Reorganising the order Bacillales through phylogenomics

Pieter De Maayer[1]*, Habibu Aliyu[2], Don A Cowan[3]

[1] School of Molecular & Cell Biology, Faculty of Science, University of the Witwatersrand, South Africa

[2]Technical Biology, Institute of Process Engineering in Life Sciences, Karlsruhe Institute of Technology, Germany

[3] Centre for Microbial Ecology and Genomics, University of Pretoria, South Africa

* Author for correspondence: pieter.demaayer@wits.ac.za

PDM: Pieter.demaayer@wits.ac.za

HA: Habibu.Aliqyu@partner.kit.edu

DAC: Don.Cowan@up.ac.za

Abstract

Bacterial classification at higher taxonomic ranks such as the class, order and family levels is currently reliant on phylogenetic analysis with a single gene, 16S rRNA, and for some taxa, the presence of shared phenotypic characteristics. However, these may not be reflective of the true genotypic and phenotypic relationships of taxa. This is evident in the order Bacillales, members of which are broadly defined as aerobic, spore-forming and rod-shaped bacteria. However, some of the taxa in this order are anaerobic, do not produce spores and are coccoid in morphology. 16S rRNA gene phylogeny has also not been able to elucidate the taxonomic positions of a number of families *incertae sedis* within this order. Recently developed phylogenetic approaches based on the whole genome may provide a more accurate means to resolve higher taxonomic levels. Here we have applied a suite of phylogenomic approaches to re-evaluate the taxonomy of 80 representative taxa of eight families (and six family *incertae sedis* taxa) within the order Bacillales. This showed several anomalies within the existing family and order level classifications including the existence of four and two distinct *Bacillaceae* and *Paenibacillaceae* "family" clades, respectively. The analysis also supported the movement of the *Staphylococcaceae* and *Listeriaceae* to the sister order Lactobacillales. Finally, we propose a consensus phylogenomic approach which may diminish algorithmic biases associated with single phylogenomic approaches and facilitate a more accurate classification of a broad range of bacterial taxa at the higher taxonomic levels.

Introduction

With its inception in 1872 [1], the genus *Bacillus* became a veritable "dumping ground" for aerobic, endospore-forming bacteria. The subsequent development of enhanced taxonomic methodologies, including morphological, physiological, chemical and molecular approaches, resulted in considerable taxonomic changes, with the development of novel genera, families and higher rank taxonomic delineations to incorporate members of this genus and novel isolates of aerobic endospore-forming bacilli [2, 3]. As such, the order Bacillales was proposed and subsequently validated on the basis of 16S rRNA gene sequence analysis [3-5]. Members of this order display a cosmopolitan functional and habitat distribution, including high temperature, extremely acidic, alkaline and hypersaline environments and incorporate clinical pathogens and strains of biotechnological value. One of the few consistent features of members of this order is the ability to form endospores, although exceptions do exist [4]. The order Bacillales currently comprises nine distinct families, *Alicyclobacillaceae*, *Bacillaceae*, *Listeriaceae*, *Paenibacillaceae*, *Pasteuriaceae*, *Planococcaceae*, *Sporolactobacillaceae*, *Staphylococcaceae* and *Thermoactinomycetaceae*, which incorporate a total of 133 genera. The largest of these is the family *Bacillaceae,* which includes 64 distinct genera [4, 6]. In addition, nine genera have been re-classified as families *incertae sedis* to reflect their ambiguous taxonomic delineation [4, 6]. In Bergey's Manual of Systematic Bacteriology Revised roadmap of the phylum Firmicutes, phylogenetic analysis on the basis of 16S rRNA gene sequences highlighted a number of discrepancies in the existing taxonomic scheme, with several of the families within the order Bacillales (e.g. *Bacillaceae, Paenibacillaceae, Thermoactinomycetaceae*) forming several distinct clades within the 16S rRNA gene phylogeny [4].

The poor resolving power at lower taxonomic levels, sequencing error, anomalies in sequences deposited in public nucleotide databases and the existence of multiple and often disparate copies in a single genome, have together sparked debate on the use of 16S rRNA gene sequences as a single marker for taxonomic delineation [7, 8]. However, given the paucity of comprehensive datasets of alternative taxonomic markers for polyphasic taxonomic approaches, 16S rRNA analysis remains the gold standard for higher level taxon delineation [4]. The development of next generation DNA sequencing technologies have led to a significant increase in the quantity and quality of sequence data at a greatly reduced cost [9]. The ability to utilise these improved technologies to sequence complete genomes has, more recently, led to their use for taxonomic delineation. Comparisons of genome sequences allows the establishment of a set of conserved core genes or proteins which can be used to define evolutionary relationships [10]. A subset of these proteins, the ribosomal biogenesis and maintenance proteins (RP), can also be extracted from the genome and used to construct phylogenies with far greater taxonomic resolution than is achievable using single gene phylogenies [11]. Furthermore, several phylogenomic metrics have been developed, such as average nucleotide identity (ANI) and average amino acid identity (AAI) values, digital DNA-DNA hybridization (dDDH) and tetranucleotide signature frequency correlation coefficient (TETRA) values. These phylogenomic metrics have been shown to correlate well with laboratory-based DNA-DNA hybridization (DDH) analyses which form the basis of valid species circumscription, with 95-96% ANI, 90% AAI, 70% dDDH values and 0.99 TETRA values approximating the wet-lab DDH species boundary value of 70% [12-14]. In contrast to laboratory-based DDH approaches, these tools provide highly reproducible results which can be readily validated by other researchers [15]. However, their use is largely restricted to taxonomic delineation at the species level, while limited tools are available for delineation at higher taxonomic levels. One exception is the Percentage of Conserved Proteins (POCP) approach, which has been proposed for circumscription at the genus level [16].

Here we have applied a suite of phylogenomic approaches to re-examine the taxonomic status of the order Bacillales, using a comprehensive set of strains representative of seventy-seven distinct genera and eight families within the order. Furthermore, we have adopted a consensus or polyphasic system incorporating the data from the different phylogenomic analyses to diminish algorithmic biases associated with each individual approach. Together, these data support several re-classifications which should be considered for the order Bacillales.

Materials and Methods

Bacillales genomes

The genomes of the type strains of type species of sixty-seven distinct genera within the order Bacillales (Table 1) were retrieved from the NCBI Genbank assembly database [17]. For a further nineteen genera, genomes were available for type strains of other species within the genera (Table 1). These were only incorporated in the analysis if their 16S rRNA gene sequences shared nucleotide identities with those of the type strain of the type species for the genus of > 95 %, which represents the prescribed cut-off value for incorporation in the same genus [18]. In addition, the genomes of type strains of the type genera for five families from the sister order Lactobacillales were included in the analyses, as well as the genome of *Clostridium butyricum* DSM 10702^T (order Clostridiales), which was used as outgroup for all analyses. The 16S rRNA gene sequences for each analysed strain were obtained from the NCBI nucleotide database [17]. Metadata relating to the morphology (cell shape), physiology (motility, ability to form spores), growth conditions (salt concentrations, temperature and pH range) and sources of isolation were obtained from the original species descriptions (Supplementary Table S1). All genomes were structurally annotated using Prokaryotic Genemark.hmm [19]. The resultant amino acid sequences were subjected to Orthofinder [20]. The ribosomal protein (RP) amino acid sequences from *Bacillus subtilis* 168 were extracted from the RiboDB database v 1.4.1 [21]. Local BlastP and tBlastN analyses were performed in Bioedit v. 7.2.5 [22] with this dataset to identify orthologous ribosomal proteins in each of the compared genomes.

Phylogenomic metric calculations

Average nucleotide identity values were calculated using OrthoANI (Supplementary Table S2) [23]. This algorithm calculates ANI values on the basis of orthologous fragment pairs only and is thus reported to provide more accurate values when genomes are reciprocally compared than traditional ANI calculations [23]. Average amino acid identity (AAI) values were determined using the aai.rb script in the enveomics package, using the two-way AAI option (Supplementary Table S3) [24]. Tetranucleotide usage (TETRA) pattern analyses (Supplementary Table S4) were performed with Jspecies v. 1.2.1 [25]. Using the Genome-to-Genome Distance Calculator (GGDC 2.0) with formula 2, digital DNA-DNA hybridization values (dDDH) values were calculated (Supplementary Table S5) [14]. These metrics are all applied to distinguish strains at the species level. A newly developed metric, Percentage Of Conserved Proteins (POCP), can be used for genus level circumscription, where POCP values < 50% indicate organisms belong to distinct genera [16]. POCP values were calculated using the formula [(C1 + C2)/(T1 + T2) x 100, where C1 and C2 represents the number of conserved proteins (amino acid identity > 40% with alignment coverage > 50%) and T1 and T2 represent the total number of proteins encoded on each genome (Supplementary Table S6) [16].

Phylogenetic analyses

The 16S rRNA gene sequences for each of the compared genomes were aligned using the M-Coffee algorithm within the T-Coffee package, which incorporates several different multiple sequence alignment methods and combines the results in a single optimal alignment [26]. The

114 core proteins (CPs) identified with Orthofinder, and the 45 conserved ribosomal proteins (RPs) were individually aligned using M-Coffee [24]. Subsequently the CP and RP alignments were concatenated in two distinct CP and RP datasets and poorly aligned positions were eliminated using GBlocks v 0.91b [27]. The 16S rRNA (1,061 nucleotide sites) and trimmed CP (22,049 amino acid sites) and RP (5,250 amino acid sites) alignments were used to construct Maximum Likelihood (ML) phylogenies using the PhyML 3.0 server [28] with the Smart Model Selection (SMS) option to determine the optimal substitution model for the alignment on the basis of the Akaike Information Criterion [29]. For the 16S rRNA gene ML phylogeny bootstrap analysis (n = 1,000) was performed to provide branch support, while for the CP and RP ML phylogenies branch support was determined using the Shimodaira-Hasegawa approximate Likelihood Ratio Test (SH-aLRT) method [30]. The percentage (ANI, AAI, dDDH, POCP) and decimal regression (TETRA) values from the phylogenomic metric calculations were converted into distance values [1- (% similarity/100)]. The resulting distance matrices were used to generate Neighbour Joining (NJ) trees with PHYLIP v 3.695 [31]. The tree topologies of the three sequence-based (CP, RP and 16S rRNA) and five phylogenomic metric (AAI, OrthoANI, POCP, TETRA) distance-based trees were pair-wise compared using Compare2Trees [32]. This program matches up the branches between two compared trees and calculates a topological score expressed as the percentage of branches which are identical between two given trees [32]. Six phylogenies (CP, RP, 16S rRNA, OrthoANI, AAI and POCP) which shared an overall topology score of >50% were selected and used to generate a consensus dendrogram using the Consense algorithm in PHYLIP [31].

Results

Constructing a consensus phylogenomic framework for the order Bacillales

Representatives of 86 out of 144 genera within the order Bacillales for which genome sequences are available were analysed using core protein (CP) and ribosomal protein (RP) phylogenies as well as a range of phylogenomic metrics. Initial comparison of the CP and 16S rRNA phylogenies (Figure 1) showed substantial congruence between the two phylogenies (tree topology score: 77.6%). Differences between these phylogenies can largely be attributed to swapping of branches among the deeper nodes, although there is some swapping at the higher taxonomic ranks, particularly when considering the family *incertae sedis* strains (Figure 1). Superior congruence was observed between the CP and RP phylogenies (Figure 1 and

Supplementary Figures S1), with a tree topology score of 91.2% (Figure 2). Here, family *incertae sedis* members also clustered differently between the two phylogenies. The NJ trees constructed on the basis of the phylogenomic metrics AAI, POCP and OrthoANI (Supplementary Figures S2, S3 and S4) shared tree topology scores ranging from 67.6 to 81.5% among them and the CP, RP and 16S rRNA phylogenies (Figure 2). By contrast, the TETRA and dDDH tree topologies (Supplementary Figures S5 and S6) were incongruent with those of the other phylogenomic metric and alignment-based methodologies, with average tree topology scores of 40.4 and 25.3%, respectively, suggesting that these approaches are not suitable for delineation of higher taxonomic ranks (Figure 2). The other phylogenomic metric- and sequence alignment-based phylogenies may also be subject to biases resulting in artefactual groupings which may be responsible for the incongruencies observed among the tree topologies. These include compositional biases where phylogenetically disparate taxa are grouped together due to similar nucleotide or protein compositions, heterotachy linked to variations in the evolutionary rates of particular amino acids and nucleotides in a given protein or gene, as well as long-branch attraction due to the higher evolutionary rates that may occur in some unrelated taxa [10, 33]. To minimise the effects of these biases, a consensus tree was constructed on those phylogenomic metric and alignment-based phylogenies which shared tree topology scores of >65.0%, namely the CP, RP, 16S rRNA, AAI, OrthoANI and POCP. The resultant cladogram (Figure 3) shared tree topology scores of >78.3% with these six analyses. In particular, the topology of the consensus tree is most congruent with the CP (92.7%) and RP (91.2%) phylogenies. Although this cladogram based on the majority rule consensus of the different phylogenies does not consider the evolutionary distances between taxa, it gives a robust indication of the clustering patterns of the different taxa with reduced biasing effects compared to the individual phylogenies. The consensus tree also shares 78.9% tree topology score with the 16S rRNA gene phylogeny. As such, the family-level delineations in the consensus tree largely agree with the classifications at the family level on the basis of 16S rRNA gene sequences [4]. This is true for several families, including the *Listeriaceae*, *Staphylococcaceae*, *Planococcaceae*, *Thermoactinomycetaceae* and *Alicyclobacillaceae*, as well as the order Lactobacillales which was included as outgroup. However, several discrepancies can be observed both in the consensus tree as well as the individual sequence alignment- and phylogenomic metric-based analyses.

The family *Bacillaceae* comprises four distinct "family" clades

In the "revised road map to the phylum Firmicutes", it was shown that on the basis of 16S rRNA gene phylogeny, the family *Bacillaceae* occurred in two distinct clades, where the first clade (*Bacillaeceae* 1) incorporated the type genus *Bacillus*, along with *Geobacillus* and *Anoxybacillus*, while the second (*Bacillaceae* 2) was comprised of eighteen genera [4]. Our phylogenomic analyses show that the majority of taxa belonging to this original family fall into four well-supported clades (Figure 3). The first clade, *Bacillaceae* 1 incorporates six genera, *Aeribacillus, Anoxybacillus, Caldibacillus, Geobacillus* and *Parageobacillus* along with the type genus strain *B. subtilis* DSM 10^{T}. A second well-supported clade (*Bacillaceae* 2) is comprised of twenty genera, while the third (*Bacillaceae* 3) and fourth (*Bacillaceae* 4) incorporate eight and two genera, respectively (Figure 3). Among the taxa of the *Bacillaceae* 3, *Sinobaca qinghaiensis* DSM 17008^{T} was previously described as a member of the family *Sporolactobacillaceae* [34]. The four *Bacillaceae* clades in the consensus tree are fully supported by the CP, RP and AAI trees (Figure 3; Supplementary Figures S1 and S2), with intra-clade AAI values ranging between 55.15 and 60.46% and inter-clade values between 48.43 and 52.88%. While these four clades are also evident in the OrthoANI tree (Supplementary Figure S4), the *Bacillaceae* 1 clade in this tree also incorporates *Domibacillus robiginosus* DSM 25058^{T} and *Jeotgalibacillus alimentarius* DSM 18867^{T}, which form part of the *Planocococcaceae* clade in the CP, 16S rRNA, AAI and consensus trees. In the POCP tree (Supplementary Figure S3), *D. robiginosus* DSM 25058^{T} again clusters with the *Bacillaceae* 1 clade, while the *Bacillaceae* 1 strain *C. debilis* DSM 16016^{T} clusters with the *Sporolactobacillaceae*. In the original descriptive paper, *D. robiginosus* was characterised as a borderline strain between the *Bacillaceae* and *Planococcaceae* [35], which may explain the clustering of this strain with the *Bacillaceae* 1 in the OrthoANI and POCP tree (Supplementary Figures S3 and S4) and with the *Planococcaceae* in the CP, 16S rRNA and AAI trees (Figure 1 and Supplementary Figure S1). On the basis of the consensus tree, however, *D. robiginosus* DSM 25058^{T} clusters on the outside of the *Planococcaeae* clade along with *J. alimentarius* DSM 18867^{T}, which forms a clade with the other *Planococcaceae* in five (all except the OrthoANI tree) trees, and as such, *D. robiginosus* should be considered as a member of the family *Planococcaceae*. The incorporation of additional *Domibacillus* species in the phylogenomic analysis may further clarify its taxonomic position.

