environmental microbiology reports

Environmental Microbiology Reports (2015) **7**(2), 341–350 doi:10.1111/1758-2229.12254

Low frequency of endospore-specific genes in subseafloor sedimentary metagenomes

Mikihiko Kawai,[1†] Ikuo Uchiyama,[2,3] Hideto Takami[4] and Fumio Inagaki[1,5*]

[1]*Geomicrobiology Group, Kochi Institute for Core Sample Research and* [5]*Geobiotechnology Group, Research and Development Center for Submarine Resources, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Monobe B200, Nankoku, Kochi 783-8502, Japan.*

[2]*National Institute for Basic Biology, National Institutes of Natural Sciences, Nishigonaka 38, Myodaiji, Okazaki, Aichi 444-8585, Japan.*

[3]*Department of Basic Biology, School of Life Science, The Graduate University for Advanced Studies, Nishigonaka 38, Myodaiji, Okazaki, Aichi 444-8585, Japan.*

[4]*Microbial Genome Research Group, Institute of Biogeosciences, JAMSTEC, Natsushima-cho 2-15, Yokosuka, Kanagawa 237-0061, Japan.*

Summary

Spore formation is considered to be one of the microbial strategies for long-term survival in subseafloor sedimentary habitats. However, our knowledge of the genetic and physiological characteristics of subseafloor microbes is limited. Here, we studied the distribution and frequency of genes that are related to endospore formation in 10 subseafloor sedimentary metagenomes from Site C9001 off Japan and Site 1229 off Peru. None or very low frequencies of endospore-specific genes (e.g. *dpaA*, *dpaB*, *sspA*, *spo0A*, *spoIIGA*, *spoIIM*, *spoIIIAB*, *spoIVA*, *spoIVB*, *yabP*, *yunB*, *spoVM*) were observed in the subseafloor metagenomes. Based on the number of universally conserved single copy genes, the frequency ratio of putative endospore-formers was estimated to be < 10%, which is consistent with the frequency of *Clostridia*-derived genomes (2–4%) but is lower than previous estimates based on the concentration of dipicolinic acid. Conceivable explanations for this discrepancy are as follows: the efficiency of lysis and DNA extraction of subseafloor endospore cells may have been lower than those of vegetative cells, conversion factor of dipicolinic acid content per cell may differ, and/or sporulation-related genes and other functional strategies for long-term survival in the deep subseafloor biosphere are evolutionarily distinct from known spore-forming gene repertoires.

Received 29 April, 2014; revised 4 November, 2014; accepted 23 November, 2014. *For correspondence. E-mail inagaki@jamstec.go.jp; Tel. (+81) 88 878 2204; Fax: (+81) 88 878 2192. [†]Present address: Faculty of Biology-Oriented Science and Technology, Kindai University, Nishimitani 930, Kinokawa, Wakayama 649-6493, Japan.

Introduction

Subseafloor sediments on the Pacific Ocean margins harbour a remarkable number of microbial cells extending down to over 1000 m below the seafloor (mbsf) (Parkes *et al.*, 1994; 2000; Lipp *et al.*, 2008; Morono *et al.*, 2009; Ciobanu *et al.*, 2014). Subseafloor sedimentary microbial activity is generally extraordinarily low due to the limited supply of nutrient and energy substrates (D'Hondt *et al.*, 2002; 2004; Hoehler and Jørgensen, 2013). Previous culture-independent molecular ecological surveys through scientific ocean drilling have demonstrated that subseafloor sedimentary microbial communities in the Pacific Ocean margins are composed mainly of bacteria and archaea, most of which are phylogenetically distinct from previously cultivated representatives (e.g. Inagaki *et al.*, 2003; 2006); therefore, physiology and metabolic functions of subseafloor microbial components, as well as the ecological consequence of these functions, remain largely unknown (D'Hondt *et al.*, 2007). Accordingly, the first metagenomic analysis of subseafloor microbial communities in sediments off Peru down to 50 mbsf showed that a large fraction of the genes (up to 85%) are unidentifiable with no close relatives to those from the surface biosphere that are deposited in the database (Biddle *et al.*, 2008; 2011). Nevertheless, elemental analysis of subseafloor cells using nano-scale secondary ion mass spectrometry (i.e. NanoSIMS) demonstrated that a large fraction of subseafloor cells are capable of incorporating stable isotope-labelled substrates (e.g. ^{13}C-labelled glucose, ^{15}N-labelled ammonia, ^{13}C- and ^{15}N-labelled amino acids), indicating that subseafloor microbial cells are generally alive but have extremely low metabolic activity,

even in dark and old sedimentary habitats (Jørgensen, 2011; Morono *et al*., 2011). In addition, single cell genomics has been used to describe the genomes of uncultured archaea within the miscellaneous crenarchaeota group and the marine benthic group-D (Lloyd *et al*., 2013; Meng *et al*., 2014) and uncultured bacteria within the *Chloroflexi* (Kaster *et al*., 2014), showing that anaerobic organotrophic and/or mixotrophic metabolism are employed for utilization of the organic matter buried in subseafloor sediments. However, how subseafloor microbes survive under extremely energy-limited conditions over geological timescales remains largely unknown (Hoehler and Jørgensen, 2013).

Dormancy is one of the microbial strategies for maintaining microbial diversity for long-term survival while minimizing energy requirements for metabolic activities (Jones and Lennon, 2010). Endospore numbers and vegetative cells were estimated to be equally abundant in subseafloor sediments (Langerhuus *et al*., 2012; Lomstein *et al*., 2012) based on the amount of dipicolinic acid (DPA; pyridine-2,6-dicarboxylic acid), a specific compound that composes ~ 10% dry weight of typical bacterial spores (Tinelli, 1955; Church and Halvorson, 1959; Fichtel *et al*., 2007). In addition, it has been reported that some endospores of thermophilic bacteria buried in cold marine sediments are capable of germination at high temperature (Hubert *et al*., 2009; 2010). Despite the ecological significance of endospore-forming populations in the subseafloor sedimentary biosphere, molecular characterization of the genes that mediate endospore formation has remained elusive.

