From Root to Tips: Sporulation Evolution and Specialization in *Bacillus subtilis* and the Intestinal Pathogen *Clostridioides difficile*

Paula Ramos-Silva,*,[†,1] Mónica Serrano,[2] and Adriano O. Henriques[2]

[1]Instituto Gulbenkian de Ciência, Oeiras, Portugal

[2]Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal

[†]Present address: Marine Biodiversity Group, Naturalis Biodiversity Center, Leiden, The Netherlands

*Corresponding author: E-mail: paula.srs@gmail.com.

Associate editor: Fabia Ursula Battistuzzi

All genomes included in the present study are available at NCBI. Their accession numbers are listed in supplementary table S6, Supplementary Material online. The datamatrix and phylogenetic tree supporting the conclusions of this article were uploaded to TreeBASE and can be accessed by the URL http://purl.org/phylo/treebase/phylows/study/TB2:S24158.

Abstract

Bacteria of the Firmicutes phylum are able to enter a developmental pathway that culminates with the formation of highly resistant, dormant endospores. Endospores allow environmental persistence, dissemination and for pathogens, are also infection vehicles. In both the model *Bacillus subtilis*, an aerobic organism, and in the intestinal pathogen *Clostridioides difficile*, an obligate anaerobe, sporulation mobilizes hundreds of genes. Their expression is coordinated between the forespore and the mother cell, the two cells that participate in the process, and is kept in close register with the course of morphogenesis. The evolutionary mechanisms by which sporulation emerged and evolved in these two species, and more broadly across Firmicutes, remain largely unknown. Here, we trace the origin and evolution of sporulation using the genes known to be involved in the process in *B. subtilis* and *C. difficile*, and estimating their gain-loss dynamics in a comprehensive bacterial macroevolutionary framework. We show that sporulation evolution was driven by two major gene gain events, the first at the base of the Firmicutes and the second at the base of the *B. subtilis* group and within the Peptostreptococcaceae family, which includes *C. difficile*. We also show that early and late sporulation regulons have been coevolving and that sporulation genes entail greater innovation in *B. subtilis* with many Bacilli lineage-restricted genes. In contrast, *C. difficile* more often recruits new sporulation genes by horizontal gene transfer, which reflects both its highly mobile genome, the complexity of the gut microbiota, and an adjustment of sporulation to the gut ecosystem.

Key words: sporulation, bacterial genome evolution, horizontal gene transfer, taxon-specific genes, *Bacillus subtilis*, *Clostridioides difficile*.

Introduction

Bacterial endospores (spores for simplicity) are among the most resilient cells known to us. Because of their resistance to extremes of several physical and chemical conditions (Setlow 2014), spores can spread and persist in the environment for long periods of time and are found in a variety of ecosystems (Nicholson et al. 2000) from deep into Earth's crust (Chivian et al. 2008) and the deep sea (Urios et al. 2004; Dick et al. 2006; Setlow 2014; Fang et al. 2017) to the stratosphere (Shivaji et al. 2006) and, as plant or animal pathogens, symbionts or commensals (Flint et al. 2005; Hong et al. 2009; Hutchison et al. 2014; Browne et al. 2016). A large proportion of the human gut microbiota consists of anaerobic spore formers and several species establish symbiotic or commensal associations thought to be important for health (Chase and Erlandsen 1976; Fujiya et al. 2007; Yano et al. 2015; Browne et al. 2016; Almeida et al. 2019;

Forster et al. 2019). On the other hand, *Clostridium difficile*, recently reclassified as *Clostridioides difficile* (Lawson et al. 2016), a strict anaerobe, is currently a major nosocomial pathogen, that can cause a range of intestinal diseases linked to antibiotic therapy worldwide (Bauer et al. 2011; Lessa et al. 2015; Abt et al. 2016). Spore formation by *C. difficile*, but more generally by intestinal spore formers, allows efficient dispersal of the organisms through the environment and among hosts (Chase and Erlandsen 1976; Lawley et al. 2009; Browne et al. 2016). Spore formers show a variety of morphologies and metabolic styles that include syntrophy, sulfate reduction, and phototrophy and form spores of different sizes, shapes, and numbers per sporangial cell (Abecasis et al. 2013; Hutchison et al. 2014, and references therein).

In spite of this diversity, the basic architectural plan of the spore is largely conserved. Spores have a core compartment

that encloses the DNA delimited by a membrane and surrounded by a thin layer of peptidoglycan that will serve as the cell wall of the new cell that emerges from spore germination. A second layer of modified peptidoglycan essential for dormancy, the cortex, is surrounded by several proteinaceous layers (Henriques and Moran 2007; McKenney et al. 2013). In the model organism *Bacillus subtilis* a multilayered protein coat and a glycosylated crust forms the spore surface (Henriques and Moran 2007; McKenney et al. 2013); in other species, such as *Bacillus cereus*, *Bacillus anthracis*, and *C. difficile*, the spore is further enclosed in a glycosylated exosporium (Stewart 2015). Both the structure and composition of the coat and exosporium vary greatly among species (Stewart 2015).

Most spore formers are rods, and sporulation begins with the formation of a polar division that produces two unequal cells, the forespore (or future spore) and the mother cell, each with an identical copy of the chromosome but expressing different genetic programs. At an intermediate stage, the forespore is engulfed by the mother cell and becomes isolated from the external medium. At later stages, while gene expression in the forespore prepares this cell for dormancy, the mother cell drives the assembly of the cortex, coat, and exosporium, and eventually triggers its own lysis to release the mature spore (Henriques and Moran 2007; Higgins and Dworkin 2012; McKenney et al. 2013). Entry in sporulation is triggered by the activation, through phosphorylation, of the transcriptional regulator Spo0A (Stephenson and Lewis 2005; Edwards and Mcbride 2014). Spo0A is conserved among spore formers and its presence is often used as an indicator of sporulation ability (Galperin 2013; Filippidou et al. 2016). The activity of Spo0A is required for the division that creates the mother cell and the forespore. Following division, gene expression in the forespore and the terminal mother cell is sequentially activated by cell type-specific RNA polymerase sigma factors (Saujet et al. 2014; Al-Hinai et al. 2015; Fimlaid and Shen 2015); prior to engulfment completion, gene expression is governed by σ^F in the forespore and by σ^E in the mother cell; following engulfment completion, σ^F is replaced by σ^G and σ^E is replaced by σ^K. Like Spo0A, sporulation-specific sigma factors are also conserved among spore formers (de Hoon et al. 2010; Galperin et al. 2012; Abecasis et al. 2013). Genetic studies were initially conducted in *B. subtilis*, an aerobic organism, representative of the class Bacilli, still the best characterized spore former; global genetic and transcriptomics studies in this organism have established a comprehensive catalog of the genes whose transcription is directly controlled by σ^F, σ^E, σ^G, and σ^K (Eichenberger et al. 2003, 2004; Feucht et al. 2003; Steil et al. 2005; Wang et al. 2006; Meeske et al. 2016). In turn, *C. difficile* has emerged in recent years as the anaerobic model for the process and is currently the best characterized spore former within the Clostridia (Paredes-Sabja et al. 2014; Al-Hinai et al. 2015; Zhu et al. 2018). As an obligate anaerobe, *C. difficile* relies on its sporulation ability to spread horizontally between hosts via fecal and oral transmission (Lawley et al. 2009; Janoir et al. 2013). The role of spores in infection, colonization, and persistence of the organism has motivated a number of studies on the control of sporulation. Importantly, several studies have been directed at the

identification of the genes under the control of the four cell type-specific sigma factors (Fimlaid et al. 2013; Saujet et al. 2013; Pishdadian et al. 2015), and several genes have been functionally characterized, highlighting conservancy of function of the main regulatory proteins, but also important differences relative to the model *B. subtilis* in the control of sporulation and in the assembly of key structures such as the spore surface layers, specifically the coat/crust and the exosporium (Paredes-Sabja et al. 2014; Saujet et al. 2014; Al-Hinai et al. 2015; Fimlaid and Shen 2015; Stewart 2015; Zhu et al. 2018). Based on the extensive knowledge of sporulation in *B. subtilis*, previous comparative genomics studies have led to the identification of a core (i.e., genes conserved in all spore formers), a signature (present in all spore formers but absent in nonspore-forming lineages), and Bacilli-specific sporulation genes (Onyenwoke et al. 2004; Traag et al. 2010; Galperin et al. 2012; Abecasis et al. 2013). More recently, studies on anaerobic spore formers isolated from the human gut have also led to the identification of a signature for sporulation in the gut (Browne et al. 2016). Overall, these studies have contributed to our knowledge of how conservation patterns are related to gene function. Although restricted to Firmicutes, sporulation is a highly diversified process within the phylum (Flint et al. 2005; Galperin 2013; Yutin and Galperin 2013; Hutchison et al. 2014). Still, the evolutionary mechanisms driving this diversity remain largely unknown. Here, we address this question by comparative genomics of the *B. subtilis* and *C. difficile* genetic machinery for sporulation. Special focus is given to those genes whose expression is dependent on the cell type-specific sigma factors σ^F, σ^E, σ^G, and σ^K and hence, expressed in either the mother cell or the forespore. We show that the σ regulons are highly divergent between the two species, which share in common only a small fraction of genes. We then map the presence/absence of both sets of sporulation genes in a solid and comprehensive bacterial macroevolutionary framework in order to estimate the origin and evolution of the gene families, from an ancient bacterial ancestor to the contemporary *B. subtilis* and *C. difficile*. Our analyses show common evolutionary patterns in the lineages of the two species, with two major gene gain events having expanded the cell type-specific regulons. Moreover, per each gain event, the cell type-specific regulons appear to grow in the same size-proportions, suggesting that sporulation stages have been coevolving along the bacterial phylogeny, whereas multiple independent gene losses emerge in some branches of the tree for both Bacilli and Clostridia. We also show that in the evolution of *B. subtilis*, gene gains more often involve the emergence of taxonomically restricted genes (TRGs) than in *C. difficile*, where more sporulation genes have homologs in other bacteria and were likely acquired by horizontal gene transfers (HGTs). Still, the two mechanisms are expected to occur to a greater or lesser extent in both lineages and are fully demonstrated here for the genes coding an exosporium protein of *C. difficile*, CdeA (Díaz-González et al. 2015) and for SpoIIIAF, a component of the sporulation-essential mother cell-to-forespore secretion system, the SpoIIIA-SpoIIQ complex, conserved across spore formers (Meisner et al. 2008; Fimlaid et al. 2015; Serrano et al. 2016; Morlot and Rodrigues 2018).

Lastly, our analysis supports the hypothesis that sporulation emerged at the base of Firmicutes, some 2.5 billion years ago (Ga), with a first major gene gain event possibly in response to the rise in oxygen levels. This major evolutionary step was followed by a series of intermediate gains that culminated with a second major gene gain event corresponding to the specialization of this developmental program and its adjustment to particular ecosystems.

Results

A Small Core and Large Diversity of Sporulation Genes between *B. subtilis* and *C. difficile*

B. subtilis is by far the best characterized spore former in terms of sporulation genes, their regulation, and the molecular mechanisms underlying spore morphogenesis (Fawcett et al. 2000; Eichenberger et al. 2003, 2004; Wang et al. 2006; Mäder et al. 2012). *C. difficile* is the best characterized within the class Clostridia (Fimlaid et al. 2013; Saujet et al. 2013; Paredes-Sabja et al. 2014; Pishdadian et al. 2015). For this study, we have manually compiled 726 sporulation genes for *B. subtilis* strain 168 (supplementary table S1, Supplementary Material online, BSU) and 307 sporulation genes from the *C. difficile* strain 630 (supplementary table S1, Supplementary Material online, CD). Each gene is under the direct control of one or more sporulation-specific sigma factors (fig. 1A) and thus part of at least one cell type-specific regulon. The regulons have different sizes (table 1): the σ^E regulon is the largest followed by the σ^G- and σ^K-controlled regulons (of similar sizes) with the σ^F regulon being the smallest.

Homology mapping between *B. subtilis* and *C. difficile* genomes was carried out using a bidirectional BlastP approach, followed by selection and analysis of ortholog groups containing sporulation genes. In figure 2, we show that only a small fraction of sporulation genes have candidate orthologs in the two species. Some of these genes are present in the same regulon (fig. 2, orange; see also supplementary table S2, Supplementary Material online): 2 genes in the σ^F regulon, 28 in the σ^E regulon, 11 controlled by σ^G, and 5 controlled by σ^K, whereas <1% belongs to different regulons (fig. 2, yellow; supplementary table S3, Supplementary Material online). A larger proportion of genes correspond to those having candidate orthologs in both genomes but only reported to be involved in sporulation in one species (fig. 2, light purple; supplementary table S4, Supplementary Material online). These proportions vary depending on the regulon, from 11% to 19% in *B. subtilis* and from 12% to 35% in *C. difficile*. Finally, the majority of sporulation genes have no candidate orthologs in the two species (fig. 2, dark purple; supplementary table S5, Supplementary Material online) (>70% for *B. subtilis* and >50% for *C. difficile*), highlighting the great genetic diversity of the sporulation process. In order to elucidate the emergence of this diversity in a larger context of bacterial evolution, we extended the homology analysis of the known cell type-specific sporulation genes to 258 bacterial genomes (supplementary fig. S1 and supplementary table S6, Supplementary Material online).

Distribution of Sporulation Genes across Bacteria Reveals a Small Core of Conserved Genes and Lineage Specificity

The presence/absence of sporulation genes was mapped across 258 bacterial genomes (supplementary fig. S1 and supplementary table S6, Supplementary Material online) using an optimized homology mapping approach (supplementary fig. S3, Supplementary Material online) (one vs. all and all vs. one BlastP) applied twice: First, using *B. subtilis* 168 as the reference, second using *C. difficile* 630 (supplementary fig. S2, Supplementary Material online). The results were represented in supermatrices of gene presence (1 or more, when candidate orthologs are present) and absence (0, when no candidate orthologs are present) (supplementary tables S7 and S8, Supplementary Material online), and used to calculate the distribution of sporulation genes across bacterial genomes (fig. 3). Both distributions show that only a small proportion of sporulation genes are conserved in the bacterial genomes (fig. 3A). A unimodal distribution is observed for sporulation genes from *B. subtilis* ($D = 0.024757$, P value $= 0.3179$) with a peak around 20%, whereas for *C. difficile* the distribution is bimodal ($D = 0.05604$, P value $= 7.262e-06$) with high densities around 20% and a secondary high-density peak at 42% (fig. 3B). These results suggest that *B. subtilis* has more TRGs for sporulation than *C. difficile*, whose genes appear more frequently in the bacterial genomic data set. To better elucidate the causes underlying these differences, we conducted an evolutionary analysis of sporulation gene families by posterior probabilities in a phylogenetic birth–death model.