Analysis of the available species description metadata revealed few distinguishing morphological and phenotypic characteristics for the four *Bacillaceae* clades. The *Bacillaceae* clades 1-3 incorporate both facultative and obligate aerobic taxa, while *Bacillaceae* clade 3 incorporates the strict anaerobe *A. arseniciselenatis* DSM 15340^T and clade 4 includes the microaerophilic *Tepidibacillus decaturensis* DSM 103037^T and the strict anaerobe *Vulcanibacillus modesticaldus* DSM 14931^T. The *Bacillaceae* clades 2-3 also include both motile and non-motile rod-shaped, spore-forming and non-spore forming taxa and some taxa in these clades can tolerate higher salt concentrations (Supplementary Table S1). The *Bacillaceae* 1 (with the exception of *B. subtilis* DSM 10^T) and 4 clades are composed primarily of thermophilic taxa, while the *Bacillaceae* 2 and 3 clades contain mesophilic taxa. Our phylogenomic analyses, however, clearly differentiate the *Bacillaceae* into four distinct clades, which may represent four distinct families. Comprehensive phenotypic analyses may reveal additional distinguishing features for this taxonomic delineation.

While twenty-five of the compared *Bacillaceae* are placed within the *Bacillaceae* 1-4 clades, a further four strains are placed elsewhere in the consensus tree (Figure 3). Aside from *D. robiginosus* DSM 25058^T which, as discussed above, clusters on the outside of the *Planococcaceae* clade, three other strains, namely *Viridibacillus arvi* DSM 16317^T, *Lysinibacillus boronitolerans* DSM 17140^T and *Psychrobacillus psychrodurans* DSM 11713^T consistently cluster with the *Planococcaceae* in the CP, RP, 16S rRNA, AAI, POCP and OrthoANI trees (Figure 1 and Supplementary Figures S1, S2, S3 and S4) and should thus be considered as members of this family. The incorporation of four members of the *Bacillaceae*, along with two members of the families *incertae sedis* (discussed below) in the family *Planococcaceae* suggest that this family is much larger than previously appreciated. As observed with the *Bacillaceae*, available metadata reveal no consistent phenotype that may distinguish members of this family, as it includes both obligate aerobic and facultative anaerobic, coccoid or rod-shaped, spore and non-spore forming, motile and non-motile, as well as mesophilic and some psychrophilic taxa. Further phenotypic characterization inclusive of the *Bacillaceae* and family *incertae sedis* taxa that should be reclassified to the *Planococcaceae* may identify features unique to this family.

The *Paenibacillaceae* comprise two distinct ‘family’ clades

Seven representative strains of the family *Paenibacillaceae* were included in this study. These representatives were observed to form part of two distinct clades in the consensus tree, the *Paenibacillaceae* 1 (including the type genus *Paenibacillus*) and 2, comprising of five and two strains, respectively (Figure 3). These distinct clades are spatially separated in four of the six phylogenies incorporated in the consensus tree, namely the CP, RP, 16S rRNA and AAI trees (Figure 1 and Supplementary Figures S1 and S4). In the other trees (POCP and OrthoANI), the *Paenibacillaceae* 1 and 2 strains form part of a single clade, but occur in two distinct branches within these trees. The *Paenibacillaceae* may thus be split into two distinct family clades; the *Paenibacillaceae* 1 (*Cohnella thermotolerans* DSM 17683^{T}, *Thermobacillus composti* DSM 18247^{T}, *Fontibacillus panacisegetis* DSM 28129^{T}, *Paenibacillus polymyxa* DSM 36^{T} and *Saccharibacillus sacchari* DSM 19268^{T}) and the *Paenibacillaceae* 2 (*Aneurinibacillus migulanus* DSM 2895^{T} and *Brevibacillus formosus* DSM 9885^{T}). This partially agrees with the previous observation of two monophyletic clades being formed in a 16S rRNA phylogeny, although *Brevibacillus* clustered with *Paenibacillus, Thermobacillus* and *Cohnella*, while *Aneurinibacillus* clustered with *Ammoniphilus* and *Oxalophagus* (for which no genomes are available) in the earlier study [36]. While both the *Paenibacillaceae* 1 and 2 clades incorporate aerobic, spore-forming, motile rods, their separation is supported by genomic characteristics (Table 1). The *Paenibacillaceae* 1 have genomes which are, on average, 945 kb larger, and with a mean G+C content of 6.54% below that of the *Paenibacillaceae* 2 (average genome size: 5.34 Mb; average G+C content: 51.78%).

The families *Listeriaceae* and *Staphylococcaceae* belong to the order Lactobacillales

The consensus tree shows the taxa belonging to the families *Listeriaceae* (*Brochothrix thermosphacta* DSM 20171^{T} and *Listeria monocytogenes* DSM 20600^{T}) and *Staphylococcaceae* (*Staphylococcus aureus* DSM 20231^{T}, *Aliicoccus persicus* DSM 28306^{T}, *Salinicoccus luteus* DSM 17002^{T} and *Jeotgalicoccus psychrophilus* DSM 19085^{T}) clustering with representatives of the order Lactobacillales (Figure 3). This clustering pattern is also observed in the CP and 16S rRNA phylogenies (Figure 1) and the POCP and OrthoANI trees (Supplementary Figures S3 and S4). In the RP phylogeny (Supplementary Figure S1) the *Listeriaceae* cluster with the Lactobacillales while the *Staphylococcaceae* form a separate clade closer to the *Planococcaceae.* In the AAI tree (Supplementary Figure S2), the Lactobacillales, *Listeriaceae* and *Staphlyococcaceae* form part of three distinct clades. The co-clustering of these two Bacillales families with the Lactobacillales observed in most of the trees is further supported by both phenotypic and genotypic features of members of both of these families. As with the Lactobacillales, the *Staphylococcaceae* and *Listeriaceae* families incorporate taxa with a coccoid morphology, rather than the rod-shaped morphology typical of the Bacillales. The genomes of the *Listeriaceae* (2.68 Mb average) and *Staphylococcaceae* (2.43 Mb average) are more similar in size to those of the Lactobacillales (2.26 Mb average) than the Bacillales

(3.76 Mb average) and the G+C contents of the *Listeriaceae* (37.17%) and *Staphylococcaceae* (40.18%) are more similar to those of the Lactobacillales (39.63%) than those of the Bacillales (44.0%). Taken together, these data provide support for the reclassification of the families *Listeriaceae* and *Staphylococcaceae* within the order Lactobacillales, rather than within the order Bacillales.

Taxonomic placement of the Bacillales families *incertae sedis*

The order Bacillales incorporates nine genera which have not been taxonomically classified at the family level. Six representatives of these family *incertae sedis* taxa were incorporated in the phylogenomic analyses. *Solibacillus silvestris* DSM 12223^T and *Rummeliibacillus pycnus* DSM 15030^T cluster with the *Planococcaceae* in all phylogenomic analyses. The former strain consistently clusters with *Caryophanon latum* DSM 14151^T, while the latter clusters with *Kurthia zopfii* DSM 20580^T. As such, these taxa, along with four members of the family *Bacillaceae*, should be reclassified to the family *Planococcaceae.* The family XI *incertae sedis* strain *Gemella haemolysans* ATCC 10379^T clusters on the outside of the *Staphylococcaceae* branch, where it was previously assigned [37], in the CP, RP and 16S rRNA phylogenies and the OrthoANI tree (Figure 1 and Supplementary Figures S1 and S4). This strain has a relatively small genome (1.92 Mb in size) and its cells are coccoid in shape, and as such, this strain should be reclassified along with the *Listeriaceae* and *Staphylococcaceae* as a member of the order Lactobacillales, as has previously been suggested in a broader-scale phylogenomic analysis of the phylum Firmicutes [38]. Aside from its clustering on the outside of the *Staphylococcaceae* clade, *Gemella* also clusters with the Lactobacillales taxa in the POCP tree (Supplementary Figure S3) and therefore may form a distinct family clade in the order Lactobacillales.

The taxonomic positions of the other three family *incertae sedis*, namely *Exiguobacterium aurantiacum* DSM 6208^T, *Thermicanus aegypticus* DSM 12793^T and *Desulfuribacillus alkaliarsenatis* DSM 24608^T are less clearly defined. The consensus phylogeny places *E. aurantiacum* DSM 6208^T (Family XII) with the *Sporolactobacillaceae*, a clustering pattern also observed in the CP and RP phylogenies (Figure 1 and Supplementary Figure S1), but distinct positions are observed for this strain in the POCP and OrthoANI (clusters with Lactobacillales, *Staphylococcaceae* and *Listeriaceae*), AAI (clusters separately) and 16S rRNA (clusters with the *Paenibacillaceae* 2) trees. Similarly, *T. aegypticus* DSM 12793^T (Family X) clusters with the *Alicyclobacillaceae* in the OrthoANI and AAI trees, with the *Bacillaceae* 4 in the POCP tree, on its own in the RP phylogeny and with the *Thermoactinomycetaceae* in the 16S rRNA phylogeny. The CP phylogeny show this strain to be associated with the family *Paenibacillaceae,* where it was classified before transfer of *Thermicanus* to Family X *incertae sedis* [37]. *Desulfuribacillus alkaliarsenatis* DSM 24608^{T} clusters with the *Bacillaceae* 4 strains in the AAI and POCP trees (Supplementary Figures S2 and S3). In the CP, RP, 16S rRNA phylogeny and the OrthoANI tree (Figure 1 and Supplementary Figures S1 and S4), this strain forms a distinct branch near the base of the tree, suggesting it forms part of a novel family distantly related to the other Bacillales. The disparate clustering for these family *incertae sedis* taxa between the different phylogenies and phylogenomic metric trees indicate that at least some of these techniques are prone to compositional biases and long-branch attraction which may not be effectively diluted in the consensus phylogeny, and as such their taxonomy cannot currently be accurately resolved. However, the incorporation of additional genomes, including those of other species for these genera, in this phylogenomic approach may elucidate the correct taxonomic position of these taxa.

Discussion

The accurate classification of bacterial taxa at the genus and species levels remains a difficult task. This is even more complicated at the higher taxonomic levels, including the family and order levels, where techniques such as DNA-DNA hybridization are not feasible. Furthermore, inconsistencies are frequently observed within 16S rRNA gene phylogenies and shared phenotypic characteristics between these higher level taxonomic ranks, resulting in a highly subjective classification scheme [12]. This is highlighted by the family *Bacillaceae* which was circumscribed to incorporate aerobic or facultatively anaerobic, chemo-organotrophic, endospore-forming Gram-positive rods [39]. This family was consolidated on the basis of 16S rRNA phylogeny [4], even though it incorporates taxa which are strictly anaerobic, non-spore-forming and have a coccoid morphology. As the 16S rRNA gene may not be representative of the phenotype or genotype of an organism [12], the development of a more robust system for higher taxonomic rank classification is an imperative.

Here we have employed phylogenomic methodologies, incorporating techniques that consider the genome at both the nucleotide and amino acid levels, to address the taxonomy of the order Bacillales. This genome-level analysis has revealed several taxonomic considerations, including the clustering of members of the family *Bacillaceae* in three distinct “family” clades, the existence of two *Paenibacillaceae* “family” clades and the grouping of the families *Staphylococcaceae* and *Listeriaceae* with the order Lactobacillales, rather than the order

Bacillales. Discrepancies in the results from the different phylogenomic analyses highlight the danger of using a single methodology to resolve taxonomy. For this reason, we have adopted a consensus phylogenomic approach, using polyphasic genomic methods for bacterial classification and diluting the effects of algorithmic biases associated with individual techniques. This robust, yet simple, methodology can easily be applied to resolve the classification of a broad range of bacterial taxa at all taxonomic levels.

One significant issue observed in these analyses was the taxonomic position of members of the Bacillales families *incertae sedis*, which had a profound influence on the variability of the phylogenomic analyses. The different clustering patterns observed may be due to the effects of long-branch attraction and compositional biases associated with the individual phylogenomic metrics. Their positions may be more accurately resolved by the incorporation of additional taxa within the families *incertae sedis*. By the same token, the inclusion of multiple species from the different genera incorporated in the analyses may better resolve the order Bacillales at the family, genus and species levels.

Figure and Tables

Table 1: Genomic properties of the compared Bacillales taxa. The strains incorporated in this study are listed along with the NCBI accession numbers for the genomes, number of contigs, genome sizes, average genomic G+C content and number of proteins encoded on the genome. For some genera the genome of the type strain of the type species was unavailable. The relative 16S rRNA sequence identity of a representative species type strain to the type species type strain for that genus is shown. Only those strains with 16S rRNA gene sequence identities >95% were selected for incorporation in the analyses.

Figure 1: Comparison of a core protein (CP) phylogeny and 16S rRNA gene phylogeny of the order Bacillales. Maximum likelihood phylogenies were constructed on the basis of 114 concatenated core proteins (22,049 amino acid sites) and the 16S rRNA gene (1,061 nucleotide sites) using PhyML-SMS [28, 29], with the optimal substitution models of LG+G+I and GTR+G+I, respectively. The representative taxa are coloured according to their original family designation where members of the Lactobacillales are coloured in dark blue, *Listeriaceae* in pink, *Staphylococcaceae* in blue, *Planococcaceae* in purple, *Bacillaceae* in red, *Sporolactobacillaceae* in grey, *Planococcaceae* in green, *Thermoactinomycetaceae* in maroon,

Alicyclobacillaceae in teal and families *incertae sedis* in olive. Solid connecting lines indicate where taxa cluster in the same "Family" clades in both the CP and 16S rRNA gene phylogenies, while dotted lines indicate where distinct clustering patterns are observed. Bootstrap values (n = 1,000 replicates) and Shimodaira-Hasegawa approximate Likelihood Ratio Test (SH-aLRT) values [30] are shown on the branches of the 16S rRNA and CP phylogenies, respectively.

Figure 2: Tree topology congruence scores (%) for the different phylogenomic analyses. Tree topology congruence scores were determined using Compare2Trees [32]. The dendrogram on the left was constructed on the basis of the % topology congruence converted into distance values. Phylogenomic approaches highlighted in green were utilised to construct the consensus phylogeny, while those highlighted in red were excluded.

