Arch Microbiol (2004) 182 : 182–192
DOI 10.1007/s00203-004-0696-y

ORIGINAL PAPER

Rob U. Onyenwoke · Julia A. Brill · Kamyar Farahi
Juergen Wiegel

Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (*Firmicutes*)

Received: 19 April 2004 / Revised: 1 June 2004 / Accepted: 14 June 2004 / Published online: 31 August 2004

Abstract Endospore formation is a specific property found within bacteria belonging to the Gram-type-positive low G+C mol% branch (*Firmicutes*) of a phylogenetic tree based on 16S rRNA genes. Within the Gram-type-positive bacteria, endospore-formers and species without observed spore formation are widely intermingled. In the present study, a previously reported experimental method (PCR and Southern hybridization assays) and analysis of genome sequences from 52 bacteria and archaea representing sporulating, non-spore-forming, and asporogenic species were used to distinguish non-spore-forming (void of the majority of sporulation-specific genes) from asporogenic (contain the majority of sporulation-specific genes) bacteria. Several sporulating species lacked sequences similar to those of *Bacillus subtilis* sporulation genes. For some of the genes thought to be sporulation specific, sequences with weak similarity were identified in non-spore-forming bacteria outside of the Gram-type-positive phylogenetic branch and in archaea, rendering these genes unsuitable for the intended classification into sporulating, asporogenic, and non-spore-forming species. The obtained results raise questions regarding the evolution of sporulation among the *Firmicutes*.

Keywords Phylogeny · Sporulation genes · Dipicolinic acid synthase · Small acid soluble protein · Asporogenic and non-spore-forming bacteria · Lactic acid bacteria · Genome sequences · Gene evolution

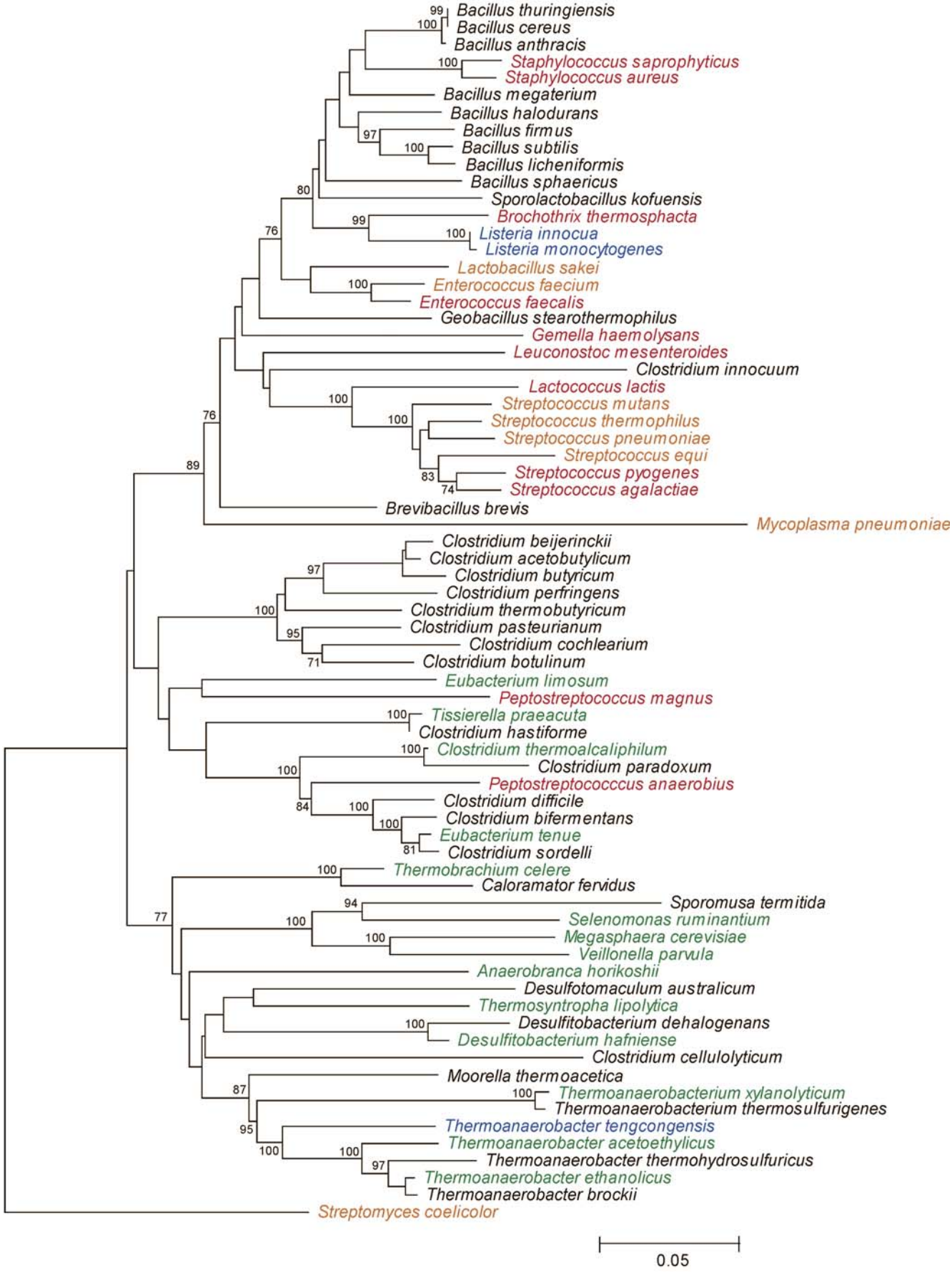
This paper is dedicated to Prof. H.G. Schlegel in honor of his 80th birthday.

R. U. Onyenwoke · J. A. Brill · K. Farahi · J. Wiegel (✉)
Department of Microbiology
and Center for Biological Resource Recovery,
University of Georgia, Athens, GA 30602-2605, USA
Tel.: +1-706-5422651,
e-mail: jwiegel@uga.edu

Introduction

Endosporulation is known to occur only among the bacteria belonging to the phylogenetic branch of Gram-type-positive bacteria (Wiegel 1981; Wiegel and Quandt 1982), also called *Firmicutes*, within the prokaryotic domain Bacteria (Gibbons and Murray 1978; Woese 1987). Garrity et al. (2003) narrowed the term "*Firmicutes*" to the status of a phylum, containing only the classes "*Clostridia*," *Mollicutes*, and "*Bacilli*," whereas other Gram-type-positive bacteria, such as *Corynebacterium*, are in the phylum "*Actinobacteria*," the second phylum containing Gram-type-positive bacteria. Being one of the most complex developmental processes in prokaryotes, endosporulation has been shown to require intricate networks of temporal and compartmental regulation in *Bacillus subtilis* and involves more than 150 different gene products, of which about 75 must act sequentially (Errington 1993; Gould 1984; Grossman 1995; Ireton and Grossman 1994; Nicholson et al. 2000; Paidhungat et al. 2001; Setlow 1995, 2001). This complexity makes the process of endosporulation vulnerable to disruptions. For instance, if not all the components that are required in the sequential process function correctly, spore formation will not be observed. Subsequently, an asporogenic phenotype can easily evolve, even though

Fig. 1 Phylogenetic tree constructed from the 16S rRNA gene with maximum likelihood correction for synonymous changes using the Fitch algorithm. Strain designations were omitted for simplicity of the tree but are included in "Materials and methods". Sporogenic species are indicated in *black*. Asporogenic species, as determined by PCR and Southern-hybridization-based assay are indicated in *green*. Assignments as asporogenic species based on genome sequence analysis are indicated in *blue*. Non-spore-forming species are indicated in *red* for PCR and Southern-hybridization-based assay and in *orange* for genome analysis. *Numbers at nodes* indicate bootstrap support values for 100 replicates. *Scale bar* denotes number of nucleotide substitutions per site. Using the neighbor-joining algorithm (figure not shown), a very similar tree was obtained, except the relative position of *Geobacillus stearothermophilus* and *Bacillus brevis* were closer to the position of *Bacillus subtilis*, whereas *Clostridium innocuum* was closer to *Mycoplasma pneumoniae* ▶