In this study, we examined the spatial distribution and frequency of sporulation-related gene homologues in 10 metagenomic assemblages from Site C9001 in the northwestern Pacific Ocean off the Shimokita Peninsula, Japan, during the *Chikyu* Shakedown Expedition CK06-06 [metagenomes from Cores 1H-1 (0.8 mbsf), 1H-4 (5.1 mbsf), 3H-2 (18.6 mbsf), 6H-3 (48.5 mbsf) and 12H-4 (107.0 mbsf); Kawai *et al*., 2014] and Site 1229 in the southeastern Pacific Ocean off Peru during the Ocean Drilling Program (ODP) Leg 201 [metagenomes from Cores 1H-1 (1 mbsf), 3H-2 (16 mbsf), 4H-5 (32 mbsf) and 7H-1 (50 mbsf); Biddle *et al*., 2008].

Results and discussion

We identified genes that are exclusively distributed among the genomes of endospore-formers based on phylogenetic profiling analysis using an orthologue table (Fig. 1). The sporulation-specific genes are conserved among all available genomes of spore-formers (Galperin *et al*., 2012), in which dipicolinate synthase genes [*dpaA* (*spoVFA*) and *dpaB* (*spoVFB*)] are essential for the DPA synthesis in many bacilli and clostridia. In addition, short acid-soluble proteins (SASPs) are known to comprise 5–10% of total spore protein (Setlow, 1975; Johnson and Tipper, 1981) and to protect intra-spore DNA (Setlow, 2007); thus, the *sspA* gene family (*sspA*, *sspB*, *sspC* and *sspD* in *Bacillus subtilis*) that encodes the α/β-type SASPs was also included in our analysis.

In this study, a total of 52 orthologous groups were analysed (Fig. 1). Genes of the orthologous groups were more or less distributed differentially among genomes of known endospore- and non-endospore-formers, as well as genomes of microorganisms the endospore-forming ability of which has not been examined extensively (e.g. genomes for uncultured microorganisms obtained through single cell genomics). Among those 52 orthologous groups, 23 were only distributed among genomes of endospore-forming bacteria within classes Bacilli and Clostridia in phylum Firmicutes, but not non-endospore-formers (Fig. 1A; group 1). In contrast, the other 29 orthologous groups were distributed among both endospore-formers and non-endospore-formers. Four genes of them were mostly distributed among Firmicutes (more than 80% of the genes were derived from genomes of Firmicutes) (Fig. 1B and Fig. S1; group 2). Among the remaining 25 genes, phylogenetic analysis (Fig. S1) showed that seven genes (Fig. 1C; group 3) formed a smaller subtree, of which genes were distributed solely among endospore-former genomes, whereas the other 18 genes (Fig. 1D) did not form clear subclusters composed of genes of endospore-formers or non-endospore-formers. In total, 34 genes within groups 1, 2 and 3 can serve as reliable genetic markers of endospore-forming bacteria. These included three genes related to the production of DPA and SASPs (i.e. *dpaA*, *dpaB* and *sspA* family), *spo0A* for the master regulator of sporulation, *sigE* and *sigK* for sporulation-specific sigma factors Sigma E and Sigma K (in mother cells), and *sigF* and *sigG for* sigma factors Sigma F and Sigma G (in developing spores), among other genes active in the post-septation stage (stage II) (*spoIIGA*, *spoIIM*, *spoIIP* and *spoIIR*; Fig. 1A), genes in post-engulfment (stages III-VI) (*spoIIAB* to *yunB* in Fig. 1A; e.g. *spoIIIAB-spoIIIAH*, *spoIIID*, *spoIVA*, *yabP* and *yunB*) and genes coding for the spore coat (*spoIVA* and *spoVM*).

Using this genomic approach, one can infer the spore-forming ability of physiologically unidentified microbes. In this study, the orthologous genes were more or less found among class Erysipelotrichi and Negativicutes of phylum Firmicutes. On the other hand, such potential endospore-forming bacteria were not found among candidate divisions and other phyla (e.g. Atribacteria, former candidate division JS1/OP9; Dodsworth *et al*., 2013). Among the 23 genes of group 1, 21 genes were also found in two genomes in class Negativicutes of phylum Firmicutes [Fig. 1A(ii)]; i.e. *Pelosinus fermentans* DSM 17108 R7 and