Figure 3: Consensus tree constructed on the basis of six alignment- and phylogenomic metric-based approaches. Taxa are coloured according to their original family designations (as in Figure 1). The consensus tree was constructed using the Consense script in PHYLIP v 3.695 [31]. Values on the branches indicate the number of times taxa nodes co-occur in branches in the six trees (CP, RP, 16S rRNA, AAI, OrthoANI, POCP) incorporated in the consensus tree. For example 4/6 denotes congruence in four out of the six approaches. Only those clades supported in ≥ 3/6 trees are shown.

Supplementary Data

Supplementary Table S1: Metadata of the compared Bacillales taxa. The metadata relating to the growth conditions (range and optimum salt concentrations, range and optimum pH, range and optimum temperature), physiology (O_2 requirement, motility, spore formation) and source of isolation were derived from the original species and genus descriptions.

Supplementary Table S2: Average nucleotide identity (ANI) values. ANI values were calculated with OrthoANI [23] and are expressed as a percentage.

Supplementary Table S3: Average amino acid identity (AAI) values. Two-way AAI values were calculated using the aai.rb script which is part of the Enveomics package [24] and are expressed as a percentage.

Supplementary Table S4: Tetranucleotide signature frequency correlation coefficient (TETRA) values. The values are expressed as a proportion and were calculated using Jspecies v 1.2.1 [25].

Supplementary Table S5: digital DNA-DNA Hydridization (dDDH) values. These values were derived through the Genome-to-Genome Distance Calculator (GGDC) 2.1 server [14] and are expressed as a percentage.

Supplementary Table S6: Percentage Of Conserved Proteins (POCP) values. Orthologous proteins conserved among pair-wise compared genomic protein datasets were used to calculated POCP by the formula (C1 + C2)/(T1 + T2), with C1 and C2 representing the number of conserved proteins and T1 and T2 the total number of proteins per genome [16].

Supplementary Figure S1: Ribosomal protein ML phylogeny. The Maximum likelihood phylogeny was constructed on the basis of 45 concatenated conserved ribosomal proteins (5,250 amino acid sites) using PhyML-SMS [28, 29], with the optimal predicted amino acid substitution model LG +G + I. Taxa are coloured according to their original family designations (as in Figure 1). SH-aLRT support values [30] are shown on the branches.

Supplementary Figure S2: NJ tree on the basis of AAI values. AAI values were converted into distance values by the formula: 1 – (% AAI/100) and used to construct a NJ tree using PHYLIP v 3.695 [31]. Taxa and branches are coloured according to their original family designations (as in Figure 1).

Supplementary Figure S3: NJ tree on the basis of POCP values. POCP values were converted into distance values by the formula: 1 – (% POCP/100) and used to construct a NJ tree using PHYLIP v 3.695 [31]. Taxa and branches are coloured according to their original family designations (as in Figure 1).

Supplementary Figure S4: NJ tree on the basis of OrthoANI values. OrthoANI values were converted into distance values by the formula: 1 – (% OrthoANI/100) and used to construct a NJ tree using PHYLIP v 3.695 [31]. Taxa and branches are coloured according to their original family designations (as in Figure 1).

Supplementary Figure S5: NJ tree on the basis of TETRA values. TETRA values were converted into distance values by the formula: 1 – TETRA score and used to construct a NJ tree using PHYLIP v 3.695 [31]. Taxa and branches are coloured according to their original family designations (as in Figure 1).

Supplementary Figure S6: NJ tree on the basis of dDDH values. dDDH values were converted into distance values by the formula: 1 – (% dDDH/100) and used to construct a NJ tree using PHYLIP v 3.695 [31]. Taxa and branches are coloured according to their original family designations (as in Figure 1).

Authors' contributions

PDM, HA and DAC conceived the study. PDM and HA performed data analyses. PDM wrote the original manuscript and all authors contributed to the final version. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

PDM acknowledges funding from the University of the Witwatersrand Faculty Research Council (FRC) and HA acknowledges the University of Pretoria for postdoctoral funding.

References

[1] Cohn, F. (1872) Untersuchungen über Bakterien, Beitr. Biol. Pflanz. 1, 127-224.

[2] Fritze, D. (2004) Taxonomy of the genus *Bacillus* and related genera: the aerobic endospore-forming bacteria. Phytopathology 94, 1245-1248.

[3] Garrity, G.M., Bell, J.A., Lilburn, T.G. (2004) Taxonomic outline of the prokaryotes. In: Bergey's Manual of Systematic Bacteriology 2nd Edition, Springer-Verlag, New York, pp. 1-399.

[4] Ludwig, W., Schleifer, K.H., Whitman, W.B. (2009) Revised road map to the phylum *Firmicutes*. In: De Vos, P., Garrity, G.M., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.H., Whitman W.B. (Eds.) Bergey's Manual of Systematic Bacteriology 3, 2nd ed., Springer-Verlag, New York, pp. 1-17.

[5] Winslow, C.E., Broadhurst, J., Buchanan, R.E., Krumwiede, C., Rogers, L.A., Smith, G.H. (1920) The families and genera of the bacteria: final report of the Committee of the Society of American Bacteriologists on Characterization and Classification of Bacterial Types. J. Bacteriol. 5, 191-229.

[6] Parte, A.C. (2018) LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. Int. J. Syst. Evol. Microbiol. 68, 1825-1829.

[7] Janda, J.M., Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic alboratory: pluses, perils and pitfalls. J. Clin. Microbiol. 45, 2761-2764.

[8] Větrovský, T., Baldrian, P. (2013) The variability of the 16S rRNA gene in bacterial genomes and its consequence for bacterial community analyses. PLoS One 8, e57923.

[9] MacLean, D., Jones, J.D., Studholme, D.J. (2009) Application of next-generation sequencing technologies to microbial genetics. Nat. Rev. Microbiol. 7, 287-296.

[10] Delsuc, F., Brinkmann, H., Philippe, H. (2005) Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Microbiol. 6, 361-375.

[11] Jolley, K.A., Bliss, D.M., Bennett, J.S., Bratcher, H.B., Brehony, C., Colles, C.M., Wimalarathna, H., Harrison, O.B., Sheppard, S.K., Cody, A.J., Maiden, M.C.J. (2012) Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 158, 1005-1015.

[12] Konstantinidis, K.T., Tiedje, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. 187, 6258-6264.

[13] Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities., Int. J. Syst. Evol. Microbiol. 57, 81-91.

[14] Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14, 1-14.

[15] Konstantinidis, K.T., Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U.S.A. 102, 2567-2572.

[16] Qin, Q.-L., Xie, B.-B., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Zhou, J., Oren, A., Zhang, Y.-Z. (2014) A proposed genus boundary for the prokaryotes based on genomic insights. J. Bacteriol. 196, 2210-2215.

[17] NCBI Resource Coordinators (2017) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45, D12-D17.

[18] Stackebrandt, E., Goebel, G.M. (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44, 846-849.

[19] Besemer, J., Lomsadze, A., Borodovsky, M. (2001) GenemarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607-2618.

[20] Emms, D.M., Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157.

[21] Jauffrit, F., Penel, S., Delmotte, S., Rey, C., de Vienne, D.M., Gouy, M., Charrier, J.P., Flandrois, J.P., Brochier-Armanet, C. (2016) RiboDB Database: A Comprehensive Resource for Prokaryotic Systematics. Mol. Biol. Evol. 33, 2170-2172.

[22] Hall, T.A. (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95-98.

[23] Lee, I., Kim, Y.O., Park, S.C., Chun, J. (2015) OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int. J. Syst. Evol. Microbiol. 66, 1100-1103.

[24] Rodriguez-R, L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints, 4: e1900v1.

[25] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic definition. Proc. Natl. Acad. Sci. U.S.A. 106, 19126-19131.

[26] Wallace, I.M., O'Sullivan, O., Higgins, D.G., Notredame, C. (2006) M-Coffee: combining multiple sequence alignment methods with T-coffee. Nucleic Acids Res. 34, 1692-1699.

[27] Talavera, G., Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564-577.

[28] Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307-321.

[29] Lefort, V., Longueville, J.-E., Gascuel, O. (2017) SMS: Smart Model Selection in PhyML. Mol. Biol. Evol. 34, 2422-2424.

[30] Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., Gascuel, O. (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of faste likelihood-based approximation schemes. Syst. Biol. 60, 685-699.

[31] Felsenstein, J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164-166.

[32] Nye, T.M.W., Liò, P., Gilks, W.R. (2006) A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. Bioinformatics 22, 117-119.

[33] Philippe, H., Lopez, P. (2001) On the conservation of protein sequences in evolution. Trends Biochem. Sci. 26, 414-416.

[34] Li, W.-J., Zhi, X.-Y., Euzéby, J.P. (2006) Proposal of *Yaniellaceae* fam. nov., *Yaniella* gen. nov. and *Sinobaca* gen. nov. as replacements for the illegitimate prokaryotic name *Yaniaceae* Li *et al.* 2005, *Yania* Li *et al.* 2004 emend Li *et al.* 2005, and *Sinococcus* Li *et al.* 2006, respectively. Int. J. Syst. Evol. Microbiol. 58, 525-527.

[35] Seiler, H., Wenning, M., Scherer, S. (2013) *Domibacillus robiginosus* gen. nov., sp. nov., isolated from a pharmaceutical clean room. Int. J. Syst. Evol. Microbiol. 63, 2054-2061.

[36] de Vos, P., Ludwig, W., Schleifer, K.H., Whitman, W.B. (2009) Family IV. Paenibacillaceae fam. nov. In: De Vos, P., Garrity, G.M., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.H., Whitman, W.B. (Eds.) Bergey's Manual of Systematic Bacteriology 3, 2nd ed., Springer-Verlag, New York, pp. 269-327.

[37] Garrity, G.M., Bell, J.A., Lilburn, T.G. (2005) The revised roadmap to the manual. In: Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M. (Eds.) Bergey's Manual of Systematic Bacteriology 2nd Edition Volume 2: The Proteobacteria Part B, The Gammaproteobacteria., Springer, New York, pp. 159-206.

[38] Zhang, W., Lu, Z. (2015) Phylogenomic evaluation of members above the species level within the phylum *Firmicutes* based on conserved proteins. Environ. Microbiol. Rep. 7, 273-281.

[39] Bergey, D.H., Harrison, F.C., Breed, R.S., Hammer, B.W., Huntoon, F.W. (1923) Bergey's Manual of Determinative Bacteriology: a key for the identification of organisms of the class *Schizomyces*. Williams & Wilkins, Baltimore, USA.

Table 1

Genomic properties of the compared Bacillales taxa. The strains incorporated in this study are listed along with the NCBI accession numbers for the genomes, number of contigs, genome sizes, average genomic G + C content and number of proteins encoded on the genome. For some genera the genome of the type strain of the type species was unavailable. The relative 16S rRNA sequence identity of a representative species type strain to the type species type strain for that genus is shown. Only those strains with 16S rRNA gene sequence identities >95% were selected for incorporation in the analyses.

Family	Strain	NCBI Acc. #	Contigs	Genome size (Mb)	G + C content (%)	# proteins encoded	Genus type strain (% 16S rRNA identity to type species)
Lactobacillales [ORDER]	*Enterococcus faecalis* DSM 20478^{T}	ASDA01000000	11	2.88	37.64	2783	Y
	Carnobacterium divergens DSM 20623^{T}	JQLO01000000	Complete (1)	2.62	35.10	2436	Y
	Streptococcus pyogenes DSM 20565^{T}	LN831034	Complete (1)	1.91	38.50	1894	Y
	Aerococcus viridans DSM 20340^{T}	CP014164	Complete (1)	2.20	39.38	2018	Y
	Lactobacillus delbrueckii DSM 20074^{T}	CP018615	Complete (1)	1.95	49.63	2055	Y
	Leuconostoc mesenteroides DSM 20343^{T}	CP000414	Complete (2)	2.08	37.67	2012	Y
Listeriaceae	*L. monocytogenes* DSM 20600^{T}	LT906436	Complete (1)	2.86	38.06	2796	Y
	B. thermosphacta DSM 20171^{T}	JHZM01000000	33	2.50	36.33	2412	Y
Staphylococcaceae	*G. haemolysans* ATCC 10379^{T}	ACDZ01000000	15	1.92	30.88	1691	Y
	S. aureus DSM 20231^{T}	CP011526	Complete (2)	2.78	32.84	2509	Y
	A. persicus DSM 28306^{T}	FOIT01000000	8	2.05	38.48	2024	Y
	S. luteus DSM 17002^{T}	JONV01000000	13	2.55	49.17	2612	96.3% (*S. roseus* DSM 5351^{T})
	J. psychrophilus DSM 19085^{T}	AUEF01000000	33	2.34	40.24	2396	98.3% (*J. halotolerans* JCM 11198^{T})
Planococcaceae	*Bhargavaea cecembensis* DSM 22132^{T}	AOFT01000000	39	3.21	54.83	3263	Y
	Sporosarcina ureae DSM 2281^{T}	AUDQ01000000	36	3.32	41.41	3265	Y
	R. pycnus DSM 15030^{T}	NJAS01000000	2	3.85	34.65	3585	97.7% (*R. stabeksii* DSM 25578^{T})
	K. zopfii DSM 20580^{T}	QFVS01000000	110	2.88	37.05	2893	Y
	V. arvi DSM 16317^{T}	LILB01000000	19	4.70	35.36	4495	Y
	S. silvestris DSM 12223^{T}	CP014609	Complete (1)	3.99	38.60	3831	Y
	C. latum DSM 14151^{T}	MATO01000000	97	3.59	42.70	3256	Y
	L. boronitolerans DSM 17140^{T}	JPVR01000000	81	4.56	37.56	4524	Y
	Psychrobacillus insolitus DSM 5^{T}	QKZI01000000	33	3.29	36.02	3230	Y
	Paenisporosarcina quisquiliarum JCM 14041^{T}	FOBQ01000000	20	4.03	35.96	3893	Y
	Planococcus rifietoensis DSM 15069^{T}	CP013659	Complete (1)	3.51	48.45	3550	99.0% (*P. citreus* DSM 20549^{T})
	Planomicrobium glaciei DSM 24857^{T}	FNDC01000000	33	3.92	46.74	3946	97.4% (*P. koreense* JCM 10704^{T})
	J. alimentarius DSM 18867^{T}	JXRQ01000000	32	3.36	43.13	3514	Y
	D. robiginosus DSM 25058^{T}	LAHL01000000	106	4.69	42.73	4771	Y

Table 1 (*Continued*)

