobacillus acidocaldarius*, and *Geobacillus thermoleovorans*). *Sulfobacillus acidophilus*, which was the fifth endospore-forming strain tested, gave no amplification in the classical PCR. Based on these results, the first four endospore-forming strains were used to fine-tune the annealing temperature conditions of a gradient PCR assay between 44 and 56°C. For primer set 1, the best annealing temperature was found to be 52°C, whereas for the second set it was 54°C. These optimal annealing temperatures were then used to test additional endospore-forming strains with the two primer sets. As shown in Table 4, 12 out of the 13 (set 1) and 14 out of the 16 (set 2) endospore-forming strains assayed resulted in a positive PCR amplification. Only *S. acidophilus* did not give a PCR product with any of the two primer sets. For the second set (spo0A655f-spo0A834r), *Desulfotomaculum reducens*, a second strain of *Paenibacillus alvei*, and *Bacillus amyloliquefaciens* were also included in the tests. The two Bacilli gave positive amplification results, whereas *D. reducens* did not.

The first tests of qPCR were also done with the same strains mentioned above. The results were similar except with primer set 1, which gave no amplification for *Clostridia* or *Brevibacillus brevis*. Although we tested different parameters, such as increased $MgCl_2$ concentration, increased primer concentration, and lower annealing temperature, none was successful for amplification of these strains. The second set with the reverse primer (spo0A834r) gave better results for the two *Clostridium* strains tested, but the amplification efficiency was still low. The melting analysis, however, highlighted that the PCR products for the *Clostridium* strains have relatively low melting temperatures (close to the one used for elongation). As a consequence, the elongation was decreased from 72 to 68°C. With this modification, a drastic increase of the amplification efficiency for the *Clostridium* strains was obtained. Concerning *B. brevis*, positive amplification results were also observed with the second primer set.