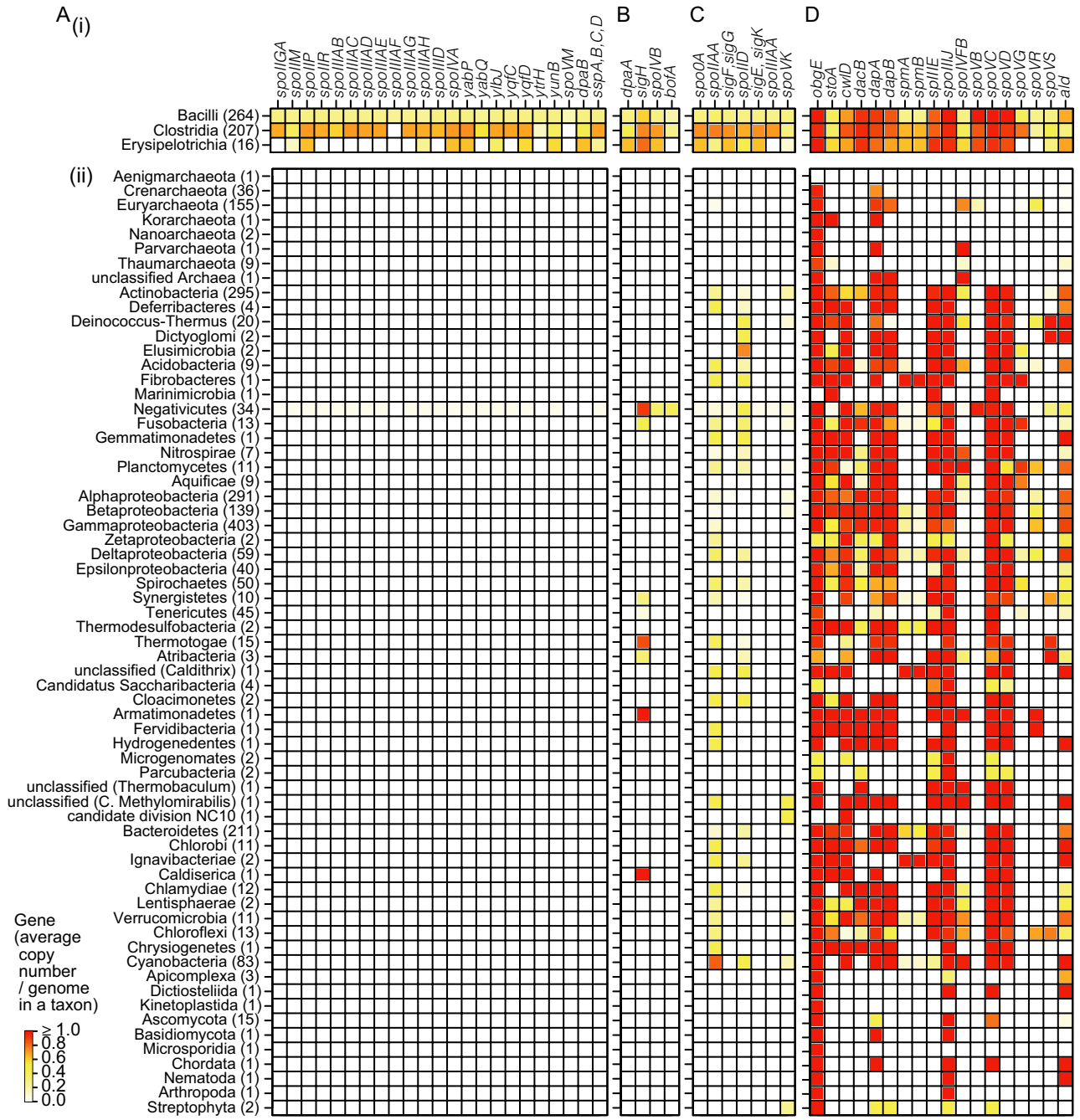

Fig. 1. Distribution of endospore-related genes among the genomes in the orthologue table. (A) Distribution of endospore-specific genes, for which homologues are only found in spore-forming bacteria (group 1). (B) Distribution of endospore-related genes, for which a few (< 20%) homologues are also found in non-spore-formers (group 2). (C) Distribution of genes that formed a subcluster composed of genes mostly from spore-forming bacteria on a phylogenetic tree (Fig. S1) (group 3). (D) Distribution of the other endospore-related genes, for which homologues were found in many non-spore-forming taxa. For genes in (C), initial orthologue clustering resulted in a larger cluster, for which homologues were also found in non-spore-forming bacteria. Subclusters only composed of genes of non-spore-forming bacteria were pruned from the initial tree. The genes used in this panel were those indicated on the initial trees in red (Fig. S1). On the other hand, the other genes in (D) did not form clear subclusters composed of genes of endospore-formers or non-spore-formers (Fig. S1). Each row represents a taxonomic classification (taxonomic rank used are class for the phylum *Firmicutes* and *Proteobacteria*, and phylum or candidate division for the others), and the number in parentheses indicates the number of genomes analysed for that taxon. Colour represents the ratio of the number of genomes that harbour a gene of an orthologous group to the total number of genomes for each taxon. Panel (i) shows taxa of endospore-forming bacteria. Panel (ii) shows taxa of other phyla and candidate divisions. The orthologue table in this study was constructed from available genomic sequences without redundant representation of species (Table S1) using the DomClust programme (Uchiyama, 2006), with parameters as previously described (Kawai *et al.*, 2014) (see Appendix S1 for details).

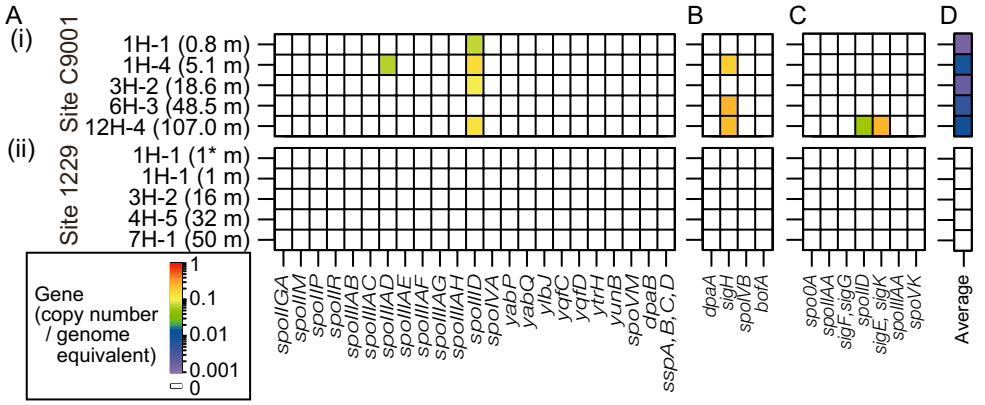

Fig. 2. Distribution of endospore-specific genes in subseafloor sedimentary metagenomes from off (i) Shimokita and (ii) Peru. (A) Group 1 genes, the same orthologous groups as in Fig. 1A. (B) Group 2 genes as in Fig. 1B. (C) Group 3 genes as in Fig. 1C. (D) Average of the estimated gene copy number per genome equivalent of the genes in (A–C) (the values 1.4×10^{-3} to 0.013 correspond to 0.14–1.3% for Site C9001; also see Table 1 as percentage of spore-former genomes). The metagenomic genes were assigned to the orthologue table using the MERGETREE programme. Colour indicates the estimated gene copy number per genome equivalent of the metagenomic pool.

Thermosinus carboxydivorans Nor1, which have most of the endospore genes (Yutin and Galperin, 2013). The present analysis of 16 incomplete genomes of Class Erysipelotrichi revealed that they have up to 10 genes of the 23 genes. Indeed, *Pelosinus fermentans* in Negativicutes and *Clostridium ramosum* in Erysipelotrichi form endospores (Alexander *et al.*, 1995; Shelobolina *et al.*, 2007).

