Published in Environmental Microbiology Reports 5, issue 6, 911-924, 2013
which should be used for any reference to this work

Stage 0 sporulation gene A as a molecular marker to study diversity of endospore-forming Firmicutes

Tina Wunderlin, Thomas Junier, Ludovic Roussel-Delif, Nicole Jeanneret and Pilar Junier*
Laboratory of Microbiology, Institute of Biology, University of Neuchâtel, Neuchâtel CH-2000, Switzerland.

Abstract

In this study, we developed and validated a culture-independent method for diversity surveys to specifically detect endospore-forming Firmicutes. The global transcription regulator of sporulation (*spo0A*) was identified as a gene marker for endospore-forming Firmicutes. To enable phylogenetic classification, we designed a set of primers amplifying a 602 bp fragment of *spo0A* that we evaluated in pure cultures and environmental samples. The amplification was positive for 35 strains from 11 genera, yet negative for strains from *Alicyclobacillus* and *Sulfobacillus*. We also evaluated various DNA extraction methods because endospores often result in reduced yields. Our results demonstrate that procedures utilizing increased physical force improve DNA extraction. An optimized DNA extraction method on biomass pre-extracted from the environmental sample source (indirect DNA extraction) followed by amplification with the aforementioned primers for *spo0A* was then tested in sediments from two different sources. Specifically, we validated our culture-independent diversity survey methodology on a set of 8338 environmental *spo0A* sequences obtained from the sediments of Lakes Geneva (Switzerland) and Baikal (Russia). The phylogenetic affiliation of the environmental sequences revealed a substantial number of new clades within endospore-formers. This novel culture-independent approach provides a significant experimental improvement that enables exploration of the diversity of endospore-forming Firmicutes.

Introduction

Endospore-formers are Gram-positive bacteria from the phylum Firmicutes, although not all species in this phylum can form endospores. In culture collections, Firmicutes represent the second most abundant bacterial phylum known (Klenk and Goker, 2010). For the endospore-forming species, the resilient outer cortex of the endospores and the small acid-soluble proteins stabilizing their DNA (Driks, 2002; Onyenwoke *et al*., 2004; Yudkin and Clarkson, 2005) facilitate their dispersion and capacity to colonize every habitat on Earth (Staley and Gosink, 1999). Indeed, endospore formers have been found in a wide range of environments on Earth's surface and subsurface (Nicholson *et al*., 2000; Nicholson, 2002). Although molecular biology techniques have greatly contributed to the general comprehension of microbial diversity, investigation of the diversity of endospore- and exospore-forming bacteria remains problematic and warrants improved methodology. In a recent phylogenetic assessment of microbial communities in a diverse set of environments, a surprisingly small number of known microbial groups containing spore-formers were observed (von Mering *et al*., 2007). Although the frequency of endospore-formation in Firmicutes varies significantly among four different environments, one explanation for their underrepresentation in genomic analyses is that spores can resist the protocols used for extracting DNA from vegetative cells.

Previous studies have identified a number of common genetic elements for endospore formation (Arcuri *et al*., 2000; Onyenwoke *et al*., 2004; Paredes *et al*., 2005; Dubey *et al*., 2009). Additionally, recent work in comparative genomics yielded a comprehensive set of the genetic elements involved in forming a minimal sporulation core (Galperin *et al*., 2012; Abecasis *et al*., 2013). However, this information has not yet been translated into the development of specific molecular markers for diversity surveys of endospore-forming Firmicutes in environmental samples.

The aim of this study was to develop a culture-independent approach to reveal the diversity of endospore-forming Firmicutes. To achieve this, we identified a functional marker for endospore formation from the genes involved in the sporulation pathway. Furthermore, because the suitability of different DNA extraction

methods emerged as a potential caveat for the detection of endospore-formers, the primer design was complemented by experiments testing different DNA extraction methods on cultures and on lake sediment samples. As a final step, to target the endospore-forming fraction of the bacterial community and discover its diversity, we amplified and sequenced the sporulation gene *spo0A* directly from DNA extracted from sediments. To the best of our knowledge, this is the first environmental set of *spo0A* sequences, whose subsequent analysis reveals a large diversity of endospore-forming bacteria.

Results and discussion

Identification of molecular markers for endospore-forming bacteria

From an initial data set of 59 genome sequences of endospore-forming Firmicutes, including both finished (48) and draft (11) genomes, 27 genomes were selected for the search of common orthologous sporulation genes (Supporting Information Table S1). Redundant species were reduced to a single species representation to avoid over-fitting to specific species. Furthermore, the amount of genomes was reduced given that we observed a large variation in the number of annotated sporulation-associated genes in the 59 initial genomes. Part of this variation was explained by annotation problems in the uncompleted genomes. For example, when the distribution of sporulation-associated genes was analysed in the 59 genomes, two peaks: one around 60 genes and a second at 190 genes were detected (data not shown). However, using only well-annotated 'finished' genomes, there was a shift in the distribution towards 80–90 minimal genes. Therefore, to avoid any exclusion of orthologs by annotation errors, only finished genomes with more than 60 sporulation-related genes were considered. This minimal number of genes coincides with recent results suggesting that ~60 protein-coding genes are essential for sporulation in Bacilli and Clostridia (Galperin *et al*., 2012).

We selected these 27 genomes in order to create a balance between Bacilli (12) and Clostridia (15), and to prevent biases because of phylogenetic distribution. These genomes originated from diverse habitats including soil (7), freshwater (2), sediment (2), clinical samples (7), deep surface habitats (3), hot springs (3) and others (3). The majority of the genomes (20) correspond to mesophilic microorganisms; six are thermophilic and one psychrotolerant. Additionally, one of the mesophilic species (*Alkaliphilus metalliredigens* QYMF) was reported to be both halophile and alkalophile.

Orthology groups were delineated based on best reciprocal Basic Local Alignment Search Tool for Proteins (BLASTP) hits (Altschul *et al*., 1997) on the annotated sporulation genes from the 27 genomes. Each sequence in the set was BLASTPed against all sequences except those of the same species (thus avoiding paralogs). The best hit in each species was retained, and sequence pairs that were each other's best match were defined as best reciprocal hits. Putative orthology groups were defined using the algorithm used by OrthoDB (Kriventseva *et al*., 2008). In this manner, six orthologous genes (*spo0A*, *spoIVB*, *spoVAC*, *spoVAD*, *spoVT* and *gpr*) were found to be common and highly conserved among endospore-forming Firmicutes (Supporting information Table S2).

All six genes are part of the core sporulation gene set that seems to be indispensable for sporulation, appearing in both classes of endospore-forming Firmicutes: the Clostridia and Bacilli (Galperin *et al*., 2012). A phylogenetic reconstruction based on the concatenated sequences of these six genes was similar to the phylogeny inferred from the 16S rRNA gene (Supporting Information Fig. S1). In particular, the phylogeny showed a clear separation between Bacilli and Clostridia.

Based on the analysis of the phylogeny and conservation profile of the individual genes, *spo0A* was chosen as a molecular marker. The phylogenetic reconstruction based on Spo0A sequences alone (Supporting information Fig. S2) was consistent with the phylogeny based on the core genes and the 16S rRNA gene (Supporting information Fig. S1), and supports a recent report on the separation of the *Bacillus subtilis* and *Bacillus cereus* clades (Bhandari *et al*., 2013). Additionally, the conservation profile showed two highly conserved regions flanking a highly variable region covering ~ 300 bp (Supporting information Fig. S2).

The stage 0 sporulation gene A (*spo0A*) is the master regulator of sporulation. No convincing homolog of *spo0A* has been found outside the Firmicutes (Brill and Wiegel, 1997; Onyenwoke *et al*., 2004). A recent profile analysis of *spo0A* on 626 genomes found one putative orthologous sequence for each of the 46 endospore-forming genomes and one ortholog in a single non-endospore forming genome (Traag *et al*., 2013). The ability of some Firmicutes species to form endospores has not yet been experimentally confirmed, even though they contain the *spo0A* gene, and are thus defined as asporogenic (Galperin *et al*., 2012). Some asporogenic species might have truly lost the trait of sporulation in the course of evolution, but they still conserve the response regulator gene as a relic of this. Such species could trigger false-positives when using *spo0A* as a functional marker. Conversely, based on the analysis of the *spo0A* gene in some asporogenic strains, it could be that some are actual endospore-formers; however, the phenotype has not been observed (Abecasis *et al*., 2013). Because *spo0A* is one of the best-studied sporulation genes, it is often annotated automatically, leading to a rapidly growing database of *spo0A* sequences. Although the risk for false-positive

exists, *spo0A* can nonetheless be considered an ideal candidate as a molecular marker to target endospore-forming Firmicutes in environmental samples.