Bacillus thuringiensis
Bacillus cereus
Bacillus anthracis
Staphylococcus saprophyticus
Staphylococcus aureus
Bacillus megaterium
Bacillus halodurans
Bacillus firmus
Bacillus subtilis
Bacillus licheniformis
Bacillus sphaericus
Sporolactobacillus kofuensis
Brochothrix thermosphacta
Listeria innocua
Listeria monocytogenes
Lactobacillus sakei
Enterococcus faecium
Enterococcus faecalis
Geobacillus stearothermophilus
Gemella haemolysans
Leuconostoc mesenteroides
Clostridium innocuum
Lactococcus lactis
Streptococcus mutans
Streptococcus thermophilus
Streptococcus pneumoniae
Streptococcus equi
Streptococcus pyogenes
Streptococcus agalactiae
Brevibacillus brevis
Mycoplasma pneumoniae
Clostridium beijerinckii
Clostridium acetobutylicum
Clostridium butyricum
Clostridium perfringens
Clostridium thermobutyricum
Clostridium pasteurianum
Clostridium cochlearium
Clostridium botulinum
Eubacterium limosum
Peptostreptococcus magnus
Tissierella praeacuta
Clostridium hastiforme
Clostridium thermoalcaliphilum
Clostridium paradoxum
Peptostreptococccus anaerobius
Clostridium difficile
Clostridium bifermentans
Eubacterium tenue
Clostridium sordelli
Thermobrachium celere
Caloramator fervidus
Sporomusa termitida
Selenomonas ruminantium
Megasphaera cerevisiae
Veillonella parvula
Anaerobranca horikoshii
Desulfotomaculum australicum
Thermosyntropha lipolytica
Desulfitobacterium dehalogenans
Desulfitobacterium hafniense
Clostridium cellulolyticum
Moorella thermoacetica
Thermoanaerobacterium xylanolyticum
Thermoanaerobacterium thermosulfurigenes
Thermoanaerobacter tengcongensis
Thermoanaerobacter acetoethylicus
Thermoanaerobacter thermohydrosulfuricus
Thermoanaerobacter ethanolicus
Thermoanaerobacter brockii
Streptomyces coelicolor
0.05

many functional sporulation genes still will be present. Most of the processes of endosporulation investigated to date appear to be highly similar among all endospore-forming species, and thus it is usually assumed that all endospore-forming species most likely arose from the same sporulating ancestor (Errington 1993; Gerhardt and Marquis 1989; Nakamura et al. 1995; Sauer et al. 1994, 1995).

Until recently, the ability to form endospores was used as a mandatory characteristic to include novel isolates into the genera *Bacillus*, *Desulfotomaculum*, and *Clostridium* (Hippe et al. 1992; Slepecky and Hemphill 1992; Sneath 1984). However, according to phylogenetic analyses based on 16S rRNA genes, *Bacillus* and *Clostridium* are not coherent genera and are interspersed with genera partly or exclusively consisting of species for which endospore formation has not been observed (Fig. 1; Ash et al. 1991; Collins et al. 1994). Consequently, the genera *Bacillus* and *Clostridium* have been redefined and subdivided into novel families and genera (P. deVos, Universiteit Gent, in preparation; Garrity and Holt 2001; J. Wiegel, R. Tanner, and T. Rainey, 2004 in press). Several of the newly formed genera contain both endospore-forming species and species for which no endosporulation has been observed. This has led to the speculation that in a large number of species that do not form endospores, the ability to sporulate has been lost due to an interruption of the sequential sporulation process.

Brill and Wiegel (1997) attempted to develop a fast method to separate novel isolates into asporogenic and non-spore-forming species on the basis of the presence and absence of sporulation genes, respectively. [The term asporogenic has lately been used, including by Brill and Wiegel (1997), in contrast to the use in the older literature and in several recent sporulation-related publications. To avoid further confusion, we have decided to use the term in agreement with the traditional use: "asporogenic" = bacteria with an impaired sporulation process but containing the majority of sporulation genes and "non-spore-forming" = absence of sporulation specific genes such as in Gram-type-negative *Escherichia coli* or *Pseudomonas* spp.] Another term for asporogenic bacteria would be cryptic spore-formers since the restoration of one or two gene products would presumably lead to spore formation. Brill and Wiegel (1997) described a PCR and Southern-hybridization-based assay to distinguish between asporogenic and non-spore-forming by employing probes directed against three representative and at the time assumed specific sporulation genes: *spo0A*, *ssp*α/β-type, and *dpaA*/*B*. We applied this assay but also extended our study to include analysis of available (up to February 2004) genome sequences. By searching for sequences with similarity to 66 sporulation related genes from the sporulation model microorganism *B. subtilis* (Stragier 2002), we sought to validate our results that some of the asporogenic members of the low G+C group of the *Firmicutes* do not contain "sporulation specific" genes. In the present study, we analyzed genes that are similar to *B. subtilis* sporulation-related genes using complete genome sequences from a multitude of prokaryotes.

Materials and methods

Organisms and growth conditions. The tested bacteria and their source are given in Table 1 and were cultivated as previously described (Brill and Wiegel 1997).

Genomic DNA isolation, PCR, Southern hybridizations, and sequencing. Genomic DNA isolation, PCR, Southern hybridizations, and Sequencing were carried out as previously described (Brill and Wiegel 1997).

Sequence retrieval and phylogenetic analysis. The following *B. subtilis* sequences served as the query sequences when searching against the databases of complete genome sequences.

- Preseptation: *minC*, *spo0A*, *spo0B*, *spo0H* (sigma factor), *rapA*, *spoVG*, *spoIIAA* (anti-sigma factor antagonist), *spoIIAB* (anti-sigma factor), *spoIIAC* (sigma factor), *spoIIB*, *spoIID*, *spoIIE*, *spoIIGA*, *spoIIGB* (sigma factor precursor), *spoIIM*, *spoIIP*, *spoIIQ*, *spoIIR*;
- Postseptation: *gerM*, *spoIIIA (A, B, C, D, E, F, G, H)*, *spoIVA*, *spoIIID*, *spoVB*, *spoVK*, *cotE*;
- Postengulfment: *spoVID*, *gerPA*, *spoIIIG* (sigma factor), *spoIVB*, *spoIVCB* (sigma factor precursor), *spoVA (A, B, C, D, E, F)*, *spoVM*, *dpaA*, *dpaB*, *hep1*, *sspA*, *sspE*, *gerA (A, B, C)*, *gerD*, *gpr*, *splB*, *sleB*, *cwlD*, *cwlJ*, *cotD*, and *yqfC*, *yqfD*, *yabP*, *yabQ*, *spmA*, *spmB*.

These sequences were obtained by screening of the amino acid sequences of the SwissProt/TrEMBL databases (http://us.expasy.org/sprot/). Non-redundant GenBank (http://www.ncbi.nlm.nih.gov/, reference sequences are indicated for GenBank), SwissProt, and ERGO (ERGO-derived sequences are indicated by superscript 1) databases were used to obtain the gene assignments for the comparisons of the amino acid sequences. Basic Local Alignment Search Tool (BLAST) [(http://www.ncbi.nlm.nih.gov/BLAST)] searches were done on each open reading frame (ORF) using a basic protein-to-protein blast search of amino acid similarities to sequences in the GenBank, SwissProt, and ERGO non-redundant databases in February 2004. The results of all of these searches were used to provide a putative identication of probable ORF sequences with similarity to the *B. subtilis* query sequences. Probable similarity was based on the combination of identity and probability (*p*) scores, and *e*-values from BLAST analyses. Final assignments of similarity, as illustrated in Tables 2, 3, 4, 5, were based on comparisons of the *e*-value of the query *B. subtilis* sequence to the similar sequence that was delivered by the BLAST search. A (++) designation in the previously mentioned tables indicated a difference of less than 20-fold of the similar sequence *e*-value to the *B. subtilis* sequence. The (+) designation was a difference greater than 20-fold but less than or equal to 40-fold. The (+/–) designation indicated a similar sequence existed but with an *e*-value greater than 40-fold different from that of the *B. subtilis* query. The (–) designation indicated no similar sequence was present or that the similar sequence