Using the 34 defined marker genes, we evaluated the frequency of the marker genes in metagenomes based on the estimated gene copy number per genome equivalent, i.e. average coverage ratio (Kawai *et al.*, 2014). Surprisingly, we detected no homologues of *dpaA*, *dpaB*, *sspA*-family and *spo0A* genes in the metagenomes from Site C9001 off Shimokita [Fig. 2A(i)–C(i)]. Homologues of other orthologous groups were also not found, except for five genes [Fig. 2A(i)–C(i)]. Furthermore, homologues of those 34 orthologous groups were completely missing in the Peruvian metagenomes [Fig. 2A(ii)–C(ii)]. The number of reads for five genes was one (Fig. 2), except for *sigH* where two and three reads were detected from 1H-4 and 6H-3 of Site C9001, respectively. The estimated gene copy number per genome equivalent for them was in the range of 0.054–0.26. Frequency of spore-former genomes, which was calculated as the average of the estimated gene copy number per genome equivalent of the 34 genes, varied from 1.4×10^{-3} to 0.013, corresponding to 0.14–1.3% for Site C9001 and 0% for Site 1229, respectively (Fig. 2D and Table 1).

Then, we estimated the frequency of endospore-former genomes in the subseafloor sedimentary metagenomes by another approach from the observed number of endospore-related genes based on the orthologue table (Fig. 3). Here, we assume that the probability of observing n endospore-related genes among a total number of sequences, N, followed a binomial distribution: $\Pr(n, N, p) = {}_N C_n \times p^n \times (1 - p)^{N-n}$, where p is the probability of observing an endospore-related gene for each metagenomic sequence, or *probability of success*. The *probability of success* depends on (i) the frequency of endospore-related genes among genomes of endospore-formers and (ii) the ratio of the number of endospore-former genomes to the total number of genomes. Here, for simplicity, we also assume that all genomes included in each metagenomic sample have almost equal genome sizes. Then, *probability of success* can be written as $p = r \times l / s$, where r is the ratio of the number of genomes of endospore-formers to the total number of genomes, l is the sum of the nucleotide length of the 34 spore-specific orthologous groups, and s is the estimated average genome size in the target metagenomes. To determine the l value, the nucleotide length for each orthologous group was calculated by taking the average of the length of genes of the complete genomes classified to the orthologous group. The sum of the lengths of the 34 orthologous groups (i.e. the value l) was 24 kb. The s value was calculated as follows.

The sequence coverage expressed as the number of genome equivalents was estimated for each sedimentary metagenome based on the average of the sequence coverage of orthologues of the universally conserved single-copy genes. In this definition, if the metagenomic sample contains 10 genomes, each of whose coverage is 0.2, the sequence coverage is considered to be 2.0 (i.e. genome equivalents). The estimated sequence coverage in metagenomes from sediments off Shimokita and Peru was 3.8–10 and 0.8–1.5, respectively (Table 1). The average genome size s in metagenomes was calculated by dividing the amount of total sequence by the estimated sequence coverage (Table 1). Using l and s, the confi-

Table 1. Frequency of endospore-former genomes in subseafloor sedimentary metagenomes.

Drilling site	Sample code	Sediment depth (mbsf)	Amount of sequences (Mb)[d]	Sequence coverage[e]	Estimated genome size (Mb)[f]	Number of sequences	Estimated frequency of spore-former genomes (%)[h]
Site C9001[a]	1H-1	0.8	50.0	10	5.0	39 167[g]	0.14 [0.013–2.9]
Site C9001	1H-4	5.1	45.0	6.5	6.9	39 029	0.78 [0.79–7.5]
Site C9001	3H-2	18.6	42.5	8.4	5.1	38 845	0.18 [0.014–3.0]
Site C9001	6H-3	48.5	43.0	6.6	6.5	39 142	0.49 [0.42–6.0]
Site C9001	12H-4	107	42.0	3.8	11	39 140	1.3 [1.3–12]
Site 1229[b]	*1H-1[c]	1	5.17	1.0	5.2	51 104	0 [0–1.5]
Site 1229	1H-1	1	5.97	0.79	7.6	59 055	0 [0–1.9]
Site 1229	3H-2	16	6.22	1.5	4.1	61 047	0 [0–1.0]
Site 1229	4H-5	32	8.00	1.4	5.7	78 154	0 [0–1.1]
Site 1229	7H-1	50	4.30	1.4	3.1	42 127	0 [0–1.1]

a. The *Chikyu* Shakedown Expedition CK06-06 Site C9001 off the Shimokita Peninsula, Japan (41° 10.6′ N, 142° 12.1′ E); the same location of the Integrated Ocean Drilling Program (IODP) Site C0020 Hole A (Inagaki *et al.*, 2012). Metagenome sequences are available on the database as the accession numbers under the BioProject ID PRJDB1115-PRJDB1119.
b. ODP Leg 201 Site 1229 off Peru (10°58.6′ S, 77°57.5′ W). Metagenome sequences are available on the database as the accession numbers under Sequence Read Archive (SRA) ID SRA001015 with run IDs SRR001324, SRR001325, SRR001322, SRR001323 and SRR001326.
c. An asterisk denotes original, unamplified sample (Biddle *et al.*, 2008).
d. Used were sequence reads longer than 90 bp.
e. Sequence coverage was defined as the average of average coverage values of universally conserved single-copy genes.
f. The value is amount of sequences divided by sequence coverage.
g. The number of clones was used for metagenomes from Site C9001.
h. The value is the average of the estimated gene copy number per genome equivalent of the 34 genes. The value in the brackets is 95% confidence interval for the ratio of spore-former genomes to the total number of genomes. The former value is based on sequence coverage, whereas the latter confidence interval value is based on count of reads.
The sequence coverage of genome equivalent for each metagenome was calculated by averaging the average sequencing coverage of universally conserved single-copy genes, which were listed in Table S2. Average sequencing coverage was calculated as the total number of nucleotides that overlapped a gene divided by the average length of the orthologous genes identified in the reference genomes as the expected length of a gene in metagenomes based on the fact that metagenomic genes are largely truncated at the ends of sequence reads (for details, see Kawai *et al.*, 2014). Average genome size in the metagenome was estimated by dividing the sum of the nucleotide sequences of the metagenome by the coverage value of universally conserved single-copy genes. Estimation of the ratio of the number of genomes of endospore-formers to that of all genomes in a metagenomic pool was calculated by the following formula. Let s be the estimated average genome size in the pool, N be the number of reads, l be the sum of the nucleotide length of genes considered (in this study, 34 orthologous groups), n be the number of observed reads that contain the genes considered, p be the probability where the observed frequency of the genes is expected by chance (*probability of success*), and r be the ratio of the number of genomes of endospore-formers to that of all genomes in a metagenomic pool. Then p follows the binomial distribution, with parameters *number of successes* = n, *number of trials* = N and *probability of success* = $r \times l / s$. According to this formula, the 95% confidence interval was calculated using the 'binom.test' function of R statistical package (R Core Team, 2013).