Design and validation of spo0A primers

Degenerate primers for diversity studies that amplify a 602 bp sequence of the *spo0A* gene were designed. The *spo0A* genes of the 27 genomes previously mentioned were aligned using CLUSTALW (Thompson *et al*., 1994) and scanned for conserved regions. Seven degenerate forward primer regions, and 10 reverse primer regions were defined. As a first screen, these primers were tested in all combinations. Based on specificity, amplification efficiency and fragment length, the primer sequences spo0A166f (5′-GATATHATYATGCCDCATYT-3′) and spo0A748r (5′-GCNACCATHGCRATRAAYTC-3′) were selected (Fig. 1).

To validate the *spo0A* primers, amplification efficacy was determined using a collection of 53 pure cultures (Table 1). The cultured strains corresponded mainly to the class Bacilli and in particular to *Bacillus* spp., with a few strains from other genera such as *Anoxybacillus*, *Brevibacillus*, *Geobacillus*, *Halobacillus*, *Lysinibacillus* and *Paenibacillus*. A polymerase chain reaction (PCR) product with the correct size (602 bp) was obtained in 35 out of 43 endospore-forming bacterial cultures belonging to nine different genera. From the strains tested, some did not yield a PCR amplicon (e.g. *Bacillus brevis* or *Geobacilllus themoparaffinovorans*), but overall, the primers demonstrated good coverage. Three strains from the genera *Alicyclobacillus* and one *Sulfobacillus* strain were also included but did not amplify with the primers. The match between the primer sequences and the *spo0A* gene sequence in two available genomes for these genera (Supporting information Fig. S3) revealed that in the case of *Alicyclobacillus acidocaldarius* TC41, there are five mismatches with the forward primer, four of which are found in the 3′ region that could impair annealing and amplification. However in the case of *Sulfobacillus acidophilus*, one mismatch with each primer was observed, and thus, the failure of the amplification is surprising. A subsequent inhibition test suggests that the lack of amplification was probably due to a chemical remnant from the culture medium (data not shown). Only three Clostridia strains could be tested, and two of them gave a positive amplification signal (*Clostridium pasteurianum* and *Desulfotomaculum reducens*). None of the 10 non-endospore formers amplified with the primers. The non-endospore formers included three exospore-formers (Actinobacteria), one non-endospore-forming Firmicute (*Lactococcus lactis* subsp. *lactis*) and six members from outside the Firmicutes (five Proteobacteria and one Bacteroidetes). A detailed protocol of the PCR conditions is given in Appendix S1.

Comparison of DNA extraction methods on cells and endospore preparations

We conducted experiments testing different DNA extraction methods in order to examine and optimize the extractability of DNA from endospores. We used the commercially available FastDNA Spin kit for soil (MP Biomedicals, Solon, OH, USA), previously shown to produce high DNA yields and a relatively good phylogenetic distribution from soil samples and low biomass samples from the deep biosphere (Webster *et al*., 2003; Dineen *et al*., 2010). The use of commercially

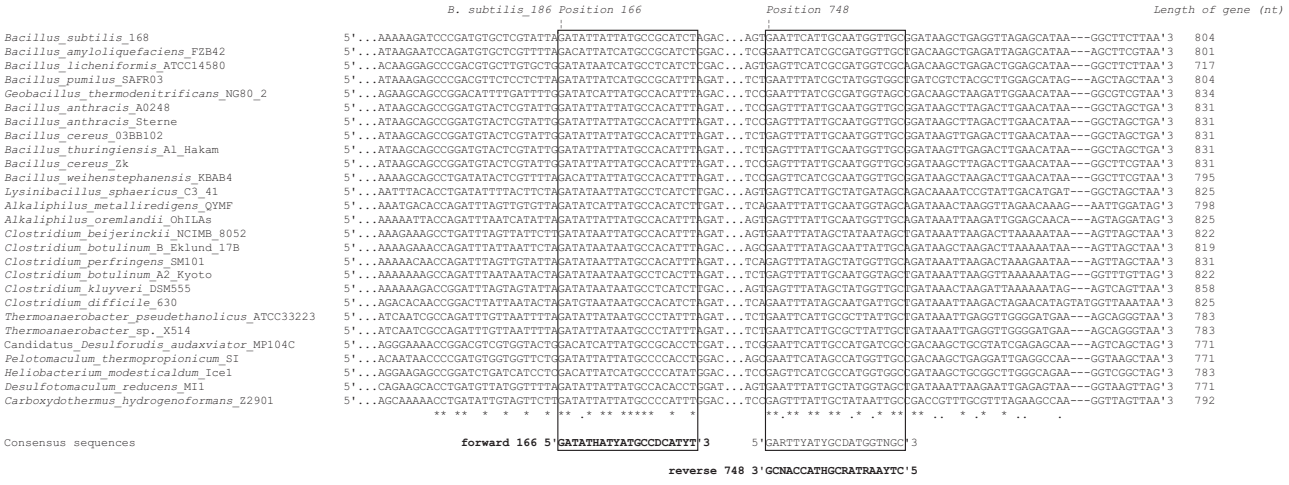

Fig. 1. Alignment of 27 *spo0A* gene sequences encompassing regions used for primer design. The position of the primers is indicated on top of the figure using the *spo0A* gene from *Bacillus subtilis* as reference. The annealing sites are marked by two squares. The consensus sequence is shown underneath (in bold). For the reverse primer, the reverse complement sequence (primer sequence) is also indicated. The degenerate positions in the primers are shown by the letters H (A, C or T), Y (C or T), D (A, G or T), R (A or G), N (A, C, G or T).

Table 1. Specificity test for the amplification of spo0A using the primers spo0A166f and spo0A748r.

Genus	Species	Optimal growth Temperature (°C)	Endospores (y/n)	Amplification of *spo0A*
Alicyclobacillus	*acidocaldarius*	55	y	–
Alicyclobacillus	*tolerans*	50	y	–
Alicyclobacillus	sp.	50	y	–
Anoxybacillus	sp.	70	y	+
Anoxybacillus	sp.	55	y	+
Bacillus	*aquimaris*	25	y	+
Bacillus	*brevis*	30	y	–
Bacillus	*cereus*	30	y	+
Bacillus	*cereus* var. *mycoïdes*	24	y	+
Bacillus	*horikoshii*	25	y	+
Bacillus	*jeotgali*	45	y	+
Bacillus	*licheniformis*	45	y	+
Bacillus	*macerans*	30	y	+
Bacillus	*niabensis*	20	y	+
Bacillus	*niacini*	45	y	+
Bacillus	*oceanisediminis*	45	y	+
Bacillus	*pallidus T*	60	y	+
Bacillus	*polymyxa*	30	y	+
Bacillus	*pumilus*	30	y	+
Bacillus	*selenatarsenatis*	37	y	+
Bacillus	*stearothermophilus*	55	y	+
Bacillus	*subtilis*	30	y	+
Bacillus	*thermoglucosidasius*	65	y	+
Bacillus	*thermoruber*	45	y	+
Bacillus	*thuringiensis*	45	y	+
Bacillus	*tusciae*	55	y	–
Bacillus	*firmus*	20	y	+
Bacillus	*vietnamensis*	37	y	+
Bacillus	sp.	45	y	+
Brevibacillus	*agri*	30	y	+
Brevibacillus	*formosus*	30	y	+
Clostridium	*pasteurianum T*	37	y	+
Clostridium	sp.	30	y	–
Desulfotomaculum	*reducens*	NA	y	+
Geobacillus	sp. A14	50	y	+
Geobacillus	*thermoleovorans*	50	y	+
Geobacillus	*thermoparaffinovorans*	70	y	–
Geobacillus	sp.	65	y	+
Halobacillus	*trueperi*	70	y	+
Lysinibacillus	*sphaericus*	30	y	+
Lysinibacillus	sp.	30	y	+
Paenibacillus	*alvei*	30	y	+
Sulfobacillus	*acidophilus*	50	y	–
Actinomyces	sp.	24	n	–
Escherichia	*coli*	37	n	–
Hymenobacter	*daecheongensis*	30	n	–
Lactococcus	*lactis subsp. lactis*	37	n	–
Streptomyces	*grisechromogenes*	30	n	–
Streptomyces	sp. nu40	30	n	–
Stenotrophomonas	*rhizophila* SMPG9	20	n	–
Comamonas	sp. lb15	20	n	–
Delftia	sp. S17	20	n	–
Pseudomonas	*fluorescens* NBRC12568	20	n	–