A Bacterial Evolutionary Backbone

Prior to the evolutionary analysis of sporulation gene families, we started by building a solid and comprehensive bacterial phylogeny based on 70 marker genes conserved across the genomic data set. Overall, the resulting tree topology (fig. 4 and supplementary fig. S4, Supplementary Material online) is similar to recent macroevolutionary bacterial trees (Hug et al. 2016; Marin et al. 2017). The phyla Dictyoglomi, Thermotogae, Aquificae, Synergistetes, and Deinococcus-Thermus form an early diverging clade. The second earliest split in the tree corresponds to the divergence of the taxon coined as hydrobacteria (which includes Proteobacteria, Bacteroidetes, Spirochaetes, and Fusobacteria) and terrabacteria (which includes Actinobacteria, Cyanobacteria, and the Firmicutes), estimated to have occurred between 2.83 and 3.54 Ga (Battistuzzi et al. 2004; Battistuzzi and Hedges 2009). Within terrabacteria, Firmicutes (monoderms, low GC, mostly gram-positive) is an early diverging phylum currently divided into seven classes: Bacilli, Clostridia, Negativicutes, Erysipelotrichia, Limnochordia, Thermolitobacteria, and Tissierellia (source: NCBI taxonomy). In our study, we included genomes from four of these classes. The major split is between Bacilli ($n = 77$) and Clostridia ($n = 95$) but other groups previously raised to the class level emerge within Clostridia, namely, the Negativicutes ($n = 11$) coding for diderm bacteria with a gram negative-type cell envelope and Erysipelotrichia ($n = 1$) (supplementary fig. S1,
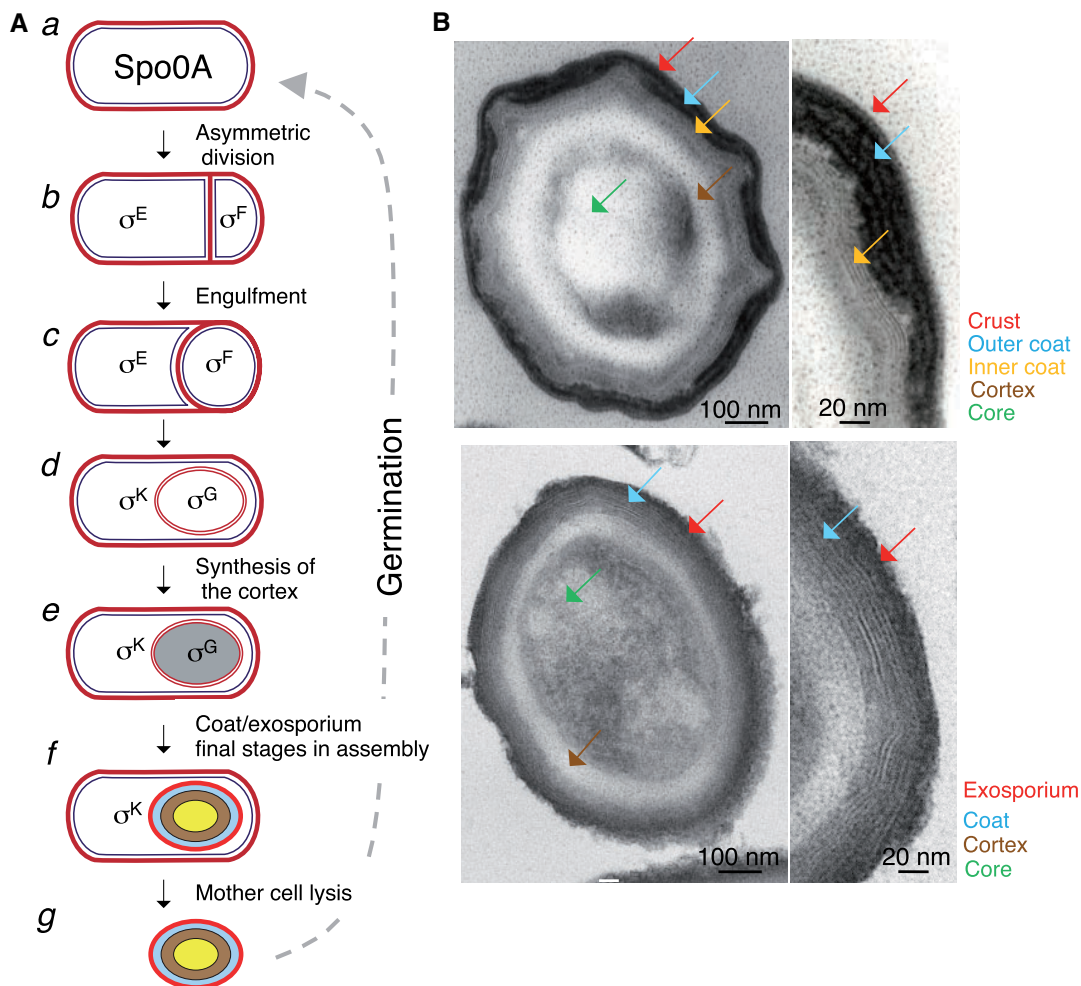
FIG. 1. Sporulation stages and spore ultrastructure. (*A*) Depicts the main stages of sporulation as described for the model organisms *B. subtilis*. At the onset of the process, the rod-shaped cells (*a*) divide asymmetrically to produce a larger mother cell and a smaller forespore (the future spore) (*b*). Asymmetric division involves PG synthesis within the septum. The mother cell then starts to engulf the forespore (*c*), eventually releasing it as a free protoplast inside its cytoplasm (*d*). Following engulfment completion, the forespore is no longer in contact with the external medium and is separated from the mother cell by a system of two membranes that derive from the asymmetric division septum, the inner and outer forespore membranes. Following engulfment completion, the forespore becomes visible as a phase dark body inside the mother cell (*e*). Synthesis of the primordial germ cell wall takes place from the forespore, whereas synthesis of the spore cortex PG layer is a function of the mother cell. Development of full spore refractility coincides with the formation of the cortex. The final stages in the assembly of the spore coat and crust are also represented (*f*). Finally, the spore is released into the environment through autolysis of the mother cell (*g*). Spores will resume vegetative growth through the process of germination. Spo0A is represented in predivisional cells; the cell in which the cell type-specific sigma factors are active in relation to the stages of sporulation is also represented. (*B*) Transmission electron microscopy (TEM) image of a thin cross section of a *B. subtilis* (top) and *C. difficile* (bottom) spore. The main spore structures are labeled in the diagram. Note that the crust layer of *B. subtilis* spores is not visible in the microscopy image, but its position, at the edge of the outer coat, is indicated in the diagram. The panels on the right show a magnification of the spore surface layers. The diagram identifies the main structures or compartments normally seen by TEM.

Table 1. Number of sporulation genes collected in the literature per regulon and described in supplementary table S1, Supplementary Material online.

Sporulation Stage	Expression	Regulon	#Genes *B. subtilis* 168	#Genes *C. dificille* 630
Initiation	Mother cell	Spo0A	65	—
Engulfment	Forespore	σ^F	81	40
Engulfment	Mother cell	σ^E	331	203
Maturation	Forespore	σ^G	156	59
Maturation	Mother cell	σ^K	147	56

Supplementary Material online). Two spore formers (supplementary fig. S1, Supplementary Material online, and fig. 4), *Symbiobacterium thermophilum* and *Thermaerobacter marianensis* (gram-positive, GC > 60%) appear as a distinct lineage from the Firmicutes and have been suggested to be an intermediate phylum between the
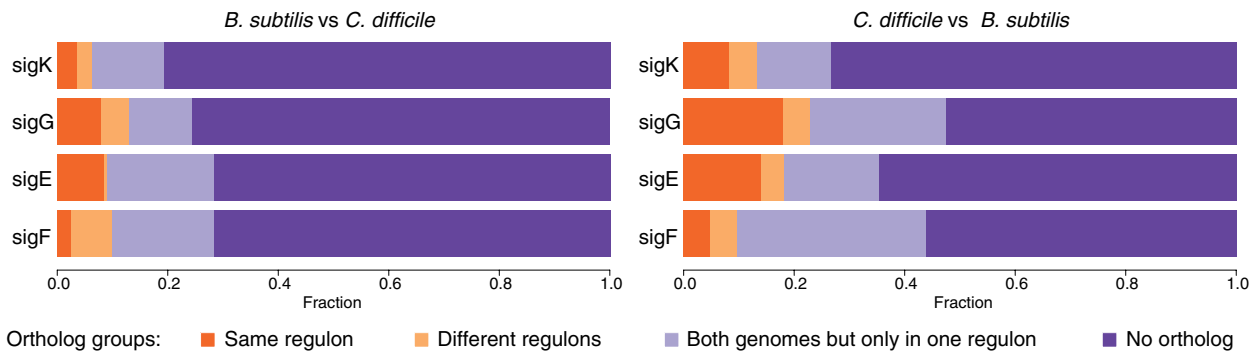
FIG. 2. Clustering of candidate orthologs from a pairwise genome comparison between *B. subtilis* 168 and *C. difficile* 630. Only clusters including genes controlled by sporulation-specific sigma factors were selected for further analyses. The fraction of orthologous genes sharing the same sporulation regulon is marked in orange. The fraction of orthologs in different regulons is in yellow. The fraction of genes without orthologs in the regulons but with orthologs in the genome is in light purple. The fraction of genes without orthologs is in dark purple.
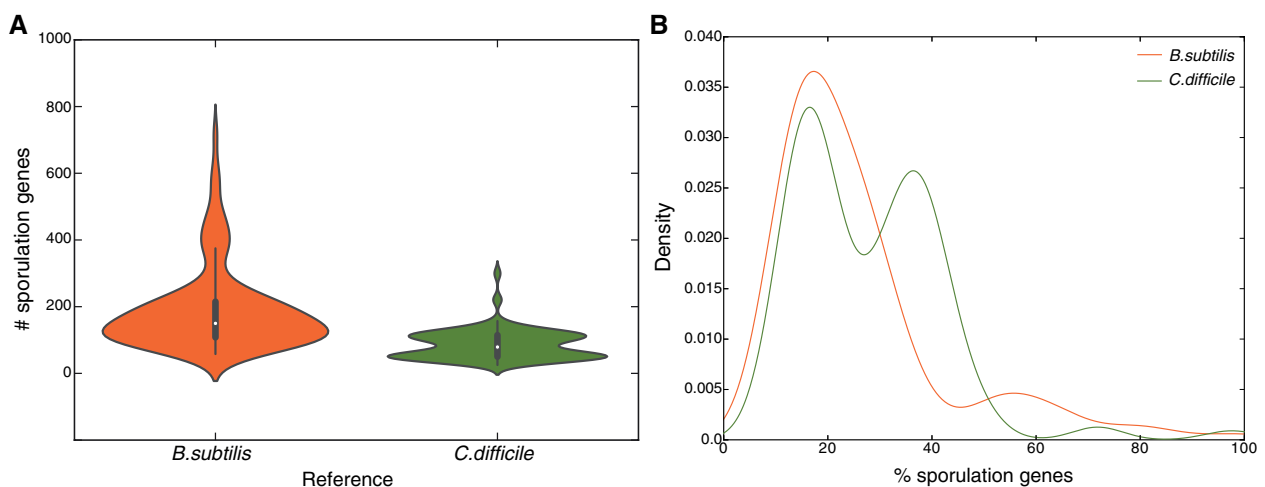

FIG. 3. Distribution of sporulation genes from *B. subtilis* and *C. difficile* across the 258 bacterial genomes sampled from NCBI. Distributions are represented in (*A*) absolute number of genes and (*B*) relative percentages. Distributions were tested for unimodality using the Hartigan's Dip test in R ($D = 0.024757$, P value $= 0.3179$ for *B. subtilis*; $D = 0.05604$, P value $= 7.262e\text{-}06$ for *C. difficile*).

high GC (Actinobacteria) and the low GC (Firmicutes) groups (Ueda et al. 2004; Han et al. 2010). Here, however, they remain as Firmicutes.

A Major Gene Gain Event at the Basis of the Firmicutes

Supermatrices of presence/absence of sporulation genes (0, 1, >1) sized 726 × 258 and 307 × 258 from reference strains *B. subtilis* 168 and *C. difficile* 630, respectively, were used in an evolutionary analysis of sporulation genes across the bacterial phylogeny (supplementary tables S7 and S8, Supplementary Material online, and fig. 4). We first focused on the gene gain events, occurred from the root to the terminal branches of *B. subtilis* (fig. 4A) and *C. difficile* (fig. 4B). Our approach estimates that throughout evolution of both species, sporulation genes were acquired in two major evolutionary steps. The first major gene gain event occurred at the base of the Firmicutes, comprising at least 166 genes (fig. 4, branch 2; see also supplementary tables S9 and S10, Supplementary Material online) including those coding for Spo0A and the four cell type-specific sigma factors. Accordingly, these genes are found in most of the species and taxonomic groups of

spore formers (supplementary fig. S5, Supplementary Material online).

The major gene gain event in branch 2 comprised other core sporulation genes, conserved among members of the phylum (supplementary tables S9 and S10, Supplementary Material online). For *B. subtilis*, gene gains in branch 2 included many of the sporulation signature genes defined by Abecasis et al. (2013) (supplementary table S9, Supplementary Material online, genes highlighted in yellow) and included those involved in the control of σ factor activation or activity, and genes coding for ancillary transcription factors (fig. 5). They also included some genes involved in the engulfment process, synthesis of the spore cortex, formation of the coat/crust as well as synthesis and transport of pyridine-2,6-dicarboxylic acid (or dipicolinic acid, DPA) (supplementary fig. S6, Supplementary Material online). The accumulation of chelates of DPA with divalent cations in spores is important for dormancy and resistance to various DNA-damaging agents (Setlow 2014). The *spoIIE* and *spoIIAB* genes, expressed in the predivisional cell under the control of Spo0A, are involved in the cell type-specific activation of σ^F,
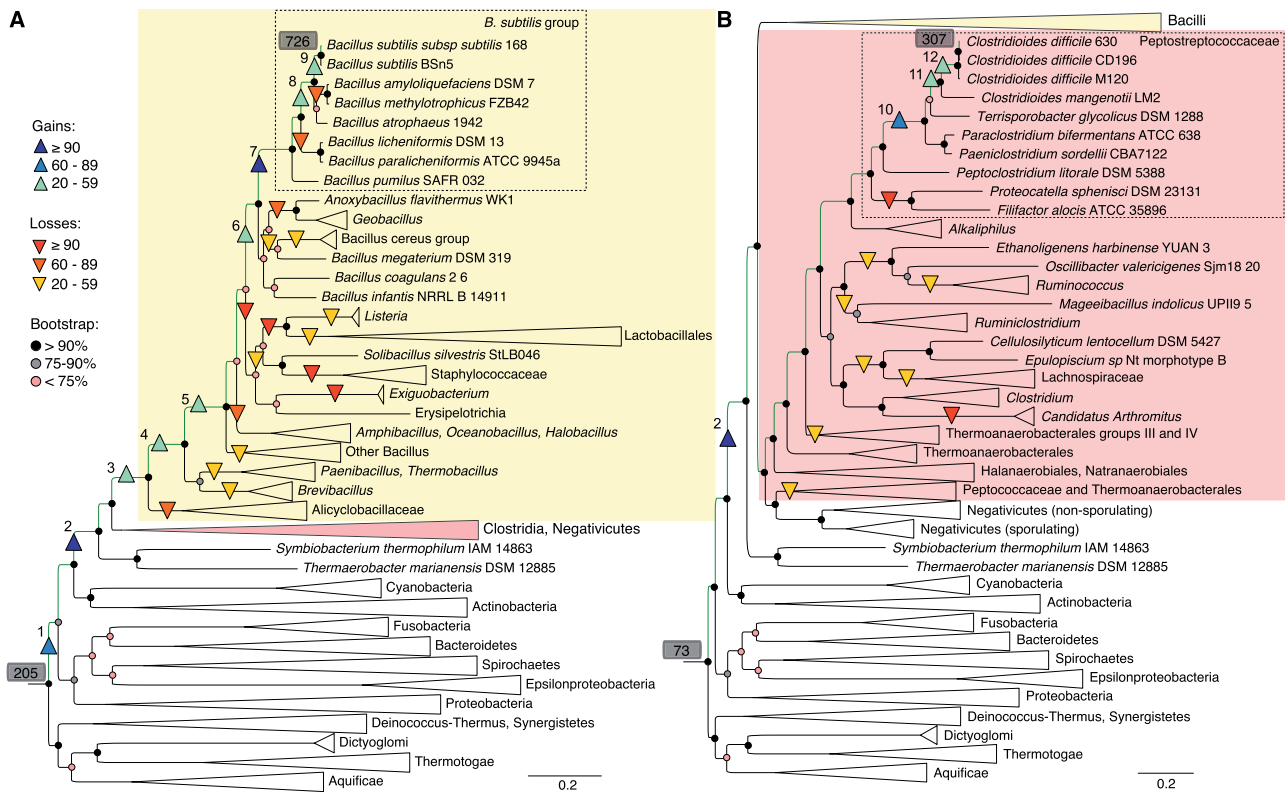
FIG. 4. Sporulation gene gain and loss events across the bacterial phylogeny for (A) Bacilli and (B) Clostridia. The total number of sporulation genes present at the root and at the referential tips is highlighted in gray. Gain events are numbered from 1 to 12. Multigene maximum likelihood (RAxML) tree inferred from an alignment of 70 orthologs and corresponding bootstraps measures in %. The PROTGAMMALG evolutionary model was used to infer the tree with branch support estimated with 100 bootstrap replicates.

whereas *spoIIR*, *spoIIT*, and *spoIIGA* are required for the activation of σ^E (Konovalova et al. 2014; Bradshaw and Losick 2015; Meeske et al. 2016; Narula et al. 2016, and references therein). The gain of *csfB* in branch 2 is interesting. This gene, lost from *C. difficile*, codes for a dual-specificity antisigma factor. CsfB is first produced under the control of σ^F (Decatur and Losick 1996) and helps preventing premature activity of σ^G, prior to engulfment completion (Chary et al. 2007; Karmazyn-Campelli et al. 2008; Rhayat et al. 2009; Serrano et al. 2011; Flanagan et al. 2016). Because of the autoregulatory nature of σ^G, several mechanisms concur with inhibition by CsfB to prevent its premature activity (Mearls et al. 2018). Later in sporulation, CsfB is produced in the mother cell, under the control of σ^K and helps preventing protracted activity of σ^E, to which it also binds an inhibits (Serrano et al. 2015; Martinez-Lumbreras et al. 2018). Thus, CsfB enforces the switch from early to late gene expression in both the forespore and the mother cell. The *spoIIIA* operon is conserved in all spore formers (Galperin et al. 2012; Morlot and Rodrigues 2018). In both *B. subtilis* and *C. difficile*, *spoIIIA* is expressed in the mother cell under the control of σ^E and codes for eight proteins, SpoIIIAA to H (Crawshaw et al. 2014; Morlot and Rodrigues 2018). Together with additional proteins, including the forespore-specific SpoIIQ (gained in branch 3) and the mother cell-specific GerM (gained in branch 2 but absent from the Clostridia), the SpoIIIAA-H proteins are required for the assembly of a transenvelope

secretion complex, also known as a channel or "feeding tube", whose main role is to maintain metabolic potential in the forespore including σ^G-dependent, transcriptional activity (Meisner et al. 2008; Camp and Losick 2009; Doan et al. 2009; Rodrigues, Henry, et al. 2016; Rodrigues, Ramirez-Guadiana, et al. 2016; Trouve et al. 2018). SpoIIIAH and SpoIIQ directly interact in the interspace and form a direct connection between the two cells (Meisner et al. 2008). The later appearance of SpoIIQ, in branch 3, suggests that initially the secretion complex did not make a direct synapse with the forespore cytoplasm. It is interesting to note two other functions of SpoIIQ. First, the SpoIIIAH-SpoIIQ zipper-like interaction also makes a contribution to engulfment (Blaylock et al. 2004). Second, a conserved membrane-embedded amino acid in SpoIIQ blocks transcription of *csfB* by σ^G maximizing the activity of this sigma factor following engulfment completion (Flanagan et al. 2016). The appearance of the σ^G-controlled *spoIVB* gene, coding for a signaling protease, is concomitant with the appearance of *bofA*, *spoIVFA*, and *spoIVFB* in the σ^K regulon (fig. 5 and supplementary table S9, Supplementary Material online). σ^K is produced as an inactive proprotein that associates with the forespore outer membrane; the pro-σ^K-processing protease, SpoIVFB, is maintained in an inactive complex in the forespore outer membrane by BofA and SpoIVFA (Cutting et al. 1990; Ricca et al. 1992; Resnekov and Losick 1998; Rudner and Losick 2002; Zhou and Kroos 2004). Release of SpoIVFB from inhibition requires production

of SpoIVB in the forespore (Cutting et al. 1991; Dong and Cutting 2003; Konovalova et al. 2014; Halder et al. 2017). This ensures that the activation of σ^K, which drives the final stages in the assembly of the spore protective structures (cortex and coat/crust) (fig. 1B) as well as mother cell lysis, is coupled to engulfment completion which marks the onset of σ^G activity in the forespore. The *bofA*, *spoIVFA*, and *spoIVFB* genes are absent from the *C. difficile* genome in which σ^K lacks a prosequence (Pereira et al. 2013), as well as from other Clostridia species (Galperin et al. 2012) (fig. 5). The gene for a second forespore-produced protease that contributes to this signaling cascade in *B. subtilis*, *ctpB* (Pan et al. 2003; Campo and Rudner 2007; Mastny et al. 2013), is present from the origin (supplementary table S9, Supplementary Material online). Also gained in branch 2 are SpoIIID, a mother cell-specific transcription factor that works with σ^E, and SpoVT, forespore-specific, which works with σ^G; SpoIIID and SpoVT both act as either activators or repressors, and define, with their cognate main transcriptional regulators, a series of type I coherent and incoherent feed-forward loops (FFLs) that result in pulses or delayed and protracted gene expression that fine tunes the genetic programs (Eichenberger et al. 2004; Wang et al. 2006; de Hoon et al. 2010).