dence interval of the ratio of spore-former genomes to total genomes, $[r_L, r_U]$, was calculated from the confidence interval of *probability of success*, i.e. $[p_L, p_U]$ satisfying $\Pr(n, N, p_L < p < p_U) = 0.95$, using the relationship $p = r \times l / s$. The upper boundary of the ratio (r_U) of spore-former genomes to total genomes for metagenomes from off Shimokita and Peru was 0.029–0.12 (3–12%) and 0.010–0.019 (1–2%) respectively (Table 1). It should be noted that these values of the estimated frequency of endospore-forming cells are the upper boundary of the confidence interval. In addition, genomes of endospore-forming cells might be derived from vegetative cells of endospore-forming bacteria: if so, endospore frequency could be even much lower than these values. The interval could even be smaller than those values if the effect of eukaryotic genomes (Orsi *et al.*, 2013) is considered, since they generally have disproportionally larger genome size and might result in the larger s value and confidence interval although this situation could impose the opposite effect because the assumption of equal genome size of the present scheme is violated. For example, it has been reported that effective genome size in seawater samples is reduced 3.6-fold for metagenomes of large fraction size (3–20 μm) only if the prokaryotic fraction of the sample is taken into account (Raes *et al.*, 2007). We also note that using only the 23 genes of the group 1 for calculation or using the 34 genes of the groups 1, 2 and 3 as described above did not make large difference in confidence values (Fig. S2).

We further examined the taxonomic distribution of universally conserved single-copy genes in the five metagenomes from off Shimokita (Fig. 4). In those sediment samples, the members of *Chloroflexi* and JS1 bacteria were the predominant bacteria [0.08–0.29 (8–29%) and 0.06–0.25 (6–25%), respectively], which was the same pattern from analysis of the much lower number of 16S rRNA gene fragments than the universally conserved single-copy genes (Kawai *et al.*, 2014), whereas the frequency of *Clostridia* within the Firmicutes was found to be 0.02–0.04. If all *Clostridia* genomes had endospore-specific genes, the frequency of *Clostridia*-derived genomes is relatively compatible with the upper boundary

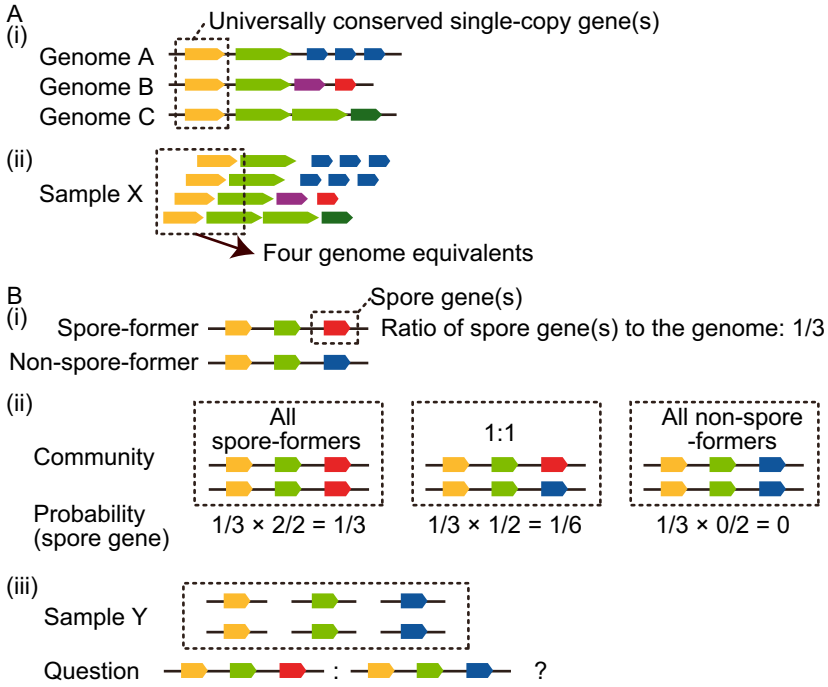

Fig. 3. Schematic explanation for the estimation of frequency of a certain type of genomes in a metagenomic pool.
(A) Calculation of the estimated average genome size *s* in a metagenomic pool. (i) Universally conserved single-copy genes are conserved in almost all of the archaeal and bacterial genomes in the orthologue table. (ii) The number of genome equivalents in a metagenomic pool is calculated by taking the average of the average sequencing coverage of the universally conserved single-copy genes. The average genome size of the pool, *s*, is estimated by dividing the number of sequences by the number of genome equivalents.
(B) Estimation of the frequency of genomes from the frequency of genes. (i) In this hypothetical situation, the ratio of the length of spore genes *l* to the length of the genome of spore-former is 1/3. (ii) The probability of taking a spore gene from a given composition of genomes follows a binomial distribution. The probability is the multiplication of the ratio *r*, the number of genomes of endospore-formers to the total number of genomes, and the ratio, *l* / *s*, of the spore genes *l* to the average genome size *s*. (iii) For a given metagenomic pool, the ratio of the number of genomes of spore-formers to that of non-spore-formers can be estimated by dividing the binomial confidence interval by the *l* / *s* value.

of estimated frequency of endospore-specific genes (0.029–0.12, that is, 3–12%). Phylogenetic classification at the class level might generally overestimate the frequency of endospore-forming organisms since there are non-endospore-forming Clostridia (Galperin *et al.*, 2012).

The estimated ratio (< 2:100, i.e. upper boundary of 0.010–0.019 for metagenomes from Site 1229 off Peru) of endospore-forming cells to total cells based on the endospore-specific genes is notably lower than the ratio of endospore cells to total cells (i.e. ~ 1:1) estimated based on muramic acid concentrations at the deep layers (< 40 mbsf) at the same Site 1229 (Lomstein *et al.*, 2012). They also showed that the muramic acid-based estimation and DPA-based estimation are consistent at another Site 1227 off Peru (Lomstein *et al.*, 2012). Conceivable explanations for this discrepancy are as follows: different efficiency of DNA extraction between endospores and vegetative cells, disproportionate DPA content, and other strategic cell forms for long-term survival.