Table 1 Bacterial species experimentally tested for the presence of sporulation-specific genes *spo0A*, *ssp*, and *dpa* (*A*/*B*). Clusters are as assigned by Collins et al. (1994) and Farrow et al. (1995). Correct lengths of PCR-products are 100 bp for *ssp* and 300 bp for *spo0A*. For sequence of the selected strains see Brill and Wiegel (1997). *ND* Not determined

Bacteria	Clusters	PCR products for *ssp* and *spo0A* and positive Southern hybridizations for all *ssp*, *spo0A*, and *dpa* probes	PCR-control Universal eubacterial 16S rRNA gene probe
Thermobrachium celere DSM 8682	Between II and III	+	+
16 isolates of *Thermobrachium celere*		+[a]	ND
Thermoanaerobacter ethanolicus DSM 2246	V	+	ND
Thermobrachium acetoethylicus ATCC 2359	V	+	ND
Thermosyntropha lipolytica DSM 11003	VIII	+[b]	+
Megasphaera cerevisiae ATCC 43254	IX	+	+
Veillonella parvula ATCC 10790	IX	+	ND
Selenomonas ruminantium ATCC 12561	IX	+	+
Selenomonas sputigena ATCC 35185	IX	+	+
Anaerobranca horikoshii DSM 11003	Closest to IX	+	ND
Desulfitobacterium dehalogenans DSM 9161	Closest to IX	+	ND
Peptostreptococcus anaerobius ATCC 27337	XI	–	+
Clostridium thermoalkaliphilum DSM 7309	XI	+	+
Eubacterium tenue ATCC 25553	XI	+	+
Tissierella praeacuta ATCC 25539	XII	+	+
Peptostreptococcus magnus ATCC 15	XIII	–	+
Eubacterium limosum ATCC 8486	XV	+[b]	+
Leuconstoc mesenteroides ssp. *mesenteroides* ATCC 8293	Lactobacillales	–	+
Lactobacillus delbrueckii ssp. *delbrueckii* ATCC 9649	Lactobacillales	–	+
Lactococcus lactis ssp. *lactis* ATCC 19435 794	Lactobacillales	–	+
Streptococcus agalactiae ATCC 13813	Lactobacillales	–	+
Streptococcus pyogenes ATCC 12344	Lactobacillales	–	+
Carnobacterium divergens ATCC 35677	Lactobacillales	–	+
Gemella haemolysans ATCC 10379	Lactobacillales	–	+
Enterococcus faecalis ATCC 19433	Bacillales	–	+
Staphylococcus saprophyticus ATCC 15305	Bacillales	–	+
Staphylococcus aureus ATCC 12600	Bacillales	–	+
Listeria monocytogenes ATCC 15313	Bacillales	–	+
Brochothrix thermosphacta ATCC 11509	Bacillales	–	+

[a]Some strains did not yield unequivocal PCR-products for *ssp1*. Others did not yield positive hybridization bands with the *dpa* probe but four out of the five tests were always positive. [b]No unequivocal PCR product for *ssp1* obtained.

had an *e*-value greater than or equal to 0.001, which was used as the cut-off value for this work. The genomes of the following bacteria and archaea were used in the comparison, superscript T indicates it is the type strain: *Staphylococcus aureus* MSSA strain 476[1], *Enterococcoccus faecalis*[1], *Enterococcus faecium* DO(JGI)[1], *Lactococcus lactis* subsp. *lactis*[1] and reference sequence: NC_002662[T], *Listeria innocua*[T], *Listeria monocytogenes* EGD-e[1] and NC_003210, *Streptococcus agalactiae* 2603V/R NC_004116, *Streptococcus equi*[1] and NC_002955, *Streptococcus mutans* UA159[1] and NC_004350 *Streptococcus pneumoniae* TIGR4[1], *Streptococcus pyogenes* SF 370-M1[1], *Streptococcus salvarius* subsp. *thermophilus*[1], *Mycoplasma pneumoniae* M129[1], *Streptomyces coelicolor* A3(2)[1], *Haemophilus actinomycetemcomitans* NC_002924 and HK 1651[1], *Caulobacter crescentus* CB15, NC_002696, *Escherichia coli* K12, NC_000913 *Escherichia coli* O157:H7, NC_002695 *Helicobacter pylori* J99[1] and NC_000921, *Myxococcus xanthus*[1] and NC_004802, *Salmonella enterica* subsp. *enterica* serovar *typhi NC_003198*, *Salmonella typhimurium* LT2[1] and NC_003197, *Borrelia burgdorferi* B31[1 T] and NC_004971, *Synechococcus* spp. PCC7002[1] and WH8102 NC_005070, *Prochlorococcus marinus* subsp. *pastoris* str. CCMP 1378 NZ_AAAW00000000, *Anabaena* spp. PCC7120[1], *Chloroflexus aurantiacus*[T], *Fusobacterium nucleatum* spp. *nucleatum* ATCC 25586[T], *Deinococcus radiodurans* R1[T], *Cytophaga hutchinsonii* DSM 4304[T], *Archaeoglobus fulgidus* DSM 4304[1 T] and NC_000917, *Halobacterium* spp. NRC-1 NC_002607, *Methanocaldococcus jannaschii* DSM 2661[T], *Pyrococcus furiosus* DSM 3638[T], *Aeropyrum pernix*[1] and K1[T], *Bacillus anthracis* strain Ames NC_003997, *Bacillus cereus* ATCC 14579[T], *Bacillus firmus*[1], *Bacillus halodurans* C-125[1] and NC_002570, *Bacillus licheniformis*[1], *Bacillus megaterium*[1], *Bacillus sphaericus*[1], *Geobacillus stearothermophilus* 10[1] and NC_002926, *Bacillus subtilis* subsp. *subtilis* strain 168[1] and NC_000964, *Bacillus thuringiensis*[1], *Bacillus thuringiensis* subsp. *israelensis* ATCC 35646 IG-59[1], *Clostridium acetobutylicum* ATCC824[T], *Clostridium botulinum* A NC_003223, *Clostridium difficile* 630[1] and NC_002933, *Clostridium perfringens* strain 13124[T], *Desulfitobacterium hafniense* NZ_AAAW00000000.

Tree Building. The Phylogeny Inference Package (PHYLIP) and Molecular Evolutionary Genetics Analysis (MEGA2) software contained all of the programs used for inferring

phylogenies. The 16S rRNA gene trees were generated using different algorithms [Fitch (Fig. 1), neighbor-joining (data not shown), and UPGMA (data not shown)] to examine differing tree constructions as a factor in determining the interrelatedness of the representative microorganisms and to generate the distance matrixes. The ClustalX alignment tool Tree Explorer was used for tree viewing.

Results and discussion

Experimentally analyzed species. Based on the assumption that sporulation evolved only once, and thus all or the majority of sporulation-specific genes exhibit sequence similarity, we applied the method of Brill and Wiegel (1997) to analyze 28 *Firmicutes* species from 22 genera and to 16 strains of *Thermobrachium celere* (Engle et al. 1996; Table 1). These asporogenic species included pairs of closely related (98% 16S rRNA sequence similarity) species representing isolates from a great variety of mesobiotic and thermobiotic environments. The selected and representative species covered eight different *Firmicutes* clusters as defined by Collins et al. (1994) as well as lactic acid bacteria and genera related to *Bacillus* (Table 1). Many other species or genera from the same or other clusters, such as clusters 16–20 (Engle et al. 1996), could have been chosen. For most of the tested strains, the method of Brill and Wiegel (1997) yielded positive results, i.e., at least three of the five assays indicated the presence of the representative sporulation genes (Table 1). The results suggested that this method provides a relatively fast experimental means to differentiate between non-spore-forming and asporogenic bacteria (Wiegel 1992), especially in species and strains for which no genome sequences are available. The data also support the notion that differentiation between some of the "clostridial" species and species of the genera described as lacking sporulation should be reevaluated. As an example, the separation of *Clostridium hastiforme* from the *Tissierella* species based on sporulation appears to be no longer justified on the basis of the high 16S rRNA sequence similarity of these bacteria (Fig. 1) combined with the presented data that the type species *Tissierella praeacuta* is asporogenic, i.e., contains sequences with similarity to sporulation genes from *B. subtilis* and other clostridia (Brill and Wiegel 1997). Thus, the genus *Tissierella* should include sporulating species such as *C. hastiforme*.

