

# Phylogeny in Aid of the Present and Novel Microbial Lineages: Diversity in *Bacillus*

Shalini Porwal<sup>1,2</sup>, Sadhana Lal<sup>1</sup>, Simrita Cheema<sup>1</sup>, Vipin Chandra Kalia<sup>1\*</sup>

**1** Microbial Biotechnology and Genomics, Institute of Genomics and Integrative Biology (IGIB), CSIR, Delhi University Campus, Delhi, India, **2** Department of Biotechnology, University of Pune, Pune, India

## Abstract

*Bacillus* represents microbes of high economic, medical and biodefense importance. *Bacillus* strain identification based on 16S rRNA sequence analyses is invariably limited to species level. Secondly, certain discrepancies exist in the segregation of *Bacillus subtilis* strains. In the RDP/NCBI databases, out of a total of 2611 individual 16S rDNA sequences belonging to the 175 different species of the genus *Bacillus*, only 1586 have been identified up to species level. 16S rRNA sequences of *Bacillus anthracis* (153 strains), *B. cereus* (211 strains), *B. thuringiensis* (108 strains), *B. subtilis* (271 strains), *B. licheniformis* (131 strains), *B. pumilus* (83 strains), *B. megaterium* (47 strains), *B. sphaericus* (42 strains), *B. clausii* (39 strains) and *B. halodurans* (36 strains) were considered for generating species-specific framework and probes as tools for their rapid identification. Phylogenetic segregation of 1121, 16S rDNA sequences of 10 different *Bacillus* species in to 89 clusters enabled us to develop a phylogenetic frame work of 34 representative sequences. Using this phylogenetic framework, 305 out of 1025, 16S rDNA sequences presently classified as *Bacillus* sp. could be identified up to species level. This identification was supported by 20 to 30 nucleotides long signature sequences and *in silico* restriction enzyme analysis specific to the 10 *Bacillus* species. This integrated approach resulted in identifying around 30% of *Bacillus* sp. up to species level and revealed that *B. subtilis* strains can be segregated into two phylogenetically distinct groups, such that one of them may be renamed.

**Citation:** Porwal S, Lal S, Cheema S, Kalia VC (2009) Phylogeny in Aid of the Present and Novel Microbial Lineages: Diversity in *Bacillus*. PLoS ONE 4(2): e4438. doi:10.1371/journal.pone.0004438

**Editor:** Jason E. Stajich, University of California, Berkeley, United States of America

**Received:** August 5, 2008; **Accepted:** December 18, 2008; **Published:** February 12, 2009

**Copyright:** © 2009 Porwal et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was financially supported by the Task Force project, Council of Scientific and Industrial Research and granted permission to publish. The funders had a role in decision to publish. The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: vckalia@igib.res.in

## Introduction

Phylogenetics, the science of estimating the evolutionary past is based on the comparison of DNA or protein sequences [1]. In this age of rapid and rampant gene sequencing, the availability of a large amount of genomic information from 639 sequenced genomes (<http://www.ncbi.nlm.nih.gov>) and 16S rDNA sequencing data of 451545 isolates (<http://rdp.cme.msu.edu/>) has given new dimensions to microbial taxonomy and is likely to lead to revision of concepts such as species, organism and evolution [2]. 16S rDNA gene sequencing is often used as an alternative method to define microbes at species level. Protein coding genes having high variability has been successfully used to differentiate taxa that cannot be identified solely on the basis of 16S rDNA sequences e.g., heat shock protein (*hsp65*) [3], *hsp70*, ATPase-β-subunit, RNA polymerases and recombinase (*recA*) [4]. In addition, partial *rpoB* sequences have been applied to classify members of the genus *Mycobacterium* [3,5]; *gyrB* gene sequences have been used to define *Acinetobacter* members [6]; *Mycobacterium* [7]; *Pseudomonas* [8] and *Shewanella* [9], *gyrA* gene for defining *Bacillus subtilis* and related taxa [10].