First, when we extract DNA from sediment samples, the efficiency of lysis of subseafloor endospore cells might be lower than that of vegetative cells. The metagenomes from both off Peru and Shimokita were extracted by the standard methods using bead beating and lytic enzyme (Biddle *et al.*, 2008; Kawai *et al.*, 2014). Even using bead beating and lytic enzymatic reactions, it might be difficult to fully extract the entire DNA present in all endospores; in fact, it has been reported that bead beating method recovered 5–13% of spores added to soil samples (Dineen *et al.*, 2010). Even for vegetative cells in subseafloor sediments, the ratio of disrupted cells to the total cell number is often significantly as low as ~ 30% or less by standard DNA extraction methods (Morono *et al.*, 2014). It has also been shown that bead beating method combined with different extraction methods, such as separation of cells before extraction or repeated extraction for environmental samples, resulted in relatively consistent estimation in the range of twofold differences of the relative gene copy number of *spo0A* to 16S rRNA genes using polymerase chain reaction (Bueche *et al.*, 2013; Wunderlin *et al.*, 2013). In addition to the lysis efficiency, relatively short reads of the study

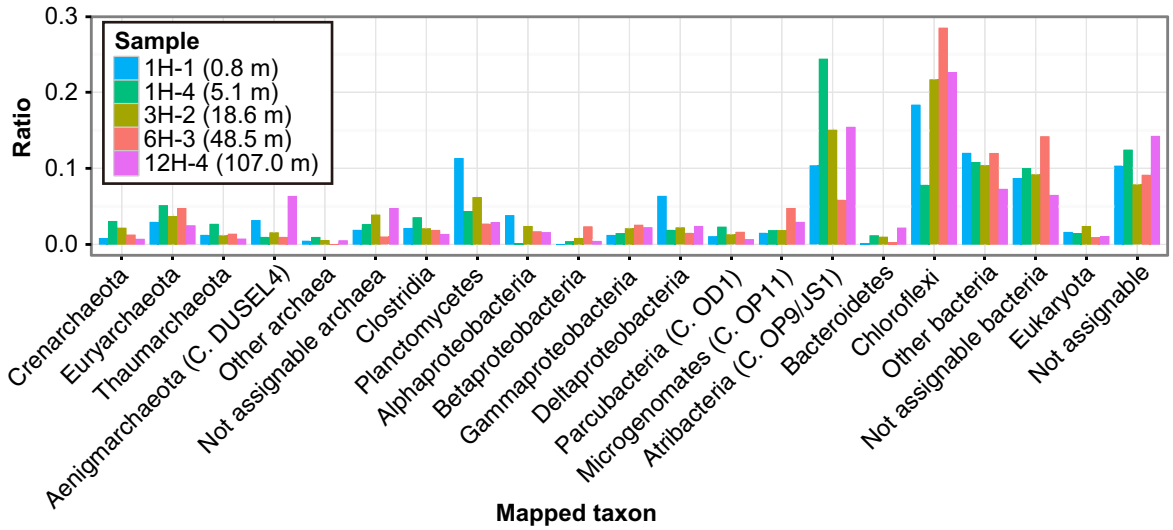

Fig. 4. Phylogenetic distribution of universally conserved single-copy genes in metagenomes at five different depths from Site C9001 off Shimokita. Colour bars represent frequency of the universally conserved single-copy genes in metagenomes for each taxon. Taxonomic information was assigned to each gene in the metagenomes by assigning the lowest common taxon of the sister subcluster on the dendrogram of an orthologous group. Taxonomic information was based on the NCBI Taxonomy database.

sites (i.e. 100 bp at Site 1229) may have missed weak homology, which might cause underestimation.

Second, the amount of DPA per cell is an important conversion factor for estimating cell biomass (Hinrichs and Inagaki, 2012). In this context, DPA content may vary over an order of magnitude or more, depending on the species and surrounding environmental conditions (Fichtel *et al*., 2007). For example, if the amount of DPA per cell in subseafloor sediments is similar to that of sulfate reducers such as *Desulfosporosinus* (1.3×10^{-15} mol cell^{-1}) (Fichtel *et al*., 2007), the ratio of the estimated number of endospores to the number of vegetative cells would differ by one tenth to one fifth from the ratio based on the amount of DPA per cell of aerobic *Bacillus* (1.4×10^{-16} mol cell^{-1}), or the average of isolates from the tidal flat (2.2×10^{-16} mol cell^{-1}). Indeed, it has been shown that thermophilic viable spores constitutes a stable component in the Arctic marine sediment with the frequency of about 0.01%, based on an exponential increase in sulfate reduction rate after germination at high temperature (Hubert *et al*., 2009). This order of sensitivity for the estimation of the fraction could not be achieved by the present metagenome sequences, but the lower boundary of the estimation [1.3×10^{-4} to 0.0079 (0.01–0.8%) for Site C9001] was reasonably consistent with their observation.

Third, sporulation-related genes in the deep subseafloor biosphere might be evolutionarily or functionally distinct from known spore-forming gene repertoires. Although the orthologous genes were more or less successfully found among genomes of likely endospore-formers in classes Erysipelotrichi and Negativicutes of phylum Firmicutes (Fig. 1), we might not be able to exclude the possibility of DPA containing other cell forms for long-term survival. For example, it has been observed that some members within the Alphaproteobacteria, such as the genera *Methylosinus* and *Methylobacter*, form exospores or cysts under the conditions of oxygen, water and energy limitation (Whittenbury *et al*., 1970), but little is known to date about molecular characteristics or genes responsible for the dormancy process.

Based on the present results, survival strategy of only < 10% of microbes was ascribed to endospore formation. Hence, it is also conceivable that subseafloor microbes may exhibit strategies for long-term survival, which differ from known mechanisms of endospore formation. Another mechanism is the 'model for life in low-energy settings' in which prolonged starvation through the stationary phase batch culture of *Escherichia coli* results in mutant cells utilizing growth advantage in stationary phase such that new cells subsist on the remains of dead predecessors (Zambrano *et al*., 1993; Finkel and Kolter, 1999; Finkel, 2006; Hoehler and Jørgensen, 2013). The significant differences in the energetic and physical conditions between laboratory batch cultures and *in situ* habitats on different timescales should also be considered. Such unknown survival strategies or yet-to-be-analysable populations in the deep subseafloor biosphere are still a mystery to be clarified in the future.