However, a few species did not yield indications for the presence of sporulation genes (Table 1). These species would now be regarded as non-spore-forming. As evident from the phylogenetic tree (Fig. 1), the non-spore-forming, Gram-type-positive bacteria identified here do not cluster all together, but belong to different sub-branches, e.g., the *Peptostreptococcus* species, or the *Enterococcus*/*Carnobacterium*/*Gemella*/*Staphylococcus* group, possibly constituting a non-spore-forming sub-branch of the otherwise spore-forming *Bacillales*/*Lactobacillales*.

Genomic analysis among Bacillus *spp.* Due to some indications that the experimental test did not give unequivocal results (E. Stackebrandt, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH-Braunschweig, Germany, personal communication) and the comparison of sequences of sporulation genes done by Stragier (2002), we used available (February 2004) genome sequences to challenge our findings and to test for the assumed similarities of sporulation genes. Fifty-two bacteria and archaea, representing diverse phyla, were selected in order to search for the occurrence of 65 assumed "spore-specific" genes that were chosen because they are a representative sampling of the genes required for each stage of sporulation: preseptation, postseptation, and postengulfment (Stragier 2002). The sequence identity matches were based on identity to the *B. subtilis* sequences (as explained in the "Materials and methods"), which were used as reference sequences. Among the chosen microorganisms were: 17 sporogenic bacteria (*Firmicutes*), 29 bacteria that do not form spores but belong to *Firmicutes*, and as negative controls Proteobacteria (representing 4 of the 5 classes), spirochetes, cyanobacteria, green non-sulfur bacteria, *Fusobacteria*, *Deinococcus*, *Bacteroides*, and *Actinobacteria*, and five archaea (see "Materials and methods"). The choice of microorganisms was somewhat random but did depend on the availability of at least greater than 95% complete genome sequence as a prerequisite. Unfortunately, no genome sequences were available (February 2004) for most of the above-described experimentally tested bacterial species (Table 1).

Occurrence and absence of sporulation specific genes in completed genome sequences. Based on the similarity of the predicted protein sequence, the analysis of the complete genome data indicated that, with a high probability, some "spore-specific" genes are in the genomes of additional *Firmicutes* that have never been shown to produce endospores (Table 4). This includes several *Streptococcus* species, two *Enterococcus* species, *Lactococcus lactis*, and most numerously in the studied *Listeria* species. *Listeria monocytogenes* and *Listeria innocua* contain 17 supposedly "spore-specific" genes (Table 4), which include many sigma-factor-related genes. In contrast to the microorganisms listed in Table 4, the analysis of several *Bacillus* spp. clearly demonstrates that the sporulating species *B. licheniformis*, *B. firmus*, *B. sphaericus*, and *B. thuringiensis* apparently have quite different or modified sporulation genes, showing no or low levels of similarity with the genes from *B. subtilis* (Table 2). This finding is in agreement with unpublished data from E. Stackebrandt (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH-Braunschweig, Germany, personal communication) using the experimental assay of Brill and Wiegel (1997). One strain of *Geobacillus stearothermophilus* revealed fewer genes with similarity to the *B. subtilis* genes than to the second analyzed strain that contained many genes with similar sequences [It needs to be pointed out that for many strains the assignment to *G. stearothermophilus* prior to the use of 16S rRNA sequence analysis is questionable (Ash et al. 1991)]. Stragier (2002), based on

Table 2 Presence and absence of sporulation genes (with sequence similarity to *Bacillus subtilis* genes) in genomes of *Bacillus* and *Geobacillus* species. Designations are based on *p*-values and *e*-scores obtained from BLAST searches to identify relative similarity of gene sequence to the *B. subtilis* query gene sequences (see "Materials and methods" section "Sequence retrieval and phylogenetic analysis" for specific details of the analysis. (++) Scores identify sequence as similar to the *B. subtilis* query sequence. (+) Some similarity over parts of the sequence exists to the *B. subtilis* query sequence. (+/–) Score is too low to allow for a definitive classification of the sequence relative to the *B. subtilis* query sequence. (–) Score indicates that little to no similarity exists relative to the *B. subtilis* query sequence. The genes that are in *bold* represent similar sequences found in species outside the *Firmicutes* (Table 5). *Asterisk* Indicates the number of genes not found were too numerous to list and include those not otherwise indicated in the other columns