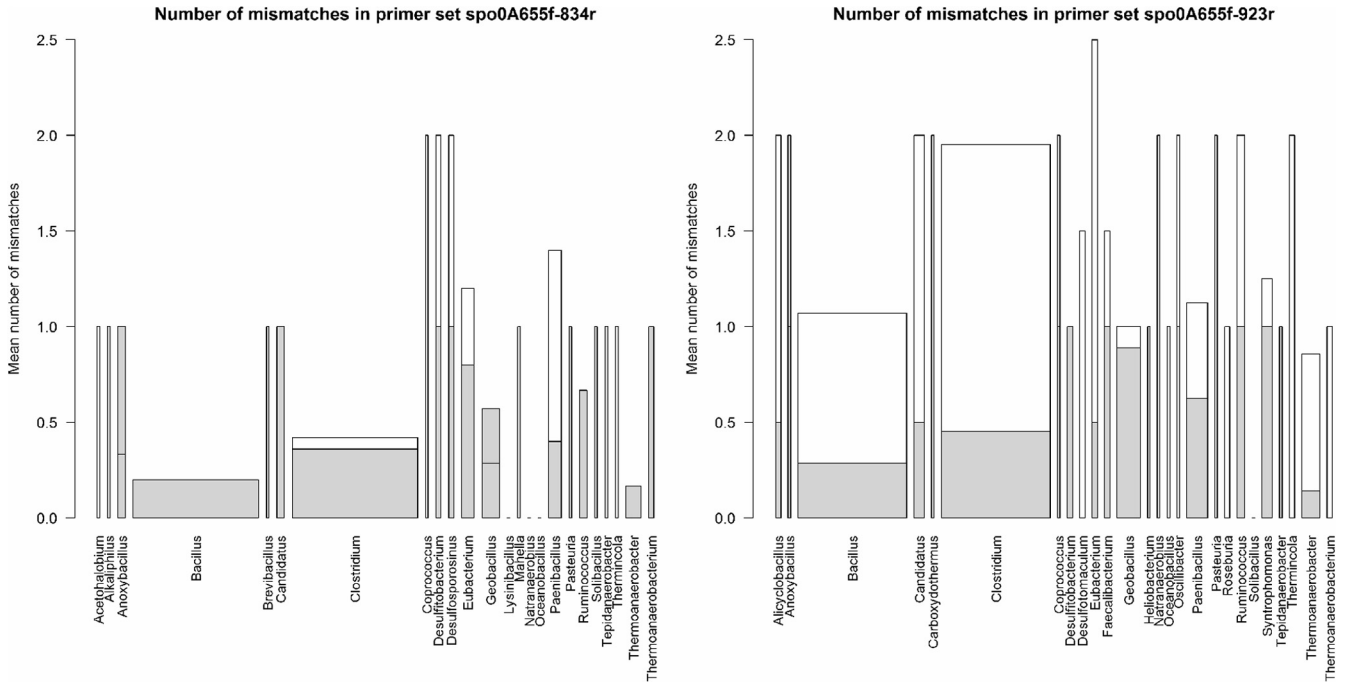


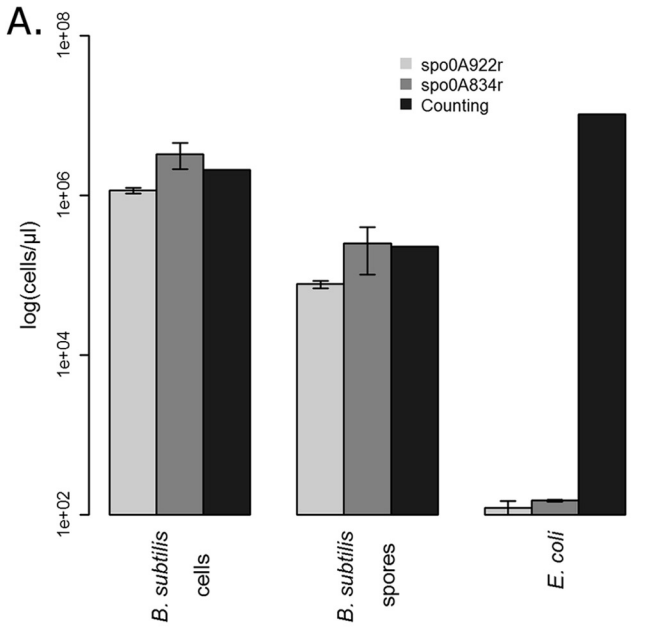
FIG 3 *In silico* test of the two primer sets assessed by NCBI primer blast. The height of the bars represents the number of mismatches detected in the alignment of the primers. The gray zone corresponds to the forward primer whereas the white zone to the reverse. The width of the bars represents the relative number of hits in the NCBI database for each genus.

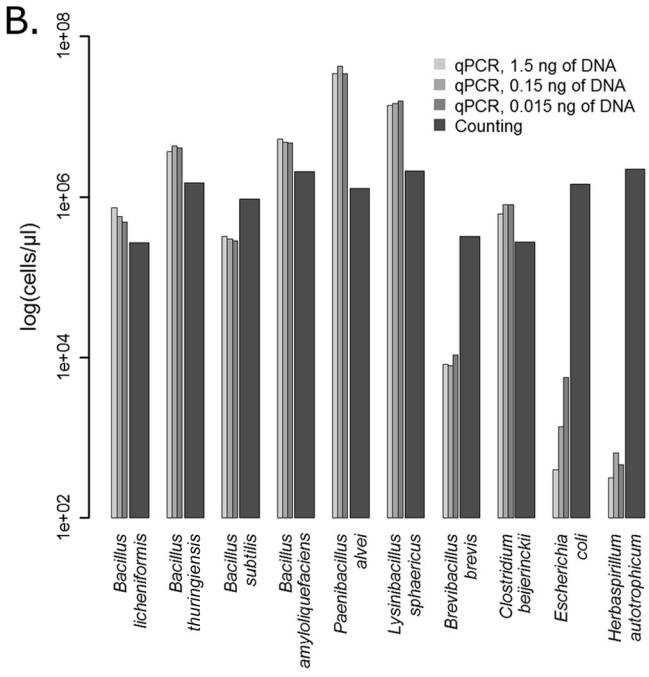
Finally, the accuracy of the *spo0A* qPCR quantification (second primer set only) was evaluated by comparison with a qPCR for the 16S rRNA gene. For most of the strains that gave a positive amplification signal (above 10^2 copies of the *spo0A* gene per μl of DNA sample), quantification of the *spo0A* gene was consistent with the 16S rRNA gene results (taking into account the number of copies of the 16S rRNA gene in each strain). This was, however, not the case for *B. brevis*, *G. thermoleovorans*, or *A. acidocaldarius*, for

TABLE 4 PCR and qPCR amplification results for the two sets of primers for the *spo0A* gene[a]