A PCR product of the correct size (602 bp) is indicated by a '+' sign, no PCR product is indicated by '–'. A total number of 53 strains were analysed, of which 43 are endospore-forming Firmicutes. DNA was extracted using the innuPREP Bacteria DNA Kit (Analytik Jena, Jena, Germany). PCR reactions were performed with 0.5 ng DNA template, 1× reaction buffer (TaKaRa Bio, Shiga, Japan), 3 mM $MgCl_2$, 10 μg bovine serum albumin (BSA; New England Biolabs, Ipswich, MA, USA), 1 U of ExTaq Polymerase (TaKaRa), 200 μM of each dNTP and 1 μM of each primer in a total reaction volume of 50 μl, completed with PCR-grade water. Reactions were performed in an Arktik Thermo Cycler (Thermo Fisher Scientific, Vantaa, Finland) with the following temperature program: initial denaturation at 94°C for 5 min; then 10 cycles of denaturation at 94°C for 30 s, touchdown annealing starting at 55°C with decrease of 0.3°C per cycle for 30 s and elongation at 72°C for 1 min; followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s and elongation at 72°C for 1 min; and a final extension at 72°C for 5 min. The size and amplification result was verified by running the products on a 1% agarose gel stained in 3× GelRed bath (Biotium, Hayward, CA, USA).

available products increases repeatability and standardization of the extraction procedure.

A DNA extraction protocol composed of three repetitive extraction cycles (to increase the total mechanical disruption by bead-beating time) was first tested on cell cultures of *Lactobacillus lactis* subsp. *lactis*, a non-spore-forming Firmicute, and on cell and endospore preparations of *Paenibacillus alvei* and *Bacillus subtilis* (Fig. 2). The total, cumulative yield of DNA isolated from cell cultures (normalized to 10^8 cells) was 198.8 ± 72.8 ng for *Lactobacillus lactis* subsp. *lactis*, 497.6 ± 36.9 ng for *B. subtilis* and 1402.3 ± 254.8 ng for *P. alvei*. The yields from endospore preparations were significantly lower ($P = 0.002$) with 86.5 ± 3.3 ng for *B. subtilis* and 83.1 ± 2.0 ng for *P. alvei.* Over three consecutive extraction cycles, the quantity of isolated DNA increased considerably; the total yield could almost be doubled when adding a second and third round of extraction. This was especially true for endospore preparations, where the overall percentage of the total DNA isolated after the first extraction step was significantly lower (average $57.8\% \pm 5.8\%$, $P = 0.009$) than that of the vegetative cells (average $76.0\% \pm 13.6\%$). After the second extraction, the percentage of isolated DNA from endospores was still significantly lower (average $81.1\% \pm 4.1\%$, $P = 0.039$) than from cells (average $92.3.0\% \pm 9.7\%$). This result agrees with previous studies showing that Gram-positive cells or endospores only lyse with harsh physical methods (More *et al*., 1994; Zhou *et al*., 1996; Kuske *et al*., 1998). In our experience, three successive cycles is the best balance of time, cost, and overall DNA yield and quality of the DNA extract.

Test of DNA extraction methods on environmental samples

Different DNA extraction protocols were then tested on sediment samples collected during a research campaign with the Mir manned submersibles in Lake Baikal (Russia) and Lake Geneva (Switzerland). Sediment cores were retrieved using a push-corer. Upon return to the surface, the core fractions 2–7 cm were immediately subsampled in the centre using sterile cut-open syringes. Samples were then stored at −20°C until DNA extraction.

Three DNA extraction protocols were tested, all based on the MP Fast DNA Spin Kit for Soil. Protocol 1 (standard) corresponded to a standard extraction with *in situ* lysis in 0.5 g sediment following the manufacturer's instructions. Protocol 2 (repeated) also corresponded to *in situ* lysis in 0.5 g sediment, but with three sequential extractions, as was used for cells and endospores (see Fig. 2). In Protocol 3 (indirect), the biomass was separated from sediment particles prior to lysis. In this case, 3 g of sediment were homogenized with 15 ml of dispersing agent (1% Na-hexa-meta-phosphate) using an Ultra-Turrax homogenizer (IKA, Staufen, Germany) at 15500 rpm for 2 min to separate cells from the sediment matrix. Coarse particles were then removed from the slurry by centrifugation at $20 \times g$ for 1 min, and the supernatant (containing the cells) was collected on a nitrocellulose membrane of 0.2 μm pore size (Whatman, Dassel, Germany). The cell separation step was then repeated. Filters were immediately frozen in liquid nitrogen and stored at −80°C. DNA was extracted directly from the membrane using Protocol 2. For the latter two protocols (2 and 3), the three individual extracts were pooled. DNA precipitated with of 0.3 M Na-acetate and 2 volumes of absolute ethanol, then washed with ethanol (70%) before being resuspended in sterile water. A detailed protocol of the DNA extraction method is provided in the Appendix S1.

For the environmental samples, the DNA yields and humic acid contamination (determined by absorbance ratio at 260/230 nm) of the different protocols varied (Table 2). The DNA yield after the repeated extractions (protocol 2 and 3) was lower than after the standard method, particularly for the indirect extraction (protocol 3). Cell lysis alone is therefore not the determining factor for DNA yield. Lower yields could be due to increased adsorption of DNA to clay particles when bead-beating for longer times (Frostegard *et al*., 1999) because of disturbance of DNA-silica binding from co-extracted humic acids or salts, or because of the exclusion of specific morphological groups by the biomass separation procedure used in the indirect extraction protocol. Overall, DNA extracts from sediments of Lake Geneva were less contaminated with humic acids, visually obvious given the colour of the extract. The DNA extracted from sediment of Lake Baikal had lower purity (brownish colour) and lower quantity.

Gene abundances of the 16S rRNA gene and the *spo0A* gene were then determined (Bueche *et al*., 2013). A detailed protocol of the qPCR method is given in Appendix S1. We observed an inverse correlation between DNA yields and gene copy numbers of 16S rRNA and the *spo0A* genes. There was a 2.6-fold increase of 16S rRNA genes and 2.9-fold increase of detection of *spo0A* gene copy numbers in the extract from Lake Geneva with the indirect method when compared with the standard protocol. In extracts from sediment of Lake Baikal, the increase was even more prominent: 2.2-fold for the 16S rRNA genes and 4.2-fold for *spo0A*. Copy numbers of extracts from the multicycle protocol were always intermediate. The percentages of *spo0A* genes relative to 16S rRNA genes were constant for samples from the same sediment, independently of the extraction protocol (Lake Geneva sediment $0.063\% \pm 0.005\%$ and Lake Baikal sediment $0.27\% \pm 0.075\%$).

