MICROBIAL GENOMICS

RESEARCH ARTICLE

Secaira-Morocho *et al.*, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000451

Diversity and evolutionary dynamics of spore-coat proteins in spore-forming species of Bacillales

Henry Secaira-Morocho[1], José A. Castillo[1,*] and Adam Driks[2]†

Abstract

Among members of the Bacillales order, there are several species capable of forming a structure called an endospore. Endospores enable bacteria to survive under unfavourable growth conditions and germinate when environmental conditions are favourable again. Spore-coat proteins are found in a multilayered proteinaceous structure encasing the spore core and the cortex. They are involved in coat assembly, cortex synthesis and germination. Here, we aimed to determine the diversity and evolutionary processes that have influenced spore-coat genes in various spore-forming species of Bacillales using an *in silico* approach. For this, we used sequence similarity searching algorithms to determine the diversity of coat genes across 161 genomes of Bacillales. The results suggest that among Bacillales, there is a well-conserved core genome, composed mainly by morphogenetic coat proteins and spore-coat proteins involved in germination. However, some spore-coat proteins are taxa-specific. The best-conserved genes among different species may promote adaptation to changeable environmental conditions. Because most of the *Bacillus* species harbour complete or almost complete sets of spore-coat genes, we focused on this genus in greater depth. Phylogenetic reconstruction revealed eight monophyletic groups in the *Bacillus* genus, of which three are newly discovered. We estimated the selection pressures acting over spore-coat genes in these monophyletic groups using classical and modern approaches and detected horizontal gene transfer (HGT) events, which have been further confirmed by scanning the genomes to find traces of insertion sequences. Although most of the genes are under purifying selection, there are several cases with individual sites evolving under positive selection. Finally, the HGT results confirm that sporulation is an ancestral feature in *Bacillus*.

DATA SUMMARY

All supporting data and methods have been provided within the article or through supplementary data files. Five supplementary tables are available with the online version of this article. A full listing of NCBI accessions for strains used in this paper is available in Table S1 (available in the online version of this article). Biopython scripts to extract significant BLASTp hits used in this study are available at GitHub – https://github.com/HSecaira/Spore_coat_proteins_BLAST_extraction

INTRODUCTION

The Bacillales order has great taxonomic and phylogenetic diversity and can thrive in many different environments [1]. Some members of this order are present in the human and mammalian gut microbiota [2, 3], while others are pathogens that cause foodborne diseases [4] or are important human pathogens [5–7]. A striking feature of the Bacillales is the ability to form a dormant cell type called the endospore or spore [8, 9].

Spores can survive a wide range of extreme environmental conditions, such as microbial predation, desiccation, heat, UV radiation and toxic chemicals [8–12]. The metabolic dormancy of spores permits them to remain in this state for

Received 28 July 2020; Accepted 16 September 2020; Published 14 October 2020

Author affiliations: [1]School of Biological Sciences and Engineering, Yachay Tech University, San Miguel de Urcuquí, Imbabura, Ecuador; [2]Department of Microbiology and Immunology, Loyola University Chicago, Chicago, IL, USA.

***Correspondence:** José A. Castillo, jcastillo@yachaytech.edu.ec

Keywords: Bacillales; *Bacillus*; horizontal gene transfer; morphogenetic coat proteins; positive/purifying selection; Spore-coat proteins.

Abbreviations: BUSTED, Branch-Site Unrestricted Statistical Test for Episodic Diversification; dN, non-synonymous substitution; dS, synonymous substitution; GEIs, genomic islands; HGT, horizontal gene transfer; ICEs, integrative and conjugative elements; MCMC, Markov Chain Monte Carlo; MEME, mixed effects model of evolution; ML, maximum likelihood.

†Deceased June 6, 2019

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary tables are available with the online version of this article,

hundreds of years [13]. In addition, the spore can sense its surrounding environment, and when growth conditions are favourable again, it germinates to generate a vegetative form of the bacteria [13–15].

To survive stress conditions, the bacterial cell undergoes an evolutionarily conserved process called sporulation to produce the spore structure. Sporulation begins in the stationary phase when nutrients begin to be scarce [16] and culminates in a mature spore composed of two external protective structures: the cortex, assembled between the inner and outer spore membranes, and the proteinaceous coat that is subjected to cross-linking [8, 16, 17]. Genomic DNA within the spore is contained in the partially dehydrated core [16].

The bacterial spore coat is a multilayered structure formed by specialized proteins. The endospore confers protection against adverse environmental conditions and contributes to spore environmental interactions, which may lead to germination to resume metabolic activity and growth [16, 17]. There is a high diversity in spore-coat morphologies among spore-forming species [8, 16]. *Bacillus subtilis* has been a major model organism to study spore-coat proteins using different approaches that include using transmission electron microscopy as well as biochemical and genetic tools [8]. The most internal layer of the spore coat is called the basement layer, which contains the proteins necessary for initiating coat assembly (SpoIVA, SpoVM, SpoVID) [8, 16]. The basement layer is followed by the inner layer, the outer coat and the crust [8]. Fig. 1 shows the positions of the four layers of the *B. subtilis* coat. Other spore-forming species, such as *Bacillus anthracis, Bacillus thuringiensis* and *Bacillus cereus* also possess an exosporium [8, 16, 18], the outermost layer that surrounds the mature spore. It is composed of fine hair-like projections that may be involved in infections by *B. anthracis* [19].

Impact Statement

Species of Bacillales can form a highly resistant cell type, called a spore, under extreme environmental conditions. The spore is surrounded by a proteinaceous coat that mediates interactions with its environment. Spore-coat synthesis, assembly, maturation and spore germination is a complex multiprotein process in which more than 80 different proteins participate. This work provides unique insight into spore-coat protein functions and occurrence during early and later stages of coat synthesis, assembly and spore germination of the most significant spore-forming Bacillales. Similarly, at *the Bacillus* genus level, a large proportion of coat genes are under positive diversifying selection and/or balancing selection, suggesting high genetic diversity that may confer unique adaptation to ensure spore survival and efficient germination. These results demonstrate the value of comparative genomics to understand evolutionary changes among spore-coat proteins, helping to identify the most conserved or common among Bacillales, as well as the selective pressures working on coat genes that allow Bacillus species-particular interactions with the surrounding environment.

Spore-coat synthesis, assembly and maturation is a complex process involving multiple proteins and requiring several

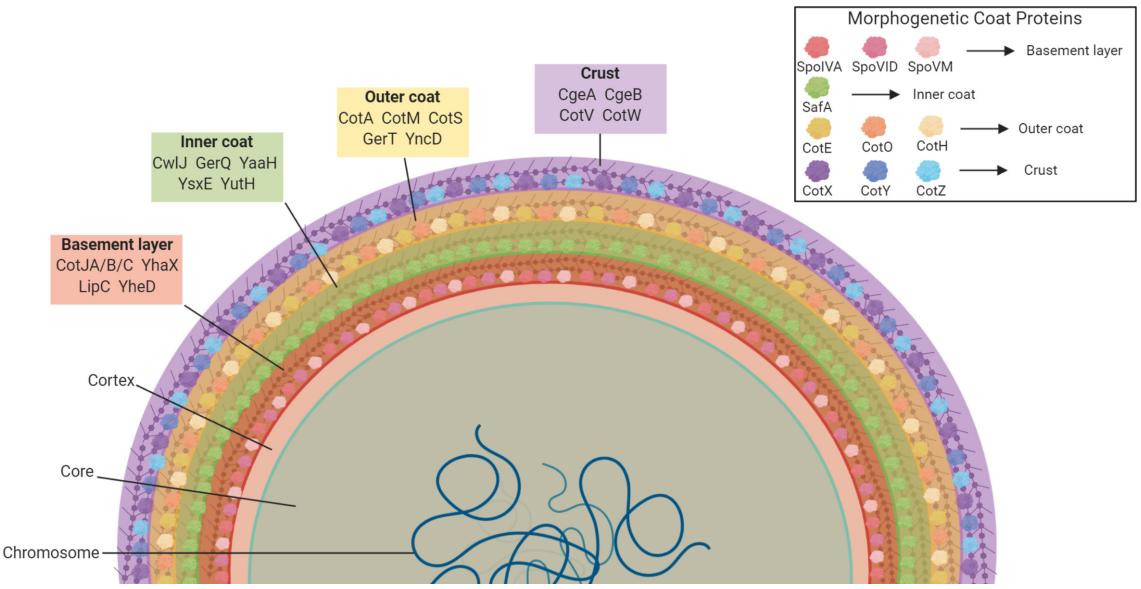

Fig. 1. Model of spore-coat structure. Assembly of each layer depend on the multimerization of a morphogenetic coat protein and its dependent individual coat proteins. Four layers with its morphogenetic and morphogenetic-dependent coat proteins are shown: basement layer (red), inner layer (green), outer layer (yellow) and crust (purple).

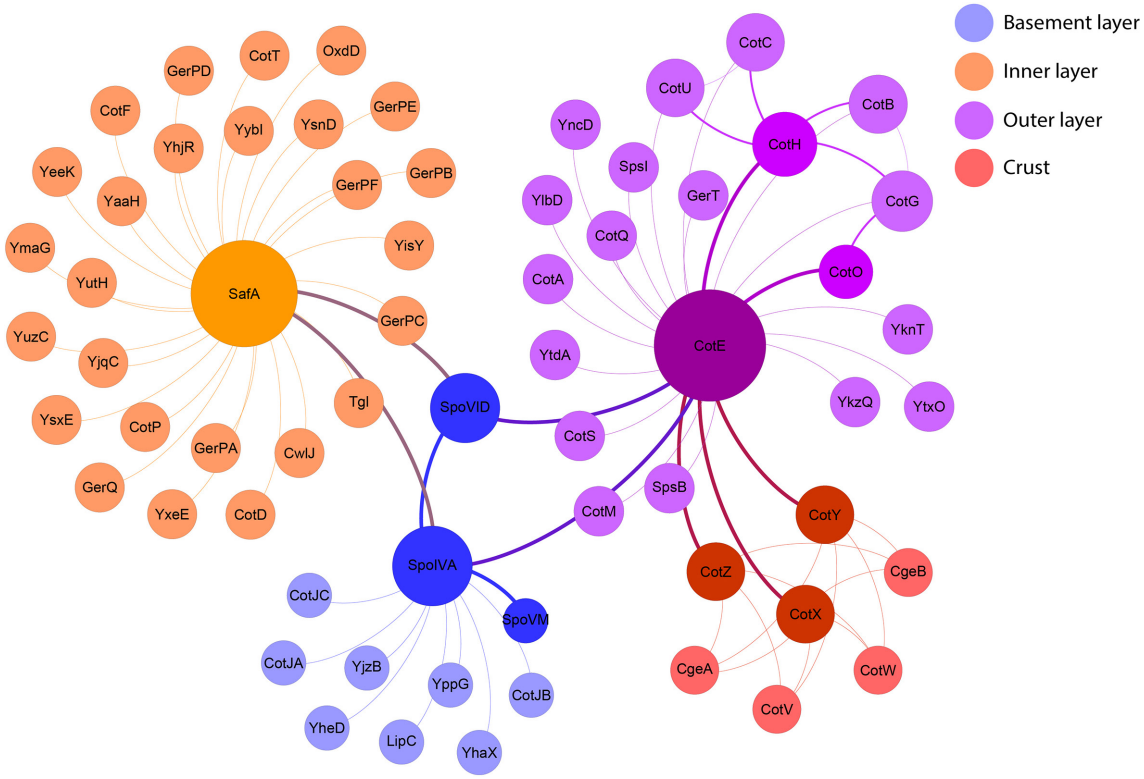

Fig. 2. Spore-coat protein interaction network in *Bacillus subtilis*. Morphogenetic and morphogenetic-dependent coat proteins interact with each other to form the four layers (basement layer, inner layer, outer layer, crust) of the spore coat. Recruitment of the morphogenetic coat proteins SafA and CotE depend on SpoIVA, whereas recruitment of CotO and CotX/Y/Z depend on CotE, the interaction network is highly hierarchical.

hours to complete [8]. Assembly of coat layers depends on morphogenetic coat proteins, such as SpoIVA, SpoVM, SpoVID, SafA, CotE, CotH, CotO, CotX, CotY, CotZ, as well as coat proteins that are dependent on these morphogenetic proteins [8, 16]. SpoIVA and SpoVM are required for spore-cortex formation, coat assembly, anchoring of the coat to the spore surface and spore encasement, whereas SpoVID is necessary for spore encasement [8, 16, 20]. CotE is the most critical protein for the assembly of the outer coat, and SafA is responsible for the assembly of the inner coat [8, 16, 21–23]. Several studies demonstrated the existence of a network of genetic interactions that consist of three independent modules: SpoIVA-dependent subnetwork, CotE-dependent subnetwork and SafA-dependent subnetwork [8, 24, 25], as shown in Fig. 2.

Despite the existence of more than 80 different spore-coat proteins, studies have demonstrated that not all of them are required for coat synthesis, assembly, maturation and spore germination [8, 16, 26, 27]. Indeed, most coat gene mutations are phenotypically silent or insignificant, except for the morphogenetic coat proteins that control the assembly of other coat proteins [8]. Similarly, external conditions, such as sporulation temperature, can affect the abundance, stability and proper function of morphogenetic and its dependent coat proteins, thus changing the structure and properties of the coat [28]. In this work, we wish to infer which coat proteins play a key role in spore-coat synthesis, assembly, maturation and environmental interactions that may promote spore germination and/or spore survival in endospore-forming species of Bacillales. We also seek to determine whether some coat proteins are better conserved within a given taxon. Likewise, we wanted to document any pattern of coat gene conservation that might indicate niche-specific adaptation, so we could discriminate among members of specific taxa that share coat proteins adapted to specific niches. Additionally, we focused on an evolutionary analysis of *Bacillus,* since in this genus we found the most complete set of spore-coat genes related to those found in our reference genome of *B. subtilis*. First, we aimed to define monophyletic groups inside the *Bacillus* genus. Using this information, we estimated the selective pressures and evolutionary histories acting upon the morphogenetic spore-coat proteins in each monophyletic group.

METHODS

Sequence data and spore-coat-protein diversity analyses

Based on a thorough literature review as of January 2019, we identified 86 genes that encode spore-coat proteins or proteins

related to sporulation or germination process in *B. subtilis*, see Table 1. Each gene sequence was downloaded from the SubtiWiki server (http://subtiwiki.uni-goettingen.de/) [29]. In parallel, 161 annotated genomes of the Bacillales order were retrieved from NCBI's FTP server (https://www.ncbi.nlm.nih.gov/genome/microbes/). This dataset is composed of 60 genomes of the genus *Bacillus* and 101 genomes of non-*Bacillus* genera, representing the greatest diversity of spore-forming genera of Bacillales known so far, see Table S1.

We employed three different strategies to determine the presence/absence of spore-coat proteins in the selected Bacillales genomes:

(1) Local BLASTp was used to search for the 86 spore-coat protein homologues in the collection of Bacillales (*Bacillus* and non-*Bacillus*) genomes. For this, we created genome databases for all the 161 genomes of Bacillales and searched for all coat proteins in these databases. We considered all hits with a Bit score ≥40 and *E*-value <0.001 as positive since these values are significant in searches of protein databases with fewer than 7000 entries [30], which occurs in Bacillales genomes that have less than 7000 different proteins.
(2) Clustering analysis of spore-coat proteins was performed using the software package Many-against-Many sequence searching (MMseqs2) [31] to group proteins from the 161 Bacillales genomes with well-known spore-coat proteins (i.e. the 86 spore-coat proteins mentioned above) with a minimum of identity and coverage of 50 and 99%, respectively.
(3) KEGG Orthology database [32] was used to search for spore-coat gene orthologues across the Bacillales genomes of Table S1.