			PCR amplification		qPCR amplification		*spo0A* vs 16S rRNA qPCR	
Strain	Species	EFF	Set 1	Set 2	Set 1	Set 2	Set 1	Set2
Neu1062	***Bacillus licheniformis***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↓ 1.38**
Neu1070	***Bacillus thuringiensis***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↑ 6.41**
Neu1121	***Bacillus subtilis***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↓ 4.55**
Neu1261	***Bacillus amyloliquefaciens***	**Yes**	**ND**	**Yes**	**ND**	**Yes**	**ND**	**↑ 9.68**
Neu91	***Paenibacillus alvei***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↑ 1.71**
Neu1005	***Paenibacillus alvei***	**Yes**	**ND**	**Yes**	**ND**	**Yes**	**ND**	**↑ 2.18**
Neu1003	***Lysinibacillus sphaericus***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↑ 5.1**
Neu1014	*Brevibacillus brevis*	Yes	Yes	Yes	No	Yes	ND	↓ 9,906
Neu1040	***Brevibacillus thermoruber***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↓ 2.92**
Neu1149	***Geobacillus thermoglucosidasius***	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**ND**	**↑ 1.2**
HFF-3	*Geobacillus thermoleovorans*	Yes	Yes	Yes	Yes	Yes	ND	↓ 176,008
B 18	*Alicyclobacillus acidocaldarius*	Yes	Yes	Yes	Yes	Yes	ND	↓ 2,210
JWO13	*Sulfobacillus acidophilus*	Yes	No	No	No	No	ND	ND
Neu28	***Clostridium beijerinckii***	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**ND**	**↑ 7.52**
Neu1155	***Clostridium pasteurianum***	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**ND**	**↑ 3.14**
MI-1	*Desulfotomaculum reducens*	Yes	ND	No	ND	No	ND	ND
Neu1021	*Escherichia coli*	No	No	No	No	No	ND	ND
Neu1037	*Lactococcus lactis* subsp. *lactis*	No	No	No	No	No	ND	ND

[a] Set 1, spo0A655f-spo0A923r; set 2, spo0A655f-spo0A834r. Sixteen EFF strains and two non-endospore formers were tested. In the qPCR amplification test, results were considered positive when more than 10^2 copies/μl were detected, which corresponded to the lowest dilution used to establish the standard curve. Strains for which data are shown in boldface are those for which quantification for *spo0A* (using the second set of primers) was consistent with that obtained for the 16S rRNA gene. To assess the consistency of results, the ratio between the qPCR quantification of the 16S rRNA gene and the *spo0A* gene was computed and then divided by the number of copies of the 16S rRNA gene found in the strain (retrieved from the rrnDB database [45]). Finally, an accurate amplification was considered when this value was between 0.1 and 10 (see Material and Methods). ↑, overestimation of the number of *spo0A* gene copies; ↓, underestimation of the number of *spo0A* gene copies. ND, not determined.





C.

	Bacillus thuringiensis	Bacillus subtilis	Bacillus amyloliquefaciens	Paenibacillus alvei	Lysinibacillus sphaericus	Brevibacillus brevis	Clostridium beijerinckii	Escherichia coli *	Herbaspirillum autotrophicum *
Bacillus licheniformis	1.11	1.19	1.31	0.88	0.92	2.05	0.81	1.35	1.34
Bacillus thuringiensis		0.84	1.02	0.97	1.04	0.97	1.03	0.95	0.87
Bacillus subtilis			1.22	1.10	1.00	1.00	0.97	1.02	1.18
Bacillus amyloliquefaciens				1.02	1.04	1.11	1.11	1.35	1.39
Paenibacillus alvei					1.20	0.99	1.01	1.00	1.04
Lysinibacillus sphaericus						1.01	0.97	1.07	1.12
Brevibacillus brevis							0.85	0.95	1.37
Clostridium beijerinckii								0.79	0.78

which the results indicated either an over- or underestimation of the number of cells.

Another reliability test of the quantification of endospore formers with these primer sets was carried out by comparing qPCR results with the quantification obtained by microscopic cell counting (Fig. 4). For vegetative cells and endospores of *B. subtilis*, the quantification by qPCR with both primer sets gave values on the same order of magnitude as those obtained by counting. The standard deviations (Fig. 4A, error bars) were much lower for the first set of primers (spo0A655f-spo0A834r) than for the second set (spo0A655f-spo0A834r). The qPCR quantification of a non-EFF, *E. coli*, gave no significant qPCR amplification signal, as expected. Additional comparisons that considered different initial DNA concentrations for eight other EFF strains (Fig. 4B) showed that the quantification by qPCR was also on the same order of magnitude as the microscopic counting, even for the lowest DNA dilution (0.015 ng of DNA). Only *P. alvei* was surprisingly overestimated by qPCR or underestimated by counting, while *B. brevis* was clearly underestimated, which is consistent with the comparison between the qPCR results for *spo0A* and the 16S rRNA gene (Table 4). The two non-EFF strains tested in Fig. 4B (*E. coli* and *H. autotrophicum*) also gave no significant qPCR signal (below 10^2 copies µl of the *spo0A* gene), as expected. Finally, the potential interference that occurs when quantifying a mixture of EFF was evaluated by combining equal amounts of DNA from two EFF strains or one EFF and one non-EFF strain (Fig. 4C). A ratio was calculated between the theoretical expected value (the sum of the individual amplifications) and the obtained measurement. Except for the combination of DNA from *B. brevis* and *B. licheniformis*, the ratio was between 0.81 and 1.31. The same range was obtained when DNA from a non-EFF strain was added, for which only the EFF contributed to the final quantification, indicating that there is no or little interference between DNA from different origins when using this *spo0A* qPCR quantification assay.