Bacillales	(++)	(+)	(+/–)	(–)
Bacillus subtilis	All genes present (reference set)			
Bacillus anthracis	***spoVG***, ***spoIIA(A, B)***, ***spoIID***, ***spoIIE***, ***spoIIGB***, *spoIIP*, *gerM*, *spoIIIAE*, *spoIVA*, *spoIIID*, ***spoVB***, *spoVID*, ***spoIIIG***, *spoIVB*, *spoVA(D, F)*, ***sspA***, *gerA(A, C)*, *gpr*, ***splB***, *yqfC*, *yqfD*, *yabP*	***spo0A***, ***spo0H***, ***spoIIAC***, *spoIIIA(B, D)*, *cotE*, *gerPA*, *spoVA(B, C)*, ***dpaB***, *cwlJ*, ***spmA***, ***spmB***	***minC***, *rapA*, *spoIIGA*, *spo0B*, *spoIIB*, *spoIIM*, ***spoIIQ***, *spoIIR*, *spoIIIA(A, C, F, G, H)*, ***spoVK***, ***spoIVCB***, *spoVA(A, E)*, ***dpaA***, *hep1*, *gerAB*, *gerD*, ***sleB***, ***cwlD***, *cotD*, *yabQ*	*spoVM*, *sspE*
Bacillus cereus	***spoVG***, ***spoIIA(A, B)***, ***spoIID***, ***spoIIE***, ***spoIIGB***, *spoIIP*, *gerM*, *spoIIIAE*, *spoIVA*, *spoIIID*, ***spoVB***, *spoVID*, *spoIIIG*, *spoIVB*, *spoVA(D, F)*, *sspA*, *gerA(A, C)*, *gpr*, ***splB***, *yqfC*, *yqfD*, *yabP*	***spo0A***, ***spo0H***, ***spoIIAC***, *spoIIIA(B, D)*, *cotE*, ***gerPA***, ***spoVA(B, C)***, *dpaB*, ***cwlJ***, *spmA*, *spmB*	*minC*, *spo0B*, *rapA*, *spoIIB*, *spoIIGA*, *spoIIM*, ***spoIIQ***, *spoIIR*, *spoIIIA(A, C, F, G, H)*, ***spoVK***, ***spoIVCB***, *spoVA(A, E)*, ***dpaA***, *hep1*, *gerAB*, *gerD*, ***sleB***, ***cwlD***, *cotD*, *yabQ*	*spoVM*, *sspE*
Bacillus firmus		*sspA*	***spoVK***	*
Bacillus halodurans	***spoVG***, ***spoIIAA***, ***spoIID***, ***spoIIE***, ***spoIIGB***, *spoIIP*, *gerM*, *spoIIIAE*, *spoIVA*, *spoIIID*, ***spoVB***, *spoVID*, *spoIVB*, *spoVAF*, *spoVM*, *sspA*, *gerAA*, *gpr*, ***splB***, *yqfD*, *yabP*	***spo0A***, ***spo0H***, ***spoIIA(B, C)***, *spoIIIA(B, D)*, *cotE*, ***spoIIIG***, *spoVA(C, D)*, ***dpaB***, *sspE*, *gerAC*, *cwlJ*, *yqfC*, ***spmA***, ***spmB***	***minC***, *spo0B*, *rapA*, *spoIIB*, *spoIIGA*, *spoIIM*, ***spoIIQ***, ***spoIIR***, *spoIIIA(A, C, F, G, H)*, ***spoVK***, ***spoIVCB***, *spoVAE*, ***dpaA***, *hep1*, *gerAB*, *gerD*, ***sleB***, ***cwlD***, *cotD*, *yabQ*	*gerPA*, *spoVA(A, B)*
Bacillus licheniformis	***spo0H***, ***spoIIA(A, B, C)***	*rapA*, *yqfC*, *yqfD*	*spoIIB*, ***spoIIE***, ***spoIIGB***, ***spoIIIG***, ***spoIVCB***, ***cwlD***	*
Bacillus megaterium	***spoVG***, ***spoIIAA***, ***spoIIE***, *spoIIP*, *spoVID*, *sspA*, *gerAA*, *gpr*, *yqfC*, *yqfD*	***spo0H***, ***spoIIA(B, C)***, *sspE*, *gerAC*	***spo0A***, *spoIIG(A, B)*, ***spoIIIG***, ***spoIVCB***, *spoVAF*, *gerAB*, ***splB***, *cwlJ*	*
Bacillus sphaericus		***spoIIAB***	***spo0A***, ***spoIIA(A, C)***, ***spoIIGB***, ***spoIIIG***, ***spoIVCB***, ***dpaA***	*
Bacillus thuringiensis	***spoIIGB***, *spoIIID*	***spo0A***	***spo0H***, *rapA*, ***spoIIAC***, *spoIIGA*, ***spoIIIG***, ***spoIVCB***, *spoVAF*	*
Bacillus thuringiensis subsp. *israelensis*	***spoVG***, ***spoIIA(A, B)***, ***spoIIE***, ***spoIIGB***, *spoIIP*, *gerM*, *spoIIIAC*, *spoIIID*, *D, E)*, ***spoVB***, *spoVID*, *spoIVB*, *spoVAC*, *sspA*, *gpr*, ***splB***, *yqfC*, *yqfD*, *yabP*	***spo0A***, *gerA(A, C)*, ***spoIIAC***, *spoIIIA(B,cotE*, *gerPA*, ***spoIIIG***, *spoVAB*, *cwlJ*, ***spmB***	***minC***, *spo0B*, ***spo0H***, *rapA*, *spoIIB*, ***spoIID***, *spoIIGA*, *spoIIM*, ***spoIIQ***, *spoIIR*, *spoIIIA(A,F, G, H)*, ***spoIVCB***, *spoVA(A, D, E, F)*, *spoVK*, ***dpaA***, ***dpaB***, *hep1*, ***sleB***, ***cwlD***, *cotD*, *yabQ*, ***spmA***, *gerAB*	*spoIVA*, ***spoVB***, *spoVM*, *sspE*, *gerA(A, C)*
Geobacillus stearothermophilus[a]	***spoIIAA***, *sspA*, *sspE*	***spo0A***, ***spoIIA(B, C)***	*rapA*, ***spoIIGB***, ***spoIIIG***, *hep1*, ***spoIVCB***, ***dpaA***	*
Geobacillus stearothermophilus 10	***spoVG***, ***spoIIA(A, B)***, ***spoIID***, ***spoIIE***, ***spoIIGB***, *spoIIP*, *gerM*, *spoIIIA(C, E)*, *spoIIID*, *spoIVA*, ***spoVB***, *spoVID*, *gerPA*, ***spoIIIG***, *spoIVB*, *spoVAF*, *gerAA*, *gpr*, ***splB***, *yqfC*, *yqfD*, *yabP*, ***spmA***	***spo0A***, ***spoIIAC***, *spoIIIA(B, D)*, *cotE*, *spoVA(B, C, D)*, *cwlJ*, ***spmB***	***minC***, *spo0B*, ***spo0H***, *rapA*, *spoIIGA*, *spoIIM*, ***spoIIQ***, *spoIIR*, *spoIIIA(A,F, G)*, ***spoVK***, ***spoIVCB***, *spoVA(A, E)*, ***dpaA***, ***dpaB***, *hep1*, *gerA(B, C)*, *gerD*, ***sleB***, ***cwlD***, *yabQ*	*spoIIB*, *spoIIIAH*, *spoVM*, *sspA*, *sspE*, *cotD*

[a]Difference between the two *Geobacillus stearothermophilus* strains refers to the source of the genome data, see "Materials and methods"

Table 3 Presence and absence of sporulation genes (with sequence similarity to *B. subtilis* genes) in genomes of *Clostridium* and *Desulfitobacterium* species. Designations are based on *p*-values and *e*-scores obtained from BLAST searches to identify relative similarity of gene sequence to the *B. subtilis* query gene sequences. (++) Scores identify sequence as similar to the *B. subtilis* query sequence. (+) Some similarity exists over parts of the sequence to the *B. subtilis* query sequence. (+/–) Score is too low to allow for a definitive classification of the sequence relative to the *B. subtilis* query sequence. (–) Score indicates that little to no similarity exists relative to the *B. subtilis* query sequence. The genes that are in *bold* represent similar sequences found in species outside the *Firmicutes* (Table 5). *Asterisk* indicates the number of genes not found were too numerous to list and include all those not otherwise indicated in the other columns

Clostridiales	(++)	(+)	(+/–)	(–)
Clostridium acetobutylicum	*spoIVA*	***spoVG***, ***spoIIE***, ***spoIIGB***, *spoIIID*, ***spoIIIG***, *sspA*, *gerA(A, C)*, *yqfC*	***minC***, ***spo0A***, ***spo0H***, *rapA*, ***spoIIA(A, B, C)***, ***spoIID***, *spoIIM*, *spoIIP*, ***spoIIQ***, *spoIIR*, *spoIIIA(A, B, C, D, E, F, G, H)*, ***spoVB***, ***spoVK***, *spoIVB*, ***spoIVCB***, *spoVA(C, D, E, F)*, *gerAB*, *gpr*, ***splB***, ***sleB***, ***cwlD***, *cwlJ*, *yqfD*, *yabP*, *yabQ*, ***spmA***, ***spmB***	*spo0B*, *spoIIB*, ***spoIIGB***, *gerM*, *cotE*, *spoVID*, *gerPA*, *spoVA(A, B)*, *spoVM*, ***dpaA***, ***dpaB***, *hep1*, *sspE*, *gerD*, *cotD*
Clostridium botulinum		*gerAC*	*spoVAF*, *gerA(A, B)*	*
Clostridium difficile	*spoIVA*	***spoVG***, ***spoIIA(A, B)***, ***spoIIE***, *spoIIIA(C, D)*, *spoIIID*, ***spoIIIG***	***minC***, ***spo0A***, ***spo0H***, *rapA*, ***spoIIAC***, ***spoIID***, *spoIIGA*, ***spoIIGB***, *spoIIM*, *spoIIP*, ***spoIIQ***, *spoIIR*, *spoIVB*, ***spoIVCB***, *spoIIIA(E, G, H)*, ***spoVB***, ***spoVK***, *spoVA(C, D, E)*, ***dpaA***, ***dpaB***, *gpr*, ***splB***, ***sleB***, ***cwlD***, *yqfC*, *yqfD*, *yabP*, *yabQ*, ***spmA***, ***spmB***	*spo0B*, *spoIIB*, *gerM*, *spoIIIA(**A, B, F**)*, *spoIVA*, *cotE*, *spoVID*, *gerPA*, *spoVA(A, B, F)*, *spoVM*, *hep1*, *sspA*, *sspE*, *gerA(A, B, C)*, *gerD*, *cwlJ*, *cotD*
Clostridium perfringens	*spoIVA*	***spoVG***, ***spoIIA(A, B)***, ***spoIID***, ***spoIIE***, ***spoIIGB***, *spoIIID*, ***spoIIIG***, *sspA*, *yqfC*	***minC***, ***spo0A***, ***spo0H***, ***spoIIAC***, *spoIIGA*, *spoIIM*, *spoIIP*, ***spoIIQ***, *spoIIR*, *spoIIIA(**A, B, C, D, E, G, H**)*, ***spoVB***, ***spoVK***, *spoIVB*, ***spoIVCB***, *spoVA(C, D, E, F)*, *gerAA*, *gpr*, ***splB***, ***sleB***, ***cwlD***, *yqfD*, *yabP*, *yabQ*, ***spmA***, ***SpmB***	*spo0B*, *rapA*, *gerM*, *spoIIB*, *spoIIIAF*, *cotE*, *spoVID*, *gerPA*, *spoVA(A, B)*, *spoVM*, ***dpaA***, ***dpaB***, *hep1*, *sspE*, *gerA(B, C)*, *gerD*, *cwlJ*, *cotD*
Desulfitobacterium hafniense		***spoVG***, *spoIVA*, *spoIIID*, ***spoIIIG***, *sspA*, *gerA(A, C)*, *yqfC*	***minC***, ***spo0A***, ***spo0H***, ***spoIIA(A, B, C)***, ***spoIID***, *spoIIG(A, **B**)*, *spoIIP*, ***spoIIQ***, *spoIIR*, *spoIIIA(**A, B, D, E, G**)*, ***spoVB***, ***spoVK***, *spoIVB*, ***spoIVCB***, *spoVA(C, D, E, F)*, ***dpaA***, ***dpaB***, *gerAB*, *gpr*, ***splB***, ***sleB***, ***cwlD***, *cwlJ*, *yqfD*, *yabP*, ***spmA***, ***spmB***	*spo0B*, *rapA*, *spoIIB*, ***spoIIE***, *spoIIM*, *gerM*, *spoIIIA(C, F, H)*, *cotE*, *spoVID*, *gerPA*, *spoVA(A, B)*, *spoVM*, *hep1*, *sspA*, *sspE*, *gerD*, *cotD*, *yabQ*