B. subtilis subsp. subtilis 168 was used as a control, since it has most of the spore-coat proteins described so far. Therefore, it is a model organism used to study the structure and functions of the coat. The asporogenous species *Bacillus beveridgei* MLTeJB and *Exiguobacterium antarcticum* B7 [33, 34] were used as negative controls. Genes with positive hits for the three methods (BLASTp, Clustering, KEGG Orthology) were recorded as highly significant and deemed as confirming of particular genes within the subject genomes. On the other hand, genes with hits for one or two methods were accepted as secondarily significant. A consensus heat map that summarizes the results provided by the three methods was created using the Seaborn data visualization library implemented in Python.

Phylogenetic reconstruction and monophyly testing

We reconstructed the phylogeny of 60 genomes of *Bacillus* using maximum-likelihood (ML) and Bayesian methods. The core protein sequences of *Bacillus* genomes were extracted using the pangenomics pipeline BPGA [35] to create an aligned sequence of 15539 amino acids. The optimal substitution model for core-protein sequences, as suggested by the SMS online server [36], was LG+Γ+I. Tree reconstruction using ML was completed in PhyML v3.0 [37] using the subtree pruning and regrafting algorithm for tree improvement and approximate likelihood ratio test (aLRT) and Shimodaira–Hasegawa to measure branch supports. Tree visualization was achieved using FigTree (Rambaut A, http://tree.bio.ed.ac.uk/software/figtree/).

Tree inference with the Bayesian method was performed using the software package BEAST v1.10.4 [38]. Initially, we performed model selection for demographic and molecular clock parameters, calculating the marginal likelihood by two approaches: 'path sampling' [39] and 'stepping-stone sampling' [40]. The marginal likelihood estimation was specified with a chain length of 150000, saving log parameters every 1000 steps and using 100 number of path steps. These two-model selection approaches allowed us to define that the Bayesian skyline plot (BSP) and strict clock are the best models for this population. Although most priors were left default, we modified the settings of the following particular priors: treeModel.rootHeight, tmrca and skyline.popSize to lognormal with mu=1.0 and sigma=1.0. We ran the Markov chains, starting from random trees for 15 million generations and sampled every 2000th generation. MCMC convergence was examined using Tracer v.1.7 [41] to ensure that the calculation had run long enough to attain stationarity.

We tested to see whether the internal phylogenetic clusters are monophyletic in the *Bacillus* tree. For this, we enforced some subpopulations of *Bacillus* (see Table S1 for strain details) to be monophyletic. This constrains the tree topology so that the *Bacillus* clustering is kept monophyletic during the course of the MCMC analysis. We used this strategy to test the following clusters: Cereus group (*B. anthracis B. bombysepticus, B. cereus, B. cytotoxicus, B. mobilis, B. mycoides, B. pseudomycoides, B. thuringiensis, B. toyonensis, B. wiedmannii, B. weihenstephanensis*); Subtilis group (*B. amyloliquefaciens, B. siamensis, B. velezensis, B. atrophaeus, B. licheniformis, B. halotolerans, B. paralicheniformis, B. sonorensis, B. subtilis, B. vallismortis, B. gibsonii, B. intestinalis, B. glycinifermentans*); Pumilus group (*B. altitudinis, B. pumilus, B. safensis, B. xiamenensis); Simplex group (B. simplex, B. butanolivorans, B. asahii, B. muralis*); Methanolicus group (*B. methanolicus, B. foraminis, B. jeotgali, B. circulans, B. infantis, B. kochii, B. oceanisediminis*); Coagulans group (*B. freudenreichii, B. lentus, B. smithii, B. thermoamylovorans, B. coagulans*); Megaterium group (*B. megaterium, B. aryabhattai, B. flexus, B. endophyticus*); Halodurans group (*B. cellulosilyticus, B. clausii, B. lehensis, B. halodurans, B. krulwichiae, B. pseudofirmus, B. beveridgei*). We used the Subtilis group as a positive control since it is a well-known internal group in the Bacillus genus and randomly selected Bacillus species belonging to different groups as a negative control (*B. cellulosilyticus*, *B. circulans, B. clausii, B. cytotoxicus, B. gibsonii, B. licheniformis, B. mycoides, B. safensis, B. weihenstephanensis, B. wiedmannii*) and included them in the pipeline for monophyly testing. We compared the tree topology of two competing models: constrained trees for the above-described clusters versus the unconstrained tree. All trees were inferred using the same settings except the enforcement for monophyly. We examined the support for the different topologies using Bayes factors [42]. For this, we

Table 1. Eighty-six spore-coat genes and their location in the genome of the model organism *B. subtilis* 168

Spore coat gene	Locus Tag	Location	Function	Domain*	References
cgeA	BSU_19780	Crust	Maturation of the outermost layer of the spore	ND	[8]
cgeB	BSU_19790	Crust	Maturation of the outermost layer of the spore	DUF3880† Glycosyl transferases group 1	[8]
cgeC	BSU_19770	ND	Maturation of the outermost layer of the spore	ND	[8]
cgeD	BSU_19760	ND	Maturation of the outermost layer of the spore	Glycosyl transferase family 2	[8]
cgeE	BSU_19750	ND	Maturation of the outermost layer of the spore	Acetyltransferase (GNAT)	[8]
cotA	BSU_06300	Outer layer	Spore pigmentation Spore resistance	Multicopper oxidase	[8]
cotB	BSU_36050	Outer layer	Spore resistance	ND	[8, 26]
cotC	BSU_17700	Outer layer	Spore resistance	ND	[8, 26]
cotD	BSU_22200	Inner layer	Spore resistance	Inner spore coat protein D	[8, 26]
cotE	BSU_17030	Outer layer	Assembly of the outer layer	Outer spore coat protein E	[8, 26]
cotF	BSU_40530	Inner layer	Spore resistance	Coat F	[8, 26, 110]
cotG	BSU_36070	Outer layer	Spore resistance	ND	[8]
cotH	BSU_36060	Outer layer	Assembly of the outer layer	CotH kinase protein	[8, 26]
cotI	BSU_30920	ND	Bacterial spore kinase Spore envelope	Phosphotransferase enzyme	[8 26]
cotJA	BSU_06890	Basement layer	ND	Spore coat associated protein JA	[8 26 110]
cotJB	BSU_06900	Basement layer	ND	CotJB protein	[8 26, 110]
cotJC	BSU_06910	Basement layer	Protection against oxidative estress	Manganese containing catalase	[8 26, 110]
cotM	BSU_17970	Outer layer	Spore resistance	ND	[8 26]
cotO	BSU_11730	Outer layer	Assembly of the outer and crust layers	Spore coat protein CotO	[8, 26 89]
cotP	BSU_05550	Inner layer	Spore resistance	Hsp20/alpha crystallin family	[8 26]
cotQ	BSU_34520	Outer layer	Spore protection	ND	[8]
cotR	BSU_34530	ND	Spore lipolytic enzyme Hydrolysis of lysophospholipids	Patatin-like phospholipase	[8]
cotS	BSU_30900	Outer layer	Bacterial spore kinase Spore resistance	ND	[8 26]
cotSA	BSU_30910	ND	Transfer of glycosyl groups	Glycosyl transferases group 1, 4	[8 26]
cotT	BSU_12090	Inner layer	Spore resistance	ND	[8]
cotU	BSU_17670	Outer layer	Spore resistance	ND	[8 26]
cotV	BSU_11780	Crust	Spore resistance	Spore Coat Protein X and V	[8]
cotW	BSU_11770	Crust	Spore resistance	ND	[8]
cotX	BSU_11760	Crust	Assembly of the crust	Spore Coat Protein X and V	[8]
cotY	BSU_11750	Crust	Assembly of the crust	Spore coat protein Z	[8 26]
cotZ	BSU_11740	Crust	Assembly of the crust	Spore coat protein Z	[8 26]

Continued

Table 1. Continued

Spore coat gene	Locus Tag	Location	Function	Domain*	References
cwlJ	BSU_02600	Inner layer	Spore cortex lytic enzyme	Cell Wall Hydrolase	[8]
gerPA	BSU_10720	Inner layer	Germination	Spore germination protein gerPA/gerPF	[8]
gerPB	BSU_10710	Inner layer	Germination	Spore germination GerPB	[8]
gerPC	BSU_10700	Inner layer	Germination	Spore germination protein GerPC	[8]
gerPD	BSU_10690	Inner layer	Germination	ND	[8]
gerPE	BSU_10680	Inner layer	Germination	Spore germination protein GerPE	[8]
gerPF	BSU_10670	Inner layer	Germination	Spore germination protein gerPA/gerPF	[8]
gerQ	BSU_37920	Inner layer	Germination CwlJ inhibitor	Spore coat protein GerQ	[8]
gerT	BSU_19490	Outer layer	Germination	ND	[8]
lipC	BSU_04110	Basement layer	Spore lipolytic enzyme	GDSL-like Lipase/Acylhydrolase family	[8 26]
oxdD	BSU_18670	Inner layer	Protection against toxic compounds	Cupin	[8]
safA	BSU_27840	Inner layer	Assembly of the inner layer	LysM	[8 26]
spoIVA	BSU_22800	ND	Spore cortex formation, coat assembly and anchoring	Stage IV sporulation protein A	[8 26]
spoVID	BSU_28110	ND	Spore encasement	LysM	[8 26]
spoVM	BSU_15810	ND	Spore cortex formation, coat assembly, spore encasement	Stage V sporulation protein family	[8 26]
spsB	BSU_37900	Outer layer	Spore polysaccharide synthesis	CDP-Glycerol:Poly (glycerophosphate) glycerophosphotransferase	[8]
spsI	BSU_37810	Outer layer	Spore polysaccharide synthesis	Nucleotidyl transferase	[8]
sscA	BSU_09958	ND	Spore assembly	ND	[8]
tasA	BSU_24620	ND	ND	Camelysin metallo-endopeptidase	[26]
tgl	BSU_31270	Inner layer	Introduction of cross-links in the coat for GerQ and SafA	ND	[8 26]
yaaH	BSU_00160	Inner layer	N-Acetylglucosaminidase Survival of ethanol stress	Glycosyl hydrolases family 18	[8 26]
ydgA	BSU_05560	ND	ND	Spore germination protein gerPA/gerPF	[8 26]
ydgB	BSU_05570	ND	ND	Spore germination protein gerPA/gerPF	[8 26]
ydhD	BSU_05710	ND	Glycosylase	Glycosyl hydrolases family 18	[8 26]
yhaX	BSU_09830	Basement layer	Spore protection	Haloacid dehalogenase-like hydrolase	[8 26]
yhbB	BSU_08920	ND	ND	Putative amidase	[8 26]
yhcQ	BSU_09180	ND	ND	Coat F	[26]
yheC	BSU_09780	ND	ND	YheC/D like ATP-grasp	[8]
yheD	BSU_09770	Basement layer	Spore protection	YheC/D like ATP-grasp	[8]
yhjQ	BSU_10600	ND	Prevention of copper toxicity	DUF326†	[8 26]
yhjR	BSU_10610	Inner layer	Spore protection	Rubrerythrin	[8 26]
yisY	BSU_10900	Inner layer	Spore protection	Alpha/beta hydrolase fold	[8, 26 110]

Continued

Table 1. Continued

Spore coat gene	Locus Tag	Location	Function	Domain*	References
yjqC	BSU_12490	Inner layer	Protection against oxidative stress	Manganese containing catalase	[8 26]
yjzB	BSU_11320	Basement layer	Spore protection	ND	[8]
yknT	BSU_14250	Outer layer	Spore protection	ND	[8 26]
ykvP	BSU_13780	ND	ND	Glycosyl transferases group 1	[8]
ykvQ	BSU_13790	ND	Glycosylase	Glycosyl hydrolases family 18	[8]
ykzQ	BSU_13789	Outer layer	ND	LysM	[8]
ylbD	BSU_14970	Outer layer	Spore protection	Putative coat protein	[26]
ymaG	BSU_17310	Inner layer	Spore protection	ND	[8 26]
yncD	BSU_17640	Outer layer	Conversion of L-Ala to D-Ala Spore protection	Alanine racemase	[8 26]
yppG	BSU_22250	Basement layer	Spore protection	YppG-like protein	[8 26]
yraD	BSU_26990	ND	ND	Coat F	[26 110]
yraF	BSU_26960	ND	ND	Coat F	[26 110]
yraG	BSU_26950	ND	ND	ND	[110]
ysnD	BSU_28320	Inner layer	Spore protection	ND	[8]
ysxE	BSU_28100	Inner layer	Bacterial spore kinase Spore protection	ND	[8 26]
ytdA	BSU_30850	Outer layer	Spore polysaccharide synthesis	Nucleotidyl transferase	[8]
ytxO	BSU_30890	Outer layer	Spore protection	ND	[8]
yutH	BSU_32270	Inner layer	Bacterial spore kinase Spore protection	ND	[8 26]
yuzC	BSU_31730	Inner layer	Spore protection	ND	[8]
ywrJ	BSU_36040	ND	ND	ND	[8 26]
yxeE	BSU_39580	Inner layer	Spore protection	ND	[8 26]
yybI	BSU_40630	Inner layer	Spore protection	ND	[8]
yeeK	BSU_06850	Inner layer	Spore protection	ND	[8 26]

ND, no data available.
*Pfam database.
†Domain of unknown function.

performed a path sampling and stepping-stone run of 150000 generations (100 steps log-likelihood sampled every 1000) from which we obtained a marginal likelihood estimate. The Bayes factor was estimated following this formula BF=ML1/ML2, where ML1 and ML2 are marginal likelihood values of unconstrained and constrained for monophyly, respectively.

Selection pressure and statistical analyses

Based on the presence/absence results of spore-coat proteins on Bacillales, we used local BLASTP to retrieve full-length spore-coat gene sequences using Biopython modules [43] from the 60 *Bacillus* species genomes. Thus, we created gene datasets (Table S2) that contained all spore-coat genes sequences for each *Bacillus* monophyletic group. Then, we carefully aligned the spore-coat genes datasets using the TranslatorX server (http://translatorx.co.uk/) [44] with MAFFT aligner and default settings.

We then applied the allele frequency summary statistic Tajima's D to detect selection pressures acting upon spore-coat genes within the different *Bacillus* groups. For this, we employed the DNASP v6.12 software [45] with nucleotide substitutions considered as segregating sites. Since DNASP requires a minimum of four aligned gene sequences to

calculate Tajima's D, spore-coat gene datasets with less than four sequences were not taken into account. Tajima's D is used to test any deviation from the standard neutral hypothesis by comparing the number of polymorphic sites observed in a set of sequences [46, 47]. Tajima's D positive values may reflect genes with an excess of common alleles that correspond to balancing selection [48]. On the other side, negative values may reflect genes with an excess of low-frequency variation, that is selective sweep and/or positive selection [46].

We used the DataMonkey webserver (http://test.datamonkey.org/), which implements the 'Branch-Site Unrestricted Statistical Test for Episodic Diversification' (BUSTED) method that is useful for detecting gene-wide positive selection by calculating the ratio (ω) of non-synonymous (d*N*) to synonymous (d*S*) on branches of the phylogeny at a gene level [49]. We also used the 'mixed effects model of evolution' (MEME) method to test whether individual sites in a proportion of branches have evolved under episodic positive selection [50]. We selected all branches of the phylogeny for the analyses.