Validation with environmental samples. A very good correlation (R^2, 0.926) was obtained when the qPCR quantification was compared to pyrosequencing results in determining the proportion of EFF in bacterial communities found in sediments of Lake Geneva (Fig. 5). The slope of the regression function (7.44) showed that the ratio calculated by qPCR was lower than the one calculated by pyrosequencing. Four samples that contained a very large proportion of EFF, close to 50% of the sequences by pyrosequencing, corresponded to the larger ratios between qPCR for the *spo0A* gene relative to the 16S rRNA gene (6.5%). In seven samples, the ratio of EFF measured by pyrosequencing varied between 19.88% and 38.93% and corresponded to intermediate values when considering qPCR ratios (2.22% to 3.93%). Finally, in the other 13 samples considered, the fraction of EFF according to the

FIG 4 Comparison between a classical quantification with a Neubauer improved counting chamber and the *spo0A* gene qPCR assay. A. Comparison of the quantification with the two sets of primers on vegetative cells and spores of *Bacillus subtilis*, as well as with cells of *Escherichia coli* as a negative non-endospore forming control. B. Comparison of quantification for eight other EFF strains and two non-EFF strains with the second set of primers (spo0A655f-spo0A834r) and different initial concentrations of DNA. C. Ratio calculated for the quantification with equal initial DNA concentration for mixtures of two-strains. The values represent the ratio between the theoretical value (sum of the individual quantifications) and the obtained measurement. *, only the EFF contributed to the final quantification.

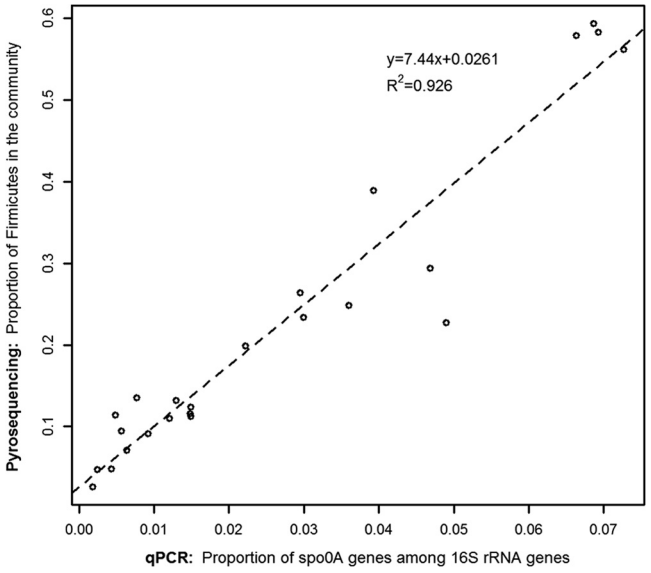


FIG 5 Correlation between two methods to quantify the ratio of endospore-formers in complex microbial communities. The proportion of endospore-forming bacteria in the community has been evaluated by two independent approaches. First (on the *y* axis): QIIME taxonomical identification of OTUs from pyrosequencing (V1-V3 region of the 16S rRNA gene) and quantification of *Firmicutes* in the overall sequences. Second (on the *x* axis): ratio of qPCR on the *spo0A* gene and qPCR on the 16S rRNA gene (V3 region). The slope of 7.44 represents approximately the bias induced by the number of 16S rRNA gene copies for both methods of quantification (see discussion section for details).

pyrosequencing data ranged from 2.65% to 13.52%. For those samples, the qPCR ratio varied from 0.18% to 0.76%.