In *B. subtilis*, engulfment requires the concerted action of a three-protein machine, formed by the σ^E-controlled SpoIID, SpoIIM, and SpoIIP proteins (Abanes-De Mello et al. 2002; Broder and Pogliano 2006; Chastanet and Losick 2007; Morlot et al. 2010; Leidal et al. 2012; Ojkic et al. 2016). These genes are conserved in *C. difficile*, although *spoIIM* is largely dispensable for engulfment (Dembek et al. 2018; Ribis et al. 2018). While *spoIID*, which codes for a lytic transglycosylase (Abanes-De Mello et al. 2002; Chastanet and Losick 2007; Gutierrez et al. 2010; Morlot et al. 2010) is present from the origin, and was later recruited to the σ^E regulon, the *spoIIM* and *spoIIP* genes appeared in branch 2 (fig. 5 and supplementary table S9, Supplementary Material online). Membrane fission at the end of the engulfment process requires FisB (Doan et al. 2013), whose coding gene is also found in branch 2 (fig. 5 and supplementary table S9, Supplementary Material online).The *spoVA* operon (σ^G regulon) and *spoVV* (*ylbJ*, σ^E regulon), involved in the uptake of DPA into the forespore (Tovar-Rojo et al. 2002; Meeske et al. 2016; Ramirez-Guadiana et al. 2017), also appeared in branch 2; it is interesting to note that gain of the *spoVA* operon is concomitant with the gain of the genes coding for the DPA synthetase (*spoVFA* and *spoVFB*) into the σ^K regulon of *B. subtilis* (supplementary fig. S6 and supplementary table S9, Supplementary Material online; see also the Discussion). *spoIVA* codes for a morphogenetic ATPase required for the formation of a coat basal layer and the targeting of the coat proteins to the surface of the developing spore (Roels et al. 1992; Stevens et al. 1992; Driks et al. 1994; Ramamurthi and Losick 2008; Castaing et al. 2013). Of note, genes required for synthesis of the spore cortex, such as *murB*, *murG*, *murF*, *spoVE*, *spoVD*, and *spoVR*, are present from the origin as they have general functions in the synthesis of peptidoglycan and have a broad taxonomic distribution; they were co-opted

to the mother cell-specific line of gene expression (supplementary fig. S6 and supplementary table S9, Supplementary Material online) (Popham and Bernhards 2015). *spoVR*, for example, (ortholog group no. COG2719) (Beall and Moran 1994; Huerta-Cepas, Szklarczyk, et al. 2016), is present in Proteobacteria and Archaea. Other genes required for cortex synthesis appeared in branch 2, such as *spoVB*, and *dacB* or *pdaA*, the later required for the production of muramic δ-lactam, a sporulation-specific modification of the cortex peptidoglycan, in both organisms (Fukushima et al. 2002; Saujet et al. 2013; Popham and Bernhards 2015; Diaz et al. 2018).

In general, the presence of the σ^F activation pathway, the σ^F to σ^E, σ^E to σ^G, and σ^G to σ^K cell–cell signaling/activation pathways, and the presence of CsfB and of the ancillary factors SpoIIID and SpoVT show that key mechanisms enforcing the careful orchestration of the developmental program in the two cells and its coupling to the course of morphogenesis were established early. Also in general, most of the sporulation genes at branch 2, as expected, are also found in *C. difficile* (supplementary tables S9 and S10, Supplementary Material online). Together, this analysis indicates that the main morphological plan of sporulation and the required effector proteins were established early, at the basis of the Firmicutes.

A Second Major Gene Gain Event at Terminal Branches

A second major gene gain event occurred at terminal branches, at the base of the *B. subtilis* group (Berkeley et al. 2002) (fig. 4A; +90 genes, branch 7) as well as within members of the family Peptostreptococcaceae (Clostridium cluster XI) (fig. 4B; +73 genes, branch 10), at the branch that predates the split between the genera *Clostridioides*, *Paeniclostridium*, *Paraclostridium*, and *Terrisporobacter* (Galperin et al. 2016).

For *B. subtilis*, genes gained in branch 7 include several σ^K-controlled determinants of coat/crust assembly such as *safA*, and *cotO*, *cotW*, *cotV*, *cotM*, and *cgeABCD* (supplementary table S9, Supplementary Material online). In *B. subtilis*, the coat is differentiated into a lamellar inner coat a striated, electrondense outer coat and a crust (fig. 1B) (Henriques and Moran 2007; McKenney et al. 2013). While *cotE*, acquired in branch 3 (i.e., Bacilli specific, supplementary table S9 and supplementary fig. S6, Supplementary Material online) is required for the assembly of the outer coat (Zheng et al. 1988; Driks et al. 1994; de Francesco et al. 2012; Nunes et al. 2018), *safA* drives assembly of the inner coat (Takamatsu et al. 1999; Ozin et al. 2000; Ozin et al. 2001) and has an additional role in linking the coat to the underlying cortex (Fernandes et al. 2018; Pereira et al. 2018). Neither *cotE* nor *safA* are conserved in *C. dificile*. Several genes involved in crust assembly appeared in branch 7. *cotV* and *cotW* form a transcriptional unit in a cluster of genes (*cotVWXYZ*, with *cotO* downstream of *cotZ* and convergently oriented) required for crust assembly (McKenney et al. 2010; Imamura et al. 2011; Shuster et al. 2019). At the basis of the hierarchy for crust assembly is CotZ, required for the localization of most crust proteins including the products of the *cgeABCD* operon, which are glycosylation enzymes (Shuster et al. 2019). CotO, in turn, has a role in
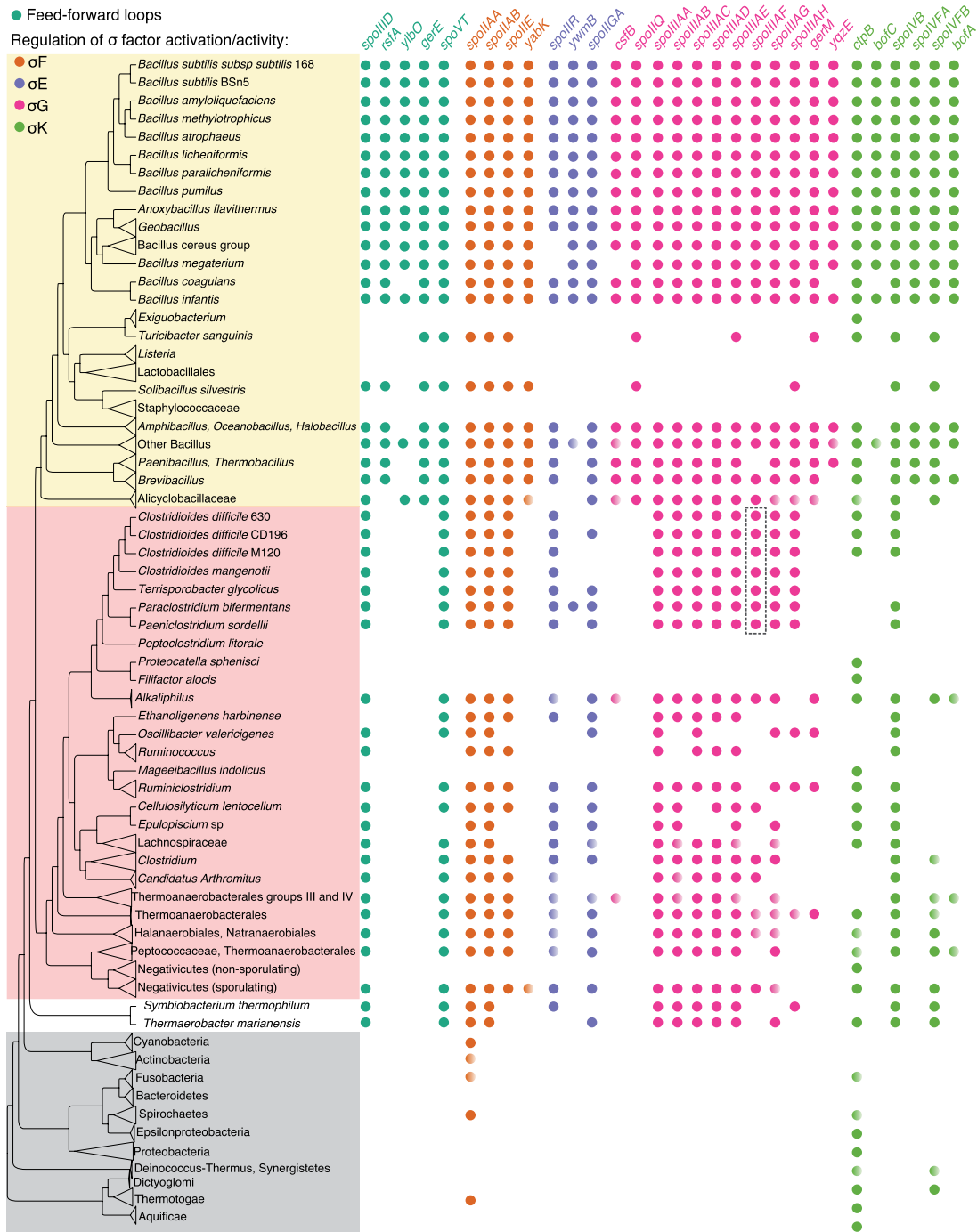
FIG. 5. Phylogenetic profile showing the conservation of gene presence in the following functional categories: FFLs (dark green) and regulation of activation and activity of σ^F (orange), σ^E (purple), σ^G (pink), and σ^K (light green). Conservation of gene presence is based on the homology mapping approach described in supplementary figure S2, Supplementary Material online, averaged for collapsed tips that included more than 1 species (e.g., *Ruminiclostridium*, Actinobacteria) and visualized with EvolView (Zhang et al. 2012). Full circles indicate gene presence above 50%. Half circles indicate gene presence at 50%. No circle indicates gene presence lower than 50% or total absence. On the top, gene reference sequences, names and functional categories were selected based on the *B. subtilis* sporulation network. Members of the ortholog cluster *spoIIIAF* are inside the rectangle.

encasement of the spore by the crust (Shuster et al. 2019). Since at least *cotX* appeared in branch 5 (an intermediate Bacilli lineage, excluding Paenibacillaceae and Alicyclobacillaceae; supplementary table S9, Supplementary Material online), the plan for crust assembly may have been laid down at this stage, and the addition of the *cotVW* and

cgeABCD modules represent a specialization of this structure within the *B. subtilis* group.

Several genes were gained at the basis of the Peptostreptococcaceae, in branch 10 (see supplementary table S10, Supplementary Material online). Among them, the genes coding for the spore germination proteins SleC, CspBA,

and CspC, these genes coding for these proteins are part of the σ^K regulon, and thus expressed late in the mother cell. *cspC* is also found in *C. perfringens* and in *C. lentocellum*, but we found no evidence that the gene was gained at the basis of the *Clostridium* clade; the most likely scenario is thus that *cspC* represents a novelty of the Peptostreptococcaceae. CspBA, on the contrary, has homologs in *Paraclostridium bifermentans*, *Paeniclostridium sordellii*, and *Terrisporobacter glycolicus* and interestingly with gut organisms belonging to the segmented filamentous bacteria, *Candidatus arthromites* sp.; the best BlastP hit outside the Peptostreptococcaceae group is with a *Candidatus* isolate, suggesting HGT from/to this or from/to highly related organisms. Importantly, the appearance of *cspC* and *cspBA* in branch 10 seems to represent a clear adaptation to the gut environment as these genes are essential for the germination of *C. difficile* spores in response to bile salts (Bhattacharjee et al. 2016; Setlow et al. 2017; Zhu et al. 2018). CspBA is subject to interdomain processing by a conserved protease, *yabG*, which is present from branch 2 and is under the control of σ^K (supplementary table S9, Supplementary Material online) (Kevorkian and Shen 2017; Shrestha et al. 2019). CspB, is a subtilisin-like serine protease involved in the activation, together with CspC, of pro-SleC (Adams et al. 2013; Francis et al. 2013; Kevorkian et al. 2016; Kevorkian and Shen 2017; Shrestha et al. 2019). SleC, in turn, is a hydrolase required for degradation of the cortex peptidoglycan during spore germination (Burns et al. 2010; Gutelius et al. 2014); production of SleC from Pro-SleC is also under the control of YabG (Kevorkian et al. 2016).

Intermediate Gains

Other intermediate gains, between branches 2 and 7/10, occurred in the evolutionary path of *B. subtilis* and *C. difficile*. For example, for *B. subtilis*, several important genes both directly involved in the morphogenesis of the spore structures and in the regulation of the transcriptional cascade, appeared in branch 3. The *cotE* gene emerged at the basis of Bacilli; this suggests that a striated electrondense outer coat-like layer evolved before the inner coat, and that clearly defined inner coat layers is a specialization of the *B. subtilis* group (supplementary table S9 and supplementary fig. S6, Supplementary Material online; see above). Also, *spoVM* appeared in branch 3; it codes for a small peptide that recognizes positive curvature and helps in the localization of SpoIVA to the spore surface, and later, together with SpoVID, promotes encasement of the spore by the proteins that form all the layers of the coat/crust, including SpoIVA (Levin et al. 1993; Ramamurthi et al. 2006; Ramamurthi and Losick 2008; Wang et al. 2009; McKenney and Eichenberger 2012). The localization of SpoIVA did not initially rely on SpoVM, a key function of which seems to be in the coordination of encasement. SpoVID interacts with both SpoIVA, CotE, and SafA (Costa et al. 2006; Mullerova et al. 2009; de Francesco et al. 2012; Qiao et al. 2012, 2013) and acts as a hub that connects the inner and outer coat modules (Nunes et al. 2018). Thus, branch 3 represents a critical point in the evolution of the spore surface layers. Moreover, two other genes coding for transcription factors that modulate the expression of genes in

the σ^E and σ^K regulons, many of which are involved in coat/exosporium assembly, also appear in branch 3. These are *ylbO*, also known as *gerR*, and *gerE*. These genes code for ancillary transcription factors that establish additional incoherent (GerR) and both coherent and incoherent (GerE) type I FFLs that with SpoIIID (above) further subdividing the σ^E and σ^K regulons (Eichenberger et al. 2004; Wang et al. 2006; de Hoon et al. 2010). Also of note, is the gain of *bofC*, coding for a negative regulator of the σ^G to σ^K cell–cell signaling pathway (Gomez and Cutting 1997; Wakeley et al. 2000; Wang et al. 2006) in branch 5 (fig. 5 and supplementary table S9, Supplementary Material online). Finally, the *yabK* gene (also known as *fin*) appeared in branch 3; *yabK* is produced under the control of σ^F and binds to the β′-subunit of RNA polymerase preventing docking of σ^F (Camp et al. 2011; Wang Erickson et al. 2017). YabK thus establishes a negative feedback loop that modulates the activity of σ^F (Wang Erickson et al. 2017). It appears that although the main mechanisms known to control the activity of the cell type-specific sigma factors were established at the basis of the Firmicutes (branch 2; see above), additional control mechanisms continued to be introduced.

Of note, SpoVM is found in branch 3 although in *C. difficile* it is not essential for spore formation (Ribis et al. 2017) whereas SipL (Entrez GI:126701193), which performs a function similar to that of SpoVID in *B. subtilis* (Putnam et al. 2013; Touchette et al. 2019), emerged later, at an intermediate branch, before the divergence of the Halanaerobiales and Natanaerobiales groups (fig. 4B). Also, CotL (Entrez GI:126698652), important for the assembly of the coat/exosporium (Alves Feliciano et al. 2019) appeared in branch 11, at base of the Clostridioides. Similarly, several of the genes specifically involved in the assembly of the exosporium layer have appeared in terminal branches from 10 to 12 (supplementary table S10, Supplementary Material online, see below).