We employed CODEML that is part of the PAML package to calculate ω (d*N*/d*S*) across spore-coat gene sequences [51] [52]. To provide the phylogeny required by CODEML, we used the PhyML programme [37] as stated above. The aligned gene sequences and phylogenetic trees were then used in CODEML. For this analysis, site and branch models were used with default settings and 'codons' as the sequence type. In the site model, we tested each gene sequence for the following nested models 'M1 nearly neutral' (ω <1; ω=1) [53, 54], 'M2 positive selection' (ω <1; ω=1; ω >1) [53, 54] and 'M7 β distribution' (ω <1; ω=1) [55], 'M8 β distribution +positive selection' (ω <1; ω=1; ω >1) [55], and we performed a 'likelihood ratio test' (LRT) to select the model that best fits the given data. Values of ω <1,=1, and >1 represent purifying, neutral, and positive selection, respectively [51] [52]. A *P*-value <0.05 was considered to validate a result as significant.

Horizontal gene transfer (HGT) analyses

To search for HGT events in sporecoat genes, we employed the software Notung v2.9 [56] that reconciles a gene tree with a species tree to infer duplication-transfer-loss (DTL) event models with a parsimony-based optimization criterion [57]. Notung analyses all event histories for temporal feasibility. We selected the 'Prefix of the gene label' option to reconcile the trees.

To infer DTL event models, Notung requires rooted trees. For this, we employed the software package BEAST v1.10.4 [38] to reconstruct the phylogeny for each spore-coat gene. The best-fit model of nucleotide substitution was inferred using the webserver SMS (http://www.atgc-montpellier.fr/sms/) [36] with a likelihood-based criterion (AIC) for spore-coat genes. The phylogenetic reconstruction was set up to a strict molecular clock and a Coalescent Bayesian Skyline tree prior. Analyses were run for 10 million and 1000 as echo state. We employed Tracer v1.7 [41] to assess the effective sample size (ESS) values of the MCMC chains produced by BEAST, and to confirm that the analysis reached a convergence. Furthermore, TreeAnnotator v1.8.4 was employed to generate a maximum clade credibility tree that summarizes the information of sampled trees produced by BEAST. For the species tree, we used the tree reconstructed using core amino acid sequences as explained above.

Notung HGT results were visualized as a donor-recipient network using Gephi v0.9.2 [58]. For this, we created 'edge tables' that contained the recipient and donor information. Then, each graph was set without edge direction (i.e. undirected) and displayed using the Force Atlas 2 algorithm with scaling=20000, stronger gravity, overlap prevention and node size ranked by the number of node connections (i.e. number of HGT events).

In order to reduce false positives, we scanned the genomes of the possible candidates of HGT events for traces of integrative, conjugative and mobile elements, based on the results provided by Notung. For this, we downloaded a region of the genome of approximately ten genes upstream and downstream from the spore-coat gene subjected to HGT from the NCBI's FTP server. Then, we used the detection tool 'WU-BLAST2 search' of the web server ICEberg 2.0 [http://db-mml.sjtu.edu.cn/ICEberg/, which is a database containing information about bacterial integrative and conjugative elements (ICEs), as well as integrative and mobilizable elements (IMEs), and *cis*-mobilizable elements (CIMEs)] [59]. Furthermore, we employed the Genomic Island Prediction Software v1.1.2 (GIPSy) [60] to detect if spore-coat genes under HGT events were present on genomic islands (GEIs). For this, we analysed each *Bacillus* genome against the most representative genome within each *Bacillus* group. Hits with an *E*-value less than 0.001 and a Bit score higher than 40 were considered as valid [30].

RESULTS

Spore-coat-protein diversity across Bacillales

In order to understand the diversity of spore-coat proteins on Bacillales, we carried out three distinct methods (BLAST, KEEG Orthology and Clustering) to identify the possible existence of 86 *B. subtilis* 168 spore-coat-protein homologues and related proteins within 161 genomes of Bacillales.

Figs. 3 and 4 show which spore-coat protein homologues are present or absent across *Bacillus* and other spore-forming non-*Bacillus* species, respectively. The spore-coat proteins CotE, CotJA, CotJB, CotJC, CotR, CotSA, CwlJ, GerQ, SpoIVA, SpoVID, SpoVM and YhbB, originally found in *B. subtilis* are nearly ubiquitous among the Bacillales genomes analysed in this work. Other spore-coat proteins (GerPA, GerPB, GerPC, GerPD, GerPE and GerPF) are present in *Alkalibacillus haloalkaliphilus*, *Amphibacillus*, *Geobacillus* and *Gracibacillus, Halalkalibacillus halophilus*, *Halobacillus, Paenibacillus beijingensis, Ornithinibacillus halophilus,* some *Paenibacillus, Paraliobacillus, Paucisalibacillus globulus, Piscibacillus halophilus, Pontibacillus, Tenuibacillus multivorans, Thalassobacillus, Tuberibacillus,* some *Virgibacillus* and *Vulcanibacillus modesticaldus* (see

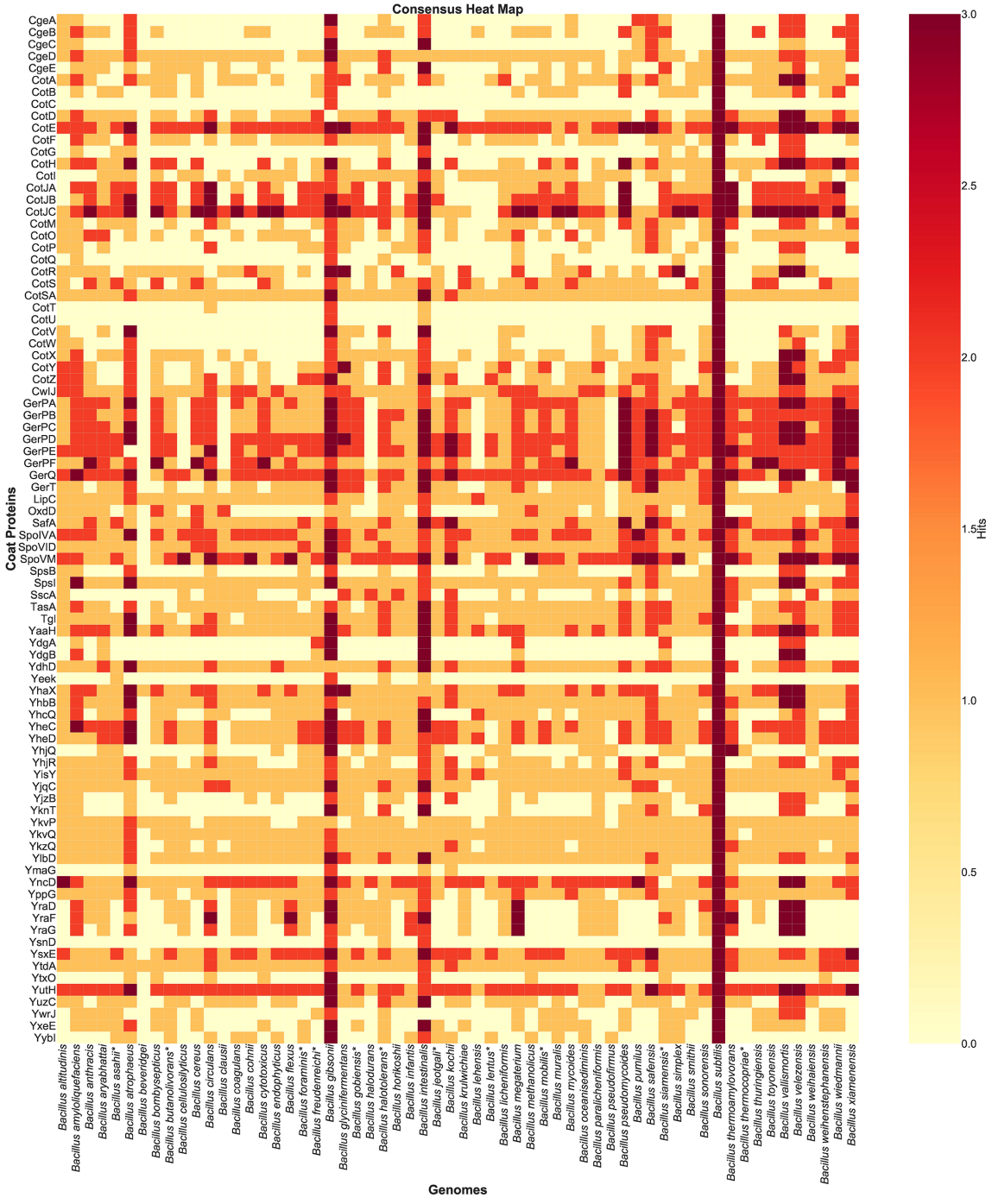

Fig. 3. Consolidated heat map of 86 spore-coat-protein homologues over 60 genomes of *Bacillus* based on three methods: BLASTp, Clustering and KEGG Orthology. Primarily significant results (dark red) have been confirmed by the three methods, whereas secondarily significant results (orange and yellow) have been confirmed by either one or two methods. *Species and proteins are missing in the KEGG database.

Fig. 4). Overall, non-*Bacillus* species contain the secondarily significant spore-coat-protein homologues (see Methods for classification of significance) CgeD, CotH, CotR, CotSA, LipC, SpsI, YaaH, YdhD, YhaX, YhcQ, YheC, YheD, YisY, YjqC, YkvP, YkvQ, YkzQ, YlbD, YncD and YtdA. Other spore-coat proteins seem to be taxa-specific, such as CgeB among the *Paenibacillus* genus or the *Geobacillus* genus that contain the spore-coat proteins CotD, CotF, TasA, YppG, YraD, YraF, YraG, YutH and YuzC (see Fig. 4).

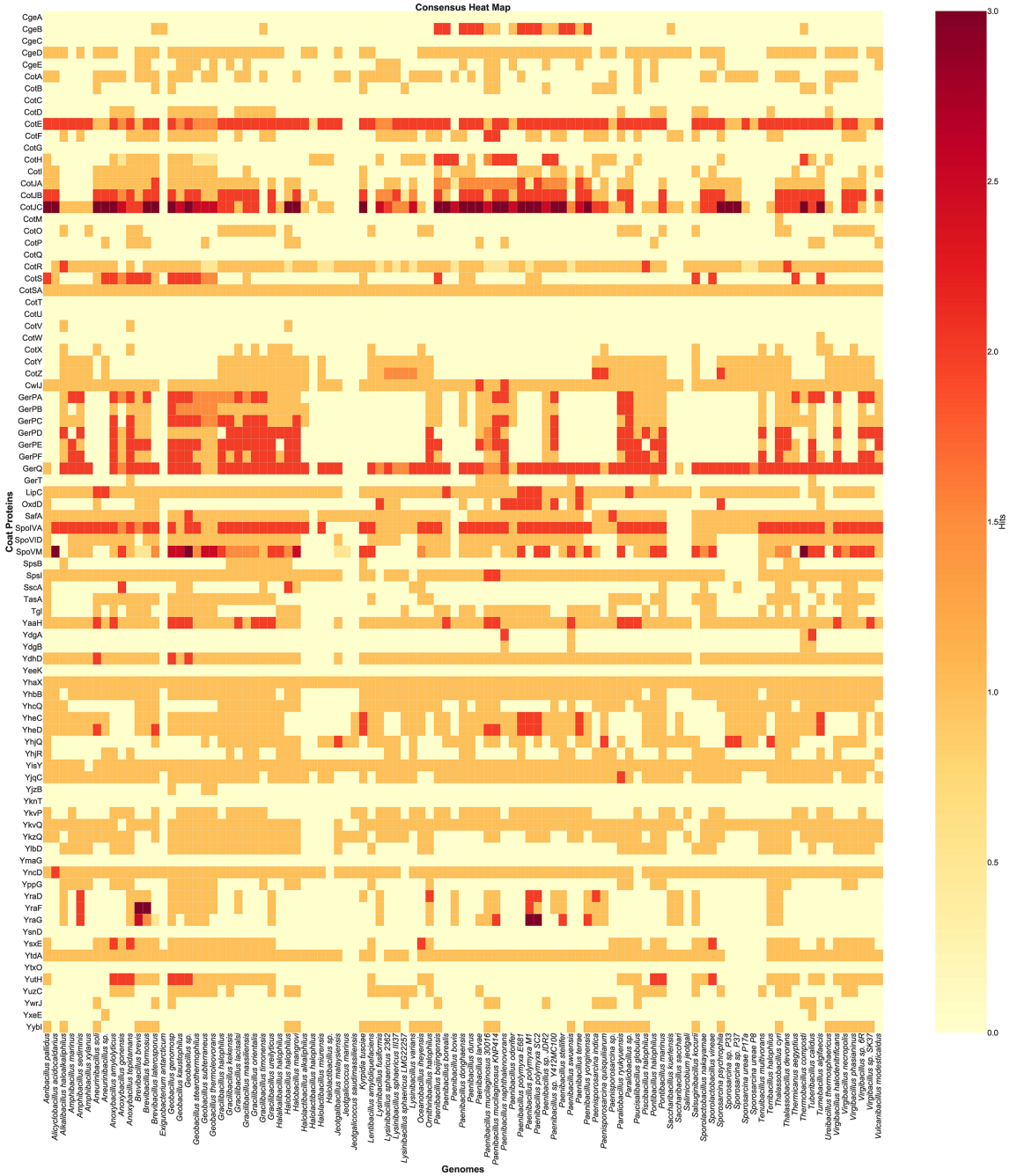

Fig. 4. Consolidated heat map of 86 spore-coat-protein homologsue over 101 genomes of non-*Bacillus* based on three methods: BLASTp, Clustering and KEGG Orthology. Primarily significant results (dark red) have been confirmed by the three methods, whereas secondarily significant results (orange and yellow) have been confirmed by either one or two methods.

The spore-coat proteins CgeA, CgeB, CgeC, CotC, CotG, CotM, CotQ, CotT, CotU, CotV, CotW, CotX, GerT, YdgA, YdgB, YeeK, YjzB, YknT, YmaG, YsnD, YtxO, YwrJ and YxeE are poorly represented in the genomes of Bacillales other than *B. subtilis* and *B. gibsonii* (see Figs. 3 and 4). For instance, it has been previously reported that CotG is not highly conserved across the *Bacillus* genus, although its role may be carried out by a non-homologous CotG-like protein that has similar structural regions to CotG [61]. Therefore, we do not rule out the possibility that non-homologous coat-like proteins with similar structural and chemical features may perform the role of poorly conserved coat proteins. As expected, *Halolactibacillus* and *Jeotgalicoccus* genomes contain few spore-coat-protein homologues, since they are non-spore-forming species [62, 63]. *Bacillus beveridgei* and *Exiguobacterium antarcticum* also do not

have spore-coat-protein homologues, as outlined by our criteria.

Since most of the *Bacillus* species harbour many coat proteins, we focused on the study of the evolutionary dynamics of these proteins in the *Bacillus* genus. To achieve this goal, we first carried out a phylogenetic analysis to test the monophyly and delimitate internal groups in *Bacillus*. This analysis allowed us to distinguish between internal monophyletic groups that were already described and new ones (see below for further details). Subsequently, we performed an analysis of presence/absence of spore-coat proteins homologous proteins at the level of each phylogenetic group within the *Bacillus* genus. Results show that the Subtilis group possesses the most conserved spore-coat proteins (morphogenetic coat proteins, basement layer, inner layer, outer layer, crust) compared to other *Bacillus* groups and non-*Bacillus* spore-forming species. CotC, CotU (outer layer) and CotT (inner layer) are only present in *B. subtilis* and *B. gibsonii*. Other spore-coat proteins, such as CotI, CotR, CotSA, YdhD, YhbB, YheC, YkvP, YkvQ and TasA, whose localization has not yet been determined, are widely distributed among members of the Subtilis group (see Fig. 3).