DISCUSSION

The seven genes evaluated here are all involved in essential steps of the endosporulation pathway. The *spo0A* gene is the major regulator for the initiation of sporulation. The *gpr* gene encodes a protease active during germination (6). Stage IV sporulation protein B (*spoIVB*) is involved in the activation of the sigma-K factor (final transcription factor expressed during the sporulation cascade in the mother cell); stage V sporulation proteins A (spo *VAC* and *spoVAD*) have been suggested to be important transmembrane proteins that are probably involved in DPA transport into the endospore (51). The stage V sporulation protein T is a global regulator activated by sigma-G (1). Finally, the gene coding for the sigma-H factor is another positive regulator of sporulation in a way similar to *spo0A* (6). The prevalence rates of *spo0A*, *spoVT*, and *spoIVB* have already been observed in a genome-wide comparative study of the evolution of the endosporulation pathway (1). Here, we showed that the same is true for *spoVAD*, *spoVAC*, and *gpr*. However, in the case of the sigma-H factor, this is clearly not the case, as this factor directs the transcription of several genes not only during the initial steps of endospore formation but also during the transition from exponential growth to stationary phase and the entry into the state of genetic competence (52). As a consequence, homologs of this gene are also found in phyla like *Thermotogae* and *Fusobacteria* (data not shown), which are not related to EFF, making this gene unsuitable for the design of endosporulation-specific primers.

The two primer sets that were designed here are promising as universal primers for the detection of endospore-forming bacteria. For example, the results of the *in silico* tests gave a positive match not only for the major groups of endospore formers (*Bacillus*, *Clostridium*, and *Geobacillus*) but also for some less-studied groups, such as *Haloanaerobiales*, *Thermoanaerobacteriales*, *Lachnospiraceae*, *Syntrophomonadaceae*, *Peptococcaceae*, *Natranaerobiaceae*, and *Heliobacteriaceae*. However, a general concern for these *in silico* studies is the bias due to the number of DNA sequences that are deposited in published online databases. As an example, the higher representation rates of *Bacillus*, *Clostridium*, *Paenibacillus*, and *Geobacillus* found in this work are probably not due to better specificities of the primers for these genera but to the higher number of related sequences in genomic databases. As a consequence, the widths of the bars in Fig. 3 give only a vague indication of the stringency of the primers.

In laboratory experiments, the two primer sets designed (set 1, spo0A655f -spo0A923r, and set 2, spo0A655f-spo0A834r) gave good *spo0A* amplification results for the major groups of *Bacillales*. However, as expected from the narrowed spectrum observed in the PCA and in the *in silico* analysis, the set spo0A655f-spo0A923r did not amplify *Clostridium* spp. efficiently, but this was improved during the design of primer set 2 (spo0A655f-spo0A834r). This was also the case for *B. brevis*, for which the lack of positive amplification with primer set 1 can be explained by the poor specificity of the reverse primer, as demonstrated in the PCA analysis (Fig. 2) and the *in silico* tests (Fig. 3). This bias was overcome by using primer set 2, which resulted in positive amplification of this strain in classical and quantitative PCRs.

No amplification was obtained for *S. acidophilus* or *D. reducens* with any of the primer sets. The full genomes of these two EFF species are available (*S. acidophilus* [53], *D. reducens* [54]), and a homolog protein structure to Spo0A was found in the TIGR02875 data set. When aligned with sequences from the refined reference database (80 *spo0A* sequences used for primer design), the sequence at the annealing site matched exactly the sequences of the primers that were designed. The PCR inhibitory effect of iron (55), which was present in the culture media of these two strains, probably explains the negative results that were obtained, and we have experimental evidence showing the inhibitory effect for these DNA extracts (data not shown).

The validation of the qPCR quantification in environmental samples showed that the ratio calculated by qPCR was lower than the one calculated by pyrosequencing. The first explanation for this is the difference in the number of copies of 16S rRNA genes compared to *spo0A*. As the ribosomal (rrn) operon is normally found in several copies, the quantification of this gene by qPCR gives values that are artificially greater than the real number of bacterial cells in the community, lowering the obtained qPCR ratio. Furthermore, the average number of rrn operon copies depends on the group of bacteria. An average value of 4.3 copies of 16S rRNA genes was found in the rrnDB in all bacterial phyla (45). This value is substantially greater (average, 7.01 copies) for the phylum *Firmicutes*. As the pyrosequencing was carried out for the 16S rRNA gene, the difference of average 16S rRNA copy numbers between bacteria and *Firmicutes* generated another bias by overestimating the ratio of EFF in the community by a factor of 7.01/4.3, or 1.63. Thus, these two bias sources combined (the overestimation of the 16S rRNA gene copy numbers an average of 4.3 times and the ratio calculated by pyrosequencing being 1.63 times greater) led to a correction factor of 7.01, which was very close to the slope of 7.44 obtained experimentally (Fig. 5).