Members of the genus *Bacillus* comprise gram-positive, spore-forming, rod-shaped, aerobic bacteria. *Bacillus* species are phenotypically and genotypically heterogenous [11,12]. *Bacillus* represents microbes of high economic, medical and biodefense importance such as bio-pesticides [13] and biofuels [14–16], pathogens [17,18]. *Bacillus thuringiensis* is currently used for the biological control of insects and in crop protection [19]. *B. subtilis*

strains produce a broad spectrum of bioactive peptides with great potential for biotechnological and biopharmaceutical applications [20]. *Bacillus licheniformis* strains also produce a variety of peptide antibiotics such as bacitracin [21,22], bacteriocin [23] and are also known to contaminate industrial processes [24–26] and cause food poisoning [27,28]. *Bacillus* spores are being used as human and animal probiotics despite the fact that studies now indicate extensive mislabeling of constituent *Bacillus* strains [29,30]. Therefore, it is becoming increasingly clear that a more rigorous selection process is required for *Bacillus* probiotic candidates [31,32]. Because of these divergent characteristics, questions arise concerning intra species diversity that could differentiate isolates of potential economic importance. It is for these reasons that 11 closely related *Bacillus* are among the 29 Bacillales sequenced so far (<http://www.ncbi.nlm.nih.gov>).

## Bacterial Systematics

Bacterial systematics began long before the discovery of DNA as the heritable material. Bacteria were originally classified largely on the basis of phenotype, morphology, ecology and other associated metabolic characteristics. Bacterial taxonomy has been a tedious, esoteric and uncertain discipline. Some excitement was brought in by the team led by Dr. Carl Woese. They provided a detailed insight into bacterial phylogeny by exploiting molecular biology in an innovative manner [33]. Genomic discoveries are posing a challenge to the classical bacterial systematics [34].

The most common molecular ecological techniques applied to assess the bacterial diversity and to analyze the genetic relationships between *Bacillus* species are based on genome fingerprinting [34]; DNA-DNA re-association studies [35,36]. Rep-PCR has been shown to be a useful technique in the subtyping of *Bacillus* species [37,38]. On the other hand, protein coding genes such as *gyrA* and *rpoB* exhibit much higher genetic variation and have been used for the classification of closely related taxa within the *B. subtilis* group [10,39]. The highly discriminating power of these genes has also been used to assess intraspecies diversity of *B. licheniformis* [40]. In general, 16S rDNA sequences are used in bacterial classification as a frame work for species delineation e.g., in *Bacillus* [41]. The genetic relationship and the phylogeny among organisms are analysed on the basis of molecular chronometer, the ribosomal operon, especially the 16S and 23S rDNA genes [42,43]. Since these genes are much conserved among microbes, they cannot be used to unambiguously discriminate at the species level [44]. With respect to the ribosomal genes, the internal transcribed spacers (ITS) between the 16S and 23S rDNA genes are hypervariable and can display polymorphisms especially in regions not implicated in rDNA maturation [43–48]. Several species resembling *Bacillus* species have been described but these new species cannot be phenotypically differentiated from the bona fide members.

Recent advances in the areas of genomics and proteomics have provided excellent opportunities to exploit genome sequence databases. Since the beginning of the genomic era the phylogenetic relationships among bacteria has been done on the basis of 16S rDNA gene Restriction Fragment Length Polymorphism (RFLP), such as for *Streptococcus aeruginosa* and *S. adjacens* [49]; *Pseudomonas aeruginosa* [50]; *Borrelia* [51]; *Lactobacillus* [52]; *Helicobacter* [53] and *Bacillus* [54]. Other structural genes that are highly conserved between species and which are useful to infer phylogenesis are the tRNA genes [55,56]. Consensus tRNA gene primers generally produce species-specific patterns, which have been successfully used to discriminate species belonging to the same genus as reported for *Acinetobacter* sp. [57]; *Streptococcus* sp. [58]; *Staphylococcus* sp. [57] and *Bacillus* sp. [59–61].