Our results reveal that morphogenetic spore-coat proteins (CotE, CotH, CotO, CotY, CotZ, SafA, SpoIVA, SpoVID and SpoVM) in the Cereus group are highly conserved. An exception is CotX, which is involved in the assembly of the crust [8]. Since coat assembly is a highly hierarchical process [8], other morphogenetic proteins present with the same role, such as CotY and CotZ, may take over the task to compensate for the absence of CotX. Nevertheless, the proteins CgeA, CotV, CotW that are part of the crust in *B. subtilis* are absent. Other spore-coat proteins (CotF, CotP, CotU, YmaG, YsnD, YuzC, YybI and YeeK) that are part of the inner layer are absent as well. Moreover, several spore-coat proteins present in the outer layer are absent despite the presence of the morphogenetic coat proteins, SpoIVA and CotE (see Fig. 3).

In the Simplex group, several spore-coat-protein homologues of the crust, inner layer and outer layer are absent. This is not surprising given the absence of the morphogenetic coat proteins CotO, CotY, CotZ that control those processes [8]. Despite the absence of some spore-coat proteins of the outer and inner layer, in the Pumilus group, the great majority of spore-coat proteins and all the morphogenetic coat proteins are present, including those of the crust. Thus, a proper assembly of the spore coat is highly conserved in this group, which is beneficial for the high spore resistance previously reported [64]. In the Methanolicus group, the morphogenetic coat proteins CotO, CotH, CotX, and other spore-coat proteins of the crust, inner and outer layer are absent (see Fig. 3).

Homologues of *B. subtilis*' morphogenetic coat proteins CotH, CotX, CotO and CotZ that are responsible for the assembly of the outer layer and the crust are absent in the Coagulans group. Similarly, the Megaterium group does not have detectable protein homologues for CotX, CotY and CotZ. As expected, several spore-coat proteins of the outer layer dependent on CotH and CotO and proteins dependent on CotX, CotY and CotZ are also absent. Thus, the crust may be absent in both groups or possibly it is composed of different proteins, as the case of *Bacillus megaterium* that possesses an exosporium as the outermost layer of the coat [17]. However, the strain *B. megaterium* QM B1551 has an exosporium composed of plasmid-borne orthologues of *B. subtilis cotW* and *cotX* genes [65]. Further studies are needed to clarify these possibilities. The Halodurans group contains a lower number of coat-protein homologues compared to other *Bacillus* monophyletic groups described here. Except for CotE, this group does not harbour the morphogenetic coat proteins responsible for the assembly of the outer coat and the crust. Hence, as expected, several spore-coat-protein homologues dependent on those morphogenetic proteins are also absent (see Fig. 3).

Monophyletic analyses

We carried out a phylogenetic analysis to test the monophyly and delimitate internal groups in *Bacillus*. For this purpose, we used a phylogenomics approach that included 60 different *Bacillus* species. The reconstructed tree allowed us to distinguish eight internal groups, many of which were already known (i.e. Subtilis group, Cereus group), but others were not described, so we named them according to the dominant species in each group (Coagulans group, Megaterium group and Methanolicus group, Fig. 5). For hypothesis testing, we enforced the internal group under analysis to be monophyletic in the tree and compared it to the non-forced best tree. Results of monophyletic testing shown that the eight internal groups resolved as monophyletic with high support within the *Bacillus* genus (Table S3).

The Subtilis group comprises a well-known species complex commonly found in soil and aquatic sediments with widespread distribution in nature. Members of this group, such as *B. subtilis*, compose the gut microflora of humans and other animals [66, 67]. This group shows valuable traits useful for biotechnological, industrial and agricultural applications [68, 69]. The Cereus group comprises human and plant pathogen species that can thrive in various environments ranging from low nutrient soil to intestinal flora of various animals [5–7, 70]. The Pumilus group was previously considered in the Subtilis group. However, the monophyly analysis shows enough robust support to consider it as a separate group from Subtilis. The Pumilus group contains species highly resistant to UV-light and H_2O_2 due to the presence of the spore-coat proteins CotA and YjqC [64]. Members of the Coagulans group have been isolated from a wide variety of environments, such as the human gut and marine sediments [3, 71]. Members of the Megaterium group have been extensively used in industrial processes because their high capacity for the production of exoenzymes and ease of cloning genes for the production of recombinant proteins. Some members also are useful in bioremediation and agriculture as plant-growth promotion agents [72, 73]. Bacteria commonly found in soil and in extreme environments compose the Halodurans group. They have industrial applications, as they produce enzymes

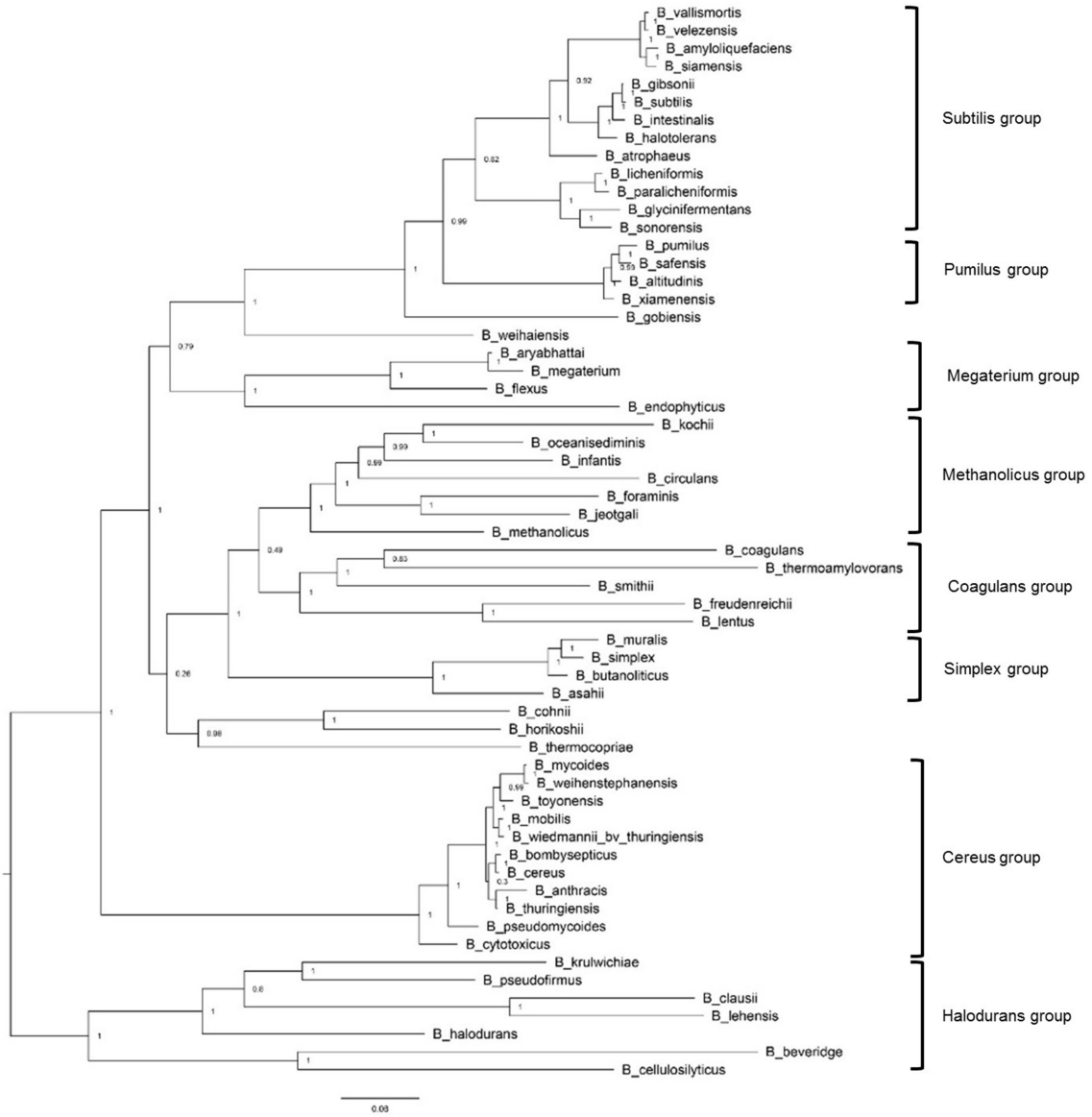

Fig. 5. Phylogenetic tree reconstruction based on 60 genomes of *Bacillus* species to evidence internal monophyletic groups.

with useful activities [74]. It has been proposed that they could be used as probiotics to improve the intestinal microbial balance [75]. The Methanolicus group is characterized by bacteria isolated from fresh or groundwater, which have industrial potential [76, 77]. However, some members were associated with urinary tract infections [78]. The Simplex group harbours environmental bacteria usually found in soil; some isolates have also been found in the intestinal tract of humans [79]. Some members of this group are useful for industrial applications focused on the remediation of organic compounds, such as fatty acids and other compounds [80, 81].

Selection pressure forces

In order to understand selection pressures acting on spore-coat genes, we employed the classical approaches of Tajima's D test and the d*N*/d*S* ratio (known also as omega, ω) as well as two new methods (BUSTED, MEME) that use modern algorithms for detecting episodic positive selection in all or a subset of branches on a phylogeny. For this, we created spore-coat-gene datasets for each *Bacillus* group, based on the results of the consensus heat map.

All significant results (P-value <0.05) of spore-coat genes displaying evidence of positive selection on different *Bacillus* groups are reported in Table 2. We successfully extracted and aligned 47 spore-coat genes for the Cereus group, 25 (53.2%) of which were found to be evolving under positive selection either by having positively selected sites (MEME), being positively selected along its entire gene sequence (BUSTED) or because of possible balancing selection (Tajima's D). Coat

Table 2. Five summary statistics (Tajima's D, BUSTED, MEME, d*N*/d*S* (branch and site models) showing positive selection across different *Bacillus* groups

Cereus group					
Coat gene	**Summary statistics**				
	Tajima's D	**BUSTED***	**MEME†**	**dN/dS (branch model)**	**dN/dS (site models)**
cgeD	0.77594	0.5	1†	0.21094	**M1:Nearly neutral** 0.2315 **M8:β distribution+positive selection** 0.2358
cotA	−0.18419	0.5	5†	0.23041	**M2:Positive selection** 0.2599 **M8:β distribution+positive selection** 0.2496
cotB	−0.18491	0.5	1†	0.24867	**M1:Nearly neutral** 0.3770 **M7:β distribution** 0.2904
cotD	0.89921	0.018†	2†	0.10672	**M1:Nearly neutral** 0.1481 **M7:β distribution** 0.1419
cotJB	2.49259‡	0.145	0	NA§	NA§
cotJC	1.12785	0.47	1†	0.02253	**M1:Nearly neutral** 0.0301 **M7:β distribution** 0.0234
cotS	−0.12103	0.5	1†	0.10342	**M1:Nearly neutral** 0.1290 **M7:β distribution** 0.1121
cotSA	0.25413	0.5	3†	0.15863	**M1:Nearly neutral** 0.1975 **M8:β distribution+positive selection** 0.1988
cotZ	0.04527	0.101	1†	0.20776	**M1:Nearly neutral** 0.2808 **M8:β distribution+positive selection** 0.2700
gerPC	0.2173	0.028*	1†	0.10771	**M1:Nearly neutral** 0.1678 **M7:β distribution** 0.1212
gerPE	0.39382	0.414	1†	0.14856	**M1:Nearly neutral** 0.1796 **M7:β distribution** 0.1621
gerQ	0.62352	0.049*	1†	0.05734	**M1:Nearly neutral** 0.1083 **M7:β distribution** 0.0670
safA	0.24669	0*	9†	0.12459	**M1:Nearly neutral** 0.1614 **M8:β distribution+positive selection** 0.1470

Continued

Table 2. Continued

spoVID	−0.08138	0.5	1†	0.15641	**M1:Nearly neutral** 0.2082 **M7:β distribution** 0.1764
tasA	0.74152	0.062	2†	0.18312	**M1:Nearly neutral** 0.3561 **M7:β distribution** 0.2056
tgl	0.28166	0.358	1†	0.087	**M1:Nearly neutral** 0.1146 **M7: β distribution** 0.0939
yaaH	0.42629	0.5	1†	NA§	NA§
ydhD	0.42629	0.495	1†	0.03899	**M1:Nearly neutral** 0.0584 **M7:β distribution** 0.0428
yhbB	0.07332	0.5	1†	NA§	NA§
yheC	0.45696	0.454	3†	0.15124	**M1:Nearly neutral** 0.2175 **M7:β distribution** 0.1732
yheD	0.45696	0.454	3†	0.15124	**M1:Nearly neutral** 0.2175 **M7:β distribution** 0.1732
yncD	−0.29741	0.447	3†	0.10343	**M1:Nearly neutral** 0.1422 **M7:β distribution** 0.1105
yppG	0.08813	0.002†	1‡	0.13435	**M1:Nearly neutral** 0.2096 **M8:β distribution+positive selection** 0.2261
ytdA	0.48066	0.106	1‡	0.07142	**M1:Nearly neutral** 0.0897 **M7:β distribution** 0.0844
yutH	−0.12103	0.5	1†	0.10342	**M1:Nearly neutral** 0.1290 **M7:β distribution** 0.1121
Coagulans group					
Coat gene			**Summary statistics**		
	Tajima's D	**BUSTED***	**MEME†**	**d*N*/d*S* (branch model)**	**d*N*/d*S* (site models)**
cgeD	2.40675‡	0.5	1†	0.3661	**M1:Nearly neutral** 0.5498 **M7:β distribution** 0.4664

Continued

Table 2. Continued

cotD	2.12158‡	0.5	0	0.16022	**M1:Nearly neutral** 0.2505 **M7:β distribution** 0.2142
cotJC	1.63432	0.5	1†	0.03028	**M1:Nearly neutral** 0.0204 **M7:β distribution** 0.0348
cotY	2.37‡	0.177	0	0.10084	**M1:Nearly neutral** 0.2236 **M7:β distribution** 0.1225
gerPA	2.42801‡	0.5	1†	0.02311	**M1:Nearly neutral** 0.1871 **M7:β distribution** 0.0415
gerPB	2.71776‡	0.5	0	0.14422	**M1:Nearly neutral** 0.4187 **M7:β distribution** 0.1980
gerPD	2.06706‡	0.5	0	0.10075	**M1:Nearly neutral** 0.1634 **M7:β distribution** 0.1105
gerPE	2.43753‡	0.5	0	0.16536	**M1:Nearly neutral** 0.2685 **M7:β distribution** 0.1991
gerQ	2.03383‡	0.5	0	0.09011	**M1:Nearly neutral** 0.3678 **M7:β distribution** 0.1817
spoIVA	1.86222‡	0.282	0	0.03558	**M1:Nearly neutral** 0.0755 **M7:β distribution** 0.0471
spsI	2.131‡	0.5	0	0.05597	**M1:Nearly neutral** 0.1365 **M7:β distribution** 0.0653
yaaH	2.04839‡	0†	1†	0.08248	**M1:Nearly neutral** 0.1914 **M7:β distribution** 0.1098
ydhD	2.36117‡	0.5	1†	0.00316	**M1:Nearly neutral** 0.1918 **M7:β distribution** 0.0680
yhbB	2.16596‡	0.5	0	0.10258	**M1:Nearly neutral** 0.3634 **M7:β distribution** 0.1723
yjqC	1.63432	0.315	1†	0.03028	**M1:Nearly neutral** 0.0482 **M7:β distribution** 0.0348