The detection limit of this new method depends on different factors. The quantity of sample used for the DNA extraction as well as the extraction methodology are probably the more important ones. DNA extraction is very important, because cells from EFF are harder to lyse than cells from other bacterial groups. This is particularly true if most of the cells are present in the form of endospores. Furthermore, in order to have quantitatively comparable results, one must pay particular attention to the repeatability of the extraction methodology in terms of yield and quality of the final product. In our experimental evaluation, the quantification was linear down to 10^2 copies/μl (quantification curve) and we could detect down to 10^4 copies of *spo0A* gene/g of wet sediment from the environmental samples. Within the range covered by the standard curve, the qPCR quantification of the *spo0A* gene in EFF strains was independent of the amount of initial DNA used, down to 0.015 ng of DNA, and it was consistent with results of direct microscopic counting of cells prior DNA extraction (Fig. 4). No significant bias in the quantification were observed when mixing DNA from different strains, confirming the selectivity of this qPCR-based method for the quantification of a complex mixture of EFF strains, like those found in environmental samples. Thus, by concentrating DNA prior to qPCR quantification or by optimizing the DNA extraction protocol, this detection limit could probably be lowered one or more orders of magnitude. In comparison, the detection limit reported in the literature for the DPA assay is on the order of 10^5 endospores/g of soil sample. Furthermore, the DPA assay does not quantify vegetative cells, only endospores. As a consequence, these two techniques can be seen as complementary in order to obtain a full overview of EFF in the environment and their morphological state.

In conclusion, the experimental evaluation suggests that the quantification of endospore-forming bacteria by qPCR using the primers developed for the *spo0A* gene gives results that are in agreement with the pyrosequencing quantification of EFF in complex microbial communities. In addition, quantification of EFF by qPCR is fast, cheap, and very reliable compared to other modern techniques, such as next-generation sequencing. Although the information obtained is not identical (pyrosequencing has the advantage of species identification), in some cases a simple quantification can be sufficiently informative. Examples of this are deep oceanic drilling projects (56), exobiology research (8) paleontology (57), study of polluted environments (61), the packaging industry (58), and sterility assessments in the food and pharmaceutical industries (59). Also, although direct clinical or industrial examples could not be tested here, the level of sensitivity obtained with environmental samples, which were notable for their high level of PCR-inhibitory substances, makes likely an even better detection potential for this approach in much cleaner anthropogenic environments.

ACKNOWLEDGMENTS

We thank Ferring Pharmaceutical and the Consulat honoraire de la Fédération de Russie à Lausanne for funding this project through elemo grant 301, as well as the EPFL for coordinating the elemo project. Funding from the Swiss National Science Foundation (grant 31003A_132358) is also acknowledged.

We thank Patricia L. Siering from Department of Biological Sciences, Humboldt State University, CA, for the strains *Geobacillus thermoleovorans* B18, *Alicyclobacillus acidocaldarius* strain JWO13, and *Sulfobacillus acidophilus*. We thank Elena Dalla-Vecchia from the Ecole Polytechnique Federale de Lausanne for providing *Desulfotomaculum reducens*.