The Ribosomal Data Project (RDP-II) provides the research community with aligned and annotated 16S rDNA gene sequences, and a phylogenetically consistent taxonomic framework for these data. The RDPII database has 451545, 16S rDNA sequences as on Nov 2007 (<http://rdp.cme.msu.edu>).

Species demarcation of some *Bacillus* isolates is quite complex, since many isolates don't provide characteristic reactions that are considered typical of that species [62,63]. Although diversity identification is based on the use of a range of different morpho-physiological and chemical parameters as well as genetic analysis to tentatively identify the isolates to the species level. For example, the API system only allows the identification of 19 of the best identified and most common species and therefore fails to identify a number of isolates [31].

The various studies have emphasized the need to reclassify certain species. Ash et al., [42] analysed 51 species of *Bacillus* and found that they fell into five distinct grouping based on 16S rDNA sequences, which they believed, "clearly represent separate genera". A few of the sporulating aerobes proved to be sequenced outliers, which might represent the nuclei of other hitherto unrecognized genera. This marked the beginning of a new era in the reorganization for the genus *Bacillus*. It is so far a continuous process, since several new genera have been defined from organisms previously known as *Bacillus* [64]: *Alicyclobacillus*, *Aneurinibacillus*, *Brevibacillus*, *Geobacillus*, *Gracibacillus*, *Paenibacillus*, *Salibacillus*, *Ureibacillus* and *Virgibacillus*. Out of the 41 *Bacillus*

species reported by Joung and Coté [54], *B. kaustophilus*, *B. sterothermophilus*, *B. thermoglucosidasius* and *B. thermoleovorans* have been transferred to the newly created genus *Geobacillus* [65]. Certain others such as *Bacillus globisporus*, *B. pasteurii* and *B. psychrophilus* have been reclassified to the genus *Sporosarcina* [66]. Similarly, *Bacillus marinus* has been reclassified as *Marinibacillus marinus* [67].

### Defining Diversity of *Bacillus*

***Bacillus cereus* group.** Three species of the *Bacillus cereus* group (*B. cereus*, *Bacillus anthracis* and *B. thuringiensis*) have a marked impact on human activity. *B. cereus* and *B. anthracis* are important pathogens of mammals, including humans, and *B. thuringiensis* is extensively used in the biological control of insects [68]. In *B. cereus* group the chromosomes of the sequenced members are extremely similar. The number of genes unique to one species is quite limited and often represents metabolic adaptations [34]. Although *B. anthracis* can be distinguished from *B. cereus* on the basis of biochemical tests [69], however, certain isolates on the periphery such as the pathogenic *B. cereus* G9241 [70] cannot be properly classified unless it is checked for the plasmid pX01, which is native to *B. anthracis* [34]. Since comparisons of chromosomal contents are not able to easily distinguish different members of *B. cereus* group - *B. cereus*, *B. anthracis* and *B. thuringiensis*, one may have to look for some more easily recognizable markers. Based on the phylogenetic homogeneity, 86 strains of *B. thuringiensis* could be closely clustered together in four different groups (Bt group I-IV) at a DNA similarity rate of 93% [54]. *B. thuringiensis* is closely related to *B. anthracis* and *Bacillus mycoides* and is regarded as a subvariant of *B. cereus* based on genotypic data [12,71,72]. Demonstration of the high genetic relatedness of *B. thuringiensis*, *B. anthracis* and *B. cereus* has led to the suggestion, that these are members of a single species of *B. cereus sensu lato* [48,72,73]. Overall genetic studies have shown that *B. cereus* and *B. thuringiensis* are essentially identical [74]. *B. anthracis* can be distinguished from *B. cereus* and *B. thuringiensis* through microbiological and biochemical tests. *B. anthracis* isolates are non-hemolytic, non-motile, penicillin sensitive, susceptible to  $\gamma$ -phage and produce a poly- $\gamma$ -D-glutamic acid capsule [70]. The classification and taxonomic separation of members of *B. cereus* group is rather difficult even with modern molecular tools.