Continued

Table 2. Continued

yppG	2.90119‡	0.5	1†	0.00759	**M1:Nearly neutral** 0.5161 **M7:β distribution** 0.2385
ytdA	2.2359‡	0*	0	0.04891	**M1:Nearly neutral** 0.1840 **M7:β distribution** 0.0573
yuzC	2.79561‡	0.5	0	0.20748	**M1:Nearly neutral** 0.4965 **M7:β distribution** 0.3580
Halodurans group					
Coat gene			**Summary statistics**		
	Tajima's D	**BUSTED***	**MEME†**	**dN/dS (branch model)**	**dN/dS (site models)**
cotE	2.33501‡	0*	0	0.04718	**M1:Nearly neutral** 0.2401 **M7:β distribution** 0.0699
cwlJ	2.22293‡	0.5	1†	0.0022	**M1:Nearly neutral** 0.0645 **M7:β distribution** 0.0033
gerQ	1.97623	0.5	1†	0.10825	**M1:Nearly neutral** 0.2912 **M8:β distribution+positive selection** 0.3657
spoIVA	2.10434‡	0.5	0	NA§	NA§
tgl	2.64696‡	0.467	0	0.14067	**M1:Nearly neutral** 0.4113 **M7:β distribution** 0.2242
yhaX	2.47692‡	0.382	1†	0.11997	**M1:Nearly neutral** 0.2107 **M7:β distribution** 0.1415
yjqC	1.46076	0.5	1†	0.09556	**M1:Nearly neutral** 0.2018 **M8:β distribution+positive selection** 0.1790
yraG	2.12556	0.5	1†	0.25053	**M1:Nearly neutral** 0.3815 **M7:β distribution** 0.3218
ytdA	2.29913‡	0.5	0	0.04657	**M1:Nearly neutral** 0.1857 **M7:β distribution** 0.0672
Megaterium group					

Continued

Table 2. Continued

Coat gene	**Summary statistics**				
	Tajima's D	**BUSTED***	**MEME†**	**d*N*/d*S* (branch model)**	**d*N*/d*S* (site models)**
gerT	1.19483	0.023*	0	0.18757	**M1:Nearly neutral** 0.3623 **M7:*β* distribution** 0.2610
spoVID	1.06769	0.5	1†	0.21812	**M1:Nearly neutral** 0.3907 **M8:*β* distribution+positive selection** 0.3623
tgl	1.45908	0.006*	0	0.04185	**M1:Nearly neutral** 0.4516 **M7:*β* distribution** 0.0569
yaaH	0.96821	0.5	1†	0.00371	**M1:Nearly neutral** 0.0448 **M7:*β* distribution** 0.0062
yncD	1.23026	0.046*	2†	0.18115	**M1:Nearly neutral** 0.3629 **M7:*β* distribution** 0.2489
ysxE	0.75614	0.5	1†	0.08323	**M1:Nearly neutral** 0.1546 **M7:*β* distribution** 0.0946
yuzC	1.73735	0.052	1†	0.06424	**M1:Nearly neutral** 0.8243 **M7:*β* distribution** 0.0789
Methanolicus group					
Coat gene	**Summary statistics**				
	Tajima's D	**BUSTED***	**MEME†**	**d*N*/d*S* (branch model)**	**d*N*/d*S* (ite models)**
cotE	1.62664	0.047*	0	0.10272	**M1:Nearly neutral** 0.1996 **M7:*β* distribution** 0.1170
cotF	2.45173‡	0.496	0	0.08622	**M1:Nearly neutral** 0.0984 **M7:*β* distribution** 0.0982
cotJA	2.05089‡	0	0	0.1059	**M1:Nearly neutral** 0.2046 **M7:*β* distribution** 0.1504
cotJB	2.10987‡	0.5	0	0.01056	**M1:Nearly neutral** 0.1407 **M7:*β* distribution** 0.0147

Continued

Table 2. Continued

cotJC	2.09006‡	0.496	1†	0.02096	**M1:Nearly neutral** 0.0269 **M7:β distribution** 0.0233
cotSA	2.6247‡	0.5	0	0.05597	**M1:Nearly neutral** 0.1369 **M7:β distribution** 0.0651
gerPA	2.20951‡	0.5	0	0.0751	**M1:Nearly neutral** 0.1302 **M7:β distribution** 0.0870
gerPB	2.58274‡	0.168	0	0.00305	**M1:Nearly neutral** 0.3324 **M7:β distribution** 0.0064
gerPD	2.52437‡	0.5	0	0.03528	**M1:Nearly neutral** 0.0866 **M7:β distribution** 0.0427
gerPE	2.34308‡	0.5	0	0.12838	**M1:Nearly neutral** 0.2792 **M7:β distribution** 0.1646
gerPF	2.16055‡	0.001†	0	0.06189	**M1:Nearly neutral** 0.1556 **M7:β distribution** 0.0677
spoIVA	2.37156‡	0.5	0	0.02067	**M1:Nearly neutral** 0.0334 **M7:β distribution** 0.1654
yaaH	2.11824‡	0.078	2†	0.05627	**M1:Nearly neutral** 0.1392 **M7:β distribution** 0.0705
ydhD	2.32379‡	0.279	2†	0.05453	**M1:Nearly neutral** 0.1255 **M8:β distribution+positive selection** 0.0832
yhaX	2.10645‡	0.5	1†	0.0897	**M1:Nearly neutral** 0.1595 **M8:β distribution+positive selection** 11.1381
yhcQ	1.92212‡	0.5	0	0.08651	**M1:Nearly neutral** 0.2220 **M7:β distribution** 0.1056
yhjR	2.19329‡	0.5	0	0.13211	**M1:Nearly neutral** 0.3260 **M7:β distribution** 0.1913
ylbD	2.39017‡	0.062	0	0.1214	**M1:Nearly neutral** 0.3662 **M7:β distribution** 0.1761

Continued

Table 2. Continued

yncD	2.29644‡	0.5	1†	0.09177	**M1:Nearly neutral** 0.2860 **M7:β distribution** 0.1239
yraF	2.20974	0.039†	0	0.0359	**M1:Nearly neutral** 0.0781 **M7:β distribution** 0.0451
yraG	2.43863‡	0.5	0	0.06799	**M1:Nearly neutral** 0.1791 **M7:β distribution** 0.0896
ytdA	2.70168‡	0.5	0	0.01055	**M1:Nearly neutral** 0.2670 **M7:β distribution** 0.1097
yutH	2.28896‡	0.5	3†	0.11558	**M1:Nearly neutral** 0.3214 **M7:β distribution** 0.1588
yuzC	2.82223‡	0.5	0	0.09933	**M1:Nearly neutral** 0.3239 **M7:β distribution** 0.1406
Pumilus group					
Coat gene			**Summary statistics**		
	Tajima's D	**BUSTED***	**MEME†**	**d*N*/d*S* (branch model)**	**d*N*/d*S* (Site models)**
cgeB	0.82607	0.044*	0	0.21246	**M1:Nearly neutral** 0.2692 **M7:β distribution** 0.2446
cotH	0.77125	0.5	1†	0.09965	**M1:Nearly neutral** 0.1270 **M7:β distribution** 0.1097
cotM	0.85448	0.187	1†	NA§	NA§
cotS	0.82748	0.5	1†	0.06266	**M1:Nearly neutral** 0.0795 **M7:β distribution** 0.0695
cwlJ	0.83023	0.06	1†	0.04195	**M1:Nearly neutral** 0.0587 **M7:β distribution** 0.0542
gerPD	−0.13219	0.03*	0	NA§	NA§
lipC	0.8556	0.04	1†	NA§	NA§
spoVID	1.21538	0.481	2†	0.19841	**M1:Nearly neutral** 0.2420 **M7:β distribution** 0.2382

Continued

Table 2. Continued

yheC	0.70157	0.5	1†	0.27466	**M1:Nearly neutral** 0.3078 **M7:β distribution** 0.2939
yisY	0.60511	0.5	1†	0.18592	**M1:Nearly neutral** 0.2201 **M7:β distribution** 0.2045
yjqC	2.41476‡	0.5	0	NA§	NA§
yutH	0.82748	0.5	1†	0.06266	**M1:Nearly neutral** 0.0795 **M7:β distribution** 0.0695
Simplex group					
Coat gene			**Summary statistics**		
	Tajima's D	**BUSTED***	**MEME†**	**dN/dS (branch model)**	**dN/dS (site models)**
cotD	0.82064	0.001*	0	0.14089	**M1:Nearly neutral** 0.2331 **M7:β distribution** 11.2858
cotH	1.02307	0.5	3†	0.10615	**M1:Nearly neutral** 0.1827 **M7:β distribution** 0.1246
cotX	0.85129	0.5	1†	0.10962	**M1:Nearly neutral** 0.1725 **M7:β distribution** 0.1319
gerPE	1.38199	0.442	1†	0.25448	**M1:Nearly neutral** 0.4100 **M7:β distribution** 0.3257
gerT	0.82125	0.5	1†	0.15016	**M1:Nearly neutral** 0.2152 **M7:β distribution** 0.1772
spoVID	1.21956	0.288	1†	0.21728	**M1:Nearly neutral** 0.3437 **M7:β distribution** 0.2686
ydhD	0.58553	0.5	1†	0.05483	**M1:Nearly neutral** 0.0752 **M7:β distribution** 0.0595
yheD	1.13466	0.187	1†	0.06729	**M1:Nearly neutral** 0.0824 **M7:β distribution** 0.0724
yisY	2.14444	0.5	1†	0.07518	**M1:Nearly neutral** 0.0921 **M7:β distribution** 0.0775

Continued

Table 2. Continued

yppG	0.6223	0.5	1†	0.12362	**M1:Nearly neutral** 0.2617 **M7:β distribution** 0.1506
Subtilis group					
Coat gene			**Summary statistics**		
	Tajima's D	**BUSTED***	**MEME†**	**d*N*/d*S* (branch model)**	**d*N*/d*S* (site models)**
cgeA	1.83274	0.5	1†	0.20934	**M1:Nearly neutral** 0.3500 **M7:β distribution** 0.2598
cgeB	1.99094‡	0.277	2†	0.19573	**M1:Nearly neutral** 0.2863 **M7:β distribution** 0.2252
cgeD	1.79061	0.5	1†	0.19155	**M1:Nearly neutral** 0.2917 **M7:β distribution** 0.2294
cgeE	2.63941‡	0.5	5†	0.18444	**M1:Nearly neutral** 0.3187 **M7:β distribution** 0.2035
cotA	2.31807‡	0.367	2†	0.10535	**M1:Nearly neutral** 0.1527 **M7:β distribution** 0.1138
cotB	1.76891	0*	4†	0.22824	**M1:Nearly neutral** 0.3445 **M7:β distribution** 0.2801
cotD	0.93573	0.5	1†	NA§	NA§
cotE	2.09262‡	0.48	1†	NA§	NA§
cotF	2.19577‡	0.133	2†	0.08357	**M1:Nearly neutral** 0.1216 **M7:β distribution** 0.0923
cotG	0.79192	0.478	2†	0.19177	**M1:Nearly neutral** 0.2831 **M7:β distribution** 0.2336
cotH	2.56746‡	0.5	2†	0.10377	**M1:Nearly neutral** 0.1657 **M7:β distribution** 0.1152
cotJA	2.1477‡	0.259	1†	0.10669	**M1:Nearly neutral** 0.1841 **M7:β distribution** 0.1277
cotJB	2.49259‡	0.339	0	0.10261	**M1:Nearly neutral** 0.1765 **M7:β distribution** 0.1148

Continued

Table 2. Continued

cotM	2.35214‡	0.403	0	0.18183	**M1:Nearly neutral** 0.3447 **M7:β distribution** 0.2176
cotO	2.26548‡	0.5	3†	0.22212	**M1:Nearly neutral** 0.4142 **M8:β distribution+positive selection** 0.4033
cotP	1.96543‡	0.5	0	NA§	NA§
cotV	1.89032	0.291	1†	0.23489	**M1:Nearly neutral** 0.2755 **M7:β distribution** 0.2484
cotW	1.96128	0.486	2†	0.20479	**M1:Nearly neutral** 0.3099 **M7:β distribution** 0.2219
cotX	1.7908	0.37	1†	0.10867	**M1:Nearly neutral** 0.1518 **M7:β distribution** 0.1157
cotY	2.0907‡	0.5	1†	0.06725	**M1:Nearly neutral** 0.1107 **M7:β distribution** 0.0735
cotZ	2.08360‡	0.376	1†	0.11551	**M1:Nearly neutral** 0.1953 **M7:β distribution** 0.1282
cwlJ	1.79177	0.5	1†	0.07168	**M1:Nearly neutral** 0.1272 **M7:β distribution** 0.0778
gerPB	2.52228‡	0.5	1†	0.14965	**M1:Nearly neutral** 0.3683 **M7:β distribution** 0.2129
gerPC	2.02921‡	0.5	1†	0.11727	**M1:Nearly neutral** 0.1948 **M7:β distribution** 0.1287
gerPD	1.95737	0.109	1†	0.13235	**M1:Nearly neutral** 0.2282 **M7:β distribution** 0.1540
gerPE	2.59394‡	0.5	1†	NA§	NA§
gerPF	1.30604	0.012*	1†	NA§	NA§
gerQ	2.02092‡	0.372	0	0.10469	**M1:Nearly neutral** 0.1790 **M8:β distribution+positive selection** 0.1383
gerT	2.55712‡	0.5	2†	0.13045	**M1:Nearly neutral** 0.2122 **M7:β distribution** 0.1576

Continued

Table 2. Continued

lipC	2.75952‡	0.5	1†	NA§	NA§
oxdD	2.83684‡	0.478	4†	0.06435	**M1:Nearly neutral** 0.1317 **M7:*β* distribution** 0.0733
safA	2.31936‡	0.005†	2†	0.16678	**M1:Nearly neutral** 0.2750 **M7:*β* distribution** 0.1930
spoIVA	2.12249‡	0.5	0	0.01660	**M1:Nearly neutral** 0.0299 **M7:*β* distribution** 0.0192
spoVID	2.66623‡	0.315	9†	0.29849	**M1:Nearly neutral** 0.5939 **M8: *β* distribution+positive selection** 0.5141
spsB	1.71279	0.5	2†	0.15996	**M1:Nearly neutral** 0.2592 **M7:β distribution** 0.1872
spsI	1.63797	0.304	1†	0.07904	**M1:Nearly neutral** 0.1207 **M7:β distribution** 0.0886
tasA	2.27556‡	0.281	2†	0.08090	**M1:Nearly neutral** 0.1036 **M7:β distribution** 0.0865
tgl	2.36389‡	0.5	4†	NA§	NA§
yaaH	2.33413‡	0.024*	5†	0.08289	**M1:Nearly neutral** 0.1270 **M7:β distribution** 0.0896
ydgB	1.80839	0.011*	0	NA§	NA§
ydhD	2.32146‡	0.5	5†	0.09731	**M1:neutral** 0.1422 **M7:β distribution** 0.1082
yhaX	2.14372‡	0.421	1†	0.06610	**M1: Nearly neutral** 0.0933 **M8: β distribution+positive selection** 0.0800
yhbB	2.26467‡	0.5	0	0.12273	**M1:Nearly neutral** 0.2253 **M7:β distribution** 0.1403
yhcQ	2.39963‡	0.013*	3†	0.07439	**M1:Nearly neutral** 0.0897 **M7:β distribution** 0.0775
yheC	2.54815‡	0.5	2†	0.10043	**M1:Nearly neutral** 0.1435 **M7:β distribution** 0.1088