the comparison of the few *Bacillus* spp. and *Clostridium* spp. genome sequences available at that time of analysis, suggested already to some extent that the inferred diversity in the sequences of sporulation genes needs to be further analyzed and quantified in the future to further elucidate the phylogeny.

Using Clostridium acetobutylicum *sequences as reference*. Due to the limitation of our analysis using *B. subtilis* gene sequences as reference, we carried out a separate analysis using the sporulation-specific genes from *C. acetobutylicum* as a reference (excluding the non-diagnostical sigma factors and related genes). However, in general, the clostridial reference sequences did not yield genes with higher similarities in the other clostridia and *Desulfitobacterium* than the *B. subtilis* reference sequences (detailed data not shown). As an example, when the sequences from *C. acetobutylicum* were used, *spo0A* and *spoIID* from *Desulfitobacterium*, *C. perfringens*, *C. botulinum*, and *C. difficile* still occurred in the (+/–)-category, except for *spoIID* from *C. perfringens* which moved from the (+)-category to the (++)-category (see "Sequence retrieval and phylogenetic analysis" section under "Materials and methods", and Table 3 for explanations).

Reclassifying assumed sporulation-specific genes as diagnostically not usable. Initial experiments suggested that the three chosen representative sporulation-specific genes could be used to differentiate (novel) species into non-spore-forming or asporogenic species (Table 1; species in green in Fig. 1). Subsequent analysis of genome sequences revealed that some sequences with similarity to assumed "sporulation-specific" genes could also be identified in non-spore-forming genera and species (Brill and Wiegel 1997). These taxa include several *Streptococcus* spp. and *Lactococcus lactis* (Table 4), and the genes include: *spo0A*, *spo0H*, *spoIIA(A, B, C)*, *spoIIE*, *spoIIGB*, *spoIIQ*, *spoIVA* (role at an early stage in the morphogenesis of the spore-coat outer layers), *spoVB*, *spoVK*, *spoVID* (required for assembly of the spore coat), *spoVG*, *minC*, *spoIIIG*,

Table 4 Gene sequences with similarity to sporulation genes observed in genomes of Gram-type-positive microorganisms that do not form endospores. (++) Scores identify a sequence as being similar to the *B. subtilis* query sequence. (+) Some similarity over parts of the sequence to the *B. subtilis* query sequence. (+/–) Score is too low to allow for a definitive classification of the sequence relative to the *B. subtilis* query sequence. The genes that are in *bold* represent similar sequences found in species outside the *Firmicutes* (Table 5). The genes *underlined* represent similar sequences not found in species outside the *Firmicutes* (Tables 2, 3, 4)

Gram-type positive	(++)	(+)	(+/–)
Staphylococcaceae			
Staphylococcus aureus		***spoVG***, *spoVID*	***spo0A***, ***spo0H***, ***spoIIA(A, B, C)***, ***spoIIGB***, ***spoIIE***, ***spoIIQ***, ***spoVB***, ***spoVK***, ***spoIIIG, spoIVCB***, ***dpaA***, ***cwlD***
Mycoplasmatales			
Mycoplasma pneumoniae			***spo0H***, *spoIIGB*, *spoVB*, *spoIIIG*
Lactobacillales			
Enterococcus faecalis			***spo0A***, ***spo0H***, ***spoIIA(B, C)***, ***spoIIGB***, ***spoIIQ***, ***spoVB***, ***spoVK***, ***spoIIIG***, ***spoIVCB***
Enterococcus faecium			***spo0A***, *spo0H*, *spoIIGB*, *spoIIQ*, *spoVB*, *cwlD*
Lactococcus lactis			***spo0A***, ***spoIIAC***, ***spoIIGB***, ***spoVB***, ***spoVK***, ***spoIIIG***, ***spoIVCB***
Listeria innocua	***spoVG***		***minC***, ***spo0A***, ***spo0H***, ***spoIIA(A, B, C)***, ***spoIIE***, ***spoIIQ***, ***spoVB***, ***spoVK***, ***spoIIIG***, ***spoIVCB***, ***dpaA***, *hep1*, ***cwlD***
Listeria monocytogenes	***spoVG***		***minC***, ***spo0A***, ***spo0H***, ***spoIIA(A, B, C)***, ***spoIIE***, ***spoIIGB***, ***spoIIQ***, ***spoVB***, ***spoVK***, ***spoIIIG***, ***spoIVCB***, ***dpaA***, *hep1*, ***cwlD***
Streptococcus agalactiae			***spoIIAC***, ***spoIIGB***, ***spoVB***, ***spoVK***, ***spoIIIG***, ***spoIVCB***
Streptococcus equi			***spo0A***, ***spo0H***, ***spoIIAC***, ***spoIIGB***, ***spoIIQ***, ***spoVB***, ***spoVK***, ***spoIIIG***
Streptococcus mutans			***spo0H***, ***spoIIAC***, ***spoIIGB***, ***spoVB***, ***spoVK***, ***spoIIIG***, ***spoIVCB***
Streptococcus pneumoniae			***spo0A***, ***spo0H***, ***spoIIA(B, C)***, ***spoIIGB***, *spoIVA*, ***spoVB***, ***spoVK***, ***spoIIIG***, ***spoIVCB***
Streptococcus pyogenes	*sspA*		***spo0A***, ***spo0H***, ***spoIIAC***, ***spoIIGB***, ***spoVB***, ***spoVB***, ***spoVK***, ***spoIIIG***, ***spoIVCB***
Streptococcus salvarius subsp. *thermophilus*			***spoVB***
Actinomycetales			
Streptomyces coelicolor			***spo0A***, ***spo0H***, ***spoIIA (A, B, C)***, ***spoIIGB***, *spoIVCB*, *spl*, ***spoIIE***, ***spoIIQ***, ***spoVK***, ***spoIIIG***, ***spoIVCB***, ***splB***

spoIVCB, *dpaA*, *splB* (involved in repair of UV-radiation-induced DNA damage during spore germination), *hep1*, *sspA* (small, acid-soluble spore protein A), and *cwlD* (Stragier 2002). However, except for *spoVID*, *hep1*, *spoIVA*, and *sspA*, sequences similar to these genes were also found in some genera from other major phylogenetic branches distant to the *Firmicutes* and regarded as classical non-spore-forming bacteria, such as Proteobacteria (Table 5). Adding to this data was the surprising fact that some of these "sporulation-specific" genes were absent in tested genome sequences of the asporogenic *Firmicutes* but present in some non-*Firmicutes*: *spoIIIAA* (a stage III sporulation protein), *spoIID* (a hypothetical protein involved in stage II of sporulation), *sleB* (a spore-cortex-lytic enzyme precursor), *dpaB* (dipicolinate synthase, B chain), *spmA* (spore maturation protein A), and *spmB* (spore maturation protein B). Since these bacteria represent divergent branches among prokaryotes, it is proposed that the data suggest the aforementioned sequences should no longer be regarded as "sporulation-specific" genes. Instead, in these bacteria these genes probably function in other processes.