Coevolution of Early and Late Sporulation Stages

The cell type-specific sporulation genes belong to at least one regulon and thus it is possible to explore the effect of gene gain events in the size of the regulons as a function of the evolutionary steps for *B. subtilis* and *C. difficile*, highlighted in green in figure 4 (branches 1–12). The general pattern for *B. subtilis* (fig. 6A) and *C. difficile* (fig. 6B) is that regulons grow in similar proportions. Larger increases in the size of the regulons corresponded to the major gene gain event at the basis of the Firmicutes (fig. 4, branch 2). In branch 2, we observe a slightly larger increase of the σ^G regulon controlling late sporulation stages in the forespore,

Extensive Gene Loss in Spore Forming and Asporogenous Lineages

Besides gene gains, significant losses have also occurred in certain Bacilli and Clostridia lineages. In Bacilli, extensive gene loss (fig. 4A; >=90, red triangles) was predicted for the branch ancestral to the asporogenous groups *Listeria* and Lactobacillales, *Solibacillus* and Staphylococcaceae, and *Exiguobacterium* and Erysipelotrichia. In each of these groups, succeeding gene losses have occurred for Staphylococcaceae,
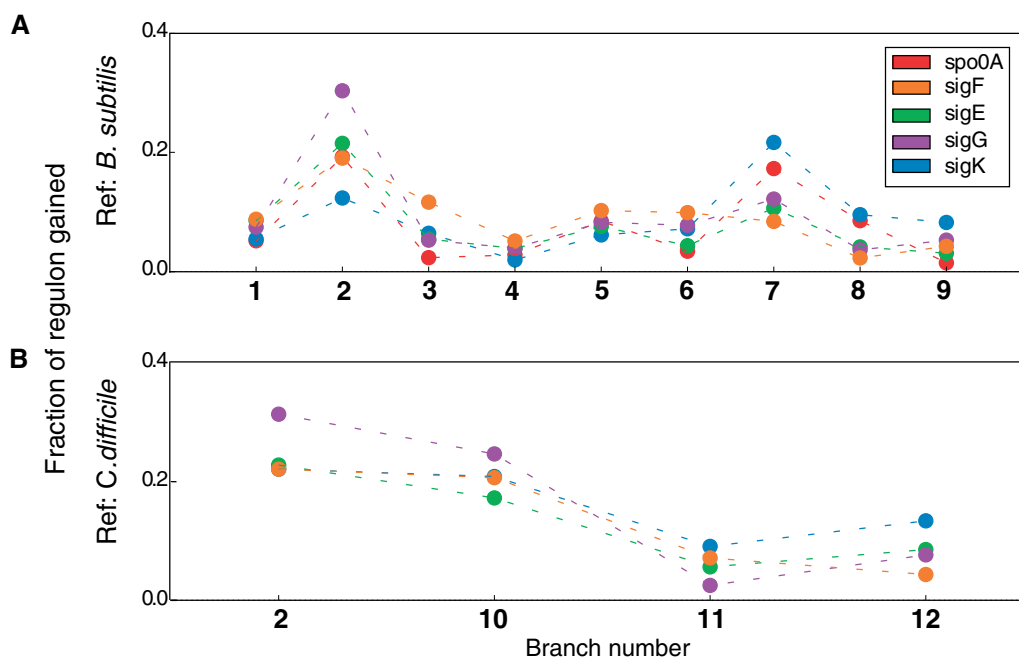
Fig. 6. Increase in size of sporulation regulons represented from root to tips for (A) *B. subtilis* and (B) *C. difficile*. Sporulation regulons have diverse sizes being σ^E the largest regulon and also the one acquiring the largest number of genes in every evolutionary step. To suppress this effect, the gains per regulon were normalized by size.

Exiguobaterium, *Listeria*, and Lactobacillales. In the extremophiles *Exiguobacterium* (Vishnivetskaya and Kathariou 2005; Vishnivetskaya et al. 2009; Moreno-Letelier et al. 2012), *spo0A* is present, but none of the genes coding for the cell-type sigma factors (supplementary fig. S5, Supplementary Material online). Moreover, the two *Exiguobacterium* species used here, *E. antarcticum* and *E. sibiricum*, have only 21% of the sporulation signature (Abecasis et al. 2013). These observations are in line with the notion that asporogenous Firmicutes have developed alternative ways to persist in diverse and sometimes extreme environments (Sauders et al. 2012; Linke et al. 2014). In addition, multiple gene losses arose independently in other groups, including spore formers, for example, at the base of Alicyclobacillaceae, of the *Geobacillus* and *Anoxybacillus* genera and in some other *Bacillus* species (fig. 4A, orange triangles).

In Clostridia, extensive gene losses have occurred independently for *Candidatus Arthromitus* and at the branch preceding the split of the genera *Filifactor* and *Proteocatella* (fig. 4B, red triangles). Other significant losses are predicted for the branches preceding the lineages Peptococcaceae and Thermoanaerobacterales groups III and IV, Thermoanaerobacterales, Lachnospiraceae, *Ruminiclostridium* and *Mageeibacillus*, and *Ruminococcus* (fig. 4B, yellow triangles). *Ruminococcus* species, in particular, have lost some of the conserved sporulation genes (fig. 4B) and have between 48% and 60% of the sporulation signature (Galperin et al. 2012; Abecasis et al. 2013) (supplementary figs. S1 and S3, Supplementary Material online), in what appears to be an intermediary stage into becoming asporogenous. Still, these organisms are likely to spread as spores or spore-like structures between hosts (Schloss et al. 2014;

Browne et al. 2016; Mukhopadhya et al. 2018), which plays as a counter evolutionary force to the emergence of asporogenous phenotypes. It should be noted, that at least *R. bromii* has most of the sporulation core genes and was shown to sporulate (Mukhopadhya et al. 2018). In another example, *Epulopiscium* morphotype B uses the sporulation program to form multiple intracellular offspring; it does not form dormant, fully resistant spores and has lost a significant proportion of the late sporulation genes (Miller et al. 2012; Hutchison et al. 2014; this work). Although reported as present, we failed to observe the *sigK* gene in the available *Epulopiscium* morphotype B genome (supplementary fig. S5, Supplementary Material online).

Significant gains and losses were also predicted for some terminal branches of Bacilli and Clostridia species/strains. However, gain/loss expectations on terminal branches are highly prone to error because false positives/negatives, or lower quality genomes, have a greater impact on the total estimates. Thus, only events occurring on internal branches were considered for discussion and presented in figure 4.

Sporulation Genes: A Landmark of Firmicutes

Evolutionary analyses of sporulation genes from *B. subtilis* and *C. difficile* estimate that 28.2% and 23.8% of their machinery was already present at the last bacterial ancestor (fig. 4 and supplementary tables S9 and S10, Supplementary Material online). These genes are conserved across early diverging bacterial clades and may have been co-opted for sporulation. Subsequent gene gains occurred later in the bacterial phylogeny, at different taxonomic levels and included TRGs, HGTs, and duplications as illustrated in figure 7. In particular, the first major gain event of sporulation genes (fig. 4, branch 2),
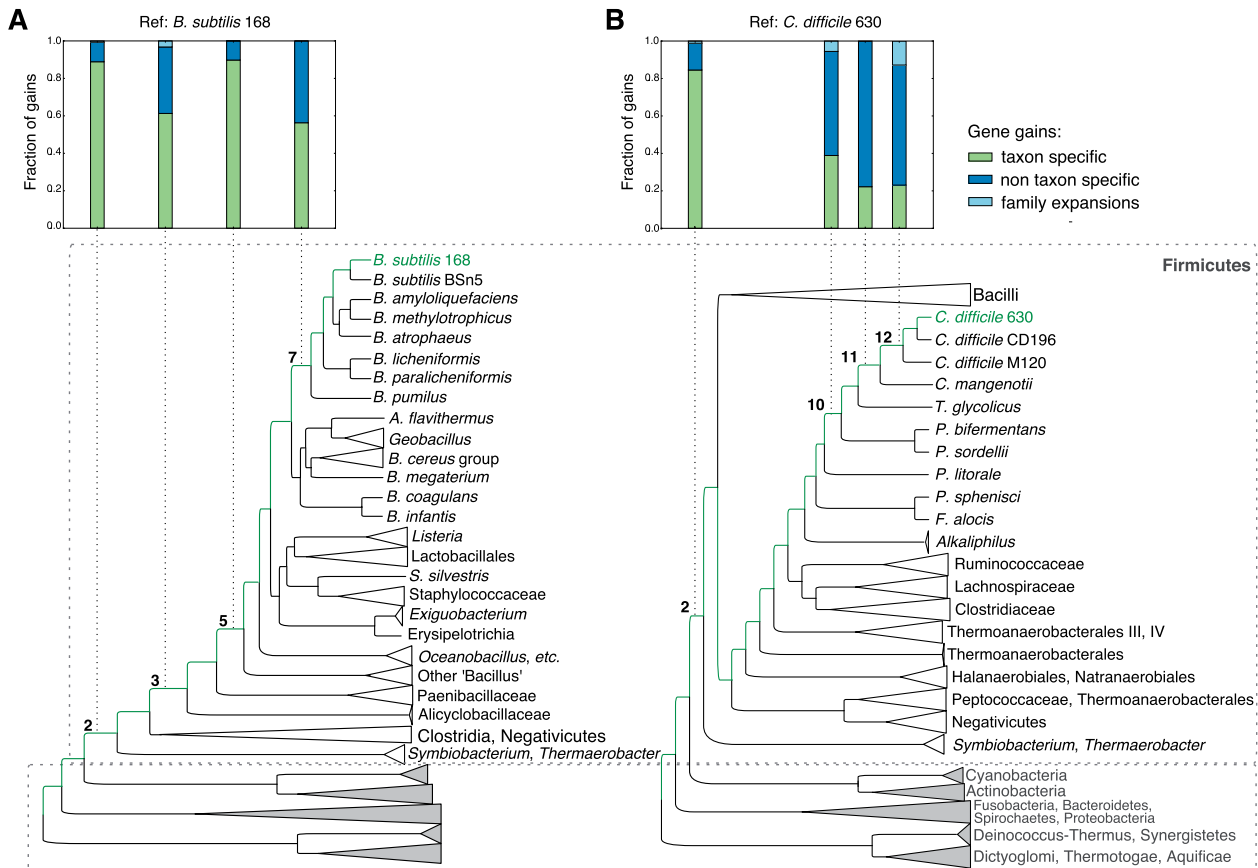
FIG. 7. Predicted categories of sporulation gene gains in four branches for (A) *B. subtilis* and (B) *C. difficile*. Gene gains were considered to be taxon specific (green) when gained at the branch, present in its descendants but not present or gained in other clades. Alternatively, gene gains were considered non taxon specific (dark blue) when gained along a certain branch but also present in other nondescendent clades. The later included genes gained by HGT in both directions (from and to the branch/clade). Family expansions (light blue) were originated by duplication events.

which correlates with the emergence of the Firmicutes as a phylum, involved a gain of at least 166 genes, from which the vast majority, 80–90%, are TRGs, that is, only found in Firmicutes members (fig. 7 and supplementary tables S9 and S10, Supplementary Material online). A much smaller number of gene gains (10.4–14.3%) are estimated to result from HGT (supplementary tables S9 and S10, Supplementary Material online) having homologs in lineages outside the Firmicutes.

B. subtilis versus *C. difficile*: Sporulation Driven by Rapid Divergence and HGT

Besides branch 2, we also estimated the possible origins of gene gains at lower taxonomic ranks. In particular, we chose branches in the phylogenetic tree for which we could strictly define descendants and outgroups in agreement with the NCBI taxonomy (release from May 2018). These ranks, shown in figure 7, included branch 3 (at the base of the class Bacilli), branch 5 (all Bacilli families excluding the early diverging Alicyclobacillaceae and Paenibacillaceae), branch 7 (at the base of the *B. subtilis* group), branch 10 (within the family Peptostreptococcaceae, a subgroup containing the genera *Clostridioides*, *Paeniclostridium*, *Paraclostridium*, and *Terrisporobacter*), branch 11 (at the base of the genus

Clostridioides), and branch 12 (at the base of the species *C. difficile*).

At the base and within Bacilli, the estimated gains of TRGs were dominant over HGTs, representing 61.3% (branch 3), 89.8% (branch 5), and 56.3% (branch 7) of the acquired genes (fig. 7A, in green). In sharp contrast, genes acquired along the clostridial lineages were more often present in other clades (fig. 7B, in blue) and more likely acquired by HGT, representing 55.6% in branch 10, 77.8% in branch 11, and 64.1% of the total gains in branch 12. Family expansions were significantly lower when compared with other mechanisms of gene acquisition with a maximum of four duplicated genes in branches 10 and 12. Lists of candidate HGTs, TRGs, and expansions by gene duplication are summarized in supplementary tables S9 for Bacilli and S10 for Clostridia, Supplementary Material online. Beyond the differences in rates of HGTs, TRGs, and family expansions, we can find evidence of both mechanisms occurring in Bacilli and Clostridia lineages.

HGT and Rapid Divergence in Clostridia

The exosporium that encases *C. difficile* spores has unique structural features and an important role in spore adhesion to host cells (Díaz-González et al. 2015; Calderón-Romero et al. 2018). Several exosporium components have been

identified in *C. difficile*, including the CdeA, CdeB, CdeC, and CdeM proteins (Díaz-González et al. 2015; Calderón-Romero et al. 2018). CdeB and CdeM are only found in *C. difficile* and represent a gene gain in branch 12 (Entrez GI:126699184, fig. 4 and supplementary table S10, Supplementary Material online). CdeC is specific to the Peptostreptococcaceae family (Calderón-Romero et al. 2018) and besides all three *C. difficile* strains, orthologs are present in the following species of our data set: *T. glycolicus*, *C. mangenotii*, and *P. sordellii* (Entrez GI:126698654, fig. 4 and supplementary table S10, Supplementary Material online). No orthologs are found in other groups suggesting that CdeC is a novelty within the Peptostreptococcaceae family.

Strikingly, CdeA is coded by a gene gained in branch 10 (Entrez GI:126699990, supplementary table S10, Supplementary Material online), which has homologs outside the Peptostreptococcaceae family, especially in species inhabiting gut ecosystems (clusters ENOG410XVF3 and HESKKIA from EggNOG 4.5.1 and OMA databases). A phylogenetic tree based on the alignment of the CdeA sequences (fig. 8A) does not reflect the species phylogeny suggesting multiple HGT events within the gut environment and through the cycling of Clostridia species among different niches, that is, mostly between sludge/sediments and the gut. Members of the *cdeA* cluster share two blocks of a four cysteine conserved domain with unknown function (InterPro Ac. no. IPR011437) (fig. 8B). CdeA thus provides a clear example of HGT, likely to have occurred in the gut ecosystem and important for spore function during infection. These results suggest that the exosporium proteins of *C. difficile* result from both gene novelty as well as HGT.

The *cotF* gene is another example of HGT in *C. difficile* (Entrez GI:126697784). Like its *B. subtilis* counterpart, CotF belongs to the coat F family (InterPro Ac. no. IPR012851) with a putative ferritin-like fold (PDB code: 2RBD). In both organisms, *cotF* is expressed in the mother cell, under the control of σ^K in *B. subtilis* and of σ^E in *C. difficile* (Cutting et al. 1991; Eichenberger et al. 2004; Steil et al. 2005; Fimlaid et al. 2013; Saujet et al. 2013). The *B. subtilis* CotF protein bears similarity to the products of four σ^G-controlled, forespore-specific genes, *yraG*, *yraF*, *yraE*, and *yraD*, part of an operon with the order *yraG-yraF-adhB-yraE-yraD* (supplementary fig. S7A, Supplementary Material online) (Wang et al. 2006). Both *yraF* and *yraD* code for proteins that resemble the C-terminal moiety of CotF; the *yraG* and *yraE* genes, in turn, code for proteins with similarity to the N-terminal part of CotF. Hence, the appearance of the *cotF* gene in branch 7, possibly results from a duplication and fusion of two *yra* genes, and is thus specific to the *Bacillus* group (supplementary table S9 and supplementary fig. S6, Supplementary Material online). The product of the *yraD* gene clusters with a single homolog in *C. difficile* coded by a *cotF-like gene*, but the genomic context of this gene is quite different from that of the *yra* genes in *B. subtilis* (supplementary fig. S7B, Supplementary Material online). This suggests HGT of an ancestral form of the *yraD* gene and its co-option for the σ^E regulon in *C. difficile*. While the

B. subtilis and *C. difficile cotF* genes may code for coat proteins (mother cell specific) in both organisms, the coat F domain also appears, at least in *B. subtilis*, to be functionally relevant in the forespore.