Analysis of large culture collections of *B. cereus*, *B. anthracis* and *B. thuringiensis* by AFLP and MLST [70,75,76] have identified a class of organisms containing toxigenic *B. cereus* and *B. thuringiensis* that are closely related to *B. anthracis*. These isolates were phylogenetically distinct from environmental *B. cereus* and *B. thuringiensis* [75] and might represent the closest ancestor *B. anthracis*.

***Bacillus subtilis* group.** Although several species resembling *B. subtilis* have been described over the last two decades, identification of *B. subtilis* like organisms has been quite difficult and laborious. Presently, the difficult part lies in confirming the significance of the sequence clusters. Such organisms have almost identical 16S rDNA sequences (99.2 to 99.6% sequence similarity) [42,77]. Comparative sequence analysis for the *gyrA* gene, which codes for DNA gyrase subunit A of 7 representatives of *B. subtilis* and allied taxa provided a frame work for their rapid and accurate classification and identification [10,65] divided *B. subtilis* in to two subspecies, namely *B. subtilis* subsp. *subtilis* and *B. subtilis* subsp. *spizizenii* on the basis of cell wall chemistry and DNA-DNA relatedness data.

***Bacillus licheniformis*.** The intra-specific diversity of *B. licheniformis* studied by means of MLEE and phenotypic analysis could distinguish these isolates into two main subgroups [78]. Despite the high phenotypic similarities among the 182 isolates of *B. licheniformis*, the DNA-DNA reassociation studies showed three very distinct groups. These were therefore regarded by Manachini

et al., [79] as genomovars of *B. licheniformis*. A 5' hypervariable region of the 16S rDNA corresponding to *B. subtilis* at positions 41–307 and similarly a *B. licheniformis* specific Taq probe 5'- FAM-GAG CTT GCT CCC TTA GGT CAG – Dab Syl – 3' were designed for targeting a section of this region corresponding to *B. subtilis* 16S rDNA numbering positions 79–99.

The *B. licheniformis* chromosome contains large regions that are co-linear with the genomes of *B. subtilis* and *Bacillus halodurans* and approximately 80% of the predicted *B. licheniformis* coding sequences have *B. subtilis* orthologs [80]. Recent taxonomic studies indicate that *B. licheniformis* is closely related to *Bacillus amyloliquefaciens* and *B. subtilis* on the basis of comparisons of 16S rDNA and 16S–23S ITS nucleotide sequence Lapidus et al., [81] and Xu and Cote [82] recently constructed a physical map of the *B. licheniformis* chromosome using a PCR approach and established a number of regions of co-linearity where gene content and organization were conserved with the *B. subtilis* genome. The close relationship between *B. licheniformis* and *B. halodurans* compared to *B. subtilis* has been shown on the basis of i) replication terminator protein (*rtf*), which is lacking in *B. licheniformis* [80] and *B. halodurans* [83]; ii) putative transposase of *B. licheniformis* shows close relation to *B. halodurans* [80,83]; iii) the 27 predicted extracellular proteins encoded by *B. licheniformis* ATCC 14580 genome that are not found in *B. subtilis* 168 [80]; iv) two gene clusters involved in cellulose degradation and utilization were discovered in *B. licheniformis* and there are no counterparts in *B. subtilis* 168. Sixty six per cent of the predicted *B. licheniformis* genes have orthologs in *B. subtilis* and 55% of the genes models are represented by orthologous sequences in *B. halodurans*, 1719 orthologs are common to all these three species. These conservations clearly support previous hypothesis [82] that *B. subtilis* and *B. licheniformis* are phylogenetically and evolutionarily closer to each other than to *B. halodurans* [80]. In our study, *B. halodurans* reference strains were very distinct from *B. licheniformis* and *B. subtilis*.