Phylogenetic ramifications. We failed to detect sequences similar to those of many sporulation genes of *B. subtilis* in genomes of sporulating species, e.g., members of the order *Clostridales* tested were shown to lack sequences similar to *sspE*. In additon, all of the *Clostridales* tested, except *Clostridium difficile* and *Desulfitobacterium hafniense*, lacked sequences similar to *dpaA*/*B* of *B. subtilis*. Even members of the order *Bacillales* were found to lack sequences similar to sporulation genes of type species *B. subtilis*: All of the *Bacillales* tested, besides *B. halodurans*, *B. megaterium*, *B. subtilis*, and *G. stearothermophilus*, lacked at least one *ssp* and all, besides *B. anthracis*, *B. cereus*, *B. halodurans*, *G. stearothermophilus* 10, *B. sub-*

Table 5 Gene sequences with similarity to sporulation genes observed in genomes of Gram-type-negative microorganisms that do not form endospores. (++) Scores identify sequence as similar to the *B. subtilis* query sequence. (+) Some similarity over parts of the sequence to the *B. subtilis* query sequence. (+/–) Score is too low to allow for a definitive classification of the sequence relative to to the *B. subtilis* query sequence

Gram-type negative	(++)	(+)	(+/–)
Proteobacteria			
Haemophilus actinomycetemcomitans			*spo0A*, *spoIIAC*, *spoIIQ*, *spoVK*, *spoIIIG*, *cwlD*
Caulobacter crescentus			*spoIIQ*, *spoIVCB*, *dpaA*, *sleB*, *cwlD*
Escherichia coli			*spo0A*, *spo0H*, *spoIIA(B, C)*, *spoIIGB*, *spoIIQ*, *spoIIIAA*, *spoVK*, *spoIIIG, spoIVCB, cwID*
Escherichia coli O157:H7			*spoIVCB*, *cwlD*, *spo0A*, *spo0H*, *spoIIAB*, *spoIIGB, spoIIQ, cwID*
Helicobacter pylori J99			*spo0A*, *spoIIAC*, *spoIIGB*, *spoIIQ*, *spoVK*, *spoIIIG*, *spoIVCB*, *dpaA*, *cwlD*
Myxococcus xanthus			*spo0A*, *spo0H*, *spoIIA(B,C)*, *spoIIGB*, *spoIIIG*, *spoIVCB*, *sleB*, *dpaA*, *cwlD*
Salmonella typhi			*spo0A*, *spo0H*, *spoIIAC*, *spoIIGB*, *spoIIQ*, *spoVK*, *spoIIIG*, *spoIVCB*, *cwlD*
Salmonella typhimurium			*spo0A*, *spo0H*, *spoIIAC*, *spoIIGB, spoIIQ*, *spoVK*, *spoIIIG*, *spoIVCB*, *cwlD*
Spirochaetes			
Borrelia burgdorferi			*spo0A*, *spo0H*, *spoIIAC*, *spoIIGB*, *spoVG*, *spoIIQ*, *spoVK*, *spoIIIG*, *spoIVCB*
Cyanobacteria			
Synechococcus spp.			*minC*, *spoIID*, *spoIIQ*, *spoIIIAA*, *spoVK, spoIVCB, cwID*
Prochlorococcus marinus			*spoIVCB*, *cwlD*, *minC*, *spoIID*, *spoIIQ*, *spoIIIAA*, *spoVK*, *spoIVCB*, *cwlD*
Anabaena spp.			*minC*, *spo0A*, *spoIIA(A, C)*, *spoIID*, *spoIIQ*, *spoIIIAA*, *spoVK*, *sleB*, *cwlD*
Chloroflexi			
Chloroflexus aurantiacus			*minC*, *spoIIAB*, *spoIIE*, *spoIIIAA*, *spoIVCB*
Fusobacteria			
Fusobacterium nucleatum		*spoVG*	*minC*, *spo0A*, *spo0H*, *spoIIAA*, *spoIID*, *spoIIE*, *spoIIQ*, *spoVK*, *spoIVCB*, *splB*, *cwlD*
Deinococcus-Thermus			
Deinococcus radiodurans			*minC*, *spoIIAB*, *spoIID*, *spoIIQ*, *spoVK*, *cwlD*
Bacteroidetes			
Cytophaga hutchinsonii			*spoIIE*, *spmA*, *spmB*
Archaea			
Euryarchaeota			
Archaeoglobus fulgidus	*spoVG*		*spo0A*, *spoVB*, *spoVK*, *dpaB*
Halobacterium spp. NRC-1			*spo0A*, *spoVK*, *dpaB*
Methanocaldococcus jannaschii			*spoVB*, *spoVK*
Pyrococcus furiosus			*spoVB*
Crenarchaeota			
Aeropyrum pernix			*spoVK*

Table 6 Spore-specific genes observed in *Bacillus* and *Clostridium* and related species

Classification	Genes
Present in some *Bacillus* spp. and some *Clostridium* spp., but no similar sequences observed in Gram-type-negative Proteobacteria or Cyanobacteria species	*rapA*, *spoIIGA*, *spoIIM*, *spoIIP*, *spoIIR*, *spoIIIA (B, C, D, E, F, G, H)*, *spoIVA*, *spoIIID*, *spoIVB*, *spoVA (C, D, E, F)*, *sspA*, *gerA (A, B, C)*, *gpr*, *cwlJ*, *yqfC*, *yqfD*, *yabP*, *yabQ*, *spo0B*
Present in some *Bacillus* spp. only and absent from *Clostridium* spp.	*spoIIB*, *gerM*, *cotE*, *spoVID*, *gerPA*, *spoVA (A, B)*, *spoVM*, *dpaA*, *hep1*, *sspE*, *gerD*, *cotD*

tilis, and *B. thuringiensis* subsp. *israelensis*, lacked *dpaA/B* (Table 2). The lack of many identifiable similar sequences in *B. firmus* and *B. licheniformis* (besides those found also in the Gram-type-negative species) was surprising since these two species are closely related to *B. subtilis* (Fig. 1). In contrast, *G. stearothermophilus* is relatively distantly related. These findings lead us to several broader issues related to phylogeny and the presence or absence of genes, as discussed below.

(1) The phylogenetic differences are relatively small (e.g., depicted by short branches), so the placement of the proposed non-spore-forming genera among the sporogenic taxa could be a product of the tree-building algorithm and the selection of included species, which also led to low boot strap values. It is possible that these taxa branched off from other ancestors of the *Bacillus/Clostridium* branches before sporulation evolved in the orders *Bacillales*, *Clostridales*, and *Thermoanaerobacteriales*, respectively (Garrity et al. 2003). However, the different tree-building methods [e.g., Fitch (Fig. 1) and neighbor-joining (data not shown)] yielded similar tree morphology making this less likely. The placement of the non-spore-forming genera such as *Streptococcus*, *Listeria*, and *Enterococcus* in the phylogenetic branch of the *Firmicutes* has recently been confirmed by using DNA-dependent RNA polymerase phylogeny (Morse et al. 2002)

(2) Heavy horizontal gene transfer or loss of the sporulation genes may have occurred among the early *Firmicutes*. However, these scenarios are doubtful as an explanation owing to the fact that genes involved in complex developmental processes are typically widely dispersed over the entire chromosome, as based on the arrangement of sporulation genes on the *B. subtilis* chromosome or the large amount and distribution of genes involved with multicellular developmental processes and regulation in the *Streptomyces coelicolor* genome (Bentley et al. 2002; Stragier 2002; Stragier and Losick 1996);

(3) Our sequence similarity searches were mainly limited to using the *B. subtilis* gene sequences as reference because it is the model organism for studying endospore formation. Even the sequences from *C. acetobutylicum*, which were used partly as a reference set, were originally identified by similarities to those from *B. subtilis*. Despite this limitation, the data imply that either sporulation developed more than once or early sporulation genes underwent major changes during the evolution to the genes of present-day species such as *B. subtilis* and *C. acetobutylicum*.