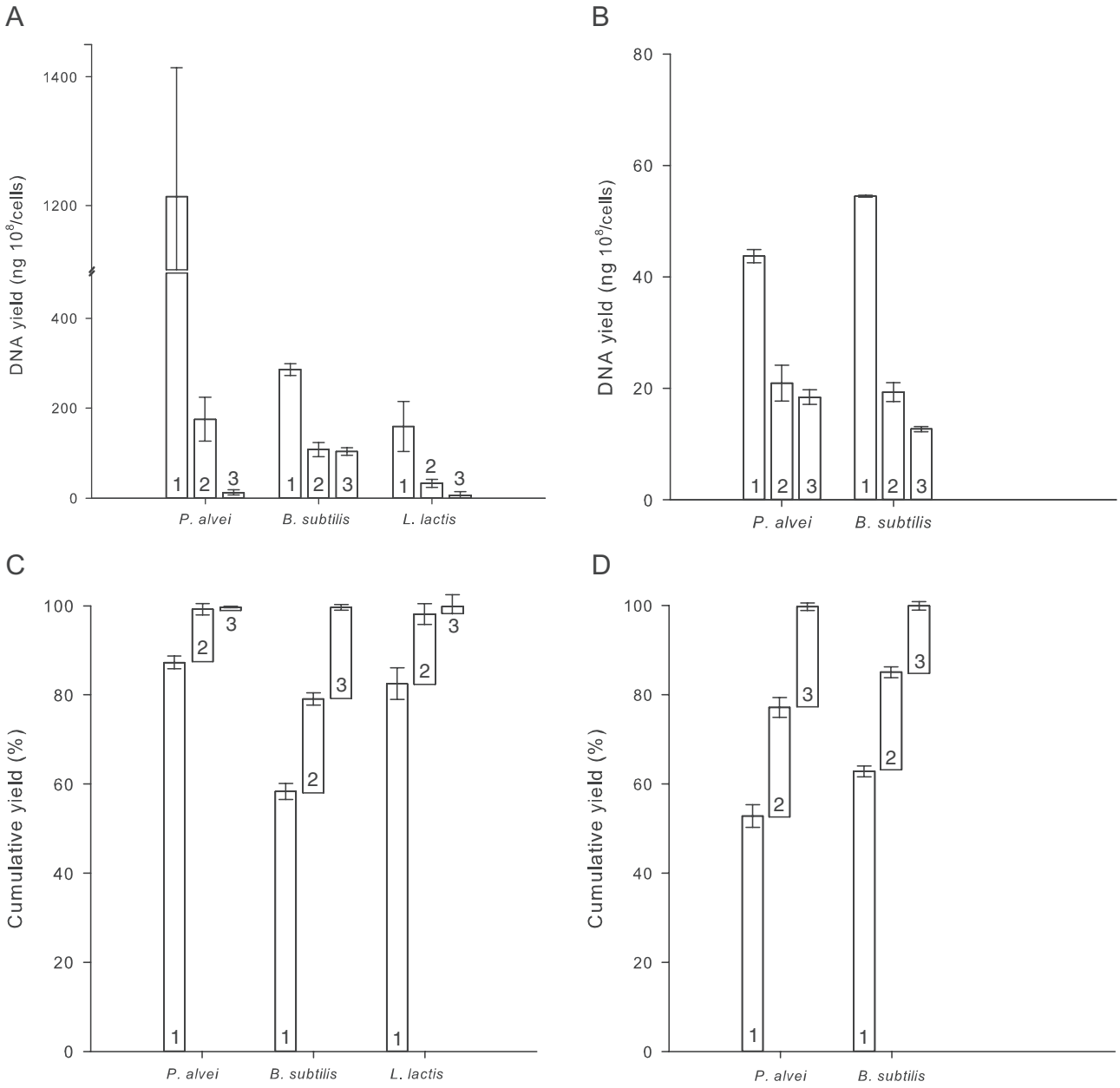

Fig. 2. DNA extraction yields (ng 10^8 cells or spores^{-1}; top panels) and cumulative yield (in percentage of the total; lower panels) obtained for each sequential extraction steps (1, 2 and 3) for cell cultures (A, C) and endospore preparations (B, D). Cell cultures of *Paenibacillus alvei* and *Bacillus subtilis* were grown in nutrient broth and *Lactococcus lactis* subsp. *lactis* in German collection of microorganisms and cell cultures (DSMZ) medium 92 at 30°C. Endospore preparations of *P. alvei* and *B. subtilis* were obtained with Schaeffer sporulation medium (Schaeffer *et al.*, 1965), by vigorous shaking until cultures were composed of > 90% spores. Numbers of cells and endospores were determined microscopically using a Neubauer counting chamber. Cell and endospore preparations were then centrifuged at 6000 × *g* for 10 min, re-suspended in phosphate-buffered saline solution (PBS) to a density of 10^8 cells/ml and 10^9 endospores/ml, respectively. Four millilitres of this preparation was collected by centrifugation at 10 000 × *g* for 10 min, and pellet was stored at −20°C until DNA extraction. DNA extraction with i*n situ* lysis and a repetitive protocol was performed by bead-beating at 50 strokes/s with the TissueLyser LT (QIAGEN, Hilden, Germany) for 10 min. The sample was then centrifuged, and 900 µl of supernatant fluid (containing free DNA) was collected in a separate tube. Lysis buffer was again added to the sample pellet before subjecting to a second round of bead-beating for 5 min, then centrifuged and supernatant fluid collected. This procedure was repeated a third time. The three supernatants were then processed separately following the standard protocol. DNA yield was measured with a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) using the Quant-iT dsDNA BR assay kit, following the manufacturer's instructions.

Table 2. Comparison of DNA extraction protocols on sediment samples from Lakes Geneva and Baikal. DNA was extracted according to standard (protocol 1) or modified extraction methods (protocol 2 –repetitive; protocol 3 –indirect).

	DNA extraction protocol	DNA yield ng[a]	A260/230	Abundance 16S rRNA gene[b]	Abundance *spo0A* gene[c]	Ratio *spo0A* vs. 16S rRNA genes (%)	No. of initial sequences	No. of curated sequences	No. of OTUs
Lake Geneva	1	7872	1.50	25.9 ± 2.3	147.8 ± 9.8	0.06	2349	804	91
	2	3888	1.30	20.4 ± 1.6	144.2 ± 24.1	0.07	2703	926	13
	3	2240	1.43	68.6 ± 1.7	428.8 ± 27.1	0.06	4210	1590	409
Lake Baikal	1	2480	0.84	17.5 ± 0.2	339.4 ± 30.3	0.19	3737	1470	212
	2	2176	0.87	24.3 ± 2.6	611.0 ± 41.3	0.25	4759	1720	71
	3	616	0.89	38.0 ± 1.9	1414.7 ± 56.3	0.37	4089	1828	289

DNA yield was measured using the Quant-iT dsDNA BR assay kit. DNA quality was verified by spectrophotometer absorbance at 260 and 230 nm using a NanoDrop ND-1000 (NanoDrop, Wilmington, DE, USA). Quantification of bacterial DNA in sediment extracts was carried out by real-time quantitative PCR of the V3 region of the 16S rRNA gene with primers 338f and 520r (Ovreås *et al*., 1997). The qPCR mix contained 1 μl of 10-fold diluted DNA template (1.3 to 8.4 ng/μl), 0.3 μM of each primer and 10 μl of QuantiTect SYBR® Green PCR Kit (QIAGEN). Reaction volume was brought to a total of 20 μl with PCR-grade water. The qPCR was run on a Rotor-GeneTM 6000 instrument (QIAGEN) with the program: enzyme activation at 95°C for 5 min, 40 cycles of denaturation at 95°C for 5 s, annealing at 55°C for 15 s and extension at 72°C for 20 s. Thresholds (Th), Ct values and derivatives of melting curves were determined using Rotor-Gene 6 software. All extracts were analysed in triplicates. For quantification three independent plasmid standards series with 300–3 000 000 gene copies/μl of the 16S rRNA gene of an environmental clone were included. Quantification of *spo0A* gene was done as mentioned earlier for the 16S rRNA gene but with the primers *spo0A*655f and *spo0A*923r (Bueche *et al*., 2013). The qPCR mix contained 1 μl of 10-fold diluted DNA sample (1.3–8.4 ng/μl), 0.76 μM of each primer and 1 × QuantiTect SYBR Green PCR Kit. Total reaction volume of 20 μl was reached with PCR-grade water. The program differed in an annealing at 52°C for 30 s and extension at 72°C for 30 s. For quantification, three independent plasmid standards series with 30–300 000 gene copies/μl of *spo0A* gene of *B. subtilis* were included.

a. Per gram sediment.

b. Times 10^4/ng DNA extracted per gram sediment.

c. Per nanogram DNA extracted per gram sediment.

In summary, extracts from the indirect protocols had substantially better amplification despite relatively lower DNA yields. This is most likely due to reduced co-extraction of contaminants that could inhibit the downstream PCR. The same effect is observed with direct *in situ* DNA extraction that provides high yields but lower purity (Leff *et al*., 1995), often requiring high dilution of the extracts in order to avoid amplification inhibition due to contaminants (Dineen *et al*., 2010).

Application of the spo0A primer on environmental samples

A 602 bp fragment was amplified from the environmental samples with the *spo0A* primers (described above) and sent for sequencing on a Roche GS FLX+ (Eurofins MWG Operon, Ebersberg, Germany) to assess the diversity of endospore-forming Firmicutes. A total of 21 847 sequences were obtained from the six samples. Sequences were binned according to their barcode and the corresponding sample origin (Lake and DNA extraction protocol) and filtered according to Phred (Ewing and Green, 1998) quality score (minimum of 30) and length. After curation of amplicon size and quality, a total of 8338 sequences with an average length of 625 bp remained and were translated to their amino acid sequence and checked for stop-less open reading frames. The numbers of *spo0A* sequences varied substantially between the different extraction protocols as well as between the two sediments (Table 2). We observed varying amplification efficiencies depending on sediment type, purity of extract and community composition, among others. In this study, the amount of final *spo0A* sequences was the greatest, in both sediments (Lakes Geneva and Baikal), in the samples prepared with the indirect extraction protocol. All metagenomic sequences were submitted to Sequence Read Archive of the National Center for Biotechnology Information (NCBI) under Accession Numbers SRR870694, SRR870695, SRR870696, SRR870698, SRR870699 and SRR870700.