Family	Strain	NCBI Acc. #	Contigs	Genome size (Mb)	G + C content (%)	# proteins encoded	Genus type strain (% 16S rRNA identity to type species)
Bacillaceae 1	*Caldibacillus debilis* DSM 16016^{T}	ARVR01000000	42	3.06	51.63	3223	Y
	Anoxybacillus pushchinoensis DSM 12423^{T}	FOJQ01000000	117	2.62	42.06	2792	Y
	Geobacillus stearothermophilus DSM 22^{T}	JYNWQ01000000	106	2.63	53.07	2969	Y
	Parageobacillus thermoglucosidasius DSM 2542^{T}	CP012712	Complete (1)	3.87	43.89	3923	Y
	Aeribacillus pallidus DSM 3670^{T}	CP017703	Complete (1)	4.09	39.32	4062	Y
	B. subtilis DSM 10^{T}	CP011115	Complete (1)	4.22	43.51	4231	Y
Bacillaceae 2	*Pelagirhabdus alkalitolerans* KCTC 33632^{T}	KCTC 33632	23	2.53	37.08	2390	Y
	Streptohalobacillus salinus DSM 22440^{T}	QJJR01000000	37	2.53	39.38	2319	Y
	Halolactibacillus halophilus DSM 17073^{T}	FMYI01000000	95	2.70	38.44	2630	Y
	Amphibacillus xylanus DSM 6626^{T}	AP012050	Complete (1)	2.57	35.72	2356	Y
	Gracilibacillus kekensis DSM 23178^{T}	FRCZ01000000	17	3.98	36.00	3730	96.0% (*G. halotolerans* DSM 11805^{T})
	Sediminibacillus halophilus DSM 18088^{T}	FHNF01000000	13	4.15	42.86	4137	Y
	Lentibacillus halodurans DSM 18342^{T}	FOJW01000000	31	3.67	41.57	3651	95.7% (*L. salicampi* JCM 11462^{T})
	Virgibacillus pantothenticus DSM 26^{T}	LGTO01000000	16	4.74	37.22	4343	Y
	Oceanobacillus iheyensis DSM 14371^{T}	BA000028	Complete (1)	3.63	35.68	3503	Y
	Paucisalibacillus globulus DSM 18846^{T}	AXVK01000000	50	4.23	35.79	4117	Y
	Ornithinibacillus californiensis DSM 16627^{T}	LDUE01000000	70	3.98	36.96	3993	97.2% (*O. bavariensis* DSM 15681^{T})
	Pontibacillus chungwhensis DSM 16287^{T}	AVBG01000000	40	3.87	40.77	3846	Y
	Thalassobacillus cyri DSM 21635^{T}	FNQR01000000	37	4.30	42.49	4355	97.7% (*T. devorans* DSM 16966^{T})
	Halobacillus halophilus DSM 2266^{T}	HE717023	Complete (3)	4.17	41.82	4137	Y
	Salimicrobium album DSM 20748^{T}	FNOS01000000	13	2.63	46.89	2693	Y
	Tenuibacillus multivorans NBRC 100370^{T}	FNIG01000000	18	2.95	37.69	3046	Y
	Piscibacillus halophilus DSM 21633^{T}	FOES01000000	83	2.99	36.74	3106	98.3% (*P. salipiscarius* JCM 13188^{T})
	Halalkalibacillus halophilus DSM 18494^{T}	AUHI01000000	27	2.71	37.43	2749	Y
	Salinibacillus kushneri JCM 12390^{T}	FOHJ01000000	23	3.49	37.40	3525	97.6% (*S. aidingensis* JCM 12389^{T})

Table 1 (*Continued*)

Family	Strain	NCBI Acc. #	Contigs	Genome size (Mb)	G + C content (%)	# proteins encoded	Genus type strain (% 16S rRNA identity to type species)
	Terribacillus saccharophilus DSM 21619^{T}	FOCD01000000	13	3.63	42.53	3693	Y
Sporolactobacillaceae	*Tuberibacillus calidus* DSM 17572^{T}	AUMM01000000	90	3.23	44.00	3272	Y
	Sporolactobacillus laevolacticus DSM 442^{T}	AWTC01000000	32	3.59	42.72	3556	97.2% (*S. inulinus* DSM 20348^{T})
Family *incertae sedis* XII	*Exiguobacterium aurantiacum* DSM 6208^{T}	JNIQ01000000	2	3.04	52.79	3120	Y
Bacillaceae 3	*Natribacillus halophilus* DSM 21771^{T}	FNEN01000000	47	3.30	46.80	3487	Y
	Salsuginibacillus kocurii DSM 18087^{T}	ARIV01000000	26	3.83	43.16	3689	Y
	S. qinghaiensis DSM 17008^{T}	2636416020[a]	14	3.40	44.71	3449	Y
	Alteribacillus bidgolensis DSM 25260^{T}	NJAU01000000	3	4.70	38.90	4705	Y
	Marinococcus halophilus DSM 20408^{T}	NPFA01000000	64	3.26	47.25	3406	Y
	Salipaludibacillus aurantiacus DSM 18675^{T}	AUCJ01000000	15	4.02	42.46	3961	Y
	Salisediminibacterium halotolerans DSM 21619^{T}	2684623023[a]	20	2.84	46.87	2774	Y
	Anaerobacillus arseniciselenatis DSM 15340^{T}	MLQQ01000000	58	3.95	36.09	3735	Y
Paenibacillaceae 2	*A. migulanus* DSM 2895^{T}	LGUG01000000	28	6.33	43.06	6496	99.4% (*A. aneurinilyticus* DSM 5562^{T})
	Brevibacillus brevis DSM 30^{T}	PXXZ01000000	97	6.61	47.38	6482	Y
Family *incertae sedis* X	*Thermicanus aegyptius* DSM 12793^{T}	AZNU01000000	11	3.66	48.21	3678	Y
Thermoactinomycetaceae	*Novibacillus thermophilus* KCTC 33118^{T}	CP019699	Complete (1)	3.63	50.44	3635	Y
	Thermoflavimicrobium dichotomicum DSM 44778^{T}	FORR01000000	77	3.85	42.53	3710	Y
	Lihuaxuella thermophila DSM 4670^{T}	FOCQ01000000	40	3.81	48.96	3830	Y
	Laceyella sediminis DSM 45262^{T}	PVTZ01000000	44	3.39	48.89	3432	99.6% (*L. sacchari* DSM 43356^{T})
	Thermoactinomyces vulgaris DSM 43016^{T}	2616644978[a]	15	2.56	47.94	2731	Y

Table 1 (*Continued*)

Family	Strain	NCBI Acc. #	Contigs	Genome size (Mb)	G + C content (%)	# proteins encoded	Genus type strain (% 16S rRNA identity to type species)
	Risungbinella massiliensis DSM 44691^{T}	CECI01000000	9	3.42	40.25	3405	98.0% (*R. pyongyangensis* NRRL B-59118^{T})
	Shimazuella kribbensis DSM 45090^{T}	ATZF01000000	46	4.18	38.37	4272	Y
	Seinonella peptonophila DSM 44666^{T}	FQVL01000000	41	3.84	39.16	3759	Y
	Planifilum fimeticola DSM 44946^{T}	PVNE01000000	73	3.59	57.50	3684	Y
	Melghirimyces thermohalophilus DSM 45514^{T}	FMZA01000000	35	3.19	52.91	3343	96.3% (*M. algeriensis* DSM 45474^{T})
	Kroppenstedtia eburnea DSM 45196^{T}	FTOD01000000	27	3.53	54.09	3566	Y
	Desmospora activa DSM 45169^{T}	PZZP01000000	16	3.79	49.21	3715	Y
	Marininema mesophilum DSM 45610^{T}	FNNQ01000000	42	3.33	44.80	3125	Y
Bacillaceae 4	*T. decaturensis* DSM 103037^{T}	LSKU01000000	3	2.78	36.09	2881	95.8% (*T. fermentans* DSM 23802^{T})
	V. modesticaldus DSM 14931^{T}	MIJF01000000	100	2.22	33.61	2240	Y
Paenibacillaceae 1	*C. thermotolerans* DSM 17683^{T}	AUCP01000000	156	5.04	58.30	4944	Y
	T. composti DSM 18247^{T}	CP003255	Complete (2)	4.36	60.12	4245	97.4% (*T. xylanilyticus* CNCM I-1017^{T})
	P. polymyxa DSM 36^{T}	AFOX01000000	65	5.90	44.93	5430	Y
	F. panacisegetis DSM 28129^{T}	FNBG01000000	76	5.26	42.84	4823	97.0% (*F. aquaticus* DSM 17643^{T})
	S. sacchari DSM 19268^{T}	JFBU01000000	25	6.07	52.72	5443	Y
Alicyclobacillaceae	*Alicyclobacillus acidocaldarius* DSM 446^{T}	CP001727	Complete (4)	3.21	61.89	3231	Y
	Kyrpidia tusciae DSM 2912^{T}	CP002017	Complete (1)	3.38	59.11	3365	Y
	Effusibacillus lacus DSM 27172^{T}	BDUF01000000	127	3.90	49.68	4069	Y
	Tumebacillus permanentifrigoris DSM 18773^{T}	QGGL01000000	61	4.69	54.91	4474	Y
Family *incertae sedis*	*D. alkaliarsenatis* DSM 2408^{T}	MIJE01000000	36	3.11	37.52	2912	Y
Clostridiales [ORDER]	*C. butyricum* DSM 10702^{T}	AQQF01000000	40	4.60	28.50	4084	Y

[a] Denotes those strains for which the genomes were derived from the Joint Genome Institute Integrated Microbial Genomes (JGI-IMG) database. The IMG accession numbers are indicated.

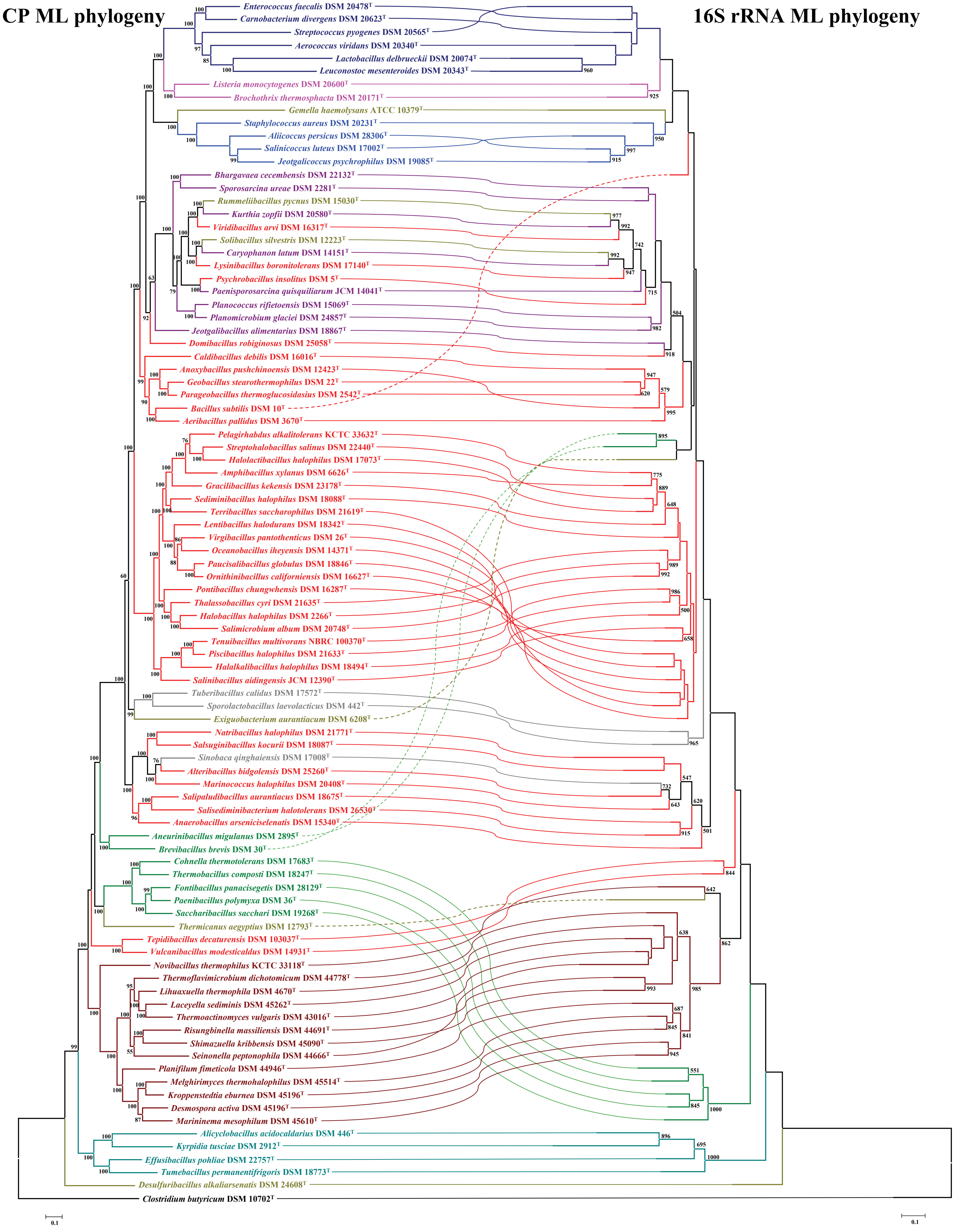
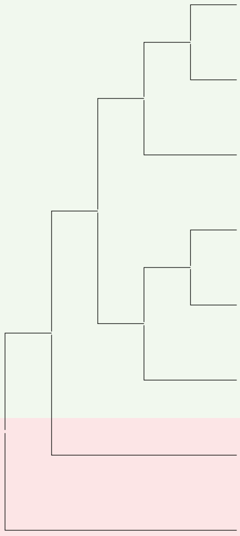