***Bacillus halodurans.*** *B. halodurans* is a group of rod shaped gram positive, aerobic or anaerobic bacterium. An alkaliphilic bacterium, strain C-125 (JCM9153), isolated in 1975, and was reidentified as *B. halodurans* based on 16S rDNA sequence DNA-DNA hybridization analysis. Out of 11 factors which belong to the extracytoplasmic function family, 10 are unique to *B. halodurans*. One hundred and twelve CDSs in *B. halodurans* genome showed significant similarity to the transposases or recombinases from various species such as *Anabaena* sp., *Rhodobacter capsulatus*, *Lactococcus lactis*, *Enterococcus faecium*, *Clostridium beijerinckii*, *Staphylococcus aureus* and *Yersinia pseudotuberculosis* indicating that these have played an important evolutionary role in HGT and also in internal rearrangement of the genome [83].

*B. halodurans* and *B. subtilis* similarities: genome sequence comparisons between *B. halodurans* and *B. subtilis* reveal that among the total CDSs; 8.8% match sequences of proteins found only in *B. subtilis*. The Shine-Dalgarno (SD) sequence was complementary to the one found at the 3' end of 16S rDNA (UCU UUC UCC ACU AG...) of alkaliphilic *B. halodurans* C-125 [83]; is the same as that of *B. subtilis*. *B. halodurans* C-125 is quite similar to *B. subtilis* in terms of genome size, G+C content of the genomic DNA and the physiological properties used for taxonomical identification, except the alkaliphilic phenotype [84]. Also, the phylogenetic placement of *B. halodurans* C-125 based on 16S rDNA sequence analysis indicates that this organism is more closely related to *B. subtilis* than to other members of the genus *Bacillus*. Four types of ATPases were also well conserved between *B. halodurans* and *B. subtilis*. ABC transporter genes are the most frequent class of protein coding genes found in *B. halodurans* genomes as in the case of *B. subtilis*.

***Bacillus pumilus.*** *Bacillus pumilus* is commonly isolated from a variety of environmental sources, particularly faeces of animals. *B. pumilus* grows as a smooth colony that becomes yellow with increased incubation; the organism is motile, b-hemolytic on blood agar, catalase positive, salt tolerant and penicillin susceptible and does not grow under strict anaerobic conditions [85]. *B. pumilus* has toxic properties; it has cytopathic effects in vero cells, hemolytic activity, lecithinase production, and proteolytic action on casein. Recently, From et al. [86] detected an emetic toxin that can be related to food poisoning incidents. Human infection due to *B. pumilus* is exceptional.

***Bacillus megaterium.*** *Bacillus megaterium* is a gram-positive, mainly aerobic spore forming bacterium found in widely diverse habitats from soil to seawater, sediment, rice paddies, honey, fish and dried food. *B. megaterium* has been industrially employed for more than 50 years, as it possesses some very useful and unusual enzymes and a high capacity for the production of exoenzymes. Genetic tools for this species include transducing phages and several hundred mutants covering the processes of biosynthesis, catabolism, division, sporulation, germination, antibiotic resistance, and recombination [87].

***Bacillus sphaericus.*** *Bacillus sphaericus* is an aerobic, mesophilic, spore-forming bacterium with terminal swollen sporangia and spherical spores [88]. Strains of *B. sphaericus* are toxic towards mosquito larvae and can be used as biological control agents of the important vectors of filariasis, malaria and yellow fever [89]. Most studies interested in the use of these highly toxic strains in biocontrol programmes such as in Brazil have focused on the isolation of more adapted strains in tropical conditions [90]. Strains of *B. sphaericus* were divided in to five distinct groups and group II was formed by two subgroups, IIA and IIB. All toxic strains were located in DNA homology subgroup IIA [91] but this homology group also contained non-pathogens [92]. Different techniques such as phage typing [93], serotyping [94,95], cellular fatty acid analysis [96] and MLEE on agarose gel [97] were also used in order to identify entomopathogenic *B. sphaericus*. However, only a few works report the diversity within mosquito toxic *B. sphaericus* strains [98,99]. As result, the attempt to define a new species (DNA homology group IIA) based on mosquito pathogenicity as the unique characteristic was then discarded [92]. Subsequently, using cloned toxin genes *bin* and *mtx* from *B. sphaericus* as probes resulted in segregating 30 strains into 22 groups within the DNA homology group IIA [37]. From the variation in the number and size of bands it was possible to identify similarities among the strains resulting in 5 groups in BOX-PCR and 8 groups in REP-PCR. *B. sphaericus* strains isolated from diverse habitats and geographically different locations in Brazil, were phenotypically and genetically quite heterogeneous and can be potentially useful as biological control agents against mosquito larvae [37].