Phylogenetic distribution of spo0A sequences

De novo operational taxonomic units (OTUs) from the curated sequences were defined with the pick_otus.py program (QIIME package) using the Uclust method applying a cut-off of 97% nucleotide identity based on the definition of OTUs applied for the 16S rRNA gene (Caporaso *et al*., 2010). This OTU assignment threshold has not yet been experimentally validated for *spo0A* and therefore is based solely on current knowledge defining bacterial species based on the 16S rRNA gene. The number and distribution of the OTUs varied with the extraction method (Table 2). For Lake Geneva sediment, a fourfold increase in the number of OTUs was obtained in the samples extracted with the indirect protocol over the standard protocol. In sediment from Lake Baikal, the differences were smaller, with repeated extraction representing ~25% of the number of OTUs from the indirect extraction.

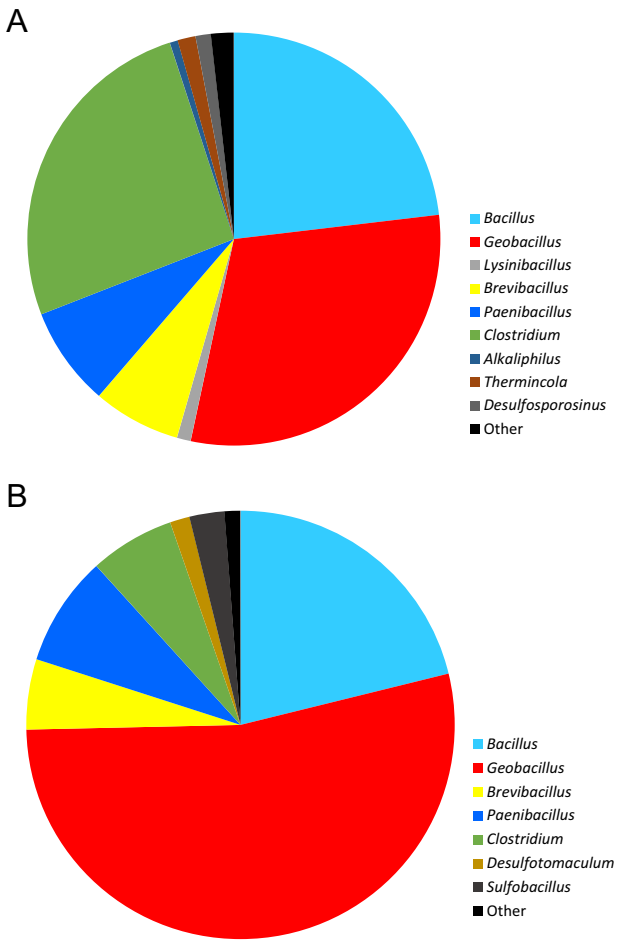

Fig. 3. Distribution of *Spo0A* OTUs from the indirect DNA extraction method classified into genera using BLASTP (Altschul *et al*., 1997) against a *Spo0A* database containing all the sequences in InterPro (Mulder *et al*., 2002). Lake Geneva (A). Lake Baikal (B).

All defined OTUs were then displayed with a phylogenetic tree (Supporting information Figs S4 and S5). The branches were collapsed according to the extraction method. Overall, the indirect method is the most promising, revealing entire clusters that do not appear in data from the other two extraction protocols. This result confirms previous research where different community profiles are detected when comparing direct or indirect DNA extraction protocols on the same soil sample (Delmont *et al*., 2011). Additionally, successive extractions can result in a shift in the community composition (Feinstein *et al*., 2009), as was observed here. However, in the case of endospore-forming Firmicutes, repetitive extractions (protocol 2) from the same sediment sample produced a poor representation of the community with groups that are either not represented at all (e.g. Bacilli in Lake Geneva, yellow branches in Supporting information Fig. S4) or underrepresented (e.g. Geobacilli in Lake Baikal, yellow branches in Supporting information Fig. S5). Separation of cells from the sediment matrix prior to DNA extraction requires additional laborious and time-consuming steps. However, with respect to the time and cost of downstream processes (sequencing, analysis and data storage), it was worth increasing the effort of applying an indirect and repetitive DNA extraction method, in particular for endospore-forming Firmicutes prone to be underrepresented, as shown in this study.

OTUs from the samples from the indirect extraction method were then compared using BLASTP (Altschul *et al*., 1997) against a *Spo0A* database containing all the sequences in InterPro (Mulder *et al*., 2002) in order to identify the closest *Spo0A* sequence belonging to a known genus (Fig. 3 and Table 3). In both samples (Lake Geneva and Lake Baikal), the OTUs could be assigned to sequences belonging to both classes; the Clostridia and Bacilli. The detection of both classes of endospore-forming Firmicutes with the primers for *spo0A* supports the results of the primer validation in pure cultures. Furthermore, we can detect a broader range of endospore-forming Firmicutes, including strains that were not present in the pure cultures. This included one additional genus of Bacilli (*Solibacillus* spp. detected in Lake Geneva) and 11 additional genera of Clostridia. Moreover, the detection of 48 OTUs related to *Sulfobacillus* spp. in Lake Baikal supports the analysis of the annealing sites (Supporting information Fig. S3) and PCR inhibition because of the composition of the culture medium to grow *S. acidophilus*.

For Lake Geneva, the OTUs could be assigned to 48 groups according to the closest *Spo0A* sequence (Table 3). The most abundant genera were *Geobacillus* (30.2% of all sequences), *Clostridium* (25.9%), and *Bacillus* (23.1%), *Paenibacillus* (7.8%) and *Brevibacillus* (6.9%). *Lysinibacillus*, *Alkaliphilus*, *Thermincola* and *Desulfosporosinus* were between 0.6% and 1.4%, and the remaining 1.8% of the sequences were assigned to seven other genera (Fig. 3A). The distribution of the groups within clusters in the cladogram was verified (Supporting information Fig. S4). For some of the species, the position in the cladogram was consistent with the affiliation by BLAST. For example, the groups J, B, D, E, I, G, H, A, F, C and M clustered together within Bacilli. Likewise, all the groups consisting of *Spo0A* related to Clostridia clustered together (clusters 27–60). However, OTUs in large groups (i.e. groups B, D, AF, C, T, V, Y and Z) appeared distributed in several clusters in the cladogram. Surprisingly, several clusters within the groups AF, L, AG and AH, which were affiliated to *Geobacillus*, *Brevibacilllus* or *Paenibacillus*, formed a third branch (indicated as undefined in the cladogram) more closely related to Clostridia than to Bacilli. For these groups, the identity levels of *Spo0A* were in some cases very low (down to 24%).

Table 3. Identification of the closest relative for *spo0A* sequences extracted with protocol 3 (indirect method) by BLAST using the translated *Spo0A* protein sequences against a reference database of 581 *Spo0A* protein sequences from InterPro (Mulder *et al.*, 2002).