REFERENCES

1. **De Hoon MJL, Eichenberger P, Vitkup D.** 2010. Hierarchical evolution of the bacterial sporulation network. Curr. Biol. **20:**R735–R745.
2. **De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K-H, Whitman WB.** 2009. Bergey's manual of systematic bacteriology, 2nd ed. Springer, Dordrecht, Netherlands.
3. **Nicholson W.** 2002. Roles of *Bacillus* endospores in the environment. Cell. Mol. Life Sci. **59:**410–416.
4. **Hageman J, Shankweiler G, Wall P, Franich K, Cowan MG, Cauble S, Grajeda J, Quinones C.** 1984. Single, chemically defined sporulation medium for *Bacillus subtilis* growth, sporulation, and extracellular protease production. J. Bacteriol. **160:**438–441.
5. **Tracy BP, Jones SW, Papoutsakis ET.** 2011. Inactivation of σ^E and σ^G in *Clostridium acetobutylicum* illuminates their roles in clostridial-cell-form biogenesis, granulose synthesis, solventogenesis, and spore morphogenesis. J. Bacteriol. **193:**1414–1426.
6. **Errington J.** 2003. Regulation of endospore formation in *Bacillus subtilis*. Nat. Rev. Microbiol. **1:**117–126.
7. **Nicholson W, Fajardo-Cavazos P, Rebeil R, Slieman T, Riesenman P, Law J, Xue Y.** 2002. Bacterial endospores and their significance in stress resistance. Antonie Van Leeuwenhoek **81:**27–32.
8. **Yung PT, Kempf MJ, Ponce A.** 2006. A rapid single spore enumeration assay, p 751–763. Aerospace Conference, 2006 IEEE. Institute of Electrical and Electronics Engineers, New York, NY. doi:10.1109/AERO.2006.1655788.
9. **Roberts MS, Cohan FM.** 1995. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. Evolution **49:**1081–1094.
10. **Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, ReysenbachA-L, Smith VH, Staley JT.** 2006. Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. **4:**102–112.
11. **Mandic-Mulec I, Prosser JI.** 2011. Diversity of endospore-forming bacteria in soil: characterization and driving mechanisms, p 31–59. *In* Logan NA, De Vos P (ed), Endospore-forming soil bacteria. Springer-Verlag, Berlin, Germany.
12. **McSpadden Gardener BB.** 2004. Ecology of *Bacillus* and *Paenibacillus* spp. in agricultural systems. Phytopathology **94:**1252–1258.
13. **Logan NA.** 2012. *Bacillus* and relatives in foodborne illness. J. Appl. Microbiol. **112:**417–429.
14. **Bibek R.** 2004. Fundamental food microbiology, 3rd ed. CRC Press LLC, Boca Raton, FL.
15. **De Clerck E, Vanhoutte T, Hebb T, Geerinck J, Devos J, De Vos P.** 2004. Isolation, characterization, and identification of bacterial contaminants in semifinal gelatin extracts. Appl. Environ. Microbiol. **70:**3664–3672.
16. **Heyndrickx M.** 2011. The importance of endospore-forming bacteria originating from soil for contamination of industrial food processing. Appl. Environ. Soil Sci. **2011:**11.
17. **Seale RB, Bremer PJ, Flint SH, McQuillan AJ.** 2010. Characterization of spore surfaces from a *Geobacillus* sp. isolate by pH dependence of surface charge and infrared spectra. J. Appl. Microbiol. **109:**1339–1348.
18. **Brown KL.** 2000. Control of bacterial spores. Br. Med. Bull. **56:**158–171.
19. **Fichtel J, Koster J, Scholz-Bottcher B, Sass H, Rullkotter J.** 2007. A highly sensitive HPLC method for determination of nanomolar concentrations of dipicolinic acid, a characteristic constituent of bacterial endospores. J. Microbiol. Methods **70:**319–327.
20. **Brandes Ammann A, Kolle L, Brandl H.** 2011. Detection of bacterial endospores in soil by terbium fluorescence. Int. J. Microbiol. **2011:**435281. doi:10.1155/2011/435281.
21. **Tabor M, Macgee J, Holland J.** 1976. Rapid determination of dipicolinic acid in spores of *Clostridium* species. Appl. Environ. Microbiol. **31:**25–28.
22. **Beverly MB, Basile F, Voorhees KJ, Hadfield TL.** 1996. A rapid approach for the detection of dipicolinic acid in bacterial spores using pyrolysis mass spectrometry. Rapid Commun. Mass Spectrom. **10:**455–458.
23. **Fichtel J, Koster J, Rullkotter J, Sass H.** 2007. Spore dipicolinic acid contents used for estimating the number of endospores in sediments. FEMS Microbiol. Ecol. **61:**522–532.
24. **Yang W-W, Ponce A.** 2011. Validation of a *Clostridium* endospore viability assay and analysis of Greenland ices and Atacama desert soils. Appl. Environ. Microbiol. **77:**2352–2358.