CP ML phylogeny
16S rRNA ML phylogeny
Enterococcus faecalis DSM 20478^{T}
Carnobacterium divergens DSM 20623^{T}
Streptococcus pyogenes DSM 20565^{T}
Aerococcus viridans DSM 20340^{T}
Lactobacillus delbrueckii DSM 20074^{T}
Leuconostoc mesenteroides DSM 20343^{T}
Listeria monocytogenes DSM 20600^{T}
Brochothrix thermosphacta DSM 20171^{T}
Gemella haemolysans ATCC 10379^{T}
Staphylococcus aureus DSM 20231^{T}
Aliicoccus persicus DSM 28306^{T}
Salinicoccus luteus DSM 17002^{T}
Jeotgalicoccus psychrophilus DSM 19085^{T}
Bhargavaea cecembensis DSM 22132^{T}
Sporosarcina ureae DSM 2281^{T}
Rummeliibacillus pycnus DSM 15030^{T}
Kurthia zopfii DSM 20580^{T}
Viridibacillus arvi DSM 16317^{T}
Solibacillus silvestris DSM 12223^{T}
Caryophanon latum DSM 14151^{T}
Lysinibacillus boronitolerans DSM 17140^{T}
Psychrobacillus insolitus DSM 5^{T}
Paenisporosarcina quisquiliarum JCM 14041^{T}
Planococcus rifietoensis DSM 15069^{T}
Planomicrobium glaciei DSM 24857^{T}
Jeotgalibacillus alimentarius DSM 18867^{T}
Domibacillus robiginosus DSM 25058^{T}
Caldibacillus debilis DSM 16016^{T}
Anoxybacillus pushchinoensis DSM 12423^{T}
Geobacillus stearothermophilus DSM 22^{T}
Parageobacillus thermoglucosidasius DSM 2542^{T}
Bacillus subtilis DSM 10^{T}
Aeribacillus pallidus DSM 3670^{T}
Pelagirhabdus alkalitolerans KCTC 33632^{T}
Streptohalobacillus salinus DSM 22440^{T}
Halolactibacillus halophilus DSM 17073^{T}
Amphibacillus xylanus DSM 6626^{T}
Gracilibacillus kekensis DSM 23178^{T}
Sediminibacillus halophilus DSM 18088^{T}
Terribacillus saccharophilus DSM 21619^{T}
Lentibacillus halodurans DSM 18342^{T}
Virgibacillus pantothenticus DSM 26^{T}
Oceanobacillus iheyensis DSM 14371^{T}
Paucisalibacillus globulus DSM 18846^{T}
Ornithinibacillus californiensis DSM 16627^{T}
Pontibacillus chungwhensis DSM 16287^{T}
Thalassobacillus cyri DSM 21635^{T}
Halobacillus halophilus DSM 2266^{T}
Salimicrobium album DSM 20748^{T}
Tenuibacillus multivorans NBRC 100370^{T}
Piscibacillus halophilus DSM 21633^{T}
Halalkalibacillus halophilus DSM 18494^{T}
Salinibacillus aidingensis JCM 12390^{T}
Tuberibacillus calidus DSM 17572^{T}
Sporolactobacillus laevolacticus DSM 442^{T}
Exiguobacterium aurantiacum DSM 6208^{T}
Natribacillus halophilus DSM 21771^{T}
Salsuginibacillus kocurii DSM 18087^{T}
Sinobaca qinghaiensis DSM 17008^{T}
Alteribacillus bidgolensis DSM 25260^{T}
Marinococcus halophilus DSM 20408^{T}
Salipaludibacillus aurantiacus DSM 18675^{T}
Salisediminibacterium halotolerans DSM 26530^{T}
Anaerobacillus arseniciselenatis DSM 15340^{T}
Aneurinibacillus migulanus DSM 2895^{T}
Brevibacillus brevis DSM 30^{T}
Cohnella thermotolerans DSM 17683^{T}
Thermobacillus composti DSM 18247^{T}
Fontibacillus panacisegetis DSM 28129^{T}
Paenibacillus polymyxa DSM 36^{T}
Saccharibacillus sacchari DSM 19268^{T}
Thermicanus aegyptius DSM 12793^{T}
Tepidibacillus decaturensis DSM 103037^{T}
Vulcanibacillus modesticaldus DSM 14931^{T}
Novibacillus thermophilus KCTC 33118^{T}
Thermoflavimicrobium dichotomicum DSM 44778^{T}
Lihuaxuella thermophila DSM 4670^{T}
Laceyella sediminis DSM 45262^{T}
Thermoactinomyces vulgaris DSM 43016^{T}
Risungbinella massiliensis DSM 44691^{T}
Shimazuella kribbensis DSM 45090^{T}
Seinonella peptonophila DSM 44666^{T}
Planifilum fimeticola DSM 44946^{T}
Melghirimyces thermohalophilus DSM 45514^{T}
Kroppenstedtia eburnea DSM 45196^{T}
Desmospora activa DSM 45196^{T}
Marininema mesophilum DSM 45610^{T}
Alicyclobacillus acidocaldarius DSM 446^{T}
Kyrpidia tusciae DSM 2912^{T}
Effusibacillus pohliae DSM 22757^{T}
Tumebacillus permanentifrigoris DSM 18773^{T}
Desulfuribacillus alkaliarsenatis DSM 24608^{T}
Clostridium butyricum DSM 10702^{T}
0.1
0.1

Figure 1



	CP	RP	16S rRNA	AAI	POCP	Ortho-ANI	TETRA	dDDH
CP	-							
RP	91.2	-						
16S rRNA	77.6	78.9	-					
AAI	81.4	78.8	71.4	-				
POCP	77.3	74.6	67.6	80.8	-			
OrthoANI	76.9	75.4	70.2	81.5	76.6	-		
TETRA	41.0	41.0	39.4	40.5	38.9	41.5	-	
dDDH	25.5	25.1	25.4	25.1	24.9	25.5	22.6	-
Consensus	92.7	91.2	78.9	85.1	78.3	80.5	39.8	25.3

Figure 2

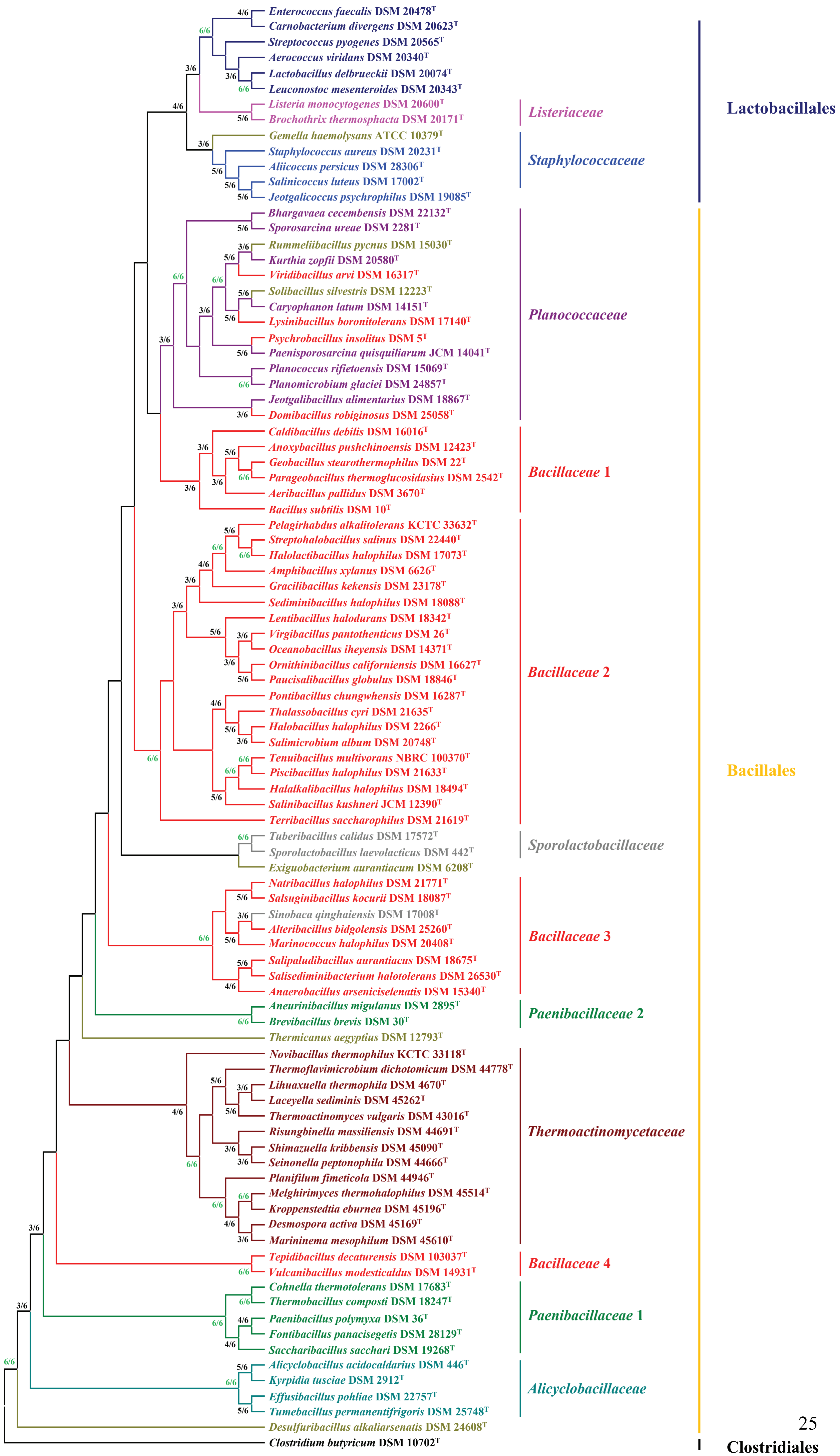

Enterococcus faecalis DSM 20478T
Carnobacterium divergens DSM 20623T
Streptococcus pyogenes DSM 20565T
Aerococcus viridans DSM 20340T
Lactobacillus delbrueckii DSM 20074T
Leuconostoc mesenteroides DSM 20343T
Listeria monocytogenes DSM 20600T
Brochothrix thermosphacta DSM 20171T
Gemella haemolysans ATCC 10379T
Staphylococcus aureus DSM 20231T
Aliicoccus persicus DSM 28306T
Salinicoccus luteus DSM 17002T
Jeotgalicoccus psychrophilus DSM 19085T
Bhargavaea cecembensis DSM 22132T
Sporosarcina ureae DSM 2281T
Rummeliibacillus pycnus DSM 15030T
Kurthia zopfii DSM 20580T
Viridibacillus arvi DSM 16317T
Solibacillus silvestris DSM 12223T
Caryophanon latum DSM 14151T
Lysinibacillus boronitolerans DSM 17140T
Psychrobacillus insolitus DSM 5T
Paenisporosarcina quisquiliarum JCM 14041T
Planococcus rifietoensis DSM 15069T
Planomicrobium glaciei DSM 24857T
Jeotgalibacillus alimentarius DSM 18867T
Domibacillus robiginosus DSM 25058T
Caldibacillus debilis DSM 16016T
Anoxybacillus pushchinoensis DSM 12423T
Geobacillus stearothermophilus DSM 22T
Parageobacillus thermoglucosidasius DSM 2542T
Aeribacillus pallidus DSM 3670T
Bacillus subtilis DSM 10T
Pelagirhabdus alkalitolerans KCTC 33632T
Streptohalobacillus salinus DSM 22440T
Halolactibacillus halophilus DSM 17073T
Amphibacillus xylanus DSM 6626T
Gracilibacillus kekensis DSM 23178T
Sediminibacillus halophilus DSM 18088T
Lentibacillus halodurans DSM 18342T
Virgibacillus pantothenticus DSM 26T
Oceanobacillus iheyensis DSM 14371T
Ornithinibacillus californiensis DSM 16627T
Paucisalibacillus globulus DSM 18846T
Pontibacillus chungwhensis DSM 16287T
Thalassobacillus cyri DSM 21635T
Halobacillus halophilus DSM 2266T
Salimicrobium album DSM 20748T
Tenuibacillus multivorans NBRC 100370T
Piscibacillus halophilus DSM 21633T
Halalkalibacillus halophilus DSM 18494T
Salinibacillus kushneri JCM 12390T
Terribacillus saccharophilus DSM 21619T
Tuberibacillus calidus DSM 17572T
Sporolactobacillus laevolacticus DSM 442T
Exiguobacterium aurantiacum DSM 6208T
Natribacillus halophilus DSM 21771T
Salsuginibacillus kocurii DSM 18087T
Sinobaca qinghaiensis DSM 17008T
Alteribacillus bidgolensis DSM 25260T
Marinococcus halophilus DSM 20408T
Salipaludibacillus aurantiacus DSM 18675T
Salisediminibacterium halotolerans DSM 26530T
Anaerobacillus arseniciselenatis DSM 15340T
Aneurinibacillus migulanus DSM 2895T
Brevibacillus brevis DSM 30T
Thermicanus aegyptius DSM 12793T
Novibacillus thermophilus KCTC 33118T
Thermoflavimicrobium dichotomicum DSM 44778T
Lihuaxuella thermophila DSM 4670T
Laceyella sediminis DSM 45262T
Thermoactinomyces vulgaris DSM 43016T
Risungbinella massiliensis DSM 44691T
Shimazuella kribbensis DSM 45090T
Seinonella peptonophila DSM 44666T
Planifilum fimeticola DSM 44946T
Melghirimyces thermohalophilus DSM 45514T
Kroppenstedtia eburnea DSM 45196T
Desmospora activa DSM 45169T
Marininema mesophilum DSM 45610T
Tepidibacillus decaturensis DSM 103037T
Vulcanibacillus modesticaldus DSM 14931T
Cohnella thermotolerans DSM 17683T
Thermobacillus composti DSM 18247T
Paenibacillus polymyxa DSM 36T
Fontibacillus panacisegetis DSM 28129T
Saccharibacillus sacchari DSM 19268T
Alicyclobacillus acidocaldarius DSM 446T
Kyrpidia tusciae DSM 2912T
Effusibacillus pohliae DSM 22757T
Tumebacillus permanentifrigoris DSM 25748T
Desulfuribacillus alkaliarsenatis DSM 24608T
Clostridium butyricum DSM 10702T
Listeriaceae
Staphylococcaceae
Planococcaceae
Bacillaceae 1
Bacillaceae 2
Sporolactobacillaceae
Bacillaceae 3
Paenibacillaceae 2
Thermoactinomycetaceae
Bacillaceae 4
Paenibacillaceae 1
Alicyclobacillaceae
Lactobacillales
Bacillales
Clostridiales