			Lake Geneva sample				Lake Baikal sample			
Genus	species	Group	Seq #	Id %	E-value	Cluster	Seq #	Id %	E-value	Cluster
Bacillus	*Bacillus pseudofirmus* OF4	J	7	79–80	1.E-115	19				
	Bacillus methanolicus	B	244	75–90	3.E-134	2,4,6,7	171	77–90	4.E-130	6,7,9,10,12,28
	Bacillus cellulosilyticus ATCC 21833	AN					1	66	3.E-93	35
	Bacillus megaterium	AI	1	99	2.E-138	NDT	1	84	5.E-119	1
	Bacillus megaterium DSM 319	AK					1	94	8.E-113	3
	Bacillus cereus	U	1	81	1.E-117	34	3	81–84	6.E-119	NDT
	Bacillus cereus subsp. *cytotoxis* NVH 391-98	D	28	76–93	5.E-136	5,10,18,22	35	77–85	8.E-120	NDT
	Bacillus thuringiensis	E	10	84	2.E-121	9,23	55	81–86	5.E-123	5,14
	Bacillus mycoides	I	39	80–100	4.E-148	17	94	77–84	4.E-122	13
	Bacillus atrophaeus 942	G	4	86	2.E-126	15				
	Bacillus subtilis	H	1	99	5.E-150	16	1	29	6.E+00	NDT
	Bacillus amyloliquefaciens	A	3	75–85	9.E-120	1,21	1	29	8.E-02	NDT
	Bacillus licheniformis	AJ	4	75–100	6.E-149	NDT	23	65–75	4.E-108	2
	Bacillus anthracis	AV	2	25	3.E-01	NDT				
	Bacillus pumilus	AW	4	99–100	1.E-151	NDT				
	Bacillus pumilus SAFR-032	AX	15	98–99	1.E-149	NDT				
	Bacillus weihenstephanensis BAB4	AY	1	98	8.E-144	NDT				
Geobacillus	*Geobacillus thermoleovorans*	AZ	1	93	1.E-108	NDT				
	Geobacillus thermodenitrificans NG80-2	F	5	98–100	6.E-144	12				
	Geobacillus sp. Y412MC10	AF	429	60–93	9.E-137	61,64,66,69,71,73,75	930	61–90	2.E-132	18,21,23,27,29,31
	Geobacillus sp. WCH70	C	41	70–94	1.E-138	3,8,11,13,14	47	73–94	3.E-137	4,8,11
Anoxybacillus	*Anoxybacillus flavithermus* DSM 21510 / WK1	AU	2	78–87	3.E-124	NDT				
Lysinibacillus	*Lysinibacillus sphaericus* C3-41	K	17	68–82	1.E-120	20				
Solibacillus	*Solibacillus silvestris* tLB046	BB	2	86–97	5.E-146	NDT				
Brevibacillus	*Brevibacillus laterosporus*	L	93	24–93	1.E-134	24,26,58,63,65,68	60	72–80	5.E-115	15,30,32,34
	Brevibacillus brevis 47	M	15	69–93	1.E-137	25	37	78–93	2.E-136	16
Paenibacillus	*Paenibacillus polymyxa*	AG	89	65–97	7.E-147	62,67,70	8	74–85	4.E-121	25
	Paenibacillus polymyxa E681	BA	3	98	3.E-146	NDT				
	Paenibacillus mucilaginosus KNP414	AH	29	84–90	2.E-130	72,74	118	71–88	1.E-125	17,20,24,26
	Paenibacillus sp. JDR-2	AM	1	90	1.E-131	NDT	26	58–91	1.E-131	22
Clostridium	*Clostridium clariflavum* DSM 19732	T	179	67–93	4.E-138	33,35,37,39	5	72–81	4.E-116	51
	Clostridium thermocellum	V	149	57–89	1.E-132	36,38,40	5	52–79	2.E-114	52
	Clostridium cellulolyticum DSM 5812	W	42	59–89	2.E-143	41	4	87–90	4.E-143	53
	Clostridium botulinum	Y	11	68–84	4.E-127	42,44,48,50,52,55,59	21	73–84	5.E-125	44,46,49

Table 3. *cont.*

			Lake Geneva sample				Lake Baikal sample			
Genus	species	Group	Seq #	Id %	E-value	Cluster	Seq #	Id %	E-value	Cluster
	Clostridium botulinum Kyoto / Type A2	AT					1	75	1.E-110	NDT
	Clostridium botulinum B Eklund 17B	BC	5	79–80	1.E-122	NDT				
	Clostridium novyi NDT	BD	3	80–84	1.E-126	NDT				
	Clostridium ljungdahlii DSM 13528	Z	11	77–86	1.E-131	43,46,49,51	49	69–89	3.E-137	43,45
	Clostridium perfringens	AA	3	100	4.E-155	47				
	Clostridium sporogenes	AB	5	75–76	1.E-112	53,54	18	74–76	9.E-113	48
	Clostridium haemolyticum	AL					1	48	2.E-27	19
	Clostridium butyricum	AR					4	79	7.E-120	42
	Clostridium kluyveri DSM 555	AS					9	83–84	1.E-126	47
Alkaliphilus	*Alkaliphilus metalliredigens* QYMF	AC	3	63–76	1.E-109	56				
	Alkaliphilus oremlandii OhILAs	AD	7	66–84	2.E-125	57	4	26–77	4.E-114	50
Thermincola	*Thermincola potens* JR	N	22	69–72	7.E-99	27	5	68–71	3.E-98	NDT
Candidatus Desulforudis	*Candidatus Desulforudis audaxviator* MP104C	AQ					8	70–72	3.E-98	39
Desulfotomaculum	*Desulfotomaculum acetoxidans* DSM 771	O	4	68–69	1.E-96	28	27	69–70	7.E-99	38,40
	Desulfotomaculum kuznetsovii DSM 6115	AE	1	66	7.E-90	60				
Syndtrophobotulus	*Syndtrophobotulus glycolicus DSM 8271*	R	2	75	1.E-112	31				
Desulfosporosinus	*Desulfosporosinus orieNDTis DSM 765*	P	19	71–81	4.E-125	29	5	79–81	3.E-124	41
Desulfitobacterium	*Desulfitobacterium hafniense*	BE	4	31–75	2.E-110					
Heliobacterium	*Heliobacterium modesticaldum* ATCC 51547	S	6	75–77	4.E-110	32				
Sulfobacillus	*Sulfobacillus acidophilus* TPY	AO					10	65–69	3.E-99	36
	Sulfobacillus acidophilus DSM 10332	AP					38	62–70	5.E-101	37
Thermosediminibacter	*Thermosediminibacter oceani* DSM 16646	Q	6	40	5.E-01	30				
Thermaerobacter	*Thermaerobacter marianensis* DSM 12885	BF	1	54	6.E+00	NDT				
	Abundance		1567				1827			
	Richness		48				35			

The identity range is listed as % of amino acid similarity. E-values are the minimum values for each closest relative. NDT = not displayed in the cladogram (Supporting information Figs S4 and S5).

Finally, the groups U (cluster 34), L (cluster 58) and Q (cluster 30) likely reflect annotation errors in the reference sequences, as they were placed within consistent clusters from a different phylogenetic affiliation.

In samples from Lake Baikal, 35 groups were assigned (Table 3). The most abundant genus was *Geobacillus* with 53.5%, followed by 21% corresponding to the genus *Bacillus*. Contrary to Lake Geneva (25.9%), the genus *Clostridium* was poorly represented (6.4%). The remaining composition consisted of *Paenibacillus* (8.3%), *Brevibacillus* (5.3%), *Sulfobacillus* (2.6%) and *Desulfotomaculum* (1.5%), with the final 1.2% of the sequences assigned to four other genera (Fig. 3B). Clusters 1–28 corresponded to different OTUs for which the closest relative species belongs to the class Bacilli. As for Lake Geneva, groups containing a large number of OTUs did not cluster together (e.g. groups B and E). OTUs related to species from the class Clostridia corresponded to clusters 35–53, although the closest relative to group AN (cluster 35) is likely a wrongly annotated *Bacillus*-like *Spo0A* sequence. The grouping of some OTUs related to *Geobacillus* (group AF), *Brevibacillus* (cluster L) and *Paenibacillus* (AH) species was closer to Clostridia than to Bacilli (indicated as Undefined in Supporting information Fig. S5).

Obtaining *spo0A* sequences directly from the environment opens the possibility of studying the patterns of distribution of endospore-forming Firmicutes. Although the *spo0A* reported here represent the first environmental sequences reported in literature, already an interesting pattern could be observed for the two sediments studied. In both sediments, *Geobacillus* represented the dominant group. Members of the genus *Geobacillus* have been traditionally isolated from environments with high temperatures, as part of the community of thermophilic Firmicutes growing with temperature optima ranging from 45 to > 70°C (Nazina *et al*., 2001). According to this, a previous study characterizing the community of Gram-positive bacteria in marine sediments at an intermediate depth (500 m) between the sediments studied here (284 and 1597 m deep) found a diverse community of Actinobacteria and Firmicutes, but no isolate was affiliated with Geobacilli (Gontang *et al*., 2007). In contrast, various species of *Geobacillus* have been isolated from cold soils (Marchant *et al*., 2011), and several publications have shown the isolation of thermophilic endospore-forming Firmicutes from cold marine sediments (Bartholomew and Paik, 1966; Hubert *et al*., 2010; de Rezende *et al*., 2013). These results suggest that endospores are in most cases allochthonous and have been deposited at the time of sedimentation, but several metabolic activity tests indicate that these microorganisms do not thrive in temperatures below 20°C. While the activity and the origin of the Geobacilli found in the present study were not assessed, it is an aspect that will be further studied.

In contrast with Geobacilli, the Clostridia sequence abundance differs greatly between the two sediments studied. While Clostridia represented 26% of the sequences in Lake Geneva sediments, in Lake Baikal, their abundance was only 6.4%. An interesting ecological feature within the group of endospore-forming Firmicutes is that there exist three ecotypes: aerobic, facultative anaerobic and strictly anaerobic. With some exceptions, aerobic types cluster among the class Bacilli and anaerobes cluster mostly in the class Clostridia (Schleifer, 2009). Very recently, we have found a correlation between an increase in the abundance of Clostridia and lake eutrophication (Wunderlin *et al*., 2013) or the pollution associated with treated wastewater disposal (Sauvain *et al*., 2013) in other areas of Lake Geneva. We postulate that the larger fraction of Clostridia found in the sediment of Lake Geneva reflects an increasing effect of human activities there; however, this needs to be verified.