Although the current classification of species within the genus *Bacillus* and correlated genera is well established and is based on a combination of numerous approaches. The present study aims at determining whether or not a new set of parameters be deduced to develop a method, which could be informative enough to be useful for the more effective classification of *Bacillus* species and other microbes. Since 16S rDNA sequence repository is quite large, we focused our studies on this gene.

The genus *Bacillus* contain a heterogeneous assembly of aerobic or facultative anaerobic bacteria, widely distributed in the environment. The phenotypic protocols though important need a supplementation of molecular approaches. Molecular approaches based on DNA sequence minimize problems associated with typability and reproducibility and enables assembly of large

reference databases [100]. Sequence specific primers for 16S rDNA gene have proved to be gold standards for the identification of pure cultures of *Bacillus* sp. such a *B. subtilis* [101], *B. cereus* and *B. thuringiensis* [102], *Paenibacillus alvei* (formerly *Bacillus alvei*) [103]. Genus specific primer have been successfully developed for *Lactobacillus*, *Mycoplasma*, *Bifidobacterium*, *Pandorea*, *Clostridium* and recently for certain *Bacillus* strains [104].

## Results

During the investigation, 1121 individual 16S rDNA sequences belonging to the genus *Bacillus* (from RDP/NCBI sites: <http://rdp.cme.msu.edu/>; <http://www.ncbi.nlm.nih.gov/>) were analyzed for generating species-specific framework and probes as a rapid tool for their identification. These sequences represented a total of 10 different species. *Bacillus cereus* group comprising of *B. anthracis*, *B. cereus* and *B. thuringiensis* was represented by 472 strains i.e. 21.9% of the total 16S rDNA sequences collected and analysed here in this study. A few other species, for which a fairly large number of 16S rDNA sequences are available, belong to *B. subtilis* (271 strains), *B. licheniformis* (131 strains), *B. pumilus* (83 strains), *B. megaterium* (47 strains), *B. sphaericus* (42 strains), *B. clausii* (39 strains) and *B. halodurans* (36 strains). These later seven species constituted 30.24% (649/2146) of the total sequences studied (Table 1).

### Phylogeny of *Bacillus* species

***Bacillus cereus* group.** (i) ***B. anthracis*:** Phylogenetic tree based on the 16S rDNA sequences from 153 strains of *B. anthracis* revealed 6 clusters (BAI to BAVI) (Figure S1). These different clusters were represented by 5 to 87 strains. Twelve strains could not be segregated clearly into any of these clusters. All the sequences from each of the cluster when subjected to multiple alignments showed that the strains were fairly similar over a large region. A visual scan of the profiles in each of the clusters (BAI to BAVI), showed that the level of genetic variability is quite low, since 84.2% of the total sequence i.e. 1309/1554 nucleotides (nts) were deemed to be indistinguishable. Further a comparison of the total length of each of the groups support the limited genetic diversity between cluster BAI - BAIIV on one hand and BAV - BAVI on the other. These clusters possess nearly identical similarity rate and were considered to be identical. Hence, the

153 *B. anthracis* 16S rDNA sequences were in fact belonging to 4 representative groups varying in length from 1309 nts to 1554 nts. The variability among four clusters was found to extend on either side of the core region, 78 nts upstream and 169 nts downstream.