Selection of genes involved solely in sporulation processes. In seeking to tabulate those genes that have one singular involvement in cells, namely sporulation, we arrived at a significantly modified table (Table 6) than the one presented by Stragier (2002). Due to the much larger database of available genome sequences, our table is both drastically more restricted and smaller. It is unclear at this stage whether the observed division into genes present in *Bacillus* and genes present in *Clostridum* is artificial or due to phylogenetic differences. The presented analysis demonstrates clearly that there is a need to obtain a detailed analysis of sets of "sporulation-specific" genes from aerobic and anaerobic species lacking sequences with unequivocal similarities to the *B. subtilis* genes. It has been speculated that the anaerobic clostridial lineage is more ancient than the one of the *Bacillales*. Thus, the *Bacillus* and *Clostridium* species might have quite differently modified sporulation genes. Currently, there is not enough knowledge about the sporulation processes in clostridia to attempt such analysis, which should also include unusual sporulation systems such as the ones from *Epulopiscium* and *Metabacterium* (Angert et al. 1996; Siunov et al. 1999).

Acknowledgments We thank Mary Ann Moran, for the universal eubacterial primers, and Phil Youngman for help with the *spo0A* primers. We are indebted to Phyllis Pienta from the ATCC for support of the early stages of this research and P. Stragier for providing us with a copy of his manuscript. We thank Ross Overbeek for the access to sequences in ERGO, and Erko Stackebrandt for sharing with us his unpublished results on our assay.

References

Angert ER, Brooks AE, Pace NR (1996) Phylogenetic analysis of *Metabacterium polyspora*: clues to the evolutionary origin of daughter cell production in *Epulopiscium* species, the largest bacteria. J Bacteriol 178:1451–1456

Ash C, Farrow JAE, Wallbanks S, Collins MD (1991) Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small subunit RNA sequences. Lett Appl Microbiol 13:202–206

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147

Brill JA, Wiegel J (1997) Differentiation between sporeforming and asporogenic bacteria using a PCR and Southern hybridization based method. J Microbiol Methods 31:29–36

Collins MD, Lawson PA, Willems A, Cordoba JJ, Fernandez-Garayzabal J, Garcia P, Cai J, Hippe H, Farrow JA (1994) The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int J Syst Bacteriol 44:812–826

Errington J (1993) *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. Microbiol Rev 57:1–33

Engle M, Li Y, Rainey F, DeBlois S, Mai V, Reichert A, Mayer F, Messner P, Wiegel J (1996) *Thermobrachium celere* gen nov, sp nov, a rapidly growing thermophilic, alkalitolerant, and proteolytic obligate anaerobe. Int J Syst Bacteriol 46:1025–1033

Farrow JA, Lawson PA, Hippe H, Gauglitz U, Collins MD (1995) Phylogenetic evidence that the gram-negative nonsporulating bacterium *Tissierella* (*Bacteroides*) *praeacuta* is a member of the *Clostridium* subphylum of the gram-positive bacteria and description of *Tissierella creatinini* sp. nov. Int J Syst Bacteriol 45:436-440

Garrity GM, Holt JG (2001) Taxonomic outline of the *Archaea* and *Bacteria*. In: Boone DR, Castenholz RW (eds) Bergey's manual of systematic bacteriology, vol 1, 2nd edn. Springer, Berlin Heidelberg New York, pp 119–166

Garrity GM, Bell JA, Lilburn TG (2003) Taxonomic outline of the prokaryotes. In: Bergey's manual of systematic bacteriology, 2nd edn release 4.0 DOI 10.1007/bergeysoutline200310

Gerhardt P, Marquis RE (1989) Spore thermoresistance mechanisms. In: Smith I, Slepecky RA, Setlow P (eds) Regulation of prokaryotic development. American Society for Microbiology, Washington, DC, pp 43–63

Gibbons NE, Murray RGE (1978) Proposals concerning the higher taxa of the bacteria. Int J Syst Bacteriol 28:1–6

Gould GW (1984) Mechanisms of resistance and dormancy. In: Hurst A, Gould GW (eds) The bacterial spore, vol 2. Academic, London, pp 173–209

Grossman AD (1995) Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. Ann Rev Gen 29:477–508

Hippe H, Andreesen JR, Gottschalk G (1992) The genus *Clostridium* nonmedical. In: Balows A (ed) The prokaryotes, vol 2. Springer, Berlin Heidelberg New York, pp 1800–1866

Ireton K, Grossman A (1994) DNA-related conditions controlling the initiation of sporulation in *Bacillus subtilis*. Cell Mol Biol Res 40:193–198

Morse R, O'Hanlon K, Collins MD (2002) Phylogenetic, amino acid content and indel analyses of the beta subunit of DNA-dependent RNA polymerase of Gram-positive and Gram-negative bacteria. Int J Syst Evol Microbiol 52:1477–1488

Nakamura K, Hashizume E, Shibata T, Nakamura Y, Mala S, Yamane K (1995) Small cytoplasmic RNA (scRNA) gene from *Clostridium perfringens* can replace the gene for the *Bacillus subtilis* scRNA in both growth and sporulation. Microbiology 141:2965–2975

Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P (2000) Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. Microbiol Mol Biol Rev 64:548–572

Paidhungat M, Ragkousi K, Setlow P (2001) Genetic requirements for induction of germination of spores of *Bacillus subtilis* by Ca(2+)-dipicolinate. J Bacteriol 183:4886–4893

Sauer U, Treuner A, Buchholz M, Santangelo JD, Dürre P (1994) Sporulation and primary sigma factor homologous genes in *Clostridium acetobutylicum*. J Bacteriol 176:6572–6582

Sauer U, Santangelo JD, Treuner A, Buchholz M, Dürre P (1995) Sigma-factor and sporulation genes in *Clostridium*. FEMS Microbiol Rev 17:331–340

Setlow P (1995) Mechanisms for the prevention of damage to DNA in spores of *Bacillus* species. Annu Rev Microbiol 49: 29–54

Setlow P (2001) Resistance of spores of *Bacillus* species to ultraviolet light. Environ Mol Mutagen 38:97–104

Siunov AV, Nikitin DV, Suzina NE, Dmitriev VV, Kuzmin NP, Duda VI (1999) Phylogenetic status of *Anaerobacter polyendosporus*, an anaerobic, polysporogenic bacterium. Int J Syst Bacteriol 49:1119–1124

Slepecky RA, Hemphill HE (1992) The genus *Bacillus* – nonmedical. In: Balows A (ed) The prokaryotes, vol 2. Springer, Berlin Heidelberg New York, pp 1663–1698

Sneath PHA (1984) Endospore-forming Gram-positive rods and cocci. In: Sneath PHA (ed) Bergey's manual of systematic bacteriology, vol 2. Williams & Wilkins, Baltimore, pp 1104–1207

Stragier P (2002) A gene odyssey: exploring the genomes of endospore-forming bacteria. In: Sonenshein AL (ed) *Bacillus subtilis* and its closest relatives: from genes to cells. ASM, Washington DC, pp 519–525

Stragier P, Losick R (1996) Molecular genetics of sporulation in *Bacillus subtilis*. Annu Rev Genet 30:297–341

Wiegel J (1981) Distinction between the Gram reaction and the Gram-type of bacteria. Int J Syst Bacteriol 31:88

Wiegel J (1992) The anaerobic thermophilic bacteria. In: Kristjansson JK (ed) Thermophilic bacteria. CRC, Boca Raton, pp 105–184

Wiegel J, Quandt L (1982) Determination of the Gram-type using the reaction between polymyxin B and lipopolysaccharides of the outer cell wall of whole bacteria. J Gen Microbiol 128: 2261–2270

Wiegel J, Tanner R, Rainey FA (2004) In: Dworkin M (ed) The prokaryotes: an evolving electronic resource for the microbiological community, 3rd edn. Springer, New York, http://www.springeronline.com

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271