Figure 3

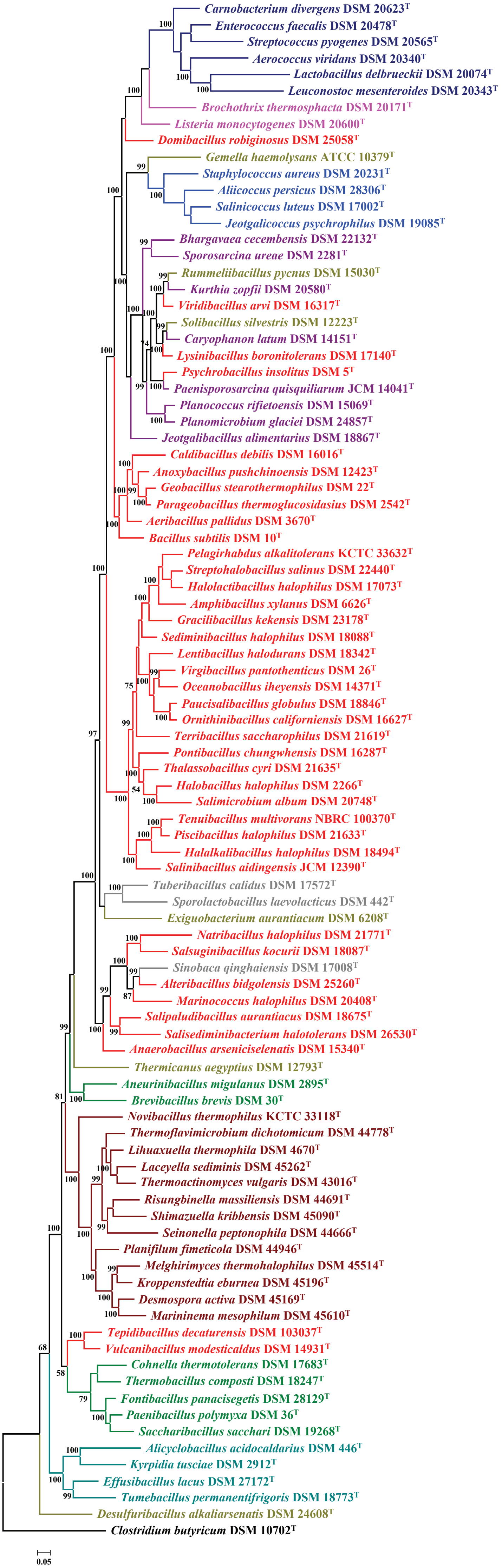

Carnobacterium divergens DSM 20623^{T}
Enterococcus faecalis DSM 20478^{T}
Streptococcus pyogenes DSM 20565^{T}
Aerococcus viridans DSM 20340^{T}
Lactobacillus delbrueckii DSM 20074^{T}
Leuconostoc mesenteroides DSM 20343^{T}
Brochothrix thermosphacta DSM 20171^{T}
Listeria monocytogenes DSM 20600^{T}
Domibacillus robiginosus DSM 25058^{T}
Gemella haemolysans ATCC 10379^{T}
Staphylococcus aureus DSM 20231^{T}
Aliicoccus persicus DSM 28306^{T}
Salinicoccus luteus DSM 17002^{T}
Jeotgalicoccus psychrophilus DSM 19085^{T}
Bhargavaea cecembensis DSM 22132^{T}
Sporosarcina ureae DSM 2281^{T}
Rummeliibacillus pycnus DSM 15030^{T}
Kurthia zopfii DSM 20580^{T}
Viridibacillus arvi DSM 16317^{T}
Solibacillus silvestris DSM 12223^{T}
Caryophanon latum DSM 14151^{T}
Lysinibacillus boronitolerans DSM 17140^{T}
Psychrobacillus insolitus DSM 5^{T}
Paenisporosarcina quisquiliarum JCM 14041^{T}
Planococcus rifietoensis DSM 15069^{T}
Planomicrobium glaciei DSM 24857^{T}
Jeotgalibacillus alimentarius DSM 18867^{T}
Caldibacillus debilis DSM 16016^{T}
Anoxybacillus pushchinoensis DSM 12423^{T}
Geobacillus stearothermophilus DSM 22^{T}
Parageobacillus thermoglucosidasius DSM 2542^{T}
Aeribacillus pallidus DSM 3670^{T}
Bacillus subtilis DSM 10^{T}
Pelagirhabdus alkalitolerans KCTC 33632^{T}
Streptohalobacillus salinus DSM 22440^{T}
Halolactibacillus halophilus DSM 17073^{T}
Amphibacillus xylanus DSM 6626^{T}
Gracilibacillus kekensis DSM 23178^{T}
Sediminibacillus halophilus DSM 18088^{T}
Lentibacillus halodurans DSM 18342^{T}
Virgibacillus pantothenticus DSM 26^{T}
Oceanobacillus iheyensis DSM 14371^{T}
Paucisalibacillus globulus DSM 18846^{T}
Ornithinibacillus californiensis DSM 16627^{T}
Terribacillus saccharophilus DSM 21619^{T}
Pontibacillus chungwhensis DSM 16287^{T}
Thalassobacillus cyri DSM 21635^{T}
Halobacillus halophilus DSM 2266^{T}
Salimicrobium album DSM 20748^{T}
Tenuibacillus multivorans NBRC 100370^{T}
Piscibacillus halophilus DSM 21633^{T}
Halalkalibacillus halophilus DSM 18494^{T}
Salinibacillus aidingensis JCM 12390^{T}
Tuberibacillus calidus DSM 17572^{T}
Sporolactobacillus laevolacticus DSM 442^{T}
Exiguobacterium aurantiacum DSM 6208^{T}
Natribacillus halophilus DSM 21771^{T}
Salsuginibacillus kocurii DSM 18087^{T}
Sinobaca qinghaiensis DSM 17008^{T}
Alteribacillus bidgolensis DSM 25260^{T}
Marinococcus halophilus DSM 20408^{T}
Salipaludibacillus aurantiacus DSM 18675^{T}
Salisediminibacterium halotolerans DSM 26530^{T}
Anaerobacillus arseniciselenatis DSM 15340^{T}
Thermicanus aegyptius DSM 12793^{T}
Aneurinibacillus migulanus DSM 2895^{T}
Brevibacillus brevis DSM 30^{T}
Novibacillus thermophilus KCTC 33118^{T}
Thermoflavimicrobium dichotomicum DSM 44778^{T}
Lihuaxuella thermophila DSM 4670^{T}
Laceyella sediminis DSM 45262^{T}
Thermoactinomyces vulgaris DSM 43016^{T}
Risungbinella massiliensis DSM 44691^{T}
Shimazuella kribbensis DSM 45090^{T}
Seinonella peptonophila DSM 44666^{T}
Planifilum fimeticola DSM 44946^{T}
Melghirimyces thermohalophilus DSM 45514^{T}
Kroppenstedtia eburnea DSM 45196^{T}
Desmospora activa DSM 45169^{T}
Marininema mesophilum DSM 45610^{T}
Tepidibacillus decaturensis DSM 103037^{T}
Vulcanibacillus modesticaldus DSM 14931^{T}
Cohnella thermotolerans DSM 17683^{T}
Thermobacillus composti DSM 18247^{T}
Fontibacillus panacisegetis DSM 28129^{T}
Paenibacillus polymyxa DSM 36^{T}
Saccharibacillus sacchari DSM 19268^{T}
Alicyclobacillus acidocaldarius DSM 446^{T}
Kyrpidia tusciae DSM 2912^{T}
Effusibacillus lacus DSM 27172^{T}
Tumebacillus permanentifrigoris DSM 18773^{T}
Desulfuribacillus alkaliarsenatis DSM 24608^{T}
Clostridium butyricum DSM 10702^{T}
0.05

Figure S1

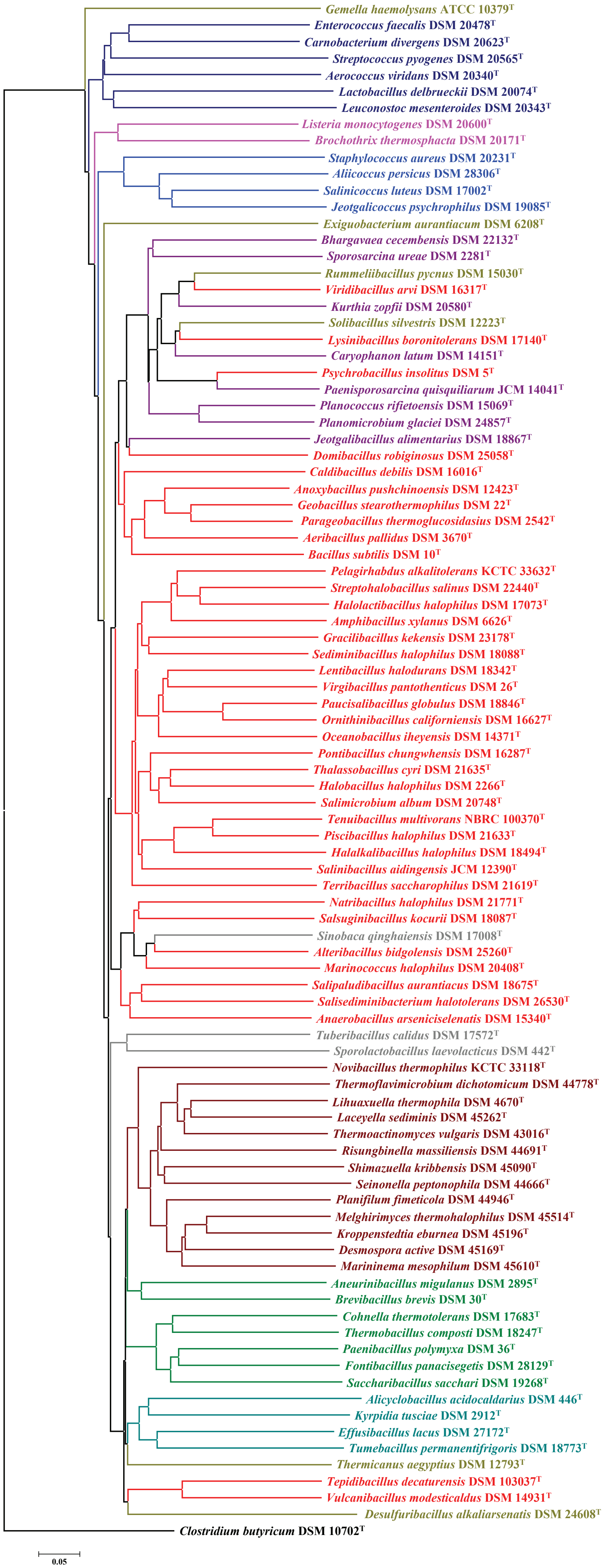

Gemella haemolysans ATCC 10379^T
Enterococcus faecalis DSM 20478^T
Carnobacterium divergens DSM 20623^T
Streptococcus pyogenes DSM 20565^T
Aerococcus viridans DSM 20340^T
Lactobacillus delbrueckii DSM 20074^T
Leuconostoc mesenteroides DSM 20343^T
Listeria monocytogenes DSM 20600^T
Brochothrix thermosphacta DSM 20171^T
Staphylococcus aureus DSM 20231^T
Aliicoccus persicus DSM 28306^T
Salinicoccus luteus DSM 17002^T
Jeotgalicoccus psychrophilus DSM 19085^T
Exiguobacterium aurantiacum DSM 6208^T
Bhargavaea cecembensis DSM 22132^T
Sporosarcina ureae DSM 2281^T
Rummeliibacillus pycnus DSM 15030^T
Viridibacillus arvi DSM 16317^T
Kurthia zopfii DSM 20580^T
Solibacillus silvestris DSM 12223^T
Lysinibacillus boronitolerans DSM 17140^T
Caryophanon latum DSM 14151^T
Psychrobacillus insolitus DSM 5^T
Paenisporosarcina quisquiliarum JCM 14041^T
Planococcus rifietoensis DSM 15069^T
Planomicrobium glaciei DSM 24857^T
Jeotgalibacillus alimentarius DSM 18867^T
Domibacillus robiginosus DSM 25058^T
Caldibacillus debilis DSM 16016^T
Anoxybacillus pushchinoensis DSM 12423^T
Geobacillus stearothermophilus DSM 22^T
Parageobacillus thermoglucosidasius DSM 2542^T
Aeribacillus pallidus DSM 3670^T
Bacillus subtilis DSM 10^T
Pelagirhabdus alkalitolerans KCTC 33632^T
Streptohalobacillus salinus DSM 22440^T
Halolactibacillus halophilus DSM 17073^T
Amphibacillus xylanus DSM 6626^T
Gracilibacillus kekensis DSM 23178^T
Sediminibacillus halophilus DSM 18088^T
Lentibacillus halodurans DSM 18342^T
Virgibacillus pantothenticus DSM 26^T
Paucisalibacillus globulus DSM 18846^T
Ornithinibacillus californiensis DSM 16627^T
Oceanobacillus iheyensis DSM 14371^T
Pontibacillus chungwhensis DSM 16287^T
Thalassobacillus cyri DSM 21635^T
Halobacillus halophilus DSM 2266^T
Salimicrobium album DSM 20748^T
Tenuibacillus multivorans NBRC 100370^T
Piscibacillus halophilus DSM 21633^T
Halalkalibacillus halophilus DSM 18494^T
Salinibacillus aidingensis JCM 12390^T
Terribacillus saccharophilus DSM 21619^T
Natribacillus halophilus DSM 21771^T
Salsuginibacillus kocurii DSM 18087^T
Sinobaca qinghaiensis DSM 17008^T
Alteribacillus bidgolensis DSM 25260^T
Marinococcus halophilus DSM 20408^T
Salipaludibacillus aurantiacus DSM 18675^T
Salisediminibacterium halotolerans DSM 26530^T
Anaerobacillus arseniciselenatis DSM 15340^T
Tuberibacillus calidus DSM 17572^T
Sporolactobacillus laevolacticus DSM 442^T
Novibacillus thermophilus KCTC 33118^T
Thermoflavimicrobium dichotomicum DSM 44778^T
Lihuaxuella thermophila DSM 4670^T
Laceyella sediminis DSM 45262^T
Thermoactinomyces vulgaris DSM 43016^T
Risungbinella massiliensis DSM 44691^T
Shimazuella kribbensis DSM 45090^T
Seinonella peptonophila DSM 44666^T
Planifilum fimeticola DSM 44946^T
Melghirimyces thermohalophilus DSM 45514^T
Kroppenstedtia eburnea DSM 45196^T
Desmospora active DSM 45169^T
Marininema mesophilum DSM 45610^T
Aneurinibacillus migulanus DSM 2895^T
Brevibacillus brevis DSM 30^T
Cohnella thermotolerans DSM 17683^T
Thermobacillus composti DSM 18247^T
Paenibacillus polymyxa DSM 36^T
Fontibacillus panacisegetis DSM 28129^T
Saccharibacillus sacchari DSM 19268^T
Alicyclobacillus acidocaldarius DSM 446^T
Kyrpidia tusciae DSM 2912^T
Effusibacillus lacus DSM 27172^T
Tumebacillus permanentifrigoris DSM 18773^T
Thermicanus aegyptius DSM 12793^T
Tepidibacillus decaturensis DSM 103037^T
Vulcanibacillus modesticaldus DSM 14931^T
Desulfuribacillus alkaliarsenatis DSM 24608^T
Clostridium butyricum DSM 10702^T
0.05