The distribution and affiliation of the environmental *spo0A* sequences also raised some questions regarding the taxonomy of endospore-forming Firmicutes. The amino acid sequence identities for *Alkaliphilus*, *Thermincola*, *Desulfotomaculum* and *Desulfitobacterium*, to name only a few, are considerably lower (in the range of 70%) than the identities for most of the well-known *Bacillus* and *Clostridium* species (mostly between 80% up to 100%). This difference could be due to underrepresentation of the former genera in the databases. The taxonomical distribution of the environmental *spo0A* sequences could also reveal problems with the annotation, or more importantly, the potential detection of a yet unknown group of endospore-forming bacteria.

Even though Firmicutes are the second most abundant bacterial phylum in terms of culture representatives (Klenk and Goker, 2010), many of the environmental *spo0A* OTUs obtained in this study were only distantly related to reference strains. Therefore, a significant effort will be required in order to evaluate the diversity of endospore-forming Firmicutes in environmental samples, including a precise characterization of species belonging to the undefined clusters related to *Geobacillus*, *Brevibacillus* and *Paenibacillus*.

Here, we demonstrate how an improved DNA extraction protocol increases the diversity of endospore-forming Firmicutes retrieved from environmental samples. This is a clear example of how specific methods must be considered by those in the microbial community where traditional molecular microbial ecology methods are inadequate. We designed and validated a primer set for the *spo0A* gene that is specific for endospore-forming bacteria, thus enabling detection of endospore-forming Firmicutes by molecular methods. Environmental

sequencing of this gene has opened, for the first time, a window into the diversity of endospore-forming bacteria by culture-independent methods. Additionally, using a targeted sequencing approach for a functional subgroup, the higher resolution and sequence coverage revealed a very diverse community and potentially uncharacterized groups of endospore-forming Firmicutes. Future studies using other environmental samples will likely clarify the environmental relevance and biogeographical distribution patterns of endospore-forming Firmicutes in nature.

Acknowledgements

We acknowledge the elemo project (elemo.ch) for the sampling campaigns using the MIR submersibles. We are grateful for the help of the MIR team and colleagues from the elemo project as well as the help from Manon Brenier in the laboratory. We thank Sevasti Filippidou and Matthieu Bueche for the inhibition test in *Sulfobacillus acidophilus*. Furthermore, we thank Patricia Siering from Humboldt State University for providing endospore-forming strains for the validation test. We would like to thank Tom J. Petty from University of Geneva for his comments. This work was supported by Swiss National Science Foundation grant no. 31003A-132358/1.

References

Abecasis, A.B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J.B., and Henriques, A.O. (2013) A genomic signature and the identification of new sporulation genes. *J Bacteriol* **195:** 2101–2115.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Arcuri, E.F., Wiedmann, M., and Boor, K.J. (2000) Phylogeny and functional conservation of sigma(E) in endospore-forming bacteria. *Microbiology* **146:** 1593–1603.

Bartholomew, J.W., and Paik, G. (1966) Isolation and identification of obligate thermophilic sporeforming bacilli from ocean basin cores. *J Bacteriol* **92:** 635–638.

Bhandari, V., Ahmod, N.Z., Shah, H.N., and Gupta, R.S. (2013) Molecular signatures for *Bacillus* species: demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus Bacillus. *Int J Syst Evol Microbiol* **63:** 2712–2726.

Brill, J.A., and Wiegel, J. (1997) Differentiation between spore-forming and asporogenic bacteria using a PCR and Southern hybridization based method. *J Microbiol Methods* **31:** 29–36.

Bueche, M., Wunderlin, T., Roussel-Delif, L., Junier, T., Sauvain, L., Jeanneret, N., and Junier, P. (2013) Quantification of endospore-forming Firmicutes by quantitative PCR with the functional gene *spo0A*. *Appl Environ Microbiol* **79:** 5302–5312.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Delmont, T.O., Robe, P., Cecillon, S., Clark, I.M., Constancias, F., Simonet, P., *et al.* (2011) Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol* **77:** 1315–1324.

Dineen, S.M., Aranda, R., Anders, D.L., and Robertson, J.M. (2010) An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *J Appl Microbiol* **109:** 1886–1896.

Driks, A. (2002) Overview: development in bacteria: spore formation in *Bacillus subtilis*. *Cell Mol Life Sci* **59:** 389–391.

Dubey, G.P., Narayan, A., Mattoo, A.R., Singh, G.P., Kurupati, R.K., Zaman, M.S., *et al.* (2009) Comparative genomic study of *spo0E* family genes and elucidation of the role of Spo0E in *Bacillus anthracis*. *Arch Microbiol* **191:** 241–253.

Ewing, B., and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8:** 186–194.

Feinstein, L.M., Sul, W.J., and Blackwood, C.B. (2009) Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl Environ Microbiol* **75:** 5428–5433.

Frostegard, A., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., *et al.* (1999) Quantification of bias related to the extraction of DNA directly from soils. *Appl Environ Microbiol* **65:** 5409–5420.

Galperin, M.Y., Mekhedov, S.L., Puigbo, P., Smirnov, S., Wolf, Y.I., and Rigden, D.J. (2012) Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol* **14:** 2870–2890.

Gontang, E.A., Fenical, W., and Jensen, P.R. (2007) Phylogenetic diversity of gram-positive bacteria cultured from marine sediments. *Appl Environ Microbiol* **73:** 3272–3282.

Hubert, C., Arnosti, C., Bruchert, V., Loy, A., Vandieken, V., and Jorgensen, B.B. (2010) Thermophilic anaerobes in Arctic marine sediments induced to mineralize complex organic matter at high temperature. *Env Microbiol* **12:** 1089–1104.

Klenk, H.P., and Goker, M. (2010) En route to a genome-based classification of Archaea and Bacteria? *Syst Appl Microbiol* **33:** 175–182.

Kriventseva, E.V., Rahman, N., Espinosa, O., and Zdobnov, E.M. (2008) OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res* **36:** D271–D275.

Kuske, C.R., Banton, K.L., Adorada, D.L., Stark, P.C., Hill, K.K., and Jackson, P.J. (1998) Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Appl Environ Microbiol* **64:** 2463–2472.

Leff, L.G., Dana, J.R., McArthur, J.V., and Shimkets, L.J. (1995) Comparison of methods of DNA extraction from stream sediments. *Appl Environ Microbiol* **61:** 1141–1143.

Marchant, R., Banat, I.M., and Franzetti, A. (2011) Thermophilic bacteria in cool soils: metabolic activity and mechanisms of dispersal. In *Biogeography of Microscopic Organisms*. Fontaneto, D. (ed.). Cambridge, UK: Cambridge University Press, pp. 43–57.

von Mering, C., Hugenholtz, P., Raes, J., Tringe, S.G., Doerks, T., Jensen, L.J., *et al.* (2007) Quantitative