Continued

Table 2. Continued

yheD	2.57861‡	0.34	4†	0.13014	**M1:Nearly neutral** 0.2168 **M7:β distribution** 0.1471
yhjQ	1.74733	0.492	1†	NA§	NA§
yhjR	2.72348‡	0.199	2†	NA§	NA§
yisY	1.97894‡	0.453	2†	0.11089	**M1:Nearly neutral** 0.1518 **M7:β distribution** 0.1182
yjqC	2.68998‡	0.196	1†	NA§	NA§
yjzB	2.59674‡	0.478	1†	NA§	NA§
yknT	2.47907‡	0.5	5†	0.20519	**M1:Nearly neutral** 0.3589 **M7:β distribution** 0.2389
ylbD	2.10676‡	0.494	1†	NA§	NA§
yncD	1.46348	0.5	1†	0.13442	**M1:Nearly neutral** 0.1868 **M7:β distribution** 0.1452
yppG	2.31307‡	0.5	0	0.23211	**M1:Nearly neutral** 0.4261 **M7:β distribution** 0.3108
yraD	2.42924‡	0.499	1†	NA§	NA§
yraG	1.47157	0.498	1†	0.09399	**M1:Nearly neutral** 0.1273 **M7:β distribution** 0.1089
ysxE	2.27601‡	0.5	1†	0.14151	**M1:Nearly neutral** 0.2254 **M7:β distribution** 0.1583
ytdA	2.28755‡	0*	0	0.03920	**M1:Nearly neutral** 0.0683 **M8:β distribution+positive selection** 0.0532
yutH	2.36638‡	0.391	3†	0.11151	**M1:Nearly neutral** 0.1723 **M8:β distribution+positive selection** 0.1548
yuzC	2.63657‡	0.359	1†	0.22296	**M1:Nearly neutral** 0.3048 **M7:β distribution** 0.2438
ywrJ	1.89712	0.5	1†	0.14905	**M1:Nearly neutral** 0.2068 **M7:β distribution** 0.1696
yybI	0.53608	0.338	1†	0.20922	**M1:Nearly neutral** 0.2922 **M7:β distribution** 0.2464

Continued

Table 2. Continued

*P value provided by BUSTED. A *P* value <0.05 indicates evidence of positive selection of the gene
†Number of significant sites under positive selection by MEME.
‡Significant at a *P* value <0.05.
§dN/dS values could not be computed in CodeML due to small branch size.

genes of the basement layer (*cotJB, cotJC, spoVID, yheD, yppG*) account for 20% of positively selected genes. Similarly, the coat genes of the inner layer (*cotD, gerPC, gerPE, gerQ, safA, tgl, yaaH* and *yutH*) represent 32%, whereas the outer layer genes (*cotA, cotB, cotS, yncD, ytdA*) represent 20%. Other coat genes (*cgeD, tasA, cotSA, ydhD, yhbB* and *yheC*) whose protein products have unknown localization, make up 24% of positively selected genes. Moreover, the morphogenetic coat genes *cotZ, spoVID* and *safA* seem to be under positive selection.

In the Coagulans group, the coat genes *gerPC, gerPF, gerT, lipC, spoVID, yhaX, yhcQ, yheC, yheD, ylbD, yncD, ysxE,* and *yutH* were highly divergent, except at conserved domains, and could not be properly aligned. Therefore, we discarded those genes and analysed the remaining 23 well-aligned spore coat genes, 18 (78.3%) of which were found to be under positive selection. Coat genes of the basement layer (*cotJC, spoIVA, yppG*) account for 16.6% of positively selected genes. Likewise, *cotD, gerPA, gerPB, gerPD, gerPE, gerQ, yaaH, yjqC,* and *yuzC* (inner layer), *spsI, ytdA* (outer layer), *cgeD, ydhD,* and *yhbB,* (localization class unknown) make up 50 11.1 and 16.6%, respectively, of coat genes under positive selection. Interestingly, *cotY,* the only coat gene of the crust present in this group, is under positive balancing selection (or population contraction), according to Tajima's D. The great majority of extracted coat genes (*cotA, cotF, cotJC, cotSA, cotX, lipC, safA, spoVID, spsI, yaaH, ydhD, yhbB, yhcQ, ylbD, yncD, yraD, yraF, ysxE,* and *yutH*) in the Halodurans group were highly divergent outside conserved domains and could not be properly aligned. Therefore, only 10 spore coat genes (Table S2) were analysed, 9 (90%, see Table 2) of which are under positive selection and the rest of genes are under neutral or negative selection (Table S4). The morphogenetic coat gene *spoIVA* and *yhaX* are the only coat genes of the basement layer evolving under positive selection. Our results show that other coat genes, such as *cwlJ, gerQ, tgl,* and *yjqC* (inner layer), *cotE, ytdA* (outer layer), and *yraG* seem to be under positive selection detected either by Tajima's D, MEME, or BUSTED.

In the Megaterium group, we extracted and aligned 35 coat genes, 7 (20%) of which show traces of positive selection. Coat genes of the inner layer (*tgl, yaaH, ysxE,* and *yuzC*) account for the majority of positively selected genes, whereas only two genes (*gerT* and *yncD*) of the outer layer are under positive selection. Additionally, *spoVID* is the only the morphogenetic coat gene evolving under positive selection in this group.

Methanolicus group coat genes with sequences that were highly diverged from reference genes (*cotD, cotM, tasA, cotP, cotS, cotY, cotZ, gerPC, gerT, lipC, spoVID, spsI, tgl, yhbB, yheC, yheD, yjqC, yppG, ysxE*, and *yybI*), were not further analysed. However, we successfully aligned 29 spore coat genes in this group, 24 (82.8%) of which show evidence of positive selection according to Tajima's D, MEME, or BUSTED. The majority of positively selected genes belong to the inner layer of the coat (*cotF, gerPA, gerPB, gerPD, gerPE, gerPF, yaaH, yhjR, yutH,* and *yuzC*), accounting for 41.6% of positively selected genes. Genes of the basement (*cotJA, cotJB, cotJC, spoIVA, yhaX*) and outer layer (*cotE, ylbD, yncD,* and *ytdA*) account for 20.8 and 16.6% of genes under positive selection, respectively. Coat genes corresponding to proteins whose localization has not been determined contribute to 20.8% of positively selected genes.

In the Pumilus group, we extracted and analysed 55 coat genes, 12 (21.8%) of which were found to be under positive selection, either along the entire gene sequence or at individual sites. In this group, spore coat genes of the crust are highly conserved and *cgeB* seems to be positively selected along its entire gene sequence. Coat genes of the basement (*lipC, spoVID*), inner (*cwlJ, gerPD, yisY, yjqC, yutH*) and outer layer (*cotH, cotM, cotS*) also show evidence of positive selection. On the other hand, in the Simplex group, we retrieved and analysed 40 spore coat genes, 10 (25%) of which are under positive selection. The morphogenetic coat genes *cotH, cotX,* and *spoVID* of the outer, crust, and basement layer are under positive selection. It is worth mentioning that *cotX* is the only coat gene belonging to the crust present in this group. The proteins present in the crust are critical for interaction with the environment. Thus the ability to adhere to and survive on variable surface structures could be a key factor that promotes diversity in coat structure and composition [20]. Furthermore, coat genes of the basement layer (*spoVID*, *yheD*, and *yppG*), inner layer (*cotD*, *gerPE*, and *yisY*), outer layer (*cotH*, and *gerT*), crust (*cotX*), and *ydhD* (localization not determined) represent 30, 30, 20, 10, and 10% of the total positively selected genes, respectively.

The Subtilis group possess the most conserved core of spore coat proteins compared to other groups analysed in this work. This is expected, since all analyses performed here used *B. subtilis* as a reference to determine the abundance and diversity of spore coat proteins (see Discussion section for further comments). We extracted, aligned, and analysed 77 coat genes, 63 (81.8%) of which show significant evidence of positive selection detected by Tajima's D, MEME, and/or BUSTED. Nearly all morphogenetic coat protein genes of the basement (except *spoVM*), inner, outer layer, and crust are positively selected or show sites under positive selection. For instance, coat genes of the basement layer, inner layer, outer layer, crust, and coat genes of localization not determined account for 14.3, 31.7, 22.2, 11.1, and 20.6% of the

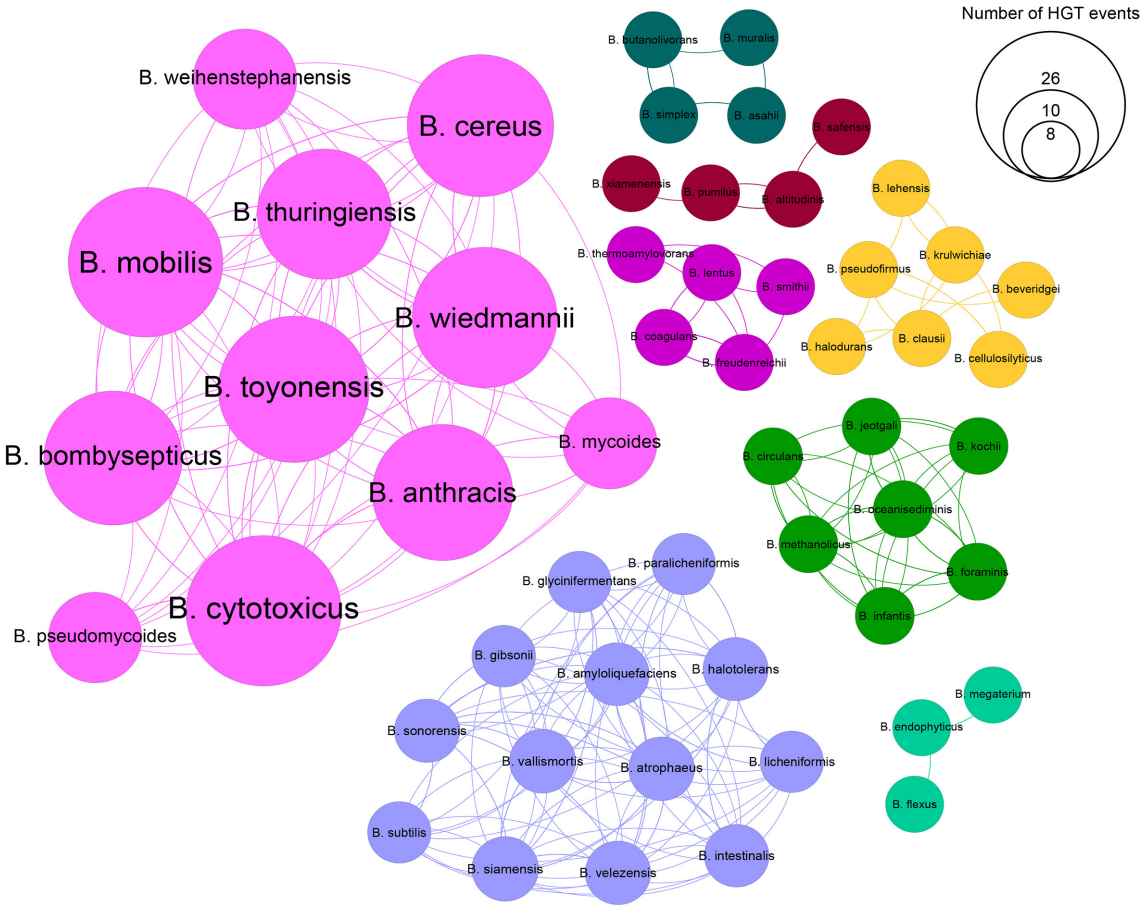

Fig. 6. Spore coat genes under HGT events as donor-recipient networks in the Cereus (pink), Coagulans (magenta), Halodurans (yellow), Methanolicus (green), Pumilus (dark red), Simplex (navy blue), and Subtilis (blue). Edges, nodes and size of nodes represent HGT events, genomes and number of HGT events per genome respectively.

total positively selected genes, respectively (see Table 2). In addition, coat genes not included in Table 2, are under purifying selection (ω <1), according to CodeML site and branch models (see Table S4).

Horizontal gene transfer (HGT)

HGT events can be detected by phylogenetic incongruences [82]. Additionally, traces of the mechanism of transfer, such as independently conjugative plasmids, integrated prophages, integrative transposons, GEIs, and other unclassified mobile genetic elements may further confirm HGT events [82–84].

Spore coat genes that displayed evidence of HGT are shown as donor-recipient networks in Fig. 6 for the eight monophyletic groups in *Bacillus*. Most spore coat genes have been recently transferred, since HGT events are displayed at or near the branch tips of their reconciled phylogenetic trees (not shown) unless otherwise stated. The Cereus group has 37 spore coat genes that have undergone HGT events, according to Notung. Spore coat genes of this group, such as *cotD, cotJA, cotY, gerPD, gerPE, yncD* have undergone HGT events near the bottom of their reconciled phylogenetic trees. The morphogenetic coat genes *safA* and *spoVID* have also undergone HGT events. The Coagulans group has 13 spore coat genes that were laterally transferred between species of this group. According to our results, *cotY* is the only morphogenetic coat gene showing evidence of a recent HGT event. The Halodurans group has six coat genes that have undergone HGT events. The Methanolicus group harbours 19 coat genes that show evidence for HGT events. *spoVM* is the only morphogenetic coat gene that has been laterally transferred in the Halodurans and Methanolicus groups.

In the Megaterium, Pumilus, and Simplex groups, 2, 10, and 14 coat genes, respectively, have been laterally transferred (Fig. 6). The morphogenetic coat genes that control the assembly of the crust, *cotX* and *cotY*, are the only morphogenetic coat proteins under HGT events in the Pumilus group. On the other hand, most HGT events of the Simplex group occur between *B. butanolivorans* and *B. simplex*.

In the Subtilis group, about half of its coat genes (33) have undergone HGT events. Most of the HGT events in this group occur near the tips of the reconciled phylogenetic

trees. However, the coat genes *cotD, yjqC, yraF, yraG*, and *ytdA* show evidence of HGT near the bottom of reconciled phylogenetic trees, according to Notung (Fig. 6), suggesting an ancient transfer of the genes. All these HGT events have been further confirmed by ICEs (Integrative and Conjugative Elements) using WU-BLAST2 of the webserver ICEBerg, see Table S5. Analysis to detect the presence of spore coat genes in genomic islands shows their complete absence in these genomic elements.

DISCUSSION

In this work, we reported the existence of several spore coat protein homologs across one hundred sixty-one genomes of spore-forming species of the Bacillales order. The most conserverd proteins are those concerned with the development and assembly of coat and spore germination. Spore coat proteins that directly depend on these morphogenetic and germinant proteins are also preserved. However, some minor spore coat proteins seem to be taxa-specific and/or may confer a unique spore coat morphology and the ability to occupy different ecological niches, as previously suggested [8, 16, 23, 26, 27, 85–87]. Nevertheless, it is important to mention that the methods used in our diversity analysis are only able to identify homologs of *B. subtilis* coat proteins across the set of genomes analysed here. This imposes a limitation in the diversity of spore coat proteins described in Bacillales because coat proteins not present in *B. subtilis* and coat-like proteins that share structural and chemical features to *B. subtilis* coat proteins cannot be considered using the methodologies of this study. Moreover, homologs of coat proteins with enzymatic activity (e.g. transferases) found across Bacillales are only putative spore coat proteins. Further studies must characterize these proteins to determine if they can be classified as true spore coat proteins. On the other hand, the lack of evidence for spore coat gene homologues in *Hallolactobacillus*, *Jeotgalicoccus* and *B. beveridgei* suggests that a major loss of genes occurred during their evolutionary history, as previously found for the *Exiguobacterium* genus. This may explain why they do not produce spores [88].