Figure S2

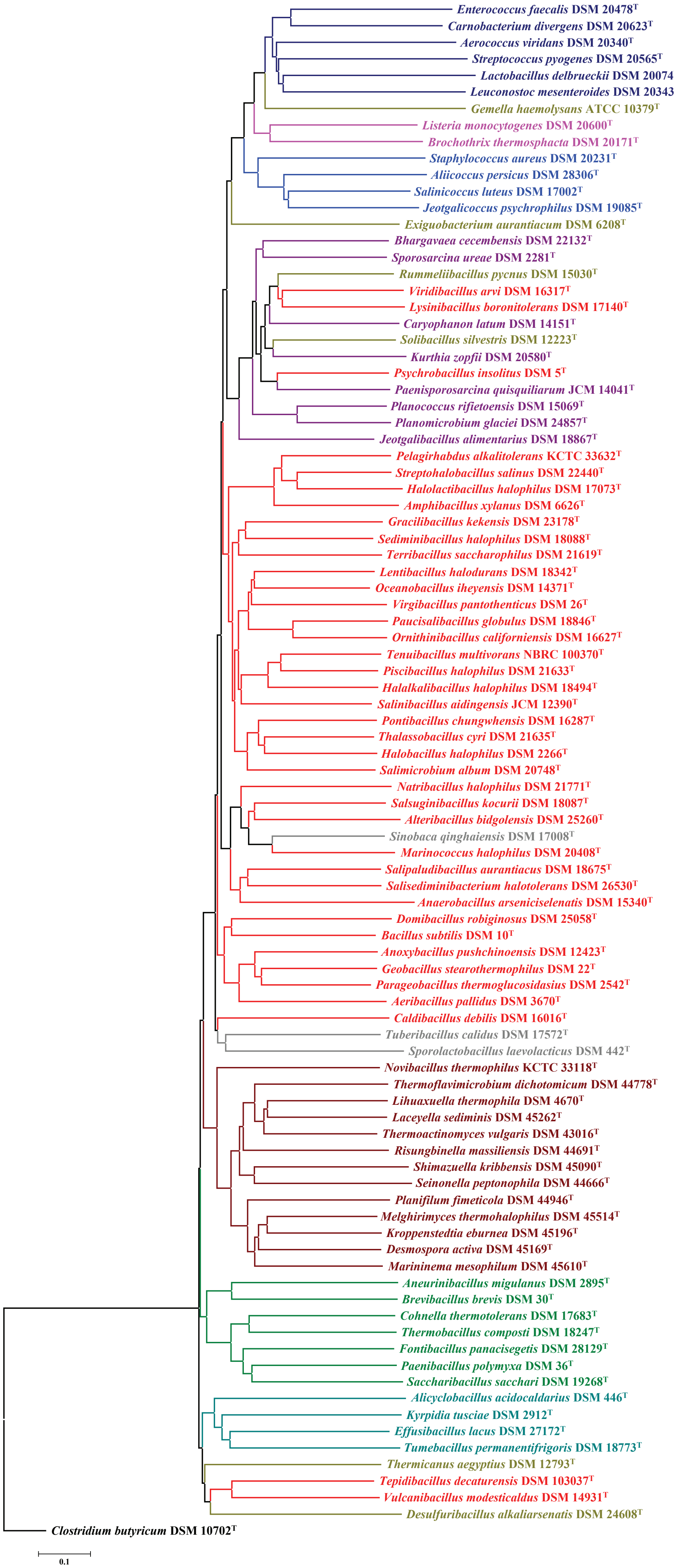

Enterococcus faecalis DSM 20478^T
Carnobacterium divergens DSM 20623^T
Aerococcus viridans DSM 20340^T
Streptococcus pyogenes DSM 20565^T
Lactobacillus delbrueckii DSM 20074^T
Leuconostoc mesenteroides DSM 20343^T
Gemella haemolysans ATCC 10379^T
Listeria monocytogenes DSM 20600^T
Brochothrix thermosphacta DSM 20171^T
Staphylococcus aureus DSM 20231^T
Aliicoccus persicus DSM 28306^T
Salinicoccus luteus DSM 17002^T
Jeotgalicoccus psychrophilus DSM 19085^T
Exiguobacterium aurantiacum DSM 6208^T
Bhargavaea cecembensis DSM 22132^T
Sporosarcina ureae DSM 2281^T
Rummeliibacillus pycnus DSM 15030^T
Viridibacillus arvi DSM 16317^T
Lysinibacillus boronitolerans DSM 17140^T
Caryophanon latum DSM 14151^T
Solibacillus silvestris DSM 12223^T
Kurthia zopfii DSM 20580^T
Psychrobacillus insolitus DSM 5^T
Paenisporosarcina quisquiliarum JCM 14041^T
Planococcus rifietoensis DSM 15069^T
Planomicrobium glaciei DSM 24857^T
Jeotgalibacillus alimentarius DSM 18867^T
Pelagirhabdus alkalitolerans KCTC 33632^T
Streptohalobacillus salinus DSM 22440^T
Halolactibacillus halophilus DSM 17073^T
Amphibacillus xylanus DSM 6626^T
Gracilibacillus kekensis DSM 23178^T
Sediminibacillus halophilus DSM 18088^T
Terribacillus saccharophilus DSM 21619^T
Lentibacillus halodurans DSM 18342^T
Oceanobacillus iheyensis DSM 14371^T
Virgibacillus pantothenticus DSM 26^T
Paucisalibacillus globulus DSM 18846^T
Ornithinibacillus californiensis DSM 16627^T
Tenuibacillus multivorans NBRC 100370^T
Piscibacillus halophilus DSM 21633^T
Halalkalibacillus halophilus DSM 18494^T
Salinibacillus aidingensis JCM 12390^T
Pontibacillus chungwhensis DSM 16287^T
Thalassobacillus cyri DSM 21635^T
Halobacillus halophilus DSM 2266^T
Salimicrobium album DSM 20748^T
Natribacillus halophilus DSM 21771^T
Salsuginibacillus kocurii DSM 18087^T
Alteribacillus bidgolensis DSM 25260^T
Sinobaca qinghaiensis DSM 17008^T
Marinococcus halophilus DSM 20408^T
Salipaludibacillus aurantiacus DSM 18675^T
Salisediminibacterium halotolerans DSM 26530^T
Anaerobacillus arseniciselenatis DSM 15340^T
Domibacillus robiginosus DSM 25058^T
Bacillus subtilis DSM 10^T
Anoxybacillus pushchinoensis DSM 12423^T
Geobacillus stearothermophilus DSM 22^T
Parageobacillus thermoglucosidasius DSM 2542^T
Aeribacillus pallidus DSM 3670^T
Caldibacillus debilis DSM 16016^T
Tuberibacillus calidus DSM 17572^T
Sporolactobacillus laevolacticus DSM 442^T
Novibacillus thermophilus KCTC 33118^T
Thermoflavimicrobium dichotomicum DSM 44778^T
Lihuaxuella thermophila DSM 4670^T
Laceyella sediminis DSM 45262^T
Thermoactinomyces vulgaris DSM 43016^T
Risungbinella massiliensis DSM 44691^T
Shimazuella kribbensis DSM 45090^T
Seinonella peptonophila DSM 44666^T
Planifilum fimeticola DSM 44946^T
Melghirimyces thermohalophilus DSM 45514^T
Kroppenstedtia eburnea DSM 45196^T
Desmospora activa DSM 45169^T
Marininema mesophilum DSM 45610^T
Aneurinibacillus migulanus DSM 2895^T
Brevibacillus brevis DSM 30^T
Cohnella thermotolerans DSM 17683^T
Thermobacillus composti DSM 18247^T
Fontibacillus panacisegetis DSM 28129^T
Paenibacillus polymyxa DSM 36^T
Saccharibacillus sacchari DSM 19268^T
Alicyclobacillus acidocaldarius DSM 446^T
Kyrpidia tusciae DSM 2912^T
Effusibacillus lacus DSM 27172^T
Tumebacillus permanentifrigoris DSM 18773^T
Thermicanus aegyptius DSM 12793^T
Tepidibacillus decaturensis DSM 103037^T
Vulcanibacillus modesticaldus DSM 14931^T
Desulfuribacillus alkaliarsenatis DSM 24608^T
Clostridium butyricum DSM 10702^T
0.1

Figure S3

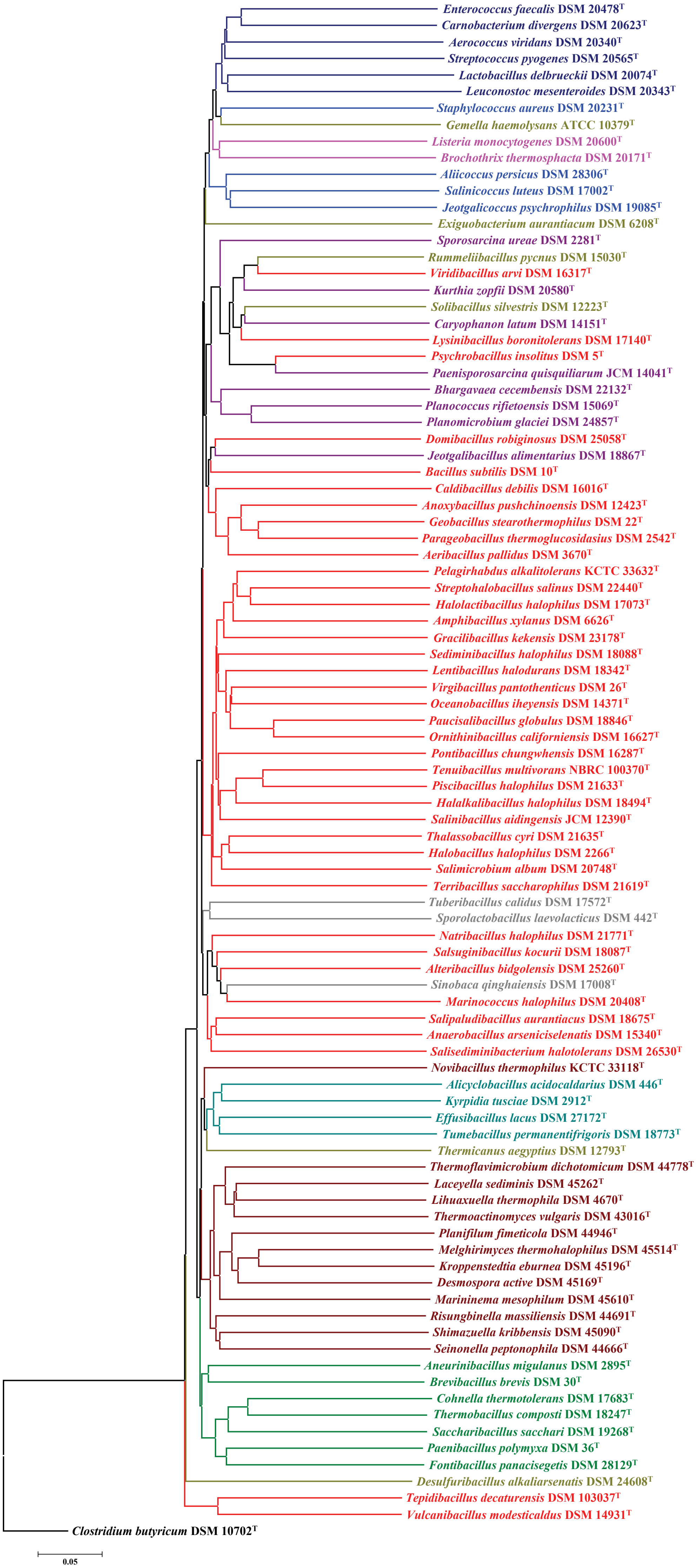

Enterococcus faecalis DSM 20478^{T}
Carnobacterium divergens DSM 20623^{T}
Aerococcus viridans DSM 20340^{T}
Streptococcus pyogenes DSM 20565^{T}
Lactobacillus delbrueckii DSM 20074^{T}
Leuconostoc mesenteroides DSM 20343^{T}
Staphylococcus aureus DSM 20231^{T}
Gemella haemolysans ATCC 10379^{T}
Listeria monocytogenes DSM 20600^{T}
Brochothrix thermosphacta DSM 20171^{T}
Aliicoccus persicus DSM 28306^{T}
Salinicoccus luteus DSM 17002^{T}
Jeotgalicoccus psychrophilus DSM 19085^{T}
Exiguobacterium aurantiacum DSM 6208^{T}
Sporosarcina ureae DSM 2281^{T}
Rummeliibacillus pycnus DSM 15030^{T}
Viridibacillus arvi DSM 16317^{T}
Kurthia zopfii DSM 20580^{T}
Solibacillus silvestris DSM 12223^{T}
Caryophanon latum DSM 14151^{T}
Lysinibacillus boronitolerans DSM 17140^{T}
Psychrobacillus insolitus DSM 5^{T}
Paenisporosarcina quisquiliarum JCM 14041^{T}
Bhargavaea cecembensis DSM 22132^{T}
Planococcus rifietoensis DSM 15069^{T}
Planomicrobium glaciei DSM 24857^{T}
Domibacillus robiginosus DSM 25058^{T}
Jeotgalibacillus alimentarius DSM 18867^{T}
Bacillus subtilis DSM 10^{T}
Caldibacillus debilis DSM 16016^{T}
Anoxybacillus pushchinoensis DSM 12423^{T}
Geobacillus stearothermophilus DSM 22^{T}
Parageobacillus thermoglucosidasius DSM 2542^{T}
Aeribacillus pallidus DSM 3670^{T}
Pelagirhabdus alkalitolerans KCTC 33632^{T}
Streptohalobacillus salinus DSM 22440^{T}
Halolactibacillus halophilus DSM 17073^{T}
Amphibacillus xylanus DSM 6626^{T}
Gracilibacillus kekensis DSM 23178^{T}
Sediminibacillus halophilus DSM 18088^{T}
Lentibacillus halodurans DSM 18342^{T}
Virgibacillus pantothenticus DSM 26^{T}
Oceanobacillus iheyensis DSM 14371^{T}
Paucisalibacillus globulus DSM 18846^{T}
Ornithinibacillus californiensis DSM 16627^{T}
Pontibacillus chungwhensis DSM 16287^{T}
Tenuibacillus multivorans NBRC 100370^{T}
Piscibacillus halophilus DSM 21633^{T}
Halalkalibacillus halophilus DSM 18494^{T}
Salinibacillus aidingensis JCM 12390^{T}
Thalassobacillus cyri DSM 21635^{T}
Halobacillus halophilus DSM 2266^{T}
Salimicrobium album DSM 20748^{T}
Terribacillus saccharophilus DSM 21619^{T}
Tuberibacillus calidus DSM 17572^{T}
Sporolactobacillus laevolacticus DSM 442^{T}
Natribacillus halophilus DSM 21771^{T}
Salsuginibacillus kocurii DSM 18087^{T}
Alteribacillus bidgolensis DSM 25260^{T}
Sinobaca qinghaiensis DSM 17008^{T}
Marinococcus halophilus DSM 20408^{T}
Salipaludibacillus aurantiacus DSM 18675^{T}
Anaerobacillus arseniciselenatis DSM 15340^{T}
Salisediminibacterium halotolerans DSM 26530^{T}
Novibacillus thermophilus KCTC 33118^{T}
Alicyclobacillus acidocaldarius DSM 446^{T}
Kyrpidia tusciae DSM 2912^{T}
Effusibacillus lacus DSM 27172^{T}
Tumebacillus permanentifrigoris DSM 18773^{T}
Thermicanus aegyptius DSM 12793^{T}
Thermoflavimicrobium dichotomicum DSM 44778^{T}
Laceyella sediminis DSM 45262^{T}
Lihuaxuella thermophila DSM 4670^{T}
Thermoactinomyces vulgaris DSM 43016^{T}
Planifilum fimeticola DSM 44946^{T}
Melghirimyces thermohalophilus DSM 45514^{T}
Kroppenstedtia eburnea DSM 45196^{T}
Desmospora active DSM 45169^{T}
Marininema mesophilum DSM 45610^{T}
Risungbinella massiliensis DSM 44691^{T}
Shimazuella kribbensis DSM 45090^{T}
Seinonella peptonophila DSM 44666^{T}
Aneurinibacillus migulanus DSM 2895^{T}
Brevibacillus brevis DSM 30^{T}
Cohnella thermotolerans DSM 17683^{T}
Thermobacillus composti DSM 18247^{T}
Saccharibacillus sacchari DSM 19268^{T}
Paenibacillus polymyxa DSM 36^{T}
Fontibacillus panacisegetis DSM 28129^{T}
Desulfuribacillus alkaliarsenatis DSM 24608^{T}
Tepidibacillus decaturensis DSM 103037^{T}
Vulcanibacillus modesticaldus DSM 14931^{T}
Clostridium butyricum DSM 10702^{T}
0.05