(ii) ***B. cereus*:** *B. cereus*, another member of the *B. cereus* group was represented by 211 different strains. A phylogenetic tree based on the 16S rDNA sequences showed 7 clusters – BCI to BCVII (Figure S2). These 7 clusters consisted of 12 to 48 strains. Three strains could not be segregated clearly into any of these clusters. Multiple alignments of all the strains revealed the conserved region in each of the 7 clusters. A visual scan of the profiles in each of the clusters BCI to BCVII, reflects that there is around 83.8% (1276/1522 nts) similarity among them. This 1276 nts long sequence may represent the core region of this *Bacillus* species. The 16.22% variability present in the region flanking the core sequence stretch was found to extend up to 86 nts upstream and 142 nts downstream. However, some further similarity was recorded in two sets of clusters i) BCI, BCIV and BCVI, ii) BCV and BCVII, which were thus considered as identical. These two major groups were comprised of 77 and 87 strains, respectively. The final number of *B. cereus* clusters could thus be reduced to 4 from the 7 clusters observed in phylogenetic tree.

(iii) ***B. thuringiensis*:** The third member of the *B. cereus* group is represented by 108 strains of *B. thuringiensis*. The phylogenetic tree based on the nucleotide sequences of the 16S rDNA gene of these 108 strains were primarily represented by 12 different clusters. The number of strains in each of these 12 clusters, BTI to BTXII, varied from 3 to 17 (Figure S3). Twenty six strains could not be segregated clearly into any of these clusters. Multiple alignments of all members within each group revealed their respective conserved regions. Further alignment of the representative conserved regions from all the 12 clusters BTI to BTXII revealed that a completely conserved stretch of 1322/1516 nts equal to 81.2% was visible. This seems to represent the core region of the *B. thuringiensis*. A visual scan of the profiles of the sequences flanking the core region in each of the BTI to BTXII cluster shows that there is 100% similarity between clusters BTIII-BTX and BTV-BTXII, which thus reduced the final tally of *B. thuringiensis* phylogenetic clusters from 12 to 10. The overall genetic variability around the core region extended up to 42 nts upstream and 142 nts downstream.

The 472 strains of the *B. cereus* group could thus be segregated into 18 clusters, where the core region varied from 1276 to 1322, covering 81.2% to 84.2% of the total 16S rDNA gene length.

***B. subtilis*.** Apart from the *B. cereus* group of 472 strains, *B. subtilis* is represented by 271 different strains. All the 16S rDNA sequences of *B. subtilis* were segregated into 30 clusters (BSI to BSXXX) on the phylogenetic tree (Figure S4). Each cluster had 4 to 18 organisms. Thirty nine strains could not be segregated clearly into any of these clusters. Multiple alignments within each of the group revealed the regions which were shared by all the strains. The length of the conserved region within clusters BSI to BSXXX varied from 1190 to 1554 nts and exceptionally was 900 nts long. A further realignment of the representative conserved sequences of each of the cluster revealed a DNA stretch of 1278 nts to be common to all. It represented 82.1% of the total length of the 16S rDNA gene of *B. subtilis*. The 17.9% variability in the flanking regions was found to extend on either side of the core region, up to 180 nts upstream and 174 nts downstream. It also reflects that there is a large genetic diversity in *B. subtilis*. On the basis of a visual scan of the regions flanking the conserved sequences, the 30 BS clusters could be reduced to 26. The four clusters quite similar to each other were BSI-BSII, BSVIII-BSXXI, BSXV-BSXXV and BSXXXII-BSXXXIV. In

**Table 1.** 16S rDNA sequences of *Bacillus* species and number of sequences used in this study (<http://rdp.cme.msu.edu/>).