Some *Bacillus* species lack the morphogenetic coat proteins CotH and CotO. Several studies have reported that CotH and CotO are minor players in the assembly of the outer coat, because these two proteins are CotE-dependent [8, 16, 23, 86]. Although CotH and CotO mutants have a disorganized outer coat, the major assembly step is carried out by CotE and CotE-dependent coat proteins [23, 86]. Recent studies have found that CotO is necessary for encasement of the spore by the crust [89], thus we can expect CotO to be conserved when coat proteins of the crust are also conserved, as confirmed by our results. Likewise, CotH is a spore kinase that phosphorylates its dependent proteins CotB and CotG [90, 91]. Our results show that in genomes where CotH is absent, its substrates, CotG and CotB, are also absent [91]. Nevertheless, the role of CotG may be carried out by a non-homologous CotG-like protein with similar structural regions, as previously reported [61]. Other CotH-dependent coat proteins, such as CotC and CotU are conserved in few genomes of the Subtilis group, and they are present when CotH and CotG are present. In this case, CotG has a negative role on CotC/CotU/CotS assembly when CotH is not present (i.e. when it is not phosphorylated by its specific kinase) [92].

The morphogenetic coat proteins CotX, CotY, and CotZ are collectively known as the insoluble fraction of the spore because they influence spore hydrophobicity and accessibility of germinants [87, 89, 93]. Moreover, they are responsible for crust assembly around the spore [8, 25, 89]. CotX, CotY, and CotZ mutants have an incomplete outer coat, but resistance to heat or lysozyme is not affected [87]. Hence, the absence of these morphogenetic coat proteins and their dependent-proteins in various spore-forming species reflects overlapping functions and a spore coat protein interaction network that is highly adapted to unique environmental conditions [8, 87, 94]. Our results confirm the overlapping functions and highly hierarchical organization of morphogenetic coat proteins in the assembly of the spore coat of *B. subtilis* but also in several spore-forming species.

The morphogenetic coat proteins CotE, SpoIVA, SpoVM, SpoVID, and SafA are present in almost all genomes of spore-forming species analysed. Usually, other proteins dependent on the morphogenetic coat proteins are also well conserved. CotE controls the assembly of the outer coat layer and other coat proteins, designated as CotE-controlled proteins [8, 20]. SafA has been found to interact with SpoVID in the early stages of coat assembly [8, 20, 22] and is required for CwlJ-dependent spore germination [95]. Furthermore, previous studies report that SpoIVA and CotE, SpoVM, and SpoVID contribute to the formation of a spore coat scaffold during earlier stages of sporulation [8, 20, 21]. Similarly, CotE-controlled proteins, such as CotSA [8, 20, 21] are conserved in all spore-forming species analysed in this study.

The SpoIVA-dependent proteins CotJA, CotJB, and CotJC are also ubiquitous among the one hundred sixty-one spore-forming species analysed in this study. These proteins are necessary for the assembly of the basement layer of the spore coat [8, 96, 97]. Spore coat proteins that have a role in germination (allowing the passage of germinants) [8, 98, 99], such as the GerPA-GerPF proteins are well preserved in all spore-forming species addressed here. Another protein involved in germination and highly conserved is GerQ along with CwlJ (a cell wall hydrolase). GerQ is cross-linked in the inner layer of the spore coat and is necessary for the localization of CwlJ [8, 100, 101]. In *Bacillus* species, the spore coat protein Tgl responsible for the GerQ, YeeK, and SafA cross-linking [8, 100–102], is highly conserved.

We carried out an analysis to estimate the monophyly extent of different subgroups within the *Bacillus* genus with the main purpose of executing a detailed study of selection forces operating in these groups. The phylogenetic reconstruction allowed us to distinguish well-known groups inside *Bacillus* and also new groups. In a recent study, Patel & Gupta [103] grouped many known *Bacillus* species into distinct clades. Although various clades according to Patel and Gupta [103] coincide

with the groups found here (Subtilis, Cereus, Simplex, and Halodurans, which is named Alcalophilus clade), other clades show discordance (Firmus and Jeotgali clades) or are absent (Coagulans, Pumilus and Megaterium groups determined in this study). Under the premise that phylogenetic groups may reflect ecological fitness, we performed selection analysis to seek a relationship between the presence/absence of spore coat protein genes and selection forces operating on these genes in different phylogenetic groups within the *Bacillus* genus.

We have detected evidence of positive selection (episodic selection and/or balancing selection) in coat genes from all monophyletic groups of the *Bacillus* genus. Positively selected coat genes have an important role in the assembly of coat layers (e.g. morphogenetic coat genes) at initial and later stages and germination of the spore. The majority of spore coat genes reported in Table 2 have individual sites evolving under positive selection, according to MEME. We hypothesize that individual selected sites may play a key role in enzymatic activity or as protein-protein interaction modules during coat assembly, as suggested previously [91, 104, 105]. For example, protein-protein interactions necessary for spore assembly and germination have been described between SafA, CotE, SpoVID, GerQ, CwlJ, Tgl, YaaH, and SafA [95, 102, 104, 105]. We found that some, if not all, of these coat genes are positively selected in most monophyletic groups of the *Bacillus* genus. This emphasizes the importance of coat protein interactions. Furthermore, we found few spore coat genes under gene-wide positive selection, and they were different across *Bacillus* monophyletic groups analysed here. This different pattern of positively selected coat genes may suggest that some spore coat genes play critical roles in specific lineages.

A significant proportion of coat genes in the Subtilis, Methanolicus, Halodurans, and Coagulans have individual positively selected sites, suggesting that balancing selection may be working on these genes. The majority of coat genes of the Methanolicus, Halodurans, and Coagulans groups contained divergent sequences outside conserved domains. These results may suggest that high genetic variation is maintained through balancing selection, which in turn may provide significant survival advantages to spore survival and germination under different environmental conditions, as previously suggested [8, 25, 26, 85, 106, 107].

To reinforce our ideas about the evolutionary role of positively selected coat genes, we discuss the function and interaction of some spore coat genes under positive selection reported in Table 2. For instance, YheC and YheD are positively selected spore coat proteins that have an ATP binding domain and are part of the same operon [8]. YheD is located in the basement layer of the spore coat and is dependent on SpoIVA, whereas the localization of YheC has not yet been determined [8, 108]. During the initial stages of sporulation, YheD forms two rings that encircle the forespore [108]. In later stages of sporulation, the two rings disappear, and YheD is redistributed around the basement layer of the forespore [8, 108]. These spore coat proteins are important for the initial stages of sporulation in *B. subtilis* [8, 108] and they would also be key in the Subtilis, Cereus, Pumilus, and Simplex groups.

YutH and YsxE are bacterial spore kinase proteins located in the inner layer and are both SpoIVA- and SafA-dependent [8, 108, 109]. YutH and YsxE provide protection against lysozyme, hypochlorite, and predation to the spore [109]. Thus, these bacterial spore kinases are evolutionarily important for the survival of the spore in different environments [109]. Our selection pressure analyses revealed that these spore coat genes show positive selection at specific sites. These sites may be highly conserved motifs associated with likely enzymatic activity [109], or may exert an important function in the final protein product as interaction/binding partners. More studies are needed to test this hypothesis.

We have found that the spore kinase and morphogenetic coat gene of the outer layer, *cotH* shows evidence of positively selected individual sites in the Subtilis, Pumilus, and Simplex groups along with *cotB, cotG*, and/or *cotS*. It was previously reported that CotH phosphorylates CotB and CotG interacts with CotS, CotC, and CotU [91, 92]. The fact that genes encoding CotH and CotH-dependent proteins both have individual sites under diversifying selection highlights the importance of such sites as protein-protein interaction modules that promote adaptation to diverse environmental conditions when sporulation occurs [28].

The morphogenetic and crust genes *cotV, cotX* and *cotY, cotZ* involved in glycosylation state of the spore have been shown to share common domains and a functional dependence between them [94]. Moreover, coat genes with domains involved in glycosylation (e.g. glycosyl transferase), such as *cgeCDE*, *cgeAB* and transferases domains (e.g glycerophosphotransferase, nucleotidyltransferase), such as *spsI, spsB*, and *ytdA* influence the morphology and properties of the crust, thus affecting spore surface proteins [89, 94]. Our results show that in the Subtilis group, crust coat genes are highly conserved and have positively selected sites. Similarly, we show that several coat genes involved in the glycosylation in the outer layer of the spore have positively selected individual sites in the Simplex, Pumilus, Coagulans, Cereus, and Halodurans groups. This highlights the possibility that sequences that are necessary for assembly the crust or that influence spore surface properties, such as hydrophobicity and adhesion, are preserved. Furthermore, our selection results show that there are other coat genes (Table 2) with positively selected sites that have not been extensively studied and may exert important functions during coat assembly and spore germination necessary for spore adaptation to different environmental conditions.

Regarding the HGT results, we have found evidence of profuse HGT events of spore coat genes in all *Bacillus* monophyletic groups, except in the Megaterium, Pumilus, and Simplex groups. Thus, HGT could be involved in enabling spores of various species to better survive diverse environmental stresses. Most HGT events occurred at or near the branch tips of the reconciled gene-species phylogenetic trees, demonstrating a recent occurrence. This supports the idea

that the ability to form spores in Firmicutes (in *Bacilli* and *Clostridia*) is an ancestral feature as other researchers have stated [27, 85, 88]. Moreover, these HGT events are further confirmed by the presence of IS sequences in genomes of the recipient species.

Bacterial species that contain spore coat genes associated with HGT events may reflect a complex evolutionary history adapted to lineage-specific environmental conditions [26, 88]. This idea must be further explored by future studies on the evolutionary dynamics of these species. Nevertheless, we have found some spore coat genes that have undergone HGT events near the bottom of the reconciled phylogenetic trees. A previous study proposed that the putative coat genes *yraG* and *yraF* are present in the Subtilis group as part of the same operon and contain a domain that resemble a significant moiety of CotF. Therefore, the YraG and YraF proteins may be functionally relevant in the forespore [88]. Indeed, our HGT analyses confirm that *yraF* and *yraG* have been acquired at the bottom of the Subtilis group. Besides, the Subtilis group, *yraG* is present only in the Halodurans group. This may suggest that some coat genes not present within monophyletic groups of the *Bacillus* genus may have been lost at some point, as previously confirmed [88]. For example, *yra* genes are not present in the Pumilus group, the most closely-related group to Subtilis. Additional experiments beyond the aim of this study must explore HGT dynamics between monophyletic groups of the *Bacillus* genus.

In summary, we have found that the most conserved coat proteins are the ones with the most important function during the early and later stages of coat synthesis, assembly, and spore germination. This suggests that there is a well-conserved core of coat genes among all Bacillales, whereas other spore coat genes seem to be taxa-specific. Additionally, we found eight monophyletic groups within the *Bacillus* genus with a significant proportion of coat genes under positive diversifying selection and/or balancing selection, suggesting high genetic diversity that may confer unique adaptation to ensure spore survival and efficient germination. The spore coat genes with individual sites evolving under diversifying selection are likely to participate in protein-protein interactions during all stages of coat formation. Although most coat genes have been subjected to HGT events, they frequently occur near or at the tips of reconciled phylogenetic trees, thus supporting the idea of sporulation as an ancestral feature of *Bacillus*.

Funding information
The authors received no specific grant from any funding agency.

Acknowledgements
The authors wish to thank Dr Jean T. Greenberg for her help to improve the original manuscript, Drs Ezio Ricca and Patrick Eichenberger for technical examination and Dr Richard Losick for the contacts with specialists in spore coat proteins. Likewise, the authors are grateful to Dr Catherine Putonti for her valuable comments and advice on the initial design of the analyses.

Author contributions
Conceptualization: A. D., J. A. C. Data curation: J. A. C., H. S. M. Formal analysis: J. A. C., H. S. M., A. D. Investigation: J. A. C., H. S. M. Methodology: J. A. C., H. S. M., A. D. Software: J. A.C., H. S. M. Supervision: J. A. C. Validation: J. A. C. Writing – original draft: H. S. M. Writing – review and editing: J. A. C.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
No experiments involving animals or humans were performed for this study.

References

1. Maayer PD, Aliyu H, Cowan DA. Reorganising the order *Bacillales* through phylogenomics. *Syst Appl Microbiol* 2019;42:178–189.
2. Paul C, Filippidou S, Jamil I, Kooli W, House GL *et al*. Bacterial spores, from ecology to biotechnology. *Adv Appl Microbiol* 2019;106:79–111.
3. Suitso I, Jõgi E, Talpsep E, Naaber P, Lõivukene K *et al*. Protective effect by *Bacillus smithii* TBMI12 spores of Salmonella serotype enteritidis in mice. *Benef Microbes* 2010;1:37–42.
4. Wells-Bennik MHJ, Eijlander RT, den Besten HMW, Berendsen EM, Warda AK *et al*. Bacterial spores in food: survival, emergence, and outgrowth. *Annu Rev Food Sci Technol* 2016;7:457–482.
5. Kotiranta A, Lounatmaa K, Haapasalo M. Epidemiology and pathogenesis of *Bacillus cereus* infections. *Microbes Infect* 2000;2:189–198.
6. Mock M, Fouet A. Anthrax. *Annu Rev Microbiol* 2001;55:647–671.
7. Stenfors Arnesen LP, Fagerlund A, Granum PE. From soil to gut: *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol Rev* 2008;32:579–606.
8. Driks A, Eichenberger P. The spore coat. *Microbiol Spectr* 2016;4.
9. Setlow P. Spore resistance properties. *Microbiol Spectr* 2014b;2.
10. Beladjal L, Gheysens T, Clegg JS, Amar M, Mertens J. Life from the ashes: survival of dry bacterial spores after very high temperature exposure. *Extremophiles* 2018;22:751–759.
11. Klobutcher LA, Ragkousi K, Setlow P. The *Bacillus subtilis* spore coat provides "eat resistance" during phagocytic predation by the protozoan *Tetrahymena thermophila*. *Proc Natl Acad Sci U S A* 2006;103:165–170.
12. Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol Mol Biol Rev* 2000;64:548–572.
13. Setlow P, Wang S, Li Y-Q. Germination of spores of the orders *Bacillales* and *Clostridiales*. *Annu Rev Microbiol* 2017;71:459–477.
14. Moir A, Cooper G. Spore germination. *Microbiol Spectr* 2015;3.
15. Setlow P. Germination of spores of *Bacillus* species: what we know and do not know. *J Bacteriol* 2014a;196:1297–1305.
16. McKenney PT, Driks A, Eichenberger P. The *Bacillus subtilis* endospore: assembly and functions of the multilayered coat. *Nat Rev Microbiol* 2013;11:33–44.
17. Henriques AO, Moran CP. Structure, assembly, and function of the spore surface layers. *Annu Rev Microbiol* 2007;61:555–588.
18. Waller LN, Fox N, Fox KF, Fox A, Price RL. Ruthenium red staining for ultrastructural visualization of a glycoprotein layer surrounding the spore of *Bacillus anthracis* and *Bacillus subtilis*. *J Microbiol Methods* 2004;58:23–30.
19. Bozue JA, Welkos S, Cote CK. The *Bacillus anthracis* Exosporium: What's the Big "Hairy" Deal? *Microbiol Spectr* 2015;3.
20. McKenney PT, Eichenberger P. Dynamics of spore coat morphogenesis in *Bacillus subtilis*. *Mol Microbiol* 2012;83:245–260.
21. Bauer T, Little S, Stöver AG, Driks A. Functional regions of the *Bacillus subtilis* spore coat morphogenetic protein CotE. *J Bacteriol* 1999;181:7043–7051.
22. Ozin AJ, Henriques AO, Yi H, Moran CP. Morphogenetic proteins SpoVID and SafA form a complex during assembly of the *Bacillus subtilis* spore coat. *J Bacteriol* 2000;182:1828–1833.