phylogenetic assessment of microbial communities in diverse environments. *Science* **315:** 1126–1130.
More, M.I., Herrick, J.B., Silva, M.C., Ghiorse, W.C., and Madsen, E.L. (1994) Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl Environ Microbiol* **60:** 1572–1580.
Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., *et al*. (2002) InterPro: an integrated documentation resource for protein families, domains and functional sites. *Brief Bioinform* **3:** 225–235.
Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E.V., Grigoryan, A.A., Ivanova, A.E., *et al*. (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. th*. *Int J Syst Evol Microbiol* **51:** 433–446.
Nicholson, W.L. (2002) Roles of Bacillus endospores in the environment. *Cell Mol Life Sci* **59:** 410–416.
Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J., and Setlow, P. (2000) Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol Mol Biol Rev* **64:** 548–572.
Onyenwoke, R.U., Brill, J.A., Farahi, K., and Wiegel, J. (2004) Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). *Arch Microbiol* **182:** 182–192.
Ovreås, L., Forney, L., Daae, F.L., and Torsvik, V. (1997) Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl Environ Microbiol* **63:** 3367–3373.
Paredes, C.J., Alsaker, K.V., and Papoutsakis, E.T. (2005) A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol* **3:** 969–978.
de Rezende, J.R., Kjeldsen, K.U., Hubert, C.R., Finster, K., Loy, A., and Jorgensen, B.B. (2013) Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J* **7:** 72–84.
Sauvain, L., Bueche, M., Junier, T., Masson, M., Wunderlin, T., Kohler-Milleret, R., *et al*. (2013) Bacterial communities in trace metal contaminated lake sediments are dominated by endospore-forming bacteria. *Aquat Sci* (in press).
Schaeffer, P., Millet, J., and Aubert, J.-P. (1965) Catabolic repression of bacterial sporulation. *Proc Natl Acad Sci U S A* **54:** 704–711.
Schleifer, K.H. (2009) Phylum XIII. Firmicutes Gibbons and Murray 1978, 5 (Firmacutes [sic] Gibbons and Murray 1978, 5). In *Bergey's Manual® of Systematic Bacteriology*. Vos, P., Garrity, G., Jones, D. *et al*. (eds). New York, USA: Springer, pp 19–1317. doi: 10.1007/978-0-387-68489-5_3.
Staley, J.T., and Gosink, J.J. (1999) Poles apart: biodiversity and biogeography of sea ice bacteria. *Annu Rev Microbiol* **53:** 189–215.
Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673–4680.
Traag, B.A., Pugliese, A., Eisen, J.A., and Losick, R. (2013) Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*. *J Bacteriol* **195:** 253–260.
Webster, G., Newberry, C.J., Fry, J.C., and Weightman, A.J. (2003) Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. *J Microbiol Methods* **55:** 155–164.
Wunderlin, T., Corella, J.P., Junier, T., Bueche, M., Girardclos, S., and Junier, P. (2013) Endospore-forming bacteria as new proxies to assess impact of eutrophication in Lake Geneva, (Switzerland-France). *Aquat Sci* (in press).
Yudkin, M.D., and Clarkson, J. (2005) Differential gene expression in genetically identical sister cells: the initiation of sporulation in *Bacillus subtilis*. *Mol Microbiol* **56:** 578–589.
Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62:** 316–322.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Experimental procedures.

Fig. S1. Comparative phylogenetic analysis of 16S rRNA gene sequences and six conserved sporulation-related genes (*spo0A*, *spoIVB*, *spoVAC*, *spoVAD*, *spoVT* and *gpr*)(spore proteome) for 27 spore-forming Firmicutes with a complete genome sequence reported and annotated. Alignments were constructed with MAFFT (Katoh *et al*., 2005) or Muscle (Edgar, 2004) using default parameters. Multiple-FastA alignments were converted to Phylip format with the *seqret* program from the EMBOSS package (Rice *et al*., 2000). Phylogenies were constructed from Phylip-formatted alignments with PhyML (Guindon and Gascuel, 2003), using default parameters, except the following: JTT+Γ substitution model for proteins and GTR+Γ model for nucleic acids; four classes of substitution rate categories; estimation of the shape parameter, proportion of invariants and transition/transversion ratios (for nucleotides). Trees were processed (re-rooting, extracting topology and plotting) with the Newick Utilities (Junier and Zdobnov, 2010). Bootstrap values (percentage over 1000 samplings) are shown at the nodes of the trees.

Fig. S2. Phylogenetic reconstruction (above) and conservation profiles (below) for sequences of the stage 0 sporulation protein Spo0A. Conservation plots were made with the *plotcon* program from EMBOSS. This is a sliding-window program that computes a weighted average of the similarity scores for all residue pairs in each window. We used the default window size of four residues.

Fig. S3. Alignment of *spo0A* gene of *Sulfobacillus acidophilus* and *Alicyclobacillus acidocaldarius* Tc41 against *spo0A* of *Bacillus subtilis* 168. The two regions shown correspond to the forward primer 166f (left) and the reverse primer

748r (right) described in this study. Stars indicate 100% identity. The exclamation points highlight mismatches with the primer sequence.

Fig. S4. Cladogram of *spo0A* sequences from sediment of Lake Geneva extracted with protocols 1 (blue), 2 (yellow) and 3 (red). The nucleotide sequences were then clustered into putative OTUs (identity of > 97%) with the pick_otus.py program from the QIIME package using the Uclust method (Caporaso, 2010), and a representative was used to build the phylogeny. Phylogenies were constructed from Phylip-formatted alignments with PhyML (Guindon, 2003 #88), using default parameters. The trees were re-rooted, condensed according to DNA extraction protocol, and displayed with the Newick utilities (Junier, 2010). Each branch represents a cluster of OTUs of > 97% sequence similarity. Identification of the closest relatives of the environmental sequences from the indirect extractions (protocol 3) was done by protein BLAST (Altschul *et al*., 1997), with the translated protein sequences using a reference database of 581 Spo0A protein sequences from the InterPro site (Mulder *et al*., 2002). Classes of closest relative are shown in color with indication of the identity ranges [< 65% identity (–), 65–74% (<), 75–84% (~), 85–94%(#), > 95% (+)]. A, *Bacillus amyloliquefaciens*; B, *B. methanolicus*; C, *Geobacillus* sp. (strain WCH70); D, *B. cereus* subsp. *cytotoxis* (strain NVH 391-98); E, *B. thuringiensis*; F, *Geobacillus thermodenitrificans* (strain NG80-2); G, *B. atrophaeus* (strain 1942); H, *B. subtilis*; I, *B. mycoides*; J, *B. pseudofirmus* (strain OF4); K, *Lysinibacillus sphaericus* (strain C3-41); L, *Brevibacillus laterosporus*; M, *Brevibacillus brevis* (strain 47); N, *Thermincola potens* (strain JR); O, *Desulfotomaculum acetoxidans* (strain ATCC 49208); P, *Desulfosporosinus orientis* (strain ATCC 19365); Q, *Thermosediminbacter oceani* (strain ATCC BAA-1034); R, *Syntrophobotulus glycolicus* (strain DSM 8271); S, *Heliobacterium medesticaldum* (strain ATCC 51547); T, *Clostridium clariflavum* (strain DSM 19732); U, *B. cereus*; V, *C. thermocellum*; W, *C. cellulovorans* (strain ATCC 35296); X, *C. cellulolyticum* (strain ATCC 35319); Y, *C. botulinum*; Z, *C. lijungdahlii* (strain ATCC 55383); AA, *C. perfringens*; AB, *C. sporogenes*; AC, *Alkaliphilus metalliredigens* (strain QYMF); AD, *A. oremlandii* (strain OhILAs); AE, *Desulfotomaculum kuznetsovii* (strain DSM 6115); AF, *Geobacillus* sp. (strain Y412MC10); AG, *Paenibacillus polymyxa*; AH, *P. mucilaginosus* (strain KNP414).

Fig. S5. Cladogram of *spo0A* sequences from sediment of Lake Baikal extracted with protocols 1 (blue), 2 (yellow) and 3 (red). Each branch represents a cluster of OTUs of > 97% sequence similarity. Closest relatives are shown in letters around the tree together with identity ranges [< 65% identity (–), 65–74% (<), 75–84% (~), 85–94%(#), > 95% (+)]. For classes, see legend in Fig. 4 and the following: AI, *B. megaterium*; AJ, *B. licheniformis*; AK, *B. megaterium* (strain DSM 319); AL, *C. haemolyticum*; AM, *Paenibacillus* sp. (strain JDR-2); AN, *B. cellulosilyticus* (strain ATCC 21833); AO, *Sulfobacillus acidophilus* (strain TPY); AP, *S. acidophilus* (strain ATCC 700253); AQ, *Desulforudis audaxviator* (strain MP104C); AR, *C. butyricum*; AS, *C. kluyveri* (strain ATCC 8527).

Table S1. List of genome sequences from the 27 endospore-forming Firmicutes used in this study. Complete and draft genome sequences were downloaded from the Comprehensive Microbial Resource (CMR, 24.0 data release, cmr.jcvi.org) and Integrated Microbial Genomes (IMG, 3.0, img.jgi.doe.gov) websites. Protein and nucleotide sequences of spore-related genes were obtained by search for role category/function *sporulation and germination* (CMR) and *sporulating* (IMG). Additional information on all retrieved genomes was obtained from the GenBank database (http://www.ncbi.nlm.nih.gov/genome). Clas = taxonomical classification; B = Bacilli; C = Clostridia; T° = temperature range; M = mesophile; T = thermophile; *P* = psychrophile; H = hyperthermophile; Sp. Genes = number of sporulation-related genes. The number of sporulation-related genes was retrieved from the available genome annotations.

Table S2. Orthologous genes found after bi-directional BLAST of the sporulation-related genes common to 27 genomes of endospore-forming Firmicutes. Protein lengths indicated for *Bacillus subtilis* as a reference were obtained from Stragier and Losick (1996).