In conclusion, our study of spore-specific genes in subseafloor sedimentary metagenomes from both of the two geologically distinct locations on the Pacific Ocean

margin indicates that DNA-extractable subseafloor cells rarely have 'known' endospore-specific genes. To better understand these organisms, discriminative detection and enumeration techniques (Morono *et al*., 2013) and the DNA extraction efficiency, as well as interpretation of DPA method for endospores and vegetative cells in sediment, will be needed to open a new window to examine physiological state and functions of subseafloor life.

Acknowledgements

The authors thank all crews, drilling team members, and shipboard scientists for core sampling during the *Chikyu* Shakedown Cruise CK06-06. We thank anonymous reviewers for very helpful comments on the manuscript. This study was supported by grants from the Japan Society for the Promotion of Science (JSPS) Strategic Fund for Strengthening Leading-Edge Research and Development (to JAMSTEC and F.I.), the JSPS Funding Program for Next Generation World-Leading Researchers (NEXT Program, no. GR102, to F.I., and the JSPS Grants-in-Aid for Science Research (no. 26251041, to F.I.)). Computational resources were provided in part at the supercomputer system, JAMSTEC, and at the Data Integration and Analysis Facility, National Institute for Basic Biology.

References

Alexander, C.J., Citron, D.M., Brazier, J.S., and Goldstein, E.J. (1995) Identification and antimicrobial resistance patterns of clinical isolates of *Clostridium clostridioforme*, *Clostridium innocuum*, and *Clostridium ramosum* compared with those of clinical isolates of *Clostridium perfringens*. *J Clin Microbiol* **33:** 3209–3215.

Biddle, J.F., Fitz-Gibbon, S., Schuster, S.C., Brenchley, J.E., and House, C.H. (2008) Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc Natl Acad Sci USA* **105:** 10583–10588.

Biddle, J.F., White, J.R., Teske, A.P., and House, C.H. (2011) Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. *ISME J* **5:** 1038–1047.

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.

Church, B.D., and Halvorson, H. (1959) Dependence of the heat resistance of bacterial endospores on their dipicolinic acid content. *Nature* **183:** 124–125.

Ciobanu, M.-C., Burgaud, G., Dufrensne, A., Breuker, A., Rédou, V., Maamar, S.B., *et al*. (2014) Microorganisms persist at record depths in the subseafloor of the Canterbury Basin. *ISME J* **8:** 1370–1380.

D'Hondt, S., Rutherford, S., and Spivack, A.J. (2002) Metabolic activity of subsurface life in deep-sea sediments. *Science* **295:** 2067–2070.

D'Hondt, S., Jørgensen, B.B., Miller, D.J., Batzke, A., Blake, R., Cragg, B.A., *et al*. (2004) Distributions of microbial activities in deep subseafloor sediments. *Science* **306:** 2216–2221.

D'Hondt, S., Inagaki, F., Ferdelman, T., Jørgensen, B.B., Kato, K., Kemp, P., *et al*. (2007) Exploring subseafloor life with the Integrated Ocean Drilling Program. *Sci Drill* **5:** 26–37.

Dineen, S.M., Aranda, R.T., 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.

Dodsworth, J.A., Blainey, P.C., Murugapiran, S.K., Swingley, W.D., Ross, C.A., Tringe, S.G., *et al*. (2013) Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. *Nat Commun* **4:** Article no. 1854. doi: 10.1038/ncomms2884.

Fichtel, J., Köster, J., Rullkötter, J., and Sass, H. (2007) Spore dipicolinic acid contents used for estimating the number of endospores in sediments. *FEMS Microbiol Ecol* **61:** 522–532.

Finkel, S.E. (2006) Long-term survival during stationary phase: evolution and the GASP phenotype. *Nat Rev Microbiol* **4:** 113–120.

Finkel, S.E., and Kolter, R. (1999) Evolution of microbial diversity during prolonged starvation. *Proc Natl Acad Sci USA* **96:** 4023–4027.

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.

Hinrichs, K.-U., and Inagaki, F. (2012) Downsizing the deep biosphere. *Science* **338:** 204–205.

Hoehler, T.M., and Jørgensen, B.B. (2013) Microbial life under extreme energy limitation. *Nat Rev Microbiol* **11:** 83–94.

Hubert, C., Loy, A., Nickel, M., Arnosti, C., Baranyi, C., Brüchert, V., *et al*. (2009) A constant flux of diverse thermophilic bacteria into the cold Arctic seabed. *Science* **325:** 1541–1544.

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

Inagaki, F., Suzuki, M., Takai, K., Oida, H., Sakamoto, T., Aoki, K., *et al*. (2003) Microbial communities associated with geological horizons in coastal subseafloor sediments from the Sea of Okhotsk. *Appl Envion Microbiol* **69:** 7224–7235.

Inagaki, F., Nunoura, T., Nakagawa, S., Teske, A., Lever, M., Lauer, A., *et al*. (2006) Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proc Natl Acad Sci USA* **103:** 2815–2820.

Inagaki, F., Hinrichs, K.-U., Kubo, Y., and the Expedition 337 Scientists (2012) Deep coalbed biosphere off Shimokita: microbial processes and hydrocarbon system associated with deeply buried coalbed in the ocean. *IODP Prel Rept* **337:** 1–62. doi:10.2204/iodp.pr.337.2012.