In contrast, the *spoIIIAF* gene, is an example of divergence. *spoIIIAF* is part of the *spoIIIA* operon and required for the assembly of the mother cell-to-forespore channel (Crawshaw et al. 2014; Morlot and Rodrigues 2018, see also above). The gene is conserved in all spore formers (Galperin et al. 2012; Crawshaw et al. 2014; Morlot and Rodrigues 2018). In both *B. subtilis* and *C. difficile*, *spoIIIA* is expressed in the mother cell under the control of σ^E. The requirement of the channel for sporulation is conserved in *C. difficile* (Fimlaid et al. 2015; Serrano et al. 2015). In *C. difficile*, *spoIIIAF* (Entrez GI:126698793; which corresponds to the *spoIIIAF* gene of *B. subtilis*), is the only gene of the *spoIIIA* operon that has been annotated with an extra "i" in its name for all *C. difficile* species. Genes named "*spoIIIAF*" have diverged to the point that they no longer cluster with their orthologs *spoIIIAF* genes, resulting in two distinct clusters of orthologous groups by our homology mapping approach (fig. 9A) as well as in orthology databases (Huerta-Cepas, Szklarczyk, et al. 2016; Altenhoff et al. 2018). The *spoIIIAF* cluster includes orthologs from all strains of *C. difficile*, *C. magenotii*, *Paraclostridium bifermentans*, *P. sordellii*, and *T. glycolicus* (figs. 5 and 9B). Because *spoIIIAF* is poorly conserved (fig. 9B), it is considered here as an innovation "gained" in branch 10, within the Peptostreptococcaceae family (supplementary table S10, Supplementary Material online). An alignment of SpoIIIAF orthologs from both clusters shows two possible transmembrane domains (residues 1–57) and divergence in the remaining sequence, which only shares 16% identity with its *B. subtilis* counterpart (fig. 9B). The exception are a group of hydrophobic residues (fig. 9B, residues highlighted in blue) which are part of the hydrophobic core of the protein and are associated with a conserved ring-building motif; this motif may support formation of a multimeric ring, similar to SpoIIIAG and SpoIIIAH (Levdikov et al. 2012; Zeytuni et al. 2017, 2018). Since a structure for *B. subtilis* SpoIIIAF is available (Zeytuni et al. 2018), we modeled the 3D structure of SpoIIIAF. Simulations generated five reliable models (supplementary fig. S8A, Supplementary Material online), which are structurally similar to SpoIIIAF and adopt a $\alpha_1\beta_1\beta_2\alpha_2\beta_4$ topology (fig. 9C). Still, the high divergence rates between the two orthologs is reflected in changes in the overall electrostatic potential and hydrophobicity of the protein surfaces (fig. 9C and supplementary fig. S8B, Supplementary Material online) and suggests that in the two organisms, SpoIIIAF has been under different evolutionary constraints. As the other *spoIIIA*-encoded products, *spoIIIAF* is predicted to be located in the outer forespore membrane (Doan et al. 2009; Zeytuni et al. 2017; Morlot and Rodrigues 2018). The high sequence divergence between SpoIIIAF and SpoIIIAF suggests functional links to distinct proteins or binding partners that need to be further explored both for Bacilli as well as for the members of the Peptostreptococcaceae family.
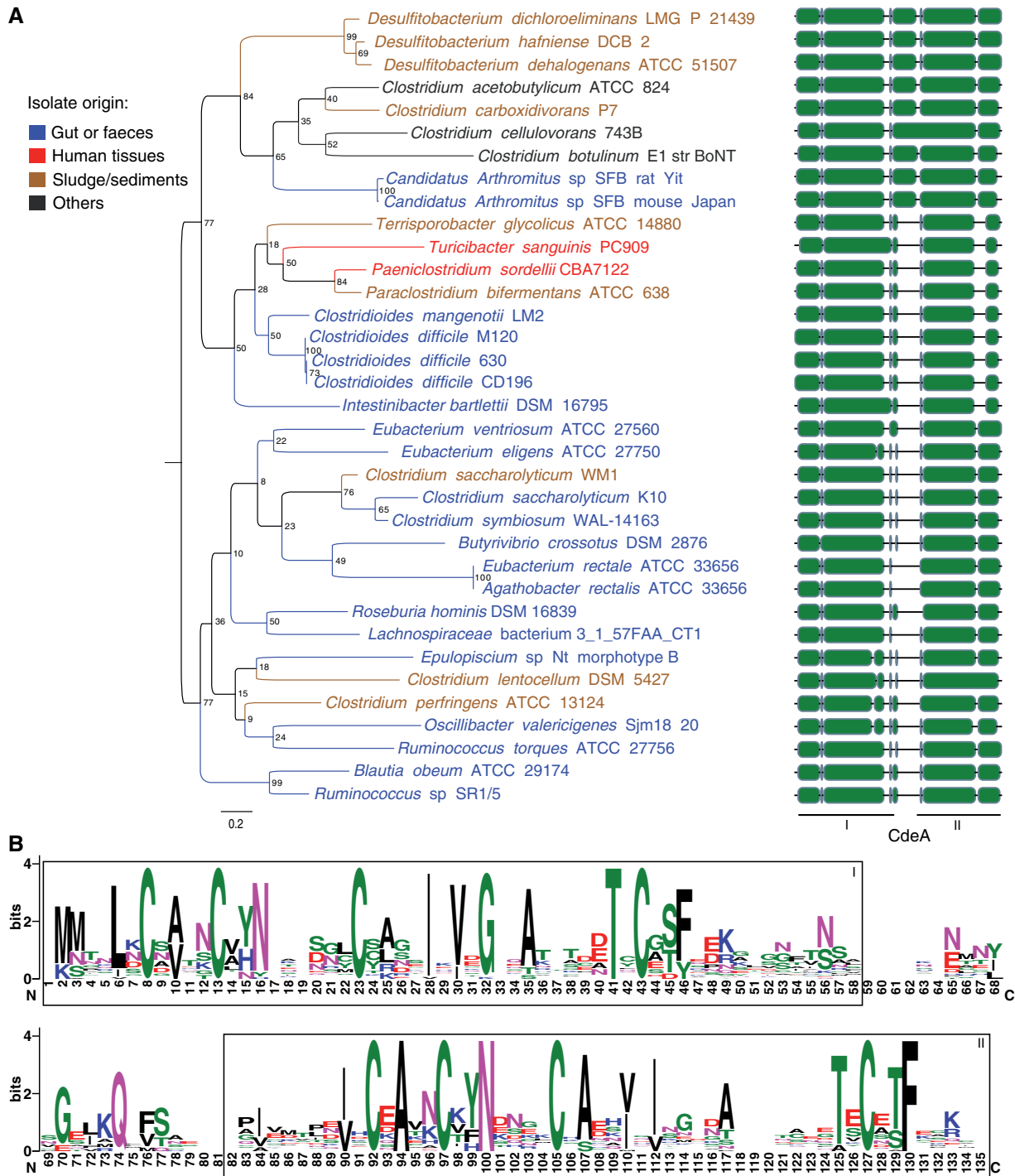
FIG. 8. Phylogenetic tree and conserved blocks of CdeA proteins. Orthologous sequences were selected based on the results of the orthology mapping approach and supplemented with orthologs from the cluster ENOG410XVF3 from EggNOG 4.5.1 (Huerta-Cepas, Szklarczyk, et al. 2016). Species are colored by environmental origin/niche. (*A*) Phylogeny with midpoint rooting and bootstrap probabilities were estimated with RAxML (Stamatakis 2014) using the best protein evolution model (LG + I) selected with ProtTest version 3 (Darriba et al. 2011), sequences were aligned using MAFFT (Katoh and Standley 2013). Protein blocks were drawn with ETE3 (Huerta-Cepas, Serra, et al. 2016). (*B*) Protein logos showing the two blocks of the conserved domain with four cysteines were estimated with WEBLOGO (Crooks et al. 2004).

Discussion

In this work, we have extended previous comparative genomics analyses based on *B. subtilis* sporulation genes (Paredes et al. 2005; de Hoon et al. 2010; Galperin et al. 2012; Abecasis et al. 2013; Galperin 2013) by including the sporulation genes

identified in the human pathogen *C. difficile*. Recent advances in *C. difficile* genetics and the ability to construct sporulation mutants in this anaerobe have enabled the identification of over 300 genes, directly regulated by the cell type-specific sigma factors of sporulation (Fimlaid et al. 2013; Saujet et al.
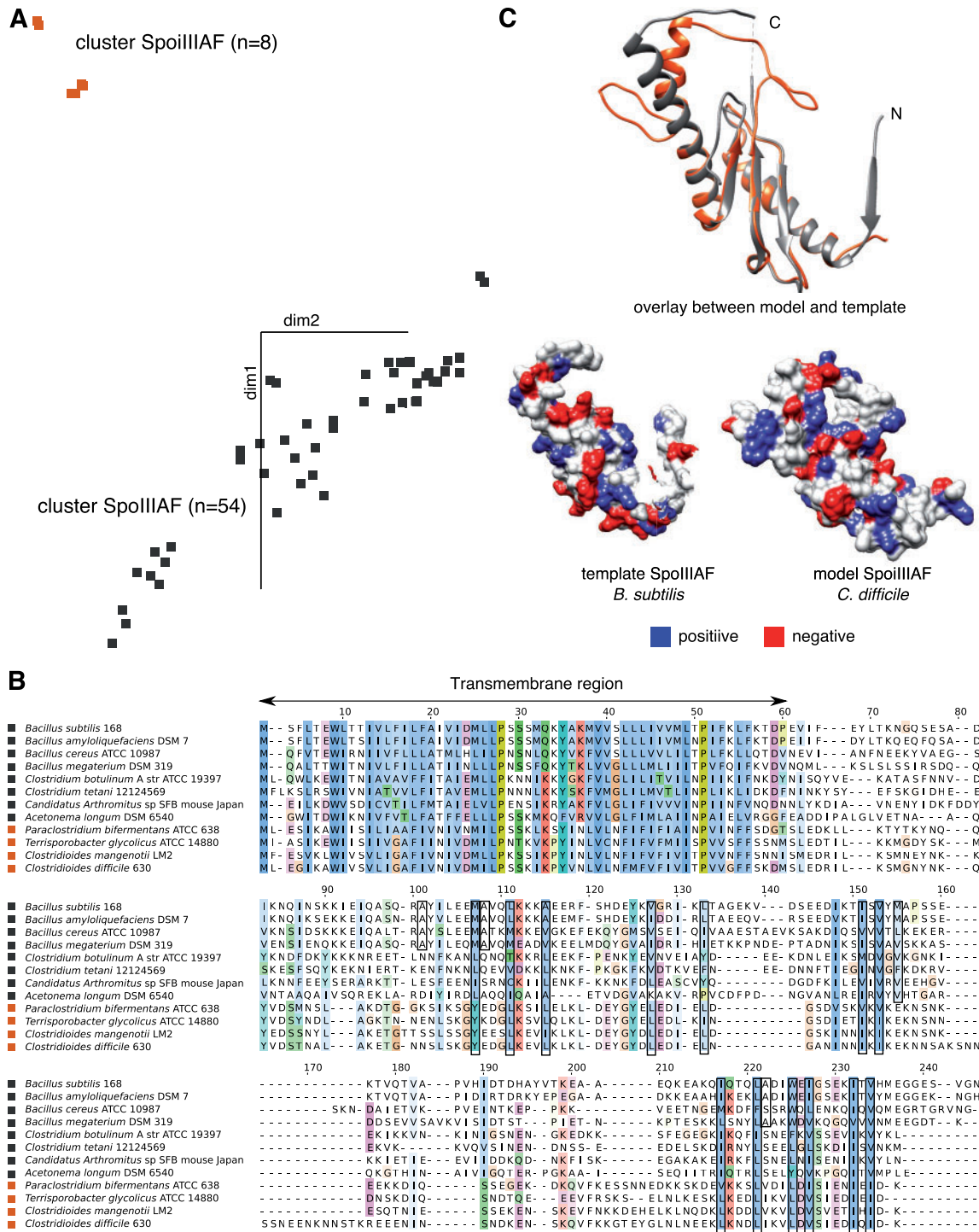
FIG. 9. Comparative analysis between sporulation proteins SpoIIIAF from *Bacillus subtilis* and SpoIIIAF from *Clostridioides difficile*. (A) Ortholog clusters of SpoIIIAF and SpoIIIAF ($n = 64$) were merged and aligned using MAFFT (Katoh and Standley 2013) followed by a principal component analysis as implemented in Jalview (Waterhouse et al. 2009) to estimate sequence divergence between clusters. (B) Selected sequences clustering with SpoIIIAF (black square) and SpoIIIAF (purple square) were aligned with T-coffee (Notredame et al. 2000). (C) Aligned structures of the monomeric forms of SpoIIIAF$_{85-206}$ (gray) and SpoIIIAF$_{84-205}$ (orange) were determined using Modeler homology-based modeling (Eswar et al. 2006) using the structure of *B. subtilis* SpoIIIAF (PDB code: 6dcs) as the template and MUSCLE (Edgar 2004) as sequence aligner. Structures and electrostatic charge distributions were visualized with Chimera (Pettersen et al. 2004).

2013; Pishdadian et al. 2015) (supplementary table S1, Supplementary Material online). By comparing the sporulation-specific regulons between the two species, we identified four main groups of genes based on their homology and role in sporulation. The largest group, representing over 50% of each regulon, consists of genes only present in one of the two reference spore formers (fig. 2 and supplementary table S5, Supplementary Material online), highlighting the

diversity and complex origins of the sporulation machinery. In some cases, genes having the same function in both species (i.e., functional equivalents) are annotated with the same name but have distinct evolutionary paths, with different ancestors, and hence belong to distinct ortholog groups. This is the case of the forespore-specific component of the mother cell-to-forespore channel, SpoIIQ, that also participates in engulfment in both *B. subtilis* (Londono-Vallejo et al. 1997; Camp and Losick 2008; Meisner et al. 2008) and in *C. difficile* (Fimlaid et al. 2015; Serrano et al. 2016). Both SpoIIQ proteins have a catalytic peptidase M23 motif (InterPro Ac. no. IPR016047) although nonfunctional in *B. subtilis* (Fimlaid et al. 2015; Serrano et al. 2016). Despite having the same name and similar roles in both species, the two *spoIIQ* genes are not orthologs; *spoIIQ* is a recognized case of a non-orthologous gene displacement (Galperin et al. 2012).

The second largest group of genes corresponds to those with candidate orthologs in the two genomes but only described as part of the sporulation regulons in one species (fig. 2 and supplementary table S4, Supplementary Material online). These genes are involved in other processes and may have been co-opted for roles in sporulation. For instance, the *dapA* and *dapB* genes, which are present in many bacterial phyla and involved in the lysine and peptidoglycan biosynthetic pathways (Mukherjee et al. 2017), have been recruited to sporulation in *B. subtilis* and form the *spoVF* operon, expressed in the mother cell and involved in the synthesis of DPA (Chen et al. 1993; Daniel and Errington 1993). In *C. difficile*, *dapA* and *dapB* are located in the *lysA-asd-dapA-dapB* operon (Mukherjee et al. 2017) but are dispensable for spore formation (Dembek et al. 2015; Donnelly et al. 2016). A study on *C. perfringens* suggests a more ancestral mechanism for DPA synthesis, involving the electron transfer flavoprotein Etfa (Orsburn et al. 2010). In the same group, the *uppS* gene (coding for undecaprenyl pyrophosphate synthetase, required for synthesis of the lipid carrier undecaprenyl phosphate, needed for peptidoglycan synthesis) is part of the σ^G regulon in *C. difficile* (Saujet et al. 2013; Dembek et al. 2015) but a role in *B. subtilis* sporulation has not yet been determined. Finally, there are two last groups of candidate orthologs in *B. subtilis* and *C. difficile*. One group corresponds to genes present in the same regulon in both species (fig. 2 and supplementary table S2, Supplementary Material online). For example, the *spmA* and *spmB* genes, involved in spore core dehydration, are under the control of σ^E in both species (Popham et al. 1995; Steil et al. 2005; Fimlaid et al. 2013; Pereira et al. 2013; Saujet et al. 2013). Other examples include the *spoIIIA* locus and *spoIID*, both controlled by σ^E or the *pdaA* gene, in the σ^G regulon. Still, despite being candidate orthologous pairs and regulated by the same sigma factors, these genes may show different modes of action in the two species through the link to nonhomologous intermediates (Fimlaid et al. 2015; Serrano et al. 2016). At last, there are genes clustering in the same ortholog group that belong to different regulons, as exemplified by the *yra/cotF* genes (fig. 2 and supplementary table S3 and supplementary fig. S7, Supplementary Material online).

We also find that some sporulation genes are conserved across several bacterial phyla, or even extended to other domains of life. For example, several genes with a role in spore cortex biogenesis established in *B. subtilis*, such as *spoVR* are present in Proteobacteria and Archaea (see above). Also, the *cotR* gene, with a role in spore coat assembly in *B. subtilis* (McKenney et al. 2013), shows a conserved hydrolase domain (InterPro Ac. no. IPR016035) and has an even broader taxonomic profile, which includes eukaryotes (ortholog group no. COG3621) (Huerta-Cepas, Szklarczyk, et al. 2016). Still, genes of broader taxonomic range constitute less than one-third of the sporulation machinery (fig. 4 and supplementary tables S9 and S10, Supplementary Material online). The great majority of the sporulation genes were acquired in the first major gain event, coincidently with the origin of the Firmicutes (fig. 4) and include the genes of the sporulation signature based on the *B. subtilis* genes (Abecasis et al. 2013) and the gut signature based on the analysis of the genomes of spore formers from the human gut microbiota (Browne et al. 2016). Since the divergence of Firmicutes is estimated to have occurred between 2.5 and 3.0 Ga, upon the emergence of terrestrial ecosystems (2.6 Ga) and after the Great Oxidation Event (2.3 Ga) (Battistuzzi et al. 2004; Battistuzzi and Hedges 2009; Marin et al. 2017), it has been suggested that sporulation is part of an adaptive process to cope with O_2 exposure (de Hoon et al. 2010; Abecasis et al. 2013).

Gene gain events occurring after the origin of sporulation (highlighted in fig. 4), correspond to lineage specializations of the program. The fact that we only find two larger gain events (≥ 90) along the evolutionary paths of *B. subtilis* and *C. difficile*, suggests that sporulation has evolved in punctual moments in the evolutionary history of these two species, only tied to certain lineages. Moreover, we find no evidence of these gains having greater influence in the size of a specific regulon since they have all coevolved proportionally (fig. 6). This and the fact that the four sporulation-specific sigma factors were present in the last common ancestor of Firmicutes (supplementary tables S9 and S10, Supplementary Material online) suggest that both early (engulfment) and late processes (synthesis of the spore cortex, coat/exosporium and preparation of the forespore for dormancy) were defined upon the emergence of sporulation. Interestingly, the single cell analysis of sporulation-specific gene expression in *C. difficile* revealed a reduced temporal segregation between the activities of the early and late sigma factors and a reduced dependency on the cell–cell signaling pathways controlling sigma factor activity or activation compared with the *B. subtilis* model (Pereira et al. 2013; Saujet et al. 2013) Possibly, this results from the loss of regulatory genes such as *csfB* (present at the basis of the Firmicutes) and the absence of additional regulatory proteins such as YabK, RsfA, and GerE (fig. 5).

Though the common macroevolutionary patterns in both lineages, our analysis to detect possible origins of gene gains, suggests distinct preferred modes of gene acquisition. While in *B. subtilis* most of the genes gained in branches 3, 5, and 7 are taxon specific (fig. 7A and supplementary table S9, Supplementary Material online), in the evolutionary path of *C. difficile*, genes were mostly acquired by HGT events, with a

consistently smaller proportion of TRGs (fig. 7B and supplementary table S10, Supplementary Material online). This finding is in line with the highly plastic genome of *C. difficile* 630, where 11% of mobile genetic elements (mainly transposons) are the main cause for acquisition of foreign genes from other species in the gut microbiota (Sebaihia et al. 2006; He et al. 2010). Genes acquired by HGT are involved in antibiotic resistance and virulence (Sebaihia et al. 2006; He et al. 2010) and as we now show, sporulation, conferring adaptive advantages to *C. difficile* in the gut ecosystem. A case study emerging from our analysis is the *cdeA* gene (fig. 8), coding for an exosporium protein, with the exosporium likely to have an important role during colonization/infection (Díaz-González et al. 2015).