23. Zilhão R, Naclerio G, Henriques AO, Baccigalupi L, Moran CP *et al*. Assembly requirements and role of CotH during spore coat formation in *Bacillus subtilis*. *J Bacteriol* 1999;181:2631–2633.
24. Krajčíková D, Forgáč V, Szabo A, Barák I. Exploring the interaction network of the *Bacillus subtilis* outer coat and crust proteins. *Microbiol Res* 2017;204:72–80.
25. McKenney PT, Driks A, Eskandarian HA, Grabowski P, Guberman J *et al*. A distance-weighted interaction map reveals a previously uncharacterized layer of the *Bacillus subtilis* spore coat. *Curr Biol* 2010;20:934–938.
26. Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI *et al*. Genomic determinants of sporulation in bacilli and clostridia: towards the minimal set of sporulation-specific genes. *Environ Microbiol* 2012;14:2870–2890.
27. Onyenwoke RU, Brill JA, Farahi K, Wiegel J. Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (*Firmicutes*). *Arch Microbiol* 2004;182:182–192.
28. Isticato R, Lanzilli M, Petrillo C, Donadio G, Baccigalupi L *et al*. *Bacillus subtilis* builds structurally and functionally different spores in response to the temperature of growth. *Environ Microbiol* 2020;22:170–182.
29. Zhu B, Stülke J. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res* 2018;46:D743–D748.
30. Pearson WR. An introduction to sequence similarity ("homology") searching. *Curr Protoc Bioinformatics* 2013;42:3.1.1–3.1.3.
31. Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. *Nat Commun* 2018;9:1–8.
32. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* 2016;44:D457–462.
33. Baesman SM, Stolz JF, Kulp TR, Oremland RS. Enrichment and isolation of *Bacillus beveridgei* sp. nov., a facultative anaerobic haloalkaliphile from Mono Lake, California, that respires oxyanions of tellurium, selenium, and arsenic. *Extremophiles* 2009;13:695–705.
34. Carneiro AR, Ramos RTJ, Dall'Agnol H, Pinto AC, de Castro Soares S *et al*. Genome sequence of *Exiguobacterium antarcticum* B7, isolated from a biofilm in ginger lake, King George Island, Antarctica. *J Bacteriol* 2012;194:6689–6690.
35. Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. *Sci Rep* 2016;6:1–10.
36. Lefort V, Longueville J-E, Gascuel O. Sms: smart model selection in PhyML. *Mol Biol Evol* 2017;34:2422–2424.
37. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
38. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ *et al*. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018;4:vey016.
39. Lartillot N, Philippe H. Computing Bayes factors using thermodynamic integration. *Syst Biol* 2006;55:195–207.
40. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol* 2011;60:150–160.
41. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol* 2018;67:901–904.
42. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995;90:773–795.
43. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ *et al*. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;25:1422–1423.
44. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res* 2010;38:W7–W13.
45. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P *et al*. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 2017;34:3299–3302.
46. Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ *et al*. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res* 2005;15:1553–1565.
47. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;123:585–595.
48. Castillo JA, Agathos SN. A genome-wide scan for genes under balancing selection in the plant pathogen *Ralstonia solanacearum*. *BMC Evol Biol* 2019;19:123.
49. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S *et al*. Gene-Wide identification of episodic selection. *Mol Biol Evol* 2015;32:1365–1371.
50. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K *et al*. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 2012;8:e1002764.
51. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 1997;13:555–556.
52. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007;24:1586–1591.
53. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 1998;148:929–936.
54. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 2005;22:1107–1118.
55. Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000;155:431–449.
56. Chen K, Durand D, Farach-Colton M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol* 2000;7:429–447.
57. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B *et al*. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 2012;28:i409–i415.
58. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *ICWSM* 2009;8:361–362.
59. Liu M, Li X, Xie Y, Bi D, Sun J *et al*. Iceberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* 2019;47:D660–D665.
60. Soares SC, Geyik H, Ramos RTJ, de Sá PHCG, Barbosa EGV *et al*. GIPSy: genomic island prediction software. *J Biotechnol* 2016;232:2–11.
61. Saggese A, Isticato R, Cangiano G, Ricca E, Baccigalupi L. CotG-like modular proteins are common among spore-forming *Bacilli*. *J Bacteriol* 2016;198:1513–1520.
62. Chen Y-G, Zhang Y-Q, Shi J-X, Xiao H-D, Tang S-K *et al*. *Jeotgalicoccus marinus* sp. nov., a marine bacterium isolated from a sea urchin. *Int J Syst Evol Microbiol* 2009;59:1625–1629.
63. Ishikawa M, Nakajima K, Itamiya Y, Furukawa S, Yamamoto Y *et al*. *Halolactibacillus halophilus* gen. nov., sp. nov. and *Halolactibacillus miurensis* sp. nov., halophilic and alkaliphilic marine lactic acid bacteria constituting a phylogenetic lineage in *Bacillus* rRNA group 1. *Int J Syst Evol Microbiol* 2005;55:2427–2439.
64. Zhang Y, Li X, Hao Z, Xi R, Cai Y *et al*. Hydrogen peroxide-resistant CotA and YjqC of *Bacillus altitudinis* spores are a promising biocatalyst for catalyzing reduction of sinapic acid and sinapine in rapeseed meal. *PLoS One* 2016;11:e0158351.
65. Manetsberger J, Ghosh A, Hall EAH, Christie G. Orthologues of *Bacillus subtilis* spore crust proteins have a structural role in the *Bacillus megaterium* QM B1551 spore exosporium. *Appl Environ Microbiol* 2018;84.
66. Fakhry S, Sorrentini I, Ricca E, De Felice M, Baccigalupi L. Characterization of spore forming *Bacilli* isolated from the human gastrointestinal tract. *J Appl Microbiol* 2008;105:2178–2186.

67. Tam NKM, Uyen NQ, Hong HA, Duc LH, Hoa TT *et al*. The intesitinal life cycle of *Bacillus subtilis* and close relatives. *J Bacteriol* 2006;188:2692–2700.
68. Jeyaram K, Romi W, Singh TA, Adewumi GA, Basanti K *et al*. Distinct differentiation of closely related species of *Bacillus subtilis* group with industrial importance. *J Microbiol Methods* 2011;87:161–164.
69. Rooney AP, Price NPJ, Ehrhardt C, Swezey JL, Bannan JD. Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. inaquosorum subsp. nov. *Int J Syst Evol Microbiol* 2009;59:2429–2436.
70. Aronson AI, Shai Y. Why *Bacillus thuringiensis* insecticidal toxins are so effective: unique features of their mode of action. *FEMS Microbiol Lett* 2001;195:1–8.
71. Bosma EF, van de Weijer AHP, Daas MJA, van der Oost J, de Vos WM *et al*. Isolation and screening of thermophilic *Bacilli* from compost for electrotransformation and fermentation: characterization of *Bacillus smithii* ET 138 as a new biocatalyst. *Appl Environ Microbiol* 2015;81:1874–1883.
72. Korneli C, David F, Biedendieck R, Jahn D, Wittmann C. Getting the big beast to work--systems biotechnology of *Bacillus megaterium* for novel high-value proteins. *J Biotechnol* 2013;163:87–96.
73. Vary PS, Biedendieck R, Fuerch T, Meinhardt F, Rohde M *et al*. *Bacillus megaterium*—From simple soil bacterium to industrial protein production host. *Appl Microbiol Biot* 2007;76:957–967.
74. Takami H, Horikoshi K. Analysis of the genome of an alkaliphilic *Bacillus* strain from an industrial point of view. *Extremophiles* 2000;4:99–108.
75. Khatri I, Sharma G, Subramanian S. Composite genome sequence of *Bacillus clausii*, a probiotic commercially available as Enterogermina®, and insights into its probiotic properties. *BMC Microbiol* 2019;19:307.
76. Schendel FJ, Bremmon CE, Flickinger MC, Guettler M, Hanson RS. L-lysine production at 50 degrees C by mutants of a newly isolated and characterized methylotrophic *Bacillus* sp. *Appl Environ Microbiol* 1990;56:963–970.
77. Tiago I, Pires C, Mendes V, Morais PV, da Costa MS *et al*. *Bacillus foraminis* sp. nov., isolated from a non-saline alkaline groundwater. *Int J Syst Evol Microbiol* 2006;56:2571–2574.
78. Alebouyeh M, Gooran Orimi P, Azimi-Rad M, Tajbakhsh M, Tajeddin E *et al*. Fatal sepsis by *Bacillus circulans* in an immunocompromised patient. *Iran J Microbiol* 2011;3:156–158.
79. Croce O, Hugon P, Lagier J-C, Bibi F, Robert C *et al*. Genome sequence of *Bacillus simplex* strain P558, isolated from a human fecal sample. *Genome Announc* 2014;2:e01241-14.
80. Kuisiene N, Raugalas J, Spröer C, Kroppenstedt RM, Chitavichius D. *Bacillus butanolivorans* sp. nov., a species with industrial application for the remediation of n-butanol. *Int J Syst Evol Microbiol* 2008;58:505–509.
81. Yumoto I, Hirota K, Yamaga S, Nodasaka Y, Kawasaki T *et al*. *Bacillus asahii* sp. nov., a novel bacterium isolated from soil with the ability to deodorize the bad smell generated from short-chain fatty acids. *Int J Syst Evol Microbiol* 2004;54:1997–2001.
82. Zhaxybayeva O, Doolittle WF. Lateral gene transfer. *Curr Biol* 2011;21:R242–R246.
83. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev* 2014;38:720–760.
84. Burrus V, Pavlovic G, Decaris B, Guédon G. Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 2002;46:601–610.
85. Galperin MY. Genome diversity of spore-forming *Firmicutes*. *Microbiol Spectr* 2013;1:TBS-0015–2012.
86. McPherson DC, Kim H, Hahn M, Wang R, Grabowski P *et al*. Characterization of the *Bacillus subtilis* spore morphogenetic coat protein CotO. *J Bacteriol* 2005;187:8278–8290.
87. Zhang J, Fitz-James PC, Aronson AI. Cloning and characterization of a cluster of genes encoding polypeptides present in the insoluble fraction of the spore coat of *Bacillus subtilis*. *J Bacteriol* 1993;175:3757–3766.
88. Ramos-Silva P, Serrano M, Henriques AO. From root to tips: sporulation evolution and specialization in *Bacillus subtilis* and the intestinal pathogen *Clostridioides difficile*. *Mol Biol Evol* 2019;36:2714–2736.
89. Shuster B, Khemmani M, Abe K, Huang X, Nakaya Y *et al*. Contributions of crust proteins to spore surface properties in *Bacillus subtilis*. *Mol Microbiol* 2019;111:825–843.
90. Freitas C, Plannic J, Isticato R, Pelosi A, Zilhão R *et al*. A protein phosphorylation module patterns the *Bacillus subtilis* spore outer coat. *Mol Microbiol* 2020;8.
91. Nguyen KB, Sreelatha A, Durrant ES, Lopez-Garrido J, Muszewska A *et al*. Phosphorylation of spore coat proteins by a family of atypical protein kinases. *Proc Natl Acad Sci U S A* 2016;113:E3482–E3491.
92. Saggese A, Scamardella V, Sirec T, Cangiano G, Isticato R *et al*. Antagonistic role of CotG and CotH on spore germination and coat formation in *Bacillus subtilis*. *PLoS One* 2014;9:e104900.
93. Krajcíková D, Lukácová M, Müllerová D, Cutting SM, Barák I. Searching for protein-protein interactions within the *Bacillus* subtilis spore coat. *J Bacteriol* 2009;191:3212–3219.
94. Bartels J, Blüher A, López Castellanos S, Richter M, Günther M *et al*. The *Bacillus subtilis* endospore crust: protein interaction network, architecture and glycosylation state of a potential glycoprotein layer. *Mol Microbiol* 2019;112:1576–1592.
95. Amon JD, Yadav AK, Ramirez-Guadiana FH, Meeske AJ, Cava F *et al*. SwsB and SafA are required for CwlJ-dependent spore germination in *Bacillus subtilis*. *J Bacteriol* 2019;202.
96. Henriques AO, Beall BW, Roland K, Moran CP. Characterization of *cotJ*, a sigma E-controlled operon affecting the polypeptide composition of the coat of *Bacillus subtilis* spores. *J Bacteriol* 1995;177:3394–3406.
97. Seyler RW, Henriques AO, Ozin AJ, Moran CP. Assembly and interactions of cotJ-encoded proteins, constituents of the inner layers of the *Bacillus subtilis* spore coat. *Mol Microbiol* 1997;25:955–966.
98. Butzin XY, Troiano AJ, Coleman WH, Griffiths KK, Doona CJ *et al*. Analysis of the effects of a *gerP* mutation on the germination of spores of *Bacillus subtilis*. *J Bacteriol* 2012;194:5749–5758.
99. Ghosh A, Manton JD, Mustafa AR, Gupta M, Ayuso-Garcia A *et al*. Proteins encoded by the *gerP* operon are localized to the inner coat in *Bacillus cereus* spores and are dependent on GerPA and SafA for assembly. *Appl Environ Microbiol* 2018;84.
100. Monroe A, Setlow P. Localization of the transglutaminase cross-linking sites in the *Bacillus subtilis* spore coat protein GerQ. *J Bacteriol* 2006;188:7609–7616.
101. Ragkousi K, Setlow P. Transglutaminase-mediated cross-linking of GerQ in the coats of *Bacillus subtilis* spores. *J Bacteriol* 2004;186:5567–5575.
102. Fernandes CG, Martins D, Hernandez G, Sousa AL, Freitas C *et al*. Temporal and spatial regulation of protein cross-linking by the pre-assembled substrates of a *Bacillus subtilis* spore coat transglutaminase. *PLoS Genet* 2019;15:e1007912.
103. Patel S, Gupta RS. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int J Syst Evol Microbiol* 2020;70:406–438.
104. Nunes F, Fernandes C, Freitas C, Marini E, Serrano M *et al*. SpoVID functions as a non-competitive hub that connects the modules for assembly of the inner and outer spore coat layers in *Bacillus subtilis*. *Mol Microbiol* 2018;110:576–595.
105. Pereira FC, Nunes F, Cruz F, Fernandes C, Isidro AL *et al*. A LysM domain intervenes in sequential protein-protein and protein-peptidoglycan interactions important for spore coat assembly in *Bacillus subtilis*. *J Bacteriol* 2019;201.

106. Abhyankar WR, Kamphorst K, Swarge BN, van Veen H, van der Wel NN *et al*. The influence of sporulation conditions on the spore coat protein composition of *Bacillus subtilis* spores. *Front Microbiol* 2016;7:1636.

107. Aronson A. Regulation of expression of a select group of *Bacillus anthracis* spore coat proteins. *FEMS Microbiol Lett* 2018;365.

108. Ooij Cvan, Eichenberger P, Losick R. Dynamic patterns of subcellular protein localization during spore coat morphogenesis in *Bacillus subtilis*. *J Bacteriol* 2004;186:4441.

109. Scheeff ED, Axelrod HL, Miller MD, Chiu H-J, Deacon AM *et al*. Genomics, evolution, and crystal structure of a new family of bacterial spore kinases. *Proteins* 2010;78:1470–1482.

110. Tirumalai MR, Rastogi R, Zamani N, Williams EO, Allen S *et al*. Candidate genes that may be responsible for the unusual resistances exhibited by *Bacillus pumilus* SAFR-032 spores. *Plos One* 2013;8:e66012.