Johnson, W.C., and Tipper, D.J. (1981) Acid-soluble spore proteins of *Bacillus subtilis*. *J Bacteriol* **146:** 972–982.
Jones, S.E., and Lennon, J.T. (2010) Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci USA* **107:** 5881–5886.
Jørgensen, B.B. (2011) Deep subseafloor microbial cells on physiological standby. *Proc Natl Acad Sci USA* **108:** 18193–18194.
Kaster, A.-K., Mayer-Blackwell, K., Pasarelli, B., and Spormann, A.M. (2014) Single cell genomic study of *Dehalococcoidetes* species from deep-sea sediments of the Peruvian Margin. *ISME J* **8:** 1831–1842.
Kawai, M., Futagami, T., Toyoda, A., Takaki, Y., Nishi, S., Hori, S., *et al*. (2014) High frequency of phylogenetically diverse reductive dehalogenase-homologous genes in deep subseafloor sedimentary metagenomes. *Front Microbiol* **5:** Article no. 80. doi:10.3389/fmicb.2014.00080.
Langerhuus, A.T., Røy, H., Lever, M.A., Morono, Y., Inagaki, F., Jørgensen, B.B., *et al*. (2012) Endospore abundance and D:L-amino acid modeling of bacterial turnover in holocene marine sediment (Aarhus Bay). *Geochim Cosmochim Acta* **99:** 87–99.
Lipp, J.S., Morono, Y., Inagaki, F., and Hinrichs, K.-U. (2008) Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454:** 991–994.
Lloyd, K.G., Schreiber, L., Petersen, D.G., Kjeldsen, K.U., Lever, M.A., Steen, A.D., *et al*. (2013) Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496:** 215–218.
Lomstein, B.A., Langerhuus, A.T., D'Hondt, S., Jørgensen, B.B., and Spivack, A.J. (2012) Endospore abundance, microbial growth and necromass turnover in deep subseafloor sediment. *Nature* **484:** 101–104.
Meng, J., Xu, J., Qin, D., He, Y., Xiao, X., and Wang, F. (2014) Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J* **8:** 650–659.
Morono, Y., Terada, T., Masui, N., and Inagaki, F. (2009) Discriminative detection and enumeration of microbial life in marine subsurface sediments. *ISME J* **3:** 503–511.
Morono, Y., Terada, T., Nishizawa, M., Ito, M., Hillion, F., Takahata, N., *et al*. (2011) Carbon and nitrogen assimilation in deep subseafloor microbial cells. *Proc Natl Acad Sci USA* **108:** 18295–18300.
Morono, Y., Terada, T., Kallmeyer, J., and Inagaki, F. (2013) An improved cell separation technique for marine subsurface sediments: applications for high-throughput analysis using flow cytometry and cell sorting. *Envion Microbiol* **15:** 2841–2849.
Morono, Y., Terada, T., Hoshino, T., and Inagaki, F. (2014) Hot-alkaline DNA extraction method for deep-subseafloor archaeal communities. *Appl Envion Microbiol* **80:** 1985–1994.
Orsi, W., Biddle, J.F., and Edgcomb, V. (2013) Deep sequencing of subseafloor eukaryotic rRNA reveals active fungi across marine subsurface provinces. *PLoS ONE* **8:** e56335. doi:10.1371/journal.pone.0056335.s009.
Parkes, R.J., Cragg, B.A., Bale, S.J., Getlifff, J.M., Goodman, K., Rochelle, P.A., *et al*. (1994) Deep bacterial biosphere in Pacific Ocean sediments. *Nature* **371:** 410–413.
Parkes, R.J., Cragg, B.A., and Wellsbury, P. (2000) Recent studies on bacterial populations and processes in subseafloor sediments: a review. *Hydrogeol J* **8:** 11–28.
R Core Team (2013) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
Raes, J., Korbel, J.O., Lercher, M.J., von Mering, C., and Bork, P. (2007) Prediction of effective genome size in metagenomic samples. *Genome Biol* **8:** R10. doi:10.1186/gb-2007-8-1-r10.
Setlow, P. (1975) Purification and properties of some unique low molecular weight basic proteins degraded during germination of *Bacillus megaterium* spores. *J Biol Chem* **250:** 8168–8173.
Setlow, P. (2007) I will survive: DNA protection in bacterial spores. *Trends Microbiol* **15:** 172–180.
Shelobolina, E.S., Nevin, K.P., Blakeney-Hayward, J.D., Johnsen, C.V., Plaia, T.W., Krader, P., *et al*. (2007) *Geobacter pickeringii* sp. nov., *Geobacter argillaceus* sp. nov. and *Pelosinus fermentans* gen. nov., sp. nov. isolated from subsurface kaolin lenses. *Int J Syst Evol Microbiol* **57:** 126–135.
Tinelli, R. (1955) Biochemistry of sporulation in *Bacillus megaterium*. I. Composition of spores obtained by deficiency of various carbonated substrates. *Ann Inst Pasteur (Paris)* **88:** 212–226.
Uchiyama, I. (2006) Hierarchical clustering algorithm for comprehensive orthologous-domain classification in multiple genomes. *Nucleic Acids Res* **34:** 647–658.
Whittenbury, R., Davies, S.L., and Davey, J.F. (1970) Exospores and cysts formed by methane-utilizing bacteria. *J Gen Microbiol* **61:** 219–226.
Wunderlin, T., Junier, T., Roussel-Delif, L., Jeanneret, N., and Junier, P. (2013) Stage 0 sporulation gene A as a molecular marker to study diversity of endospore-forming Firmicutes. *Environ Microbiol Rep* **5:** 911–924.
Yutin, N., and Galperin, M.Y. (2013) A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ Microbiol* **15:** 2631–2641.
Zambrano, M.M., Siegele, D.A., Almirón, M., Tormo, A., and Kolter, R. (1993) Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science* **259:** 1757–1760.

Supporting information

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

Fig. S1. Phylogenetic tree of orthologous groups that were composed of genes of Firmicutes and those of other phyla/candidate divisions, i.e. groups 2, 3 and 4 genes in Fig. 1B–D. For each orthologous group, genes are indicated in colour according to taxonomic assignment (upper panel) and protein domain(s) (lower panel). A frame in red for group 3 genes indicates the subcluster that was used as the reliable marker genes in Figs. 1 and 2. Genes are ordered as in Fig. 1B–D (see Appendix S1 for details).
Fig. S2. Confidence intervals estimated from different set of orthologous groups for genes in subseafloor sedimentary

metagenomes from off (A) Shimokita and (B) Peru. Genes used for the estimation of confidence interval at a given position of X-axis are those from the leftmost gene (*spoIIGA*) to the gene at the position. Groups of genes correspond to those in Figs. 1A–C and 2A–C.

Table S1. List of genomic sequences used for the construction of the orthologue table.

Table S2. Universally conserved single-copy genes in the orthologue table.

Appendix S1. Methods in detail.