Figure S4

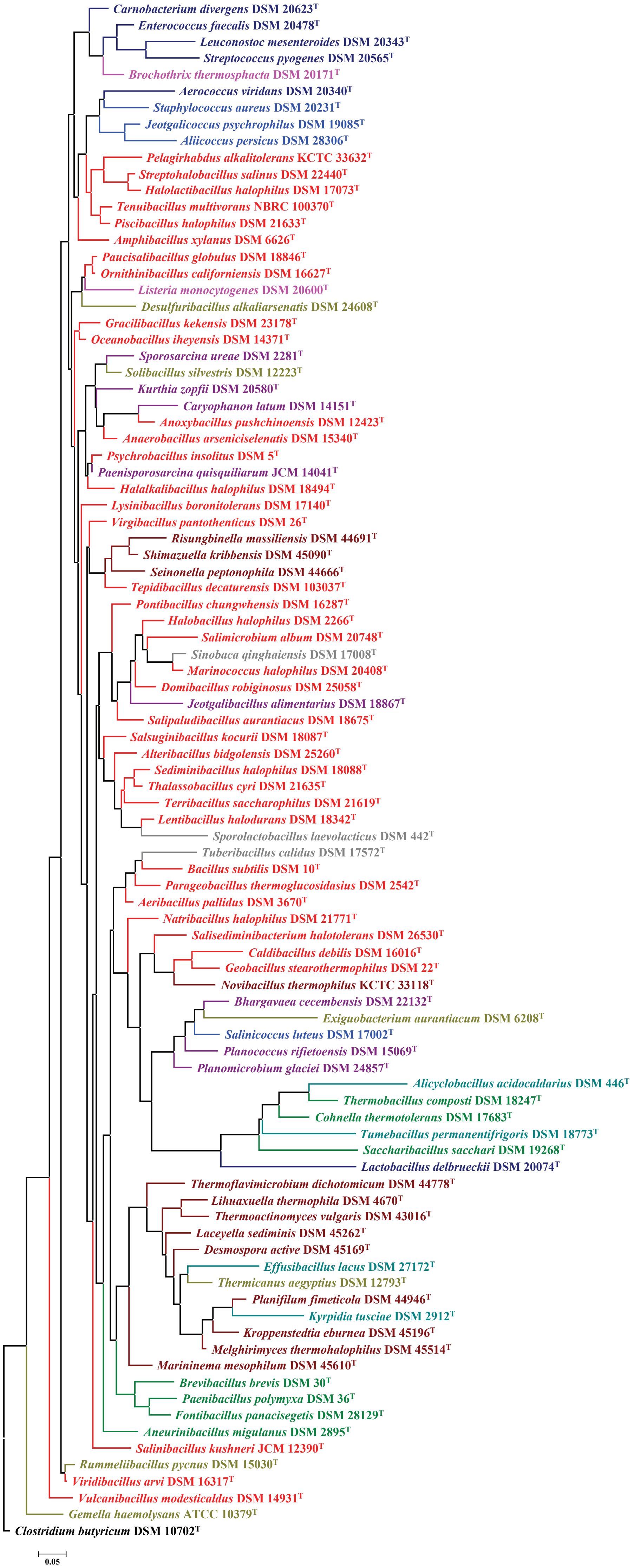

Carnobacterium divergens DSM 20623^{T}
Enterococcus faecalis DSM 20478^{T}
Leuconostoc mesenteroides DSM 20343^{T}
Streptococcus pyogenes DSM 20565^{T}
Brochothrix thermosphacta DSM 20171^{T}
Aerococcus viridans DSM 20340^{T}
Staphylococcus aureus DSM 20231^{T}
Jeotgalicoccus psychrophilus DSM 19085^{T}
Aliicoccus persicus DSM 28306^{T}
Pelagirhabdus alkalitolerans KCTC 33632^{T}
Streptohalobacillus salinus DSM 22440^{T}
Halolactibacillus halophilus DSM 17073^{T}
Tenuibacillus multivorans NBRC 100370^{T}
Piscibacillus halophilus DSM 21633^{T}
Amphibacillus xylanus DSM 6626^{T}
Paucisalibacillus globulus DSM 18846^{T}
Ornithinibacillus californiensis DSM 16627^{T}
Listeria monocytogenes DSM 20600^{T}
Desulfuribacillus alkaliarsenatis DSM 24608^{T}
Gracilibacillus kekensis DSM 23178^{T}
Oceanobacillus iheyensis DSM 14371^{T}
Sporosarcina ureae DSM 2281^{T}
Solibacillus silvestris DSM 12223^{T}
Kurthia zopfii DSM 20580^{T}
Caryophanon latum DSM 14151^{T}
Anoxybacillus pushchinoensis DSM 12423^{T}
Anaerobacillus arseniciselenatis DSM 15340^{T}
Psychrobacillus insolitus DSM 5^{T}
Paenisporosarcina quisquiliarum JCM 14041^{T}
Halalkalibacillus halophilus DSM 18494^{T}
Lysinibacillus boronitolerans DSM 17140^{T}
Virgibacillus pantothenticus DSM 26^{T}
Risungbinella massiliensis DSM 44691^{T}
Shimazuella kribbensis DSM 45090^{T}
Seinonella peptonophila DSM 44666^{T}
Tepidibacillus decaturensis DSM 103037^{T}
Pontibacillus chungwhensis DSM 16287^{T}
Halobacillus halophilus DSM 2266^{T}
Salimicrobium album DSM 20748^{T}
Sinobaca qinghaiensis DSM 17008^{T}
Marinococcus halophilus DSM 20408^{T}
Domibacillus robiginosus DSM 25058^{T}
Jeotgalibacillus alimentarius DSM 18867^{T}
Salipaludibacillus aurantiacus DSM 18675^{T}
Salsuginibacillus kocurii DSM 18087^{T}
Alteribacillus bidgolensis DSM 25260^{T}
Sediminibacillus halophilus DSM 18088^{T}
Thalassobacillus cyri DSM 21635^{T}
Terribacillus saccharophilus DSM 21619^{T}
Lentibacillus halodurans DSM 18342^{T}
Sporolactobacillus laevolacticus DSM 442^{T}
Tuberibacillus calidus DSM 17572^{T}
Bacillus subtilis DSM 10^{T}
Parageobacillus thermoglucosidasius DSM 2542^{T}
Aeribacillus pallidus DSM 3670^{T}
Natribacillus halophilus DSM 21771^{T}
Salisediminibacterium halotolerans DSM 26530^{T}
Caldibacillus debilis DSM 16016^{T}
Geobacillus stearothermophilus DSM 22^{T}
Novibacillus thermophilus KCTC 33118^{T}
Bhargavaea cecembensis DSM 22132^{T}
Exiguobacterium aurantiacum DSM 6208^{T}
Salinicoccus luteus DSM 17002^{T}
Planococcus rifietoensis DSM 15069^{T}
Planomicrobium glaciei DSM 24857^{T}
Alicyclobacillus acidocaldarius DSM 446^{T}
Thermobacillus composti DSM 18247^{T}
Cohnella thermotolerans DSM 17683^{T}
Tumebacillus permanentifrigoris DSM 18773^{T}
Saccharibacillus sacchari DSM 19268^{T}
Lactobacillus delbrueckii DSM 20074^{T}
Thermoflavimicrobium dichotomicum DSM 44778^{T}
Lihuaxuella thermophila DSM 4670^{T}
Thermoactinomyces vulgaris DSM 43016^{T}
Laceyella sediminis DSM 45262^{T}
Desmospora active DSM 45169^{T}
Effusibacillus lacus DSM 27172^{T}
Thermicanus aegyptius DSM 12793^{T}
Planifilum fimeticola DSM 44946^{T}
Kyrpidia tusciae DSM 2912^{T}
Kroppenstedtia eburnea DSM 45196^{T}
Melghirimyces thermohalophilus DSM 45514^{T}
Marininema mesophilum DSM 45610^{T}
Brevibacillus brevis DSM 30^{T}
Paenibacillus polymyxa DSM 36^{T}
Fontibacillus panacisegetis DSM 28129^{T}
Aneurinibacillus migulanus DSM 2895^{T}
Salinibacillus kushneri JCM 12390^{T}
Rummeliibacillus pycnus DSM 15030^{T}
Viridibacillus arvi DSM 16317^{T}
Vulcanibacillus modesticaldus DSM 14931^{T}
Gemella haemolysans ATCC 10379^{T}
Clostridium butyricum DSM 10702^{T}
0.05

Figure S5

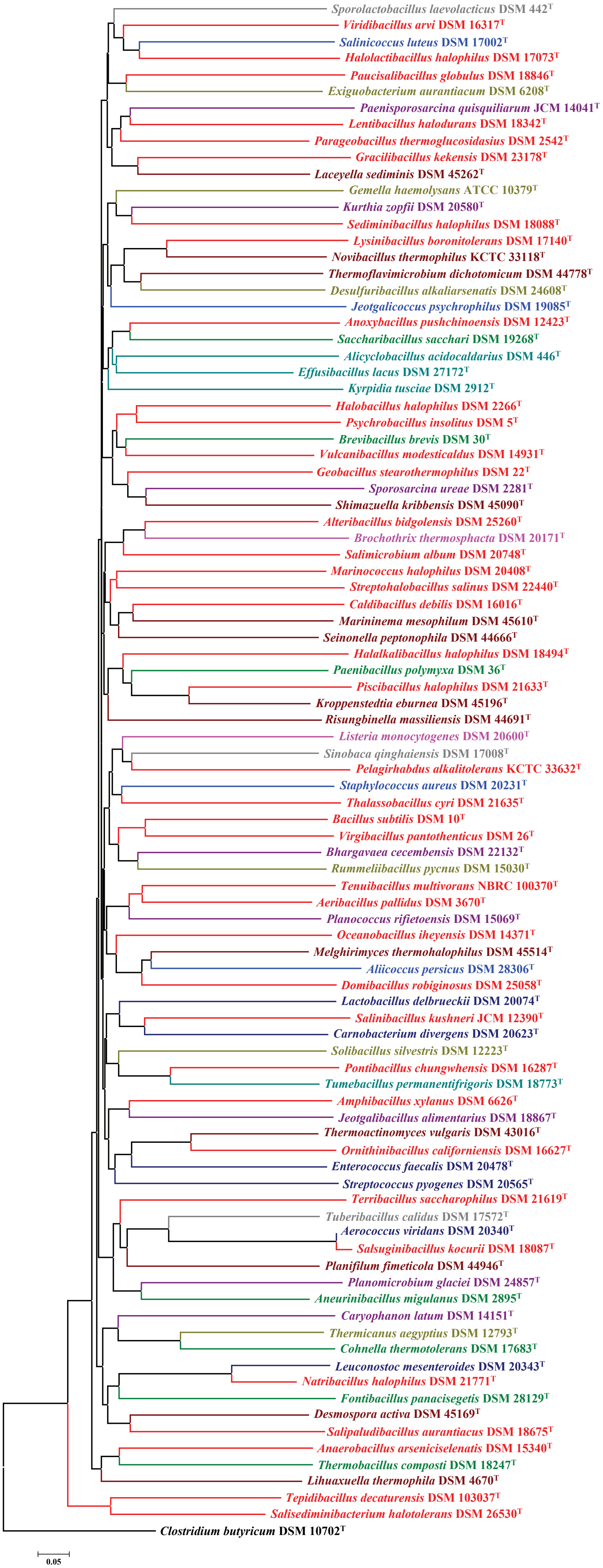

Sporolactobacillus laevolacticus DSM 442^{T}
Viridibacillus arvi DSM 16317^{T}
Salinicoccus luteus DSM 17002^{T}
Halolactibacillus halophilus DSM 17073^{T}
Paucisalibacillus globulus DSM 18846^{T}
Exiguobacterium aurantiacum DSM 6208^{T}
Paenisporosarcina quisquiliarum JCM 14041^{T}
Lentibacillus halodurans DSM 18342^{T}
Parageobacillus thermoglucosidasius DSM 2542^{T}
Gracilibacillus kekensis DSM 23178^{T}
Laceyella sediminis DSM 45262^{T}
Gemella haemolysans ATCC 10379^{T}
Kurthia zopfii DSM 20580^{T}
Sediminibacillus halophilus DSM 18088^{T}
Lysinibacillus boronitolerans DSM 17140^{T}
Novibacillus thermophilus KCTC 33118^{T}
Thermoflavimicrobium dichotomicum DSM 44778^{T}
Desulfuribacillus alkaliarsenatis DSM 24608^{T}
Jeotgalicoccus psychrophilus DSM 19085^{T}
Anoxybacillus pushchinoensis DSM 12423^{T}
Saccharibacillus sacchari DSM 19268^{T}
Alicyclobacillus acidocaldarius DSM 446^{T}
Effusibacillus lacus DSM 27172^{T}
Kyrpidia tusciae DSM 2912^{T}
Halobacillus halophilus DSM 2266^{T}
Psychrobacillus insolitus DSM 5^{T}
Brevibacillus brevis DSM 30^{T}
Vulcanibacillus modesticaldus DSM 14931^{T}
Geobacillus stearothermophilus DSM 22^{T}
Sporosarcina ureae DSM 2281^{T}
Shimazuella kribbensis DSM 45090^{T}
Alteribacillus bidgolensis DSM 25260^{T}
Brochothrix thermosphacta DSM 20171^{T}
Salimicrobium album DSM 20748^{T}
Marinococcus halophilus DSM 20408^{T}
Streptohalobacillus salinus DSM 22440^{T}
Caldibacillus debilis DSM 16016^{T}
Marininema mesophilum DSM 45610^{T}
Seinonella peptonophila DSM 44666^{T}
Halalkalibacillus halophilus DSM 18494^{T}
Paenibacillus polymyxa DSM 36^{T}
Piscibacillus halophilus DSM 21633^{T}
Kroppenstedtia eburnea DSM 45196^{T}
Risungbinella massiliensis DSM 44691^{T}
Listeria monocytogenes DSM 20600^{T}
Sinobaca qinghaiensis DSM 17008^{T}
Pelagirhabdus alkalitolerans KCTC 33632^{T}
Staphylococcus aureus DSM 20231^{T}
Thalassobacillus cyri DSM 21635^{T}
Bacillus subtilis DSM 10^{T}
Virgibacillus pantothenticus DSM 26^{T}
Bhargavaea cecembensis DSM 22132^{T}
Rummeliibacillus pycnus DSM 15030^{T}
Tenuibacillus multivorans NBRC 100370^{T}
Aeribacillus pallidus DSM 3670^{T}
Planococcus rifietoensis DSM 15069^{T}
Oceanobacillus iheyensis DSM 14371^{T}
Melghirimyces thermohalophilus DSM 45514^{T}
Aliicoccus persicus DSM 28306^{T}
Domibacillus robiginosus DSM 25058^{T}
Lactobacillus delbrueckii DSM 20074^{T}
Salinibacillus kushneri JCM 12390^{T}
Carnobacterium divergens DSM 20623^{T}
Solibacillus silvestris DSM 12223^{T}
Pontibacillus chungwhensis DSM 16287^{T}
Tumebacillus permanentifrigoris DSM 18773^{T}
Amphibacillus xylanus DSM 6626^{T}
Jeotgalibacillus alimentarius DSM 18867^{T}
Thermoactinomyces vulgaris DSM 43016^{T}
Ornithinibacillus californiensis DSM 16627^{T}
Enterococcus faecalis DSM 20478^{T}
Streptococcus pyogenes DSM 20565^{T}
Terribacillus saccharophilus DSM 21619^{T}
Tuberibacillus calidus DSM 17572^{T}
Aerococcus viridans DSM 20340^{T}
Salsuginibacillus kocurii DSM 18087^{T}
Planifilum fimeticola DSM 44946^{T}
Planomicrobium glaciei DSM 24857^{T}
Aneurinibacillus migulanus DSM 2895^{T}
Caryophanon latum DSM 14151^{T}
Thermicanus aegyptius DSM 12793^{T}
Cohnella thermotolerans DSM 17683^{T}
Leuconostoc mesenteroides DSM 20343^{T}
Natribacillus halophilus DSM 21771^{T}
Fontibacillus panacisegetis DSM 28129^{T}
Desmospora activa DSM 45169^{T}
Salipaludibacillus aurantiacus DSM 18675^{T}
Anaerobacillus arseniciselenatis DSM 15340^{T}
Thermobacillus composti DSM 18247^{T}
Lihuaxuella thermophila DSM 4670^{T}
Tepidibacillus decaturensis DSM 103037^{T}
Salisediminibacterium halotolerans DSM 26530^{T}
Clostridium butyricum DSM 10702^{T}
0.05

Figure S6