In turn, *B. subtilis* shows a stronger lineage-specific specialization of the program, with genes that are only found in closely related species (fig. 7A and supplementary table S9, Supplementary Material online). The appearance of a higher number of such genes in members of the *B. subtilis* group may be due to the higher homologous recombination rates observed for Bacilli species when compared with *C. difficile* (Vos and Didelot 2009; He et al. 2010). A key feature of *B. subtilis* and close relatives is their natural competence, that is, the ability to enter a physiological state in which cells are able to take up exogenous DNA fragments and recombine them into the chromosome (Dubnau 1991; Haijema et al. 2001; Smits et al. 2005; Chen et al. 2007). Competence has been proposed as a main driver of genome diversity within the *B. subtilis* group (Brito et al. 2018) and to facilitate homologous recombination (Vos 2009; Mell and Redfield 2014). The other mechanisms giving rise to TRGs are of high mutation rates that saturate the signal of homology. Rapid divergence between orthologs may occur driven by exposure to particular lineage-specific environment/niches and to different binding partners. This latter is clearly illustrated for the counterparts of SpoIIIAF from members of the Peptostreptococcaceae family (SpoIIIAF) (fig. 9).

Besides gene gains, there were also extensive gene losses occurring independently throughout the bacterial phylogeny (fig. 4). These losses have resulted in the emergence of asporogenous lineages (i.e., nonspore-forming Firmicutes, Onyenwoke et al. 2004) among spore formers at different taxonomic levels. Within Bacilli, we find the group including *Listeria*, Lactobacillales, Staphylococcaceae, and *Exiguobacterium*; in Clostridia, the groups including the species *Proteocatella sphenisci* and *Filifactor alocis*, members of the family Lachnospiraceae and the genus *Caldicellulosiruptor* (within Thermoanaerobacterales group III) (supplementary fig. S1, Supplementary Material online). Although the ability to form spores is often seen has an evolutionary advantage, the process is energetically costly, involving the expression of hundreds of genes and taking up several hours to complete (Fujita and Losick 2005; Paredes et al. 2005; Fimlaid et al. 2013). In nutrient-rich environments after 6.000 generations of *B. subtilis*, neutral processes but also selection were found to facilitate loss of sporulation, through the accumulation of indels or single-nucleotide substitutions (Maughan et al. 2007, 2009; Maughan and Nicholson 2011; Nicholson 2012; Maitra and Dill 2015). In nature though, sporulation loss is mostly

driven by consecutive gene losses as suggested earlier (Galperin et al. 2012; Galperin 2013) and demonstrated in figure 4. As mentioned above, in *Epulopiscium* morphotype B and *Candidatus Arthromitus*, as well as in other intestinal spore formers, sporulation has evolved as a reproductive/propagation mechanism, concomitant with some gene loss (fig. 4) and eventually the gain of new genes under distinct evolutionary forces (Miller et al. 2012; Galperin 2013; Hutchison et al. 2014).

Conclusions

In this study, we were able to combine the comprehensive knowledge of sporulation genes in *B. subtilis* with the emerging model *C. difficile*. Our analysis validates earlier studies in that sporulation emerged once, at the base of Firmicutes. From this first major gene gain event, we have identified intermediate evolutionary steps combined with a second major lineage-specific gain event that underlies the current specialization and diversity observed among *B. subtilis* and *C. difficile* sporulation programs. Furthermore, predominant mechanisms of gene acquisition and innovation differ in the lineages of the two species. While *B. subtilis* entails greater innovation with many TRGs, *C. difficile* has acquired new genes mostly through HGTs. We also show that extensive gene losses underlie the emergence of asporogenous lineages and the co-option of sporulation as a reproductive mechanism. Overall, our comparative genomics approach provides a framework with which to elucidate the evolution and diversity observed in the sporulation program across Firmicutes and unveils new evolutionary case studies, such as those described for the *cdeA*, *cotF*, and *spoIIIAF* genes.

Materials and Methods

Data Sets Construction

All known sporulation genes from *B. subtilis* strain 168 and *C. difficile* strain 630, the two species that are better characterized in terms of their sporulation machinery, were collected from the literature (Fimlaid et al. 2013; Saujet et al. 2013; Pishdadian et al. 2015; Meeske et al. 2016) and SubtiWiki 2.0 (Mäder et al. 2012). These included the genes coding for the main regulators of sporulation: *spo0A*—controlling initiation; *sigE* and *sigK*—coding for the mother cell early and late specific sigma factors; *sigF* and *sigG*—coding for the early and late forespore-specific sigma factors; and the genes under their control (supplementary table S1, Supplementary Material online). Bacterial genomes ($n = 258$) were collected from the NCBI Database (release from April 2017) including main representative species from the following classes: Actinobacteria (15), Aquificae (2), Bacteroidetes (3), Cyanobacteria (3), Dictyogiomi (2), Fusobacteria (4), Proteobacteria (30), Spirochaetes (6), Synergistetes (1), Thermotogae (5), and Deinococcus-Thermus (1) (supplementary fig. S1 and supplementary table S6, Supplementary Material online). The selection of genomes from the class Firmicutes (186) was more exhaustive in order to include all the diversity of spore-forming bacteria described in the literature (Galperin 2013; Hutchison et al. 2014; Browne

Orthology/Paralogy Mapping and Selection of Homologous Families

The predicted proteins from each genome were clustered into homologous families using the bidirectional best hits (BDBH) algorithm implemented in GET_HOMOLOGUES (Contreras-Moreira and Vinuesa 2013) (supplementary fig. S2, Supplementary Material online). The software is built on top of BLAST+ (Camacho et al. 2009). First, inparalogs, defined as intraspecific BDBHs, are identified in each genome. Second, new genomes are incrementally compared with the reference genome and their BDBHs annotated. The main script ran with BlastP default options and a minimum coverage fixed at 65% for the pairwise alignments (-C), option -s to save memory, and reporting at least clusters with two sequences (-t set at 2). The choice of the BDBH algorithm and corresponding cutoffs resulted from prior validation with sets and subsets of the bacterial genomes used in this study (supplementary fig. S3, Supplementary Material online). The method was applied twice, the first time with *B. subtilis* strain 168 (BSU) genome fixed as reference (-r) and the second time with *C. difficile* strain 630 (CD) (supplementary fig. S2, Supplementary Material online). Homology mapping resulted in over 3,500 clusters of homologous genes for references BSU and CD. Clusters corresponding to sporulation genes were selected making a total of 726 and 307 clusters for BSU and CD, respectively. The results were represented in two large-scale profile matrices of presence/absence of genes (0 for absence, 1 presence, or >1 for multiple presence). The two reference genomes were also compared between each other using the orthoMCL algorithm implemented in GET_HOMOLOGUES. The main script ran with BlastP default options, -C set at 65% and reporting all clusters (-t 0).

Bacterial Phylogenies

Seventy gene clusters of single-copy orthologs conserved across the genomic data set (*adk, alaS, aspS, coaD, cysS, efp, engA, fabG, fmt, frr, gcp, gltX, guaA, infA, infC, ksgA, miaA, mraA, mraW, mraY, murC, murD, murE, murG, obgE, pheS, pheT, prfA, pth, pyrH, rluD, rpe, rplA-F, rplI-L, rplN-P, rplR-X, rpmA, rpoA, rpsB-E, rpsH, rpsJ, rpsK, rpsM, rpsQ, rpsS, ruvB, uppS, uvrB, yabC, yabD, ylbH,* and *yyaF*) were assembled with good occupancy; allowing for a maximum of five sequences missing per cluster. Clusters were aligned independently using MAFFT (Katoh and Standley 2013) or MUSCLE (Edgar 2004) under the default options. The best alignment per cluster was selected and trimmed using trimAl (Capella-Gutiérrez et al. 2009) with the -automated1 option. The 70 alignments were concatenated into a single superalignment comprising 258 species/strains and 14,401 amino acid positions. The alignment was used to construct phylogenetic trees by maximum likelihood and Bayesian methods (supplementary fig. S4, Supplementary Material online). Calculations and tree construction were performed with the RaxML 8 (Stamatakis 2014) (-N 100) and Mr. Bayes 3.2.5 (Ronquist et al. 2012) using the best protein evolution model (LG + G) selected with

ProtTest version 3 (Darriba et al. 2011). Tree topologies generated with both methods were compared using ete-compare (Huerta-Cepas, Serra, et al. 2016) based on the normalized Robinson–Foulds distance (nRF). The RF scores are the total number of leaf bipartitions occurring in one but not in both trees. Since nonnormalized measures of phylogenetic distance typically increase with the number of leaves in the tree, we used the normalized version, nRF, which divides RF by the maximum possible RF value for *n* leaves (maxRF). For topologically identical trees, nRF will be 0. Our nRF of 0.03 (RF/maxRF = 16.0/512.0) indicates similar tree topologies.

Phylogenetic Profiles and Evolutionary Analysis of Sporulation Genes

Sporulation gene clusters were represented in two large-scale profile matrices of presence/absence, sized 726 × 258 and 307 × 258, for references BSU and CD. The profiles were combined with the bacterial phylogeny to estimate gain (0 -> 1), loss (1 -> 0), and duplication events (1 -> 2 or more) in every tree branch with the software Count v 10.04 (Csűrös 2010). First, rates were optimized using the gain–loss–duplication model implemented in Count, allowing different gain–loss and duplication–loss rates in all lineages and rate variation across gene families (1–4 gamma categories). One hundred rounds of optimization were computed. Next, a gene family history analysis was computed by posterior probabilities based on a phylogenetic birth-and-death model. Tables with the posterior probabilities based on the two profile matrices (supplementary tables S7 and S8, Supplementary Material online) were extracted for further analysis with a customized Python script available at https://github.com/PaulaRS/ProcessCount (last accessed August 2019) to estimate total number of gene gains, losses and duplications (per branch), and gene presences (per node).

Estimating Origins of Gene Gains

To estimate the source of gene gains only events with a posterior probability of ≥0.5 by Count (Csűrös 2010) were considered. Gene gains were classified as "taxon specific" (TRGs) when present at a given branch, present in its descendants but not present in other clades. Alternatively, they were considered "non specific" when gained at a certain branch, present in its descendants but also present in other clades of the phylogeny. The later included genes gained by HGT in both directions (from and to the branch/clade), since it is often challenging to identify the donor (Koonin et al. 2001), specially when dealing with large genomic sets. At last, a small fraction of sporulation genes was also gained by gene duplication and classified as "family expansions."

Since our ancestral reconstruction was limited to 258 bacterial genomes, we needed to validate genes classified as "taxon specific" in a broader taxonomic context. To do so, we performed searches against the NCBI nr database (released in January 2018, Cutoffs: E-value 10E-10, 75% query coverage with 50% identity) using BlastP and checked for the best hits outside the clade and corresponding taxonomic distribution. While using this approach we were confronted with several false positives in the NCBI nr database that were

removed from the BlastP reports either because of misidentified species, for example, a genome assigned to the species *Varibaculum timonense*, from Actinobacteria (Guilhot et al. 2017), that is in fact a Clostridia, or because of recurring genomic contaminations of spore formers in genome assemblies of other species as was shown for *Mycobacterium* (Ghosh et al. 2009; Traag et al. 2010). Genes gained in the braches 2, 3, 5, 7, 10, 11, and 12 and classified as "taxon specific" and "non specific" or expansions are listed in supplementary tables S9 and S10, Supplementary Material online, with corresponding cluster names and Entrez GI numbers.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Author Contributions

P.R.-S. designed the study, collected data sets, implemented the bioinformatics methods, performed analyses, interpreted the data, and wrote the manuscript. M.S. provided the manually curated sets of sporulation genes and contributed with scientific input for the manuscript. A.O.H. designed the study, interpreted the data, and cowrote the manuscript.

References

Abanes-De Mello A, Sun YL, Aung S, Pogliano K. 2002. A cytoskeleton-like role for the bacterial cell wall during engulfment of the *Bacillus subtilis* forespore. *Genes Dev.* 16(24):3253–3264.

Abecasis AB, Serrano M, Alves R, Quintais L, Pereira-Leal JB, Henriques AO. 2013. A genomic signature and the identification of new sporulation genes. *J Bacteriol.* 195(9):2101–2115.

Abt MC, McKenney PT, Pamer EG. 2016. *Clostridium difficile* colitis: pathogenesis and host defence. *Nat Rev Microbiol.* 14(10):609–620.

Adams CM, Eckenroth BE, Putnam EE, Doublié S, Shen A. 2013. Structural and functional analysis of the CspB protease required for *Clostridium* spore germination. *PLoS Pathog.* 9(2):e1003165.

Al-Hinai MA, Jones SW, Papoutsakis ET. 2015. The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. *Microbiol Mol Biol Rev.* 79(1):19–37.

Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD. 2019. A new genomic blueprint of the human gut microbiota. *Nature* 568(7753):499–504.

Altenhoff AM, Glover NM, Train C-M, Kaleb K, Warwick Vesztrocy A, Dylus D, de Farias TM, Zile K, Stevenson C, Long J, et al. 2018. The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* 46(D1):D477–D485.

Alves Feliciano C, Douche T, Giai Gianetto Q, Matondo M, Martin-Verstraete I, Dupuy B. 2019. CotL, a new morphogenetic spore coat protein of *Clostridium difficile*. *Environ Microbiol.* 21(3):984–1003.

Battistuzzi FU, Feijao A, Hedges SB. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol.* 4:44.

Battistuzzi FU, Hedges SB. 2009. A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol.* 26(2):335–343.

Bauer MP, Notermans DW, van Benthem BH, Brazier JS, Wilcox MH, Rupnik M, Monnet DL, van Dissel JT, Kuijper EJ. 2011. *Clostridium difficile* infection in Europe: a hospital-based survey. *Lancet* 377(9759):63–73.

Beall B, Moran CP Jr. 1994. Cloning and characterization of spoVR, a gene from *Bacillus subtilis* involved in spore cortex formation. *J Bacteriol.* 176(7):2003–2012.

Berkeley R, Heyndrickx M, Logan N, De Vos P, editors. 2002. Applications and systematics of Bacillus and relatives. Oxford: Blackwell Science Ltd.

Bhattacharjee D, McAllister KN, Sorg JA. 2016. Germinants and their receptors in Clostridia. *J Bacteriol.* 198(20):2767–2775.

Blaylock B, Jiang X, Rubio A, Moran CP Jr, Pogliano K. 2004. Zipper-like interaction between proteins in adjacent daughter cells mediates protein localization. *Genes Dev.* 18(23):2916–2928.

Bradshaw N, Losick R. 2015. Asymmetric division triggers cell-specific gene expression through coupled capture and stabilization of a phosphatase. *Elife* 4:pii: e08145.

Brito PH, Chevreux B, Serra CR, Schyns G, Henriques AO, Pereira-Leal JB. 2018. Genetic competence drives genome diversity in *Bacillus subtilis*. *Genome Biol Evol.* 10(1):108–124.

Broder DH, Pogliano K. 2006. Forespore engulfment mediated by a ratchet-like mechanism. *Cell* 126(5):917–928.

Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. 2016. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* 533(7604):543–546.

Burns DA, Heap JT, Minton NP. 2010. SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. *J Bacteriol.* 192(3):657–664.

Calderón-Romero P, Castro-Córdova P, Reyes-Ramírez R, Milano-Céspedes M, Guerrero-Araya E, Pizarro-Guajardo M, Olguín-Araneda V, Gil F, Paredes-Sabja D. 2018. *Clostridium difficile* exosporium cysteine-rich proteins are essential for the morphogenesis of the exosporium layer, spore resistance, and affect *C. difficile* pathogenesis. *PLoS Pathog.* 14(8):e1007199.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST plus: architecture and applications. *BMC Bioinformatics* 10(1):421.

Camp AH, Losick R. 2008. A novel pathway of intercellular signalling in *Bacillus subtilis* involves a protein with similarity to a component of type III secretion channels. *Mol Microbiol.* 69(2):402–417.

Camp AH, Losick R. 2009. A feeding tube model for activation of a cell-specific transcription factor during sporulation in *Bacillus subtilis*. *Genes Dev.* 23(8):1014–1024.

Camp AH, Wang AF, Losick R. 2011. A small protein required for the switch from {sigma}F to {sigma}G during sporulation in *Bacillus subtilis*. *J Bacteriol.* 193(1):116–124.

Campo N, Rudner DZ. 2007. SpoIVB and CtpB are both forespore signals in the activation of the sporulation transcription factor sigmaK in *Bacillus subtilis*. *J Bacteriol.* 189(16):6021–6027.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.

Castaing JP, Nagy A, Anantharaman V, Aravind L, Ramamurthi KS. 2013. ATP hydrolysis by a domain related to translation factor GTPases drives polymerization of a static bacterial morphogenetic protein. *Proc Natl Acad Sci U S A.* 110(2):E151–E160.

Chary VK, Xenopoulos P, Piggot PJ. 2007. Expression of the sigmaF-directed csfB locus prevents premature appearance of sigmaG activity during sporulation of *Bacillus subtilis*. *J Bacteriol*. 189(23):8754–8757.

Chase DG, Erlandsen SL. 1976. Evidence for a complex life cycle and endospore formation in the attached, filamentous, segmented bacterium from murine ileum. *J Bacteriol*. 127(1):572–583.

Chastanet A, Losick R. 2007. Engulfment during sporulation in *Bacillus subtilis* is governed by a multi-protein complex containing tandemly acting autolysins. *Mol Microbiol*. 64(1):139–152.

Chen NY, Jiang SQ, Klein DA, Paulus H. 1993. Organization and nucleotide sequence of the *Bacillus subtilis* diaminopimelate operon, a cluster of genes encoding the first three enzymes of diaminopimelate synthesis and dipicolinate synthase. *J Biol Chem*. 268(13):9448–9465.

Chen XH, Koumoutsi A, Scholz R, Eisenreich A, Schneider K, Heinemeyer I, Morgenstern B, Voss B, Hess WR, Reva O, et al. 2007. Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nat Biotechnol*. 25(9):1007–1014.

Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, DeSantis TZ, Gihring TM, Lapidus A, Lin LH, Lowry SR, et al. 2008. Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* 322(5899):275–278.

Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol*. 79(24):7696–7701.

Costa T, Isidro AL, Moran CP Jr, Henriques AO. 2006. Interaction between coat morphogenetic proteins SafA and SpoVID. *J Bacteriol*. 188(22):7731–7741.

Crawshaw AD, Serrano M, Stanley WA, Henriques AO, Salgado PS. 2014. A mother cell-to-forespore channel: current understanding and future challenges. *FEMS Microbiol Lett*. 358(2):129–136.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res*. 14(6):1188–1190.

Csurös M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26(15):1910–1912.

Cutting S, Oke V, Driks A, Losick R, Lu S, Kroos L. 1990. A forespore checkpoint for mother cell gene expression during development in *B. subtilis*. *Cell* 62(2):239–250.

Cutting S, Roels S, Losick R. 1991. Sporulation operon spoIVF and the characterization of mutations that uncouple mother-cell from forespore gene expression in *Bacillus subtilis*. *J Mol Biol*. 221(4):1237–1256.

Daniel RA, Errington J. 1993. Cloning, DNA sequence, functional analysis and transcriptional regulation of the genes encoding dipicolinic acid synthetase required for sporulation in *Bacillus subtilis*. *J Mol Biol*. 232(2):468–483.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.

de Francesco M, Jacobs JZ, Nunes F, Serrano M, McKenney PT, Chua MH, Henriques AO, Eichenberger P. 2012. Physical interaction between coat morphogenetic proteins SpoVID and CotE is necessary for spore encasement in *Bacillus subtilis*. *J Bacteriol*. 194(18):4941–4950.

de Hoon MJL, Eichenberger P, Vitkup D. 2010. Hierarchical evolution of the bacterial sporulation network. *Curr Biol*. 20(17):R735–R745.

Decatur A, Losick R. 1996. Identification of additional genes under the control of the transcription factor sigma F of *Bacillus subtilis*. *J Bacteriol*. 178(16):5039–5041.

Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF, Fagan RP. 2015. High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. *MBio* 6(2):e02383.

Dembek M, Kelly A, Barwinska-Sendra A, Tarrant E, Stanley WA, Vollmer D, Biboy J, Gray J, Vollmer W, Salgado PS. 2018. Peptidoglycan degradation machinery in *Clostridium difficile* forespore engulfment. *Mol Microbiol*. 110(3):390–410.

Diaz OR, Sayer CV, Popham DL, Shen A. 2018. Clostridium difficile lipoprotein GerS is required for cortex modification and thus spore germination. *mSphere* 3:pii: e00205-18.

Díaz-González F, Milano M, Olguin-Araneda V, Pizarro-Cerda J, Castro-Córdova P, Tzeng SC, Maier CS, Sarker MR, Paredes-Sabja D. 2015. Protein composition of the outermost exosporium-like layer of *Clostridium difficile* 630 spores. *J Proteomics* 123:1–13.

Dick GJ, Lee YE, Tebo BM. 2006. Manganese(II)-oxidizing *Bacillus* spores in Guaymas Basin hydrothermal sediments and plumes. *Appl Environ Microbiol*. 72(5):3184–3190.

Doan T, Coleman J, Marquis KA, Meeske AJ, Burton BM, Karatekin E, Rudner DZ. 2013. FisB mediates membrane fission during sporulation in *Bacillus subtilis*. *Genes Dev*. 27(3):322–334.

Doan T, Morlot C, Meisner J, Serrano M, Henriques AO, Moran CP, Rudner DZ. 2009. Novel secretion apparatus maintains spore integrity and developmental gene expression in *Bacillus subtilis*. *PLoS Genet*. 5(7):e1000566.

Dong TC, Cutting SM. 2003. SpoIVB-mediated cleavage of SpoIVFA could provide the intercellular signal to activate processing of Pro-sigmaK in *Bacillus subtilis*. *Mol Microbiol*. 49(5):1425–1434.

Donnelly ML, Fimlaid KA, Shen A. 2016. Characterization of *Clostridium difficile* spores lacking either SpoVAC or dipicolinic acid synthetase. *J Bacteriol*. 198(11):1694–1707.

Driks A, Roels S, Beall B, Moran CP Jr, Losick R. 1994. Subcellular localization of proteins involved in the assembly of the spore coat of *Bacillus subtilis*. *Genes Dev*. 8(2):234–244.

Dubnau D. 1991. Genetic competence in *Bacillus subtilis*. *Microbiol Rev*. 55(3):395–424.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.

Edwards AN, Mcbride SM. 2014. Initiation of sporulation in *Clostridium difficile*: a twist on the classic model. *FEMS Microbiol Lett*. 358(2):110–118.

Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, et al. 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol*. 2(10):e328.

Eichenberger P, Jensen ST, Conlon EM, van Ooij C, Silvaggi J, González-Pastor J-E, Fujita M, Ben-Yehuda S, Stragier P, Liu JS, et al. 2003. The σ^E regulon and the identification of additional sporulation genes in *Bacillus subtilis*. *J Mol Biol*. 327(5):945–972.

Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, Pieper U, Sali A. 2006. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5:Unit-5.6.[CVOCROSSCVO]

Fang J, Kato C, Runko GM, Nogi Y, Hori T, Li J, Morono Y, Inagaki F. 2017. Predominance of viable spore-forming piezophilic bacteria in high-pressure enrichment cultures from ~1.5 to 2.4 km-deep coal-bearing sediments below the ocean floor. *Front Microbiol*. 8:137.

Fawcett P, Eichenberger P, Losick R, Youngman P. 2000. The transcriptional profile of early to middle sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci U S A*. 97(14):8063–8068.

Fernandes CG, Moran CP Jr, Henriques AO. 2018. Auto-regulation of SafA assembly through recruitment of a protein cross-linking enzyme. *J Bacteriol*. 200(14):e00066-18.

Feucht A, Evans L, Errington J. 2003. Identification of sporulation genes by genome-wide analysis of the sigmaE regulon of *Bacillus subtilis*. *Microbiology (Reading, Engl)*. 149(Pt 10):3023–3034.

Filippidou S, Wunderlin T, Junier T, Jeanneret N, Dorador C, Molina V, Johnson DR, Junier P. 2016. A combination of extreme environmental conditions favor the prevalence of endospore-forming *Firmicutes*. *Front Microbiol*. 7:1–11.

Fimlaid KA, Bond JP, Schutz KC, Putnam EE, Leung JM, Lawley TD, Shen A. 2013. Global analysis of the sporulation pathway of *Clostridium difficile*. *PLoS Genet*. 9(8):e1003660.

Fimlaid KA, Jensen O, Donnelly ML, Siegrist MS, Shen A. 2015. Regulation of *Clostridium difficile* spore formation by the SpoIIQ and SpoIIIA proteins. *PLoS Genet*. 11(10):e1005562.

Fimlaid KA, Shen A. 2015. Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. *Curr Opin Microbiol.* 24:88–95.

Flanagan KA, Comber JD, Mearls E, Fenton C, Wang Erickson AF, Camp AH. 2016. A membrane-embedded amino acid couples the SpoIIQ channel protein to anti-sigma factor transcriptional repression during *Bacillus subtilis* sporulation. *J Bacteriol.* 198(9):1451–1463.

Flint JF, Drzymalski D, Montgomery WL, Southam G, Angert ER. 2005. Nocturnal production of endospores in natural populations of *Epulopiscium*-like surgeonfish symbionts. *J Bacteriol.* 187(21):7460–7470.

Forster SC, Kumar N, Anonye BO, Almeida A, Viciani E, Stares MD, Dunn M, Mkandawire TT, Zhu A, Shao Y, et al. 2019. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat Biotechnol.* 37(2):186–192.

Francis MB, Allen CA, Shrestha R, Sorg JA. 2013. Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. *PLoS Pathog.* 9(5):e1003356.

Fujita M, Losick R. 2005. Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev.* 19(18):2236–2244.

Fujiya M, Musch MW, Nakagawa Y, Hu S, Alverdy J, Kohgo Y, Schneewind O, Jabri B, Chang EB. 2007. The *Bacillus subtilis* quorum-sensing molecule CSF contributes to intestinal homeostasis via OCTN2, a host cell membrane transporter. *Cell Host Microbe* 1(4):299–308.

Fukushima T, Yamamoto H, Atrih A, Foster SJ, Sekiguchi J. 2002. A polysaccharide deacetylase gene (pdaA) is required for germination and for production of muramic delta-lactam residues in the spore cortex of *Bacillus subtilis*. *J Bacteriol.* 184(21):6007–6015.

Galperin MY. 2013. Genome diversity of spore-forming Firmicutes. *Microbiol Spectr.* 1:1–15.

Galperin MY, Brover V, Tolstoy I, Yutin N. 2016. Phylogenomic analysis of the family Peptostreptococcaceae (*Clostridium* cluster XI) and proposal for reclassification of *Clostridium litorale* (Fendrich et al. 1991) and *Eubacterium acidaminophilum* (Zindel et al. 1989) as *Peptoclostridium litorale* gen. nov. *Int J Syst Evol Microbiol.* 66(12):5506–5513.

Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI, Rigden DJ. 2012. Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol.* 14(11):2870–2890.

Ghosh J, Larsson P, Singh B, Pettersson BMF, Islam NM, Sarkar SN, Dasgupta S, Kirsebom LA. 2009. Sporulation in mycobacteria. *Proc Natl Acad Sci U S A.* 106(26):10781–10786.

Gomez M, Cutting SM. 1997. BofC encodes a putative forespore regulator of the *Bacillus subtilis* sigma K checkpoint. *Microbiology* 143(Pt 1):157–170.

Guilhot E, Lagier JC, Raoult D, Khelaifia S. 2017. 'Prevotella ihumii' sp. nov. and 'Varibaculum timonense' sp. nov., two new bacterial species isolated from a fresh human stool specimen. *New Microbes New Infect.* 18:3–5.

Gutelius D, Hokeness K, Logan SM, Reid CW. 2014. Functional analysis of SleC from *Clostridium difficile*: an essential lytic transglycosylase involved in spore germination. *Microbiology (Reading, Engl).* 160(Pt 1):209–216.

Gutierrez J, Smith R, Pogliano K. 2010. SpoIID-mediated peptidoglycan degradation is required throughout engulfment during *Bacillus subtilis* sporulation. *J Bacteriol.* 192(12):3174–3186.

Haijema BJ, Hahn J, Haynes J, Dubnau D. 2001. A ComGA-dependent checkpoint limits growth during the escape from competence. *Mol Microbiol.* 40(1):52–64.

Halder S, Parrell D, Whitten D, Feig M, Kroos L. 2017. Interaction of intramembrane metalloprotease SpoIVFB with substrate Pro-sigma(K). *Proc Natl Acad Sci U S A.* 114(50):E10677–E10686.

Han C, Gu W, Zhang X, Lapidus A, Nolan M, Copeland A, Lucas S, del Rio TG, Tice H, Cheng JF, et al. 2010. Complete genome sequence of *Thermaerobacter marianensis* type strain (7p75a T). *Stand Genomic Sci.* 3(3):337–345.

He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, Holt KE, Seth-Smith HMB, Quail MA, Rance R, et al. 2010. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci U S A.* 107(16):7527–7532.

Henriques AO, Moran CP. 2007. Structure, assembly, and function of the spore surface layers. *Annu Rev Microbiol.* 61:555–588.

Higgins D, Dworkin J. 2012. Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiol Rev.* 36(1):131–148.

Hong HA, Khaneja R, Tam NMK, Cazzato A, Tan S, Urdaci M, Brisson A, Gasbarrini A, Barnes I, Cutting SM. 2009. *Bacillus subtilis* isolated from the human gastrointestinal tract. *Res Microbiol.* 160(2):134–143.

Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 33(6):1635–1638.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44(D1):D286–D293.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, et al. 2016. A new view of the tree of life. *Nat Microbiol.* 1:16048.

Hutchison EA, Miller DA, Angert ER. 2014. Sporulation in bacteria: beyond the standard model. *Microbiol Spectr.* 2:1–15.

Imamura D, Kuwana R, Takamatsu H, Watabe K. 2011. Proteins involved in formation of the outermost layer of *Bacillus subtilis* spores. *J Bacteriol.* 193(16):4075–4080.

Janoir C, Denève C, Bouttier S, Barbut F, Hoys S, Caleechum L, Chapetón-Montes D, Pereira FC, Henriques AO, Collignon A, et al. 2013. Adaptive strategies and pathogenesis of *Clostridium difficile* from in vivo transcriptomics. *Infect Immun.* 81(10):3757–3769.

Karmazyn-Campelli C, Rhayat L, Carballido-Lopez R, Duperrier S, Frandsen N, Stragier P. 2008. How the early sporulation sigma factor sigmaF delays the switch to late development in *Bacillus subtilis*. *Mol Microbiol.* 67(5):1169–1180.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.

Kevorkian Y, Shen A. 2017. Revisiting the role of Csp family proteins in regulating *Clostridium difficile* spore germination. *J Bacteriol.* 199:e00266-17.

Kevorkian Y, Shirley DJ, Shen A. 2016. Regulation of *Clostridium difficile* spore germination by the CspA pseudoprotease domain. *Biochimie* 122:243–254.

Konovalova A, Sogaard-Andersen L, Kroos L. 2014. Regulated proteolysis in bacterial development. *FEMS Microbiol Rev.* 38(3):493–522.

Koonin EV, Makarova KS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol.* 55:709–742.

Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, Mastroeni P, Scott P, Raisen C, Mottram L, et al. 2009. Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infect Immun.* 77(9):3661–3669.

Lawson PA, Citron DM, Tyrrell KL, Finegold SM. 2016. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe* 40:95–99.

Leidal AM, Lee PWK, McCormick C. 2012. Viral subversion of autophagy impairs oncogene-induced senescence. *Autophagy* 8(7):1138–1140.

Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, et al. 2015. Burden of *Clostridium difficile* infection in the United States. *N Engl J Med.* 372(9):825–834.

Levdikov VM, Blagova EV, McFeat A, Fogg MJ, Wilson KS, Wilkinson AJ. 2012. Structure of components of an intercellular channel complex in sporulating *Bacillus subtilis*. *Proc Natl Acad Sci U S A.* 109(14):5441–5445.

Levin PA, Fan N, Ricca E, Driks A, Losick R, Cutting S. 1993. An unusually small gene required for sporulation by *Bacillus subtilis*. *Mol Microbiol*. 9(4):761–771.

Linke K, Rückerl I, Brugger K, Karpiskova R, Walland J, Muri-Klinger S, Tichy A, Wagner M, Stessl B. 2014. Reservoirs of *Listeria* species in three environmental ecosystems. *Appl Environ Microbiol*. 80(18):5583–5592.

Londono-Vallejo JA, Frehel C, Stragier P. 1997. SpoIIQ, a forespore-expressed gene required for engulfment in *Bacillus subtilis*. *Mol Microbiol*. 24(1):29–39.

Mäder U, Schmeisky AG, Flórez LA, Stülke J. 2012. SubtiWiki—a comprehensive community resource for the model organism *Bacillus subtilis*. *Nucleic Acids Res*. 40:D1278–D1287.

Maitra A, Dill KA. 2015. Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc Natl Acad Sci U S A*. 112(2):406–411.

Marin J, Battistuzzi FU, Brown AC, Hedges SB. 2017. The timetree of prokaryotes: new insights into their evolution and speciation. *Mol Biol Evol*. 34(2):437–446.

Martinez-Lumbreras S, Alfano C, Evans NJ, Collins KM, Flanagan KA, Atkinson RA, Krysztofinska EM, Vydyanath A, Jackter J, Fixon-Owoo S, et al. 2018. Structural and functional insights into *Bacillus subtilis* sigma factor inhibitor, CsfB. *Structure* 26:640–648.e5.

Mastny M, Heuck A, Kurzbauer R, Heiduk A, Boisguerin P, Volkmer R, Ehrmann M, Rodrigues CD, Rudner DZ, Clausen T. 2013. CtpB assembles a gated protease tunnel regulating cell–cell signaling during spore formation in *Bacillus subtilis*. *Cell* 155(3):647–658.

Maughan H, Birky CW, Nicholson WL. 2009. Transcriptome divergence and the loss of plasticity in *Bacillus subtilis* after 6,000 generations of evolution under relaxed selection for sporulation. *J Bacteriol*. 191(1):428–433.

Maughan H, Masel J, Birky CW, Nicholson WL. 2007. The roles of mutation accumulation and selection in loss of sporulation in experimental populations of *Bacillus subtilis*. *Genetics* 177(2):937–948.

Maughan H, Nicholson WL. 2011. Increased fitness and alteration of metabolic pathways during *Bacillus subtilis* evolution in the laboratory. *Appl Environ Microbiol*. 77(12):4105–4118.

McKenney PT, Driks A, Eichenberger P. 2013. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. *Nat Rev Microbiol*. 11(1):33–44.

McKenney PT, Driks A, Eskandarian HA, Grabowski P, Guberman J, Wang KH, Gitai Z, Eichenberger P. 2010. A distance-weighted interaction map reveals a previously uncharacterized layer of the *Bacillus subtilis* spore coat. *Curr Biol*. 20(10):934–938.

McKenney PT, Eichenberger P. 2012. Dynamics of spore coat morphogenesis in *Bacillus subtilis*. *Mol Microbiol*. 83(2):245–260.

Mearls EB, Jackter J, Colquhoun JM, Farmer V, Matthews AJ, Murphy LS, Fenton C, Camp AH. 2018. Transcription and translation of the sigG gene is tuned for proper execution of the switch from early to late gene expression in the developing *Bacillus subtilis* spore. *PLoS Genet*. 14(4):e1007350.

Meeske AJ, Rodrigues CDA, Brady J, Lim HC, Bernhardt TG, Rudner DZ. 2016. High-throughput genetic screens identify a large and diverse collection of new sporulation genes in *Bacillus subtilis*. *PLoS Biol*. 14(1):e1002341.

Meisner J, Wang X, Serrano M, Henriques AO, Moran CP. 2008. A channel connecting the mother cell and forespore during bacterial endospore formation. *Proc Natl Acad Sci U S A*. 105(39):15100–15105.