25. **Rueckert A, Ronimus RS, Morgan HW.** 2006. Development of a real-time PCR assay targeting the sporulation gene, spo0A, for the enumeration of thermophilic bacilli in milk powder. Food Microbiol. **23:**220–230.
26. **Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonak J, Lind K, Sindelka R, Sjoback R, Sjogreen B, Strombom L, Stahlberg A, Zoric N.** 2006. The real-time polymerase chain reaction. Mol. Aspects Med. **27:**95–125.
27. **Schnoes AM, Brown SD, Dodevski I, Babbitt PC.** 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput. Biol. **5:**e1000605. doi:10.1371/journal.pcbi.1000605.
28. **Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR, White O.** 2007. TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. Nucleic Acids Res. **35:**D260–D264.
29. **Stothard P, Wishart DS.** 2006. Automated bacterial genome analysis and annotation. Curr. Opin. Microbiol. **9:**505–510.
30. **Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC.** 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. **40:**D115–D122.
31. **Rice P, Longden I, Bleasby A.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. **16:**276–277.
32. **Katoh K, Misawa K, Kuma K, Miyata T.** 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30:**3059–3066.
33. **R Development Core Team.** 2009. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
34. **Oksanen J.** 2011. Multivariate analysis of ecological communities in R: vegan tutorial. R Foundation for Statistical Computing, Vienna, Austria.
35. **Sharma S, Radl V, Hai B, Kloos K, Fuka MM, Engel M, Schauss K, Schloter M.** 2007. Quantification of functional genes from procaryotes in soil by PCR. J. Microbiol. Methods **68:**445–452.
36. **Smith CJ, Osborn AM.** 2009. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol. Ecol. **67:**6–20.
37. **Stadhouders R, Pas SD, Anber J, Voermans J, Mes THM, Schutten M.** 2010. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay. J. Mol. Diagn. **12:**109–117.
38. **Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden T.** 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics **13:**134.
39. **Johnson D.** 1995. Selective solid media for isolating and enumerating acidophilic bacteria. J. Microbiol. Methods **23:**205–218.
40. **Zobell CE.** 1941. Studies on marine bacteria: the cultural requirements of heterotrophic aerobes. J. Mar. Res. **4:**42–75.
41. **Leighton TJ, Doi RH.** 1971. The stability of messenger ribonucleic acid during sporulation in *Bacillus subtilis*. J. Biol. Chem. **246:**3189–3195.
42. **Dineen SM, Aranda R, Anders DL, Robertson JM.** 2010. An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. J. Appl. Microbiol. **109:**1886–1896.
43. **Bakke I, De Schryver P, Boon N, Vadstein O.** 2011. PCR-based community structure studies of bacteria associated with eukaryotic organisms: a simple PCR strategy to avoid co-amplification of eukaryotic DNA. J. Microbiol. Methods **84:**349–351.
44. **Muyzer G, Dewaal E, Uitterlinden A.** 1993. Profiling of complex microbial populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction-amplified genes-coding for 16S ribosomal RNA. Appl. Environ. Microbiol. **59:**695–700.
45. **Lee ZM-P, Bussema C, Schmidt TM.** 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res. **37:**D489–D493.
46. **Li H, Zhang Y, Li D, Xu H, Chen G, Zhang C.** 2009. Comparisons of different hypervariable regions of rrs genes for fingerprinting of microbial communities in paddy soils. Soil Biol. Biochem. **41:**954–968.
47. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Tumbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **7:**335–336.
48. **McDonald D, Clemente J, Kuczynski J, Rideout J, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, Knight R, Caporaso J.** 2012. The biological observation matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaScience **1:**7.
49. **Abecasis AB, Serrano M, Alves R, Quintais L, Pereira-Leal JB, Henriques AO.** 2013. A genomic signature and the identification of new sporulation genes. J. Bacteriol. **195:**2101–2115.
50. **Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD.** 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. **38:**D473–D479.
51. **Vepachedu VR, Setlow P.** 2005. Localization of SpoVAD to the inner membrane of spores of *Bacillus subtilis*. J. Bacteriol. **187:**5677–5682.
52. **Britton RA, Eichenberger P, Gonzalez-Pastor JE, Fawcett P, Monson R, Losick R, Grossman AD.** 2002. Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis*. J. Bacteriol. **184:**4881–4890.
53. **Li B, Chen Y, Liu Q, Hu S, Chen X.** 2011. Complete genome analysis of *Sulfobacillus acidophilus* strain TPY, isolated from a hydrothermal vent in the Pacific Ocean. J. Bacteriol. **193:**5555–5556.
54. **Junier P, Junier T, Podell S, Sims DR, Detter JC, Lykidis A, Han CS, Wigginton NS, Gaasterland T, Bernier-Latmani R.** 2010. The genome of the Gram-positive metal- and sulfate-reducing bacterium *Desulfotomaculum reducens* strain MI-1. Environ. Microbiol. **12:**2738–2754.
55. **Teng F, Guan Y, Zhu W.** 2008. A simple and effective method to overcome the inhibition of Fe to PCR. J. Microbiol. Methods **75:**362–364.
56. **Lomstein BA, Langerhuus AT, D'Hondt S, Jorgensen BB, Spivack AJ.** 2012. Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. Nature **484:**101–104.
57. **Vreeland R, Rosenzweig W, Powers D.** 2000. Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. Nature **407:**897–900.
58. **Pirttijarvi TSM, Graeffe TH, Salkinoja-Salonen MS.** 1996. Bacterial contaminants in liquid packaging boards: assessment of potential for food spoilage. J. Appl. Bacteriol. **81:**445–458.
59. **Yung PT, Ponce A.** 2008. Fast sterility assessment by germinable-endospore biodosimetry. Appl. Environ. Microbiol. **74:**7669–7674.
60. **Kibbe WA.** 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. **35:**W43–W46.
61. **Sauvain L, Bueche M, Junier T, Masson M, Wunderlin T, Kohler R, Gascon Diez E, Loizeau J-L, Tercier-Waeber M-L, Junier P.** Bacterial communities in trace metals contaminated lake sediments are dominated by endospore forming bacteria. Aquat. Sci., in press.