Mell JC, Redfield RJ. 2014. Natural competence and the evolution of DNA uptake specificity. *J Bacteriol*. 196(8):1471–1483.

Miller DA, Suen G, Clements KD, Angert ER. 2012. The genomic basis for the evolution of a novel form of cellular reproduction in the bacterium *Epulopiscium*. *BMC Genomics*. 13:265.

Moreno-Letelier A, Olmedo-Alvarez G, Eguiarte LE, Souza V. 2012. Divergence and phylogeny of Firmicutes from the Cuatro Ciénegas Basin, Mexico: a window to an ancient ocean. *Astrobiology* 12(7):674–684.

Morlot C, Rodrigues C. 2018. The new kid on the block: a specialized secretion system during bacterial sporulation. *Trends Microbiol*. 26(8):663–676.

Morlot C, Uehara T, Marquis KA, Bernhardt TG, Rudner DZ. 2010. A highly coordinated cell wall degradation machine governs spore morphogenesis in *Bacillus subtilis*. *Genes Dev*. 24(4):411–422.

Mukherjee S, Barash D, Sengupta S. 2017. Comparative genomics and phylogenomic analyses of lysine riboswitch distributions in bacteria. *PLoS One* 12(9):e0184314.

Mukhopadhya I, Moraïs S, Laverde-Gomez J, Sheridan PO, Walker AW, Kelly W, Klieve AV, Ouwerkerk D, Duncan SH, Louis P, et al. 2018. Sporulation capability and amylosome conservation among diverse human colonic and rumen isolates of the keystone starch-degrader *Ruminococcus bromii*. *Environ Microbiol*. 20(1):324–336.

Mullerova D, Krajcikova D, Barak I. 2009. Interactions between *Bacillus subtilis* early spore coat morphogenetic proteins. *FEMS Microbiol Lett*. 299:74–85.

Narula J, Fujita M, Igoshin OA. 2016. Functional requirements of cellular differentiation: lessons from *Bacillus subtilis*. *Curr Opin Microbiol*. 34:38–46.

Nicholson WL. 2012. Increased competitive fitness of *Bacillus subtilis* under nonsporulating conditions via inactivation of pleiotropic regulators AlsR, SigD, and SigW. *Appl Environ Microbiol*. 78(9):3500–3503.

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(3):548–572.

Notredame C, Higgins D, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. 302(1):205–217.

Nunes F, Fernandes C, Freitas C, Marini E, Serrano M, Moran CP Jr, Eichenberger P, Henriques AO. 2018. SpoVID functions as a noncompetitive hub that connects the modules for assembly of the inner and outer spore coat layers in *Bacillus subtilis*. *Mol Microbiol*. 110(4):576–595.

Ojkic N, López-Garrido J, Pogliano K, Endres RG. 2016. Cell-wall remodeling drives engulfment during *Bacillus subtilis* sporulation. *Elife* 5:pii: e18657.

Onyenwoke RU, Brill JA, Farahi K, Wiegel J. 2004. Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). *Arch Microbiol*. 182:182–192.

Orsburn BC, Melville SB, Popham DL. 2010. EtfA catalyses the formation of dipicolinic acid in *Clostridium perfringens*. *Mol Microbiol*. 75(1):178–186.

Ozin AJ, Henriques AO, Yi H, Moran CP. 2000. Morphogenetic proteins SpoVID and SafA form a complex during assembly of the *Bacillus subtilis* spore coat. *J Bacteriol*. 182(7):1828–1833.

Ozin AJ, Samford CS, Henriques AO, Moran CP. 2001. SpoVID guides SafA to the spore coat in *Bacillus subtilis*. *J Bacteriol*. 183(10):3041–3049.

Pan Q, Losick R, Rudner DZ. 2003. A second PDZ-containing serine protease contributes to activation of the sporulation transcription factor sigmaK in *Bacillus subtilis*. *J Bacteriol*. 185(20):6051–6056.

Paredes CJ, Alsaker KV, Papoutsakis ET. 2005. A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol*. 3(12):969–978.

Paredes-Sabja D, Shen A, Sorg JA. 2014. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends Microbiol*. 22(7):406–416.

Pereira FC, Nunes F, Cruz F, Fernandes C, Isidro AL, Lousa D, Soares CM, Moran CP Jr, Henriques AO, Serrano M. 2018. A LysM domain intervenes in sequential protein–protein and protein–peptidoglycan interactions important for spore coat assembly in *Bacillus subtilis*. *J Bacteriol*. 201:pii: e00642-18.

Pereira FC, Saujet L, Tomé AR, Serrano M, Monot M, Couture-Tosi E, Martin-Verstraete I, Dupuy B, Henriques AO. 2013. The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. *PLoS Genet*. 9(10):e1003782.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* 25(13):1605–1612.

Pishdadian K, Fimlaid KA, Shen A. 2015. SpoIIID-mediated regulation of σ^K function during *Clostridium difficile* sporulation. *Mol Microbiol.* 95(2):189–208.

Popham DL, Bernhards CB. 2015. Spore peptidoglycan. *Microbiol Spectr.* 3(6):TBS-0005-2012.

Popham DL, Illades-Aguiar B, Setlow P. 1995. The *Bacillus subtilis* dacB gene, encoding penicillin-binding protein 5*, is part of a three-gene operon required for proper spore cortex synthesis and spore core dehydration. *J Bacteriol.* 177(16):4721–4729.

Putnam EE, Nock AM, Lawley TD, Shen A. 2013. SpoIVA and SipL are *Clostridium difficile* spore morphogenetic proteins. *J Bacteriol.* 195(6):1214–1225.

Qiao H, Krajcikova D, Liu C, Li Y, Wang H, Barak I, Tang J. 2012. The interactions of spore-coat morphogenetic proteins studied by single-molecule recognition force spectroscopy. *Chem Asian J.* 7(4):725–731.

Qiao H, Krajcikova D, Xing C, Lu B, Hao J, Ke X, Wang H, Barak I, Tang J. 2013. Study of the interactions between the key spore coat morphogenetic proteins CotE and SpoVID. *J Struct Biol.* 181(2):128–135.

Ramamurthi KS, Clapham KR, Losick R. 2006. Peptide anchoring spore coat assembly to the outer forespore membrane in *Bacillus subtilis*. *Mol Microbiol.* 62(6):1547–1557.

Ramamurthi KS, Losick R. 2008. ATP-driven self-assembly of a morphogenetic protein in *Bacillus subtilis*. *Mol Cell* 31(3):406–414.

Ramirez-Guadiana FH, Meeske AJ, Rodrigues CDA, Barajas-Ornelas RDC, Kruse AC, Rudner DZ. 2017. A two-step transport pathway allows the mother cell to nurture the developing spore in *Bacillus subtilis*. *PLoS Genet.* 13:e1007015.

Resnekov O, Losick R. 1998. Negative regulation of the proteolytic activation of a developmental transcription factor in *Bacillus subtilis*. *Proc Natl Acad Sci U S A.* 95(6):3162–3167.

Rhayat L, Duperrier S, Carballido-Lopez R, Pellegrini O, Stragier P. 2009. Genetic dissection of an inhibitor of the sporulation sigma factor sigma(G). *J Mol Biol.* 390(5):835–844.

Ribis JW, Fimlaid KA, Shen A. 2018. Differential requirements for conserved peptidoglycan remodeling enzymes during *Clostridioides difficile* spore formation. *Mol Microbiol.* 110(3):370–389.

Ribis JW, Ravichandran P, Putnam EE, Pishdadian K, Shen A. 2017. The conserved spore coat protein SpoVM is largely dispensable in *Clostridium difficile* spore formation. *mSphere* 2(5):pii: e00315-17.

Ricca E, Cutting S, Losick R. 1992. Characterization of bofA, a gene involved in intercompartmental regulation of pro-sigma K processing during sporulation in *Bacillus subtilis*. *J Bacteriol.* 174(10):3177–3184.

Rodrigues CDA, Henry X, Neumann E, Kurauskas V, Bellard L, Fichou Y, Schanda P, Schoehn G, Rudner DZ, Morlot C. 2016. A ring-shaped conduit connects the mother cell and forespore during sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci U S A.* 113(41):11585–11590.

Rodrigues CDA, Ramirez-Guadiana FH, Meeske AJ, Wang X, Rudner DZ. 2016. GerM is required to assemble the basal platform of the SpoIIIA-SpoIIQ transenvelope complex during sporulation in *Bacillus subtilis*. *Mol Microbiol.* 102(2):260–273.

Roels S, Driks A, Losick R. 1992. Characterization of spoIVA, a sporulation gene involved in coat morphogenesis in *Bacillus subtilis*. *J Bacteriol.* 174(2):575–585.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.

Rudner DZ, Losick R. 2002. A sporulation membrane protein tethers the pro-sigmaK processing enzyme to its inhibitor and dictates its subcellular localization. *Genes Dev.* 16(8):1007–1018.

Sauders BD, Overdevest J, Fortes E, Windham K, Schukken Y, Lembo A, Wiedmann M. 2012. Diversity of *Listeria* species in urban and natural environments. *Appl Environ Microbiol.* 78(12):4420–4433.

Saujet L, Pereira FC, Henriques AO, Martin-Verstraete I. 2014. The regulatory network controlling spore formation in *Clostridium difficile*. *FEMS Microbiol Lett.* 358(1):1–10.

Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin PV, Gelfand MS, Dupuy B, Henriques AO, Martin-Verstraete I. 2013. Genome-wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. *PLoS Genet.* 9(10):e1003756.

Schloss PD, Iverson KD, Petrosino JF, Schloss SJ. 2014. The dynamics of a family's gut microbiota reveal variations on a theme. *Microbiome* 2(1):1–13.

Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, et al. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet.* 38(7):779–786.

Serrano M, Crawshaw AD, Dembek M, Monteiro JM, Pereira FC, Pinho MG, Fairweather NF, Salgado PS, Henriques AO. 2016. The SpoIIQ-SpoIIIAH complex of *Clostridium difficile* controls forespore engulfment and late stages of gene expression and spore morphogenesis. *Mol Microbiol.* 100(1):204–228.

Serrano M, Gao J, Bota J, Bate AR, Meisner J, Eichenberger P, Moran CP Jr, Henriques AO. 2015. Dual-specificity anti-sigma factor reinforces control of cell-type specific gene expression in *Bacillus subtilis*. *PLoS Genet.* 11(4):e1005104.

Serrano M, Real G, Santos J, Carneiro J, Moran CP Jr, Henriques AO. 2011. A negative feedback loop that limits the ectopic activation of a cell type-specific sporulation sigma factor of *Bacillus subtilis*. *PLoS Genet.* 7(9):e1002220.

Setlow P. 2014. Spore resistance properties. *Microbiol Spectr.* 2(5):TBS-0003-2012.

Setlow P, Wang S, Li YQ. 2017. Germination of spores of the orders *Bacillales* and *Clostridiales*. *Annu Rev Microbiol.* 71:459–477.

Shivaji S, Chaturvedi P, Suresh K, Reddy GSN, Dutt CBS, Wainwright M, Narlikar JV, Bhargava PM. 2006. *Bacillus aerius* sp. nov., *Bacillus aerophilus* sp. nov., *Bacillus stratosphericus* sp. nov. and *Bacillus altitudinis* sp. nov., isolated from cryogenic tubes used for collecting air samples from high altitudes. *Int J Syst Evol Microbiol.* 56(7):1465–1473.

Shrestha R, Cochran AM, Sorg JA. 2019. The requirement for co-germinants during *Clostridium difficile* spore germination is influenced by mutations in yabG and cspA. *PLoS Pathog.* 15(4):e1007681.

Shuster B, Khemmani M, Abe K, Huang X, Nakaya Y, Maryn N, Buttar S, Gonzalez AN, Driks A, Sato T, et al. 2019. Contributions of crust proteins to spore surface properties in *Bacillus subtilis*. *Mol Microbiol.* 111(3):825–843.

Smits WK, Eschevins CC, Susanna KA, Bron S, Kuipers OP, Hamoen LW. 2005. Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. *Mol Microbiol.* 56(3):604–614.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Steil L, Serrano M, Henriques AO, Völker U. 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. *Microbiology (Reading, Engl).* 151(Pt 2):399–420.

Stephenson K, Lewis RJ. 2005. Molecular insights into the initiation of sporulation in Gram-positive bacteria: new technologies for an old phenomenon. *FEMS Microbiol Rev.* 29(2):281–301.

Stevens CM, Daniel R, Illing N, Errington J. 1992. Characterization of a sporulation gene, spoIVA, involved in spore coat morphogenesis in *Bacillus subtilis*. *J Bacteriol.* 174(2):586–594.

Stewart GC. 2015. The exosporium layer of bacterial spores: a connection to the environment and the infected host. *Microbiol Mol Biol Rev.* 79(4):437–457.

Takamatsu H, Kodama T, Watabe K. 1999. Assembly of the CotSA coat protein into spores requires CotS in *Bacillus subtilis*. *FEMS Microbiol Lett.* 174(1):201–206.

Touchette MH, Benito de la Puebla H, Ravichandran P, Shen A. 2019. SpoIVA-SipL complex formation is essential for *Clostridioides difficile* spore assembly. *J Bacteriol*. 201(8):e00042-19.

Tovar-Rojo F, Chander M, Setlow B, Setlow P. 2002. The products of the spoVA operon are involved in dipicolinic acid uptake into developing spores of *Bacillus subtilis*. *J Bacteriol*. 184(2):584–587.

Traag BA, Driks A, Stragier P, Bitter W, Broussard G, Hatfull G, Chu F, Adams KN, Ramakrishnan L, Losick R. 2010. Do mycobacteria produce endospores? *Proc Natl Acad Sci U S A*. 107(2):878–881.

Trouve J, Mohamed A, Leisico F, Contreras-Martel C, Liu B, Mas C, Rudner DZ, Rodrigues CDA, Morlot C. 2018. Structural characterization of the sporulation protein GerM from *Bacillus subtilis*. *J Struct Biol*. 204(3):481–490.

Ueda K, Yamashita A, Ishikawa J, Shamada M, Watsuji TO, Morimura K, Ikeda H, Hattori M, Beppu T. 2004. Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. *Nucleic Acids Res*. 32(16):4937–4944.

Urios L, Cueff V, Pignet P, Barbier G. 2004. *Tepidibacter formicigenes* sp. nov., a novel spore-forming bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent. *Int J Syst Evol Microbiol*. 54(Pt 2):439–443.

Vishnivetskaya TA, Kathariou S. 2005. Putative transposases conserved in *Exiguobacterium* isolates from ancient Siberian permafrost and from contemporary surface habitats. *Appl Environ Microbiol*. 71(11):6954–6962.

Vishnivetskaya TA, Kathariou S, Tiedje JM. 2009. The *Exiguobacterium* genus: biodiversity and biogeography. *Extremophiles* 13(3):541–555.

Vos M. 2009. Why do bacteria engage in homologous recombination? *Trends Microbiol*. 17(6):226–232.

Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. *ISME J*. 3(2):199–208.

Wakeley P, Hoa NT, Cutting S. 2000. BofC negatively regulates SpoIVB-mediated signalling in the *Bacillus subtilis* sigmaK-checkpoint. *Mol Microbiol*. 36(6):1415–1424.

Wang KH, Isidro AL, Domingues L, Eskandarian HA, McKenney PT, Drew K, Grabowski P, Chua MH, Barry SN, Guan M, et al. 2009. The coat morphogenetic protein SpoVID is necessary for spore encasement in *Bacillus subtilis*. *Mol Microbiol*. 74(3):634–649.

Wang ST, Setlow B, Conlon EM, Lyon JL, Imamura D, Sato T, Setlow P, Losick R, Eichenberger P. 2006. The forespore line of gene expression in *Bacillus subtilis*. *J Mol Biol*. 358(1):16–37.

Wang Erickson AF, Deighan P, Chen S, Barrasso K, Garcia CP, Martínez-Lumbreras S, Alfano C, Krysztofinska EM, Thapaliya A, Camp AH, et al. 2017. A novel RNA polymerase-binding protein that interacts with a sigma-factor docking site. *Mol Microbiol*. 105(4):652–662.

Waterhouse AM, Procter JB, Martin DA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191.

Yano JM, Yu K, Donaldson GP, Shastri GG, Ann P, Ma L, Nagler CR, Ismagilov RF, Mazmanian SK, Hsiao EY. 2015. Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* 161(2):264–276.

Yutin N, Galperin MY. 2013. A genomic update on clostridial phylogeny: gram-negative spore formers and other misplaced clostridia. *Environ Microbiol*. 15(10):2631–2641.

Zeytuni N, Flanagan KA, Worrall LJ, Massoni SC, Camp AH, Strynadka N. 2018. Structural and biochemical characterization of SpoIIIAF, a component of a sporulation-essential channel in *Bacillus subtilis*. *J Struct Biol*. 204(1):1–8.

Zeytuni N, Hong C, Flanagan KA, Worrall LJ, Theiltges KA, Vuckovic M, Huang RK, Massoni SC, Camp AH, Yu Z, et al. 2017. Near-atomic resolution cryoelectron microscopy structure of the 30-fold homo-oligomeric SpoIIIAG channel essential to spore formation in *Bacillus subtilis*. *Proc Natl Acad Sci U S A*. 114(34):E7073–E7081.

Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. 2012. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res*. 40:569–572.

Zheng LB, Donovan WP, Fitz-James PC, Losick R. 1988. Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. *Genes Dev*. 2(8):1047–1054.

Zhou R, Kroos L. 2004. BofA protein inhibits intramembrane proteolysis of pro-sigmaK in an intercompartmental signaling pathway during *Bacillus subtilis* sporulation. *Proc Natl Acad Sci U S A*. 101(17):6385–6390.

Zhu D, Sorg JA, Sun X. 2018. *Clostridioides difficile* biology: sporulation, germination, and corresponding therapies for *C. difficile*. *Front Cell Infect Microbiol*. 8:29.