S. No.	Organism	No. of sequences
1.	<i>Bacillus anthracis</i>	153
2.	<i>B. cereus</i>	211
3.	<i>B. thuringiensis</i>	108
4.	<i>B. subtilis</i>	271
5.	<i>B. licheniformis</i>	131
6.	<i>B. pumilus</i>	83
7.	<i>B. megaterium</i>	47
8.	<i>B. sphaericus</i>	42
9.	<i>B. clausii</i>	39
10.	<i>B. halodurans</i>	36
11.	<i>B. sp.</i>	1025
	Total	2146

doi:10.1371/journal.pone.0004438.t001

addition, to the long and variable regions, there were two unique groups, BSXVII and BSXXIII, where the highly conserved region was intercalated by a highly variable region of 105 nts and 24 nts length, respectively. These were the only two such cases comprised of 14 sequences of *B. subtilis* strains, with this unique characteristic among all the 2146 sequences studied here.

***B. licheniformis*.** In the RDP/NCBI database, *B. licheniformis* was represented by 131 different strains. These strains were distributed as 13 clusters, BLI to BLXIII on the 16S rDNA phylogenetic tree. Three to 15 organisms were present in each of the clusters (Figure S5). Multiple alignment of these 16S rDNA sequence of all the members in each group revealed the regions which were shared among them. The length of the conserved region in clusters BLI to BLXIII, varied from 1219 to 1497 nts. Further alignment of the representative conserved sequences from each of the clusters revealed a completely conserved core DNA stretch of 1139 nts i.e. at position 210 to 1348. On the basis of this high similarity between the representative core regions equivalent to 74.59% of the total length, the 13 clusters could be reduced to 10. BLI, BLIII, BLVII and BLX, BLXI were of almost indistinguishable among themselves. The long and variable regions flanking the core sequence, indicates the high genetic variability within the *B. licheniformis*. The 25.41% variability in the flanking regions was found to extend on either side of the core region, up to 183 nts upstream and 180 nts downstream.

***B. pumilus*.** A group of 83 strains represents the *B. pumilus* species in the 16S rDNA RDP database. The phylogenetic tree of 16S rDNA from *B. pumilus* strains showed 12 clusters (Figure S6). Nine strains could not be segregated clearly into any of these clusters. Each of the clusters, BPI to BPXII was represented by 4 to 13 strains. Multiple alignment of gene sequences of members of each of the cluster showed that the conserved regions vary in length from 1215 nts to 1503 nts. Further alignment of the representative 16S rDNA sequences from each of the 12 clusters showed that there is a conserved region of 1215 nts. The core region represented 79.61% of the total length of the *B. pumilus* 16S rDNA sequence. It indicates the extent of similarity within this *Bacillus* species. On the other hand, the genetic variability extends on either side of the core region, up to 103 nts upstream and 205 nts downstream. Taking into account the core region and the flanking regions, the 12 clusters BPI to BPXII did not show any redundancy and could be easily distinguished from each other.

***B. sphaericus*.** The phylogenetic tree based on the 16S rDNA gene from a small group of 42 strains of *B. sphaericus*, showed 7 clusters (Figure S7). Each of the clusters BSPI to BSPVII was represented by 2 to 12 different strains. Five strains could not be segregated clearly into any of these clusters. Multiple alignment of gene sequences of members of each of the cluster revealed that the conserved regions varied in length from 1179 nts to 1501 nts. Subsequent alignment of the representative 16S rDNA sequences from each of the 7 clusters showed that there is a conserved region of 1081 nts. It represented 72.06% of the total length of the *B. sphaericus* 16S rDNA gene. There is thus a great genetic variability among the different clusters. The variable region extended up to 248 nts upstream of the 16S rDNA core region and 174 nts downstream. When all the clusters and their core regions are taken into account, two clusters BSPI and BSPVI showed high similarity and could be considered as one. The final tally of clusters of *B. sphaericus* could thus be reduced to six.

***B. halodurans*.** *B. halodurans* group was constituted by 36 strains. Phylogenetic tree of the 16S rDNA sequences could segregate these 36 strains into 4 clusters (Figure S8). These 4 clusters, BHI to BHIV were represented by 3 to 13 strains. Three strains could not be segregated clearly into any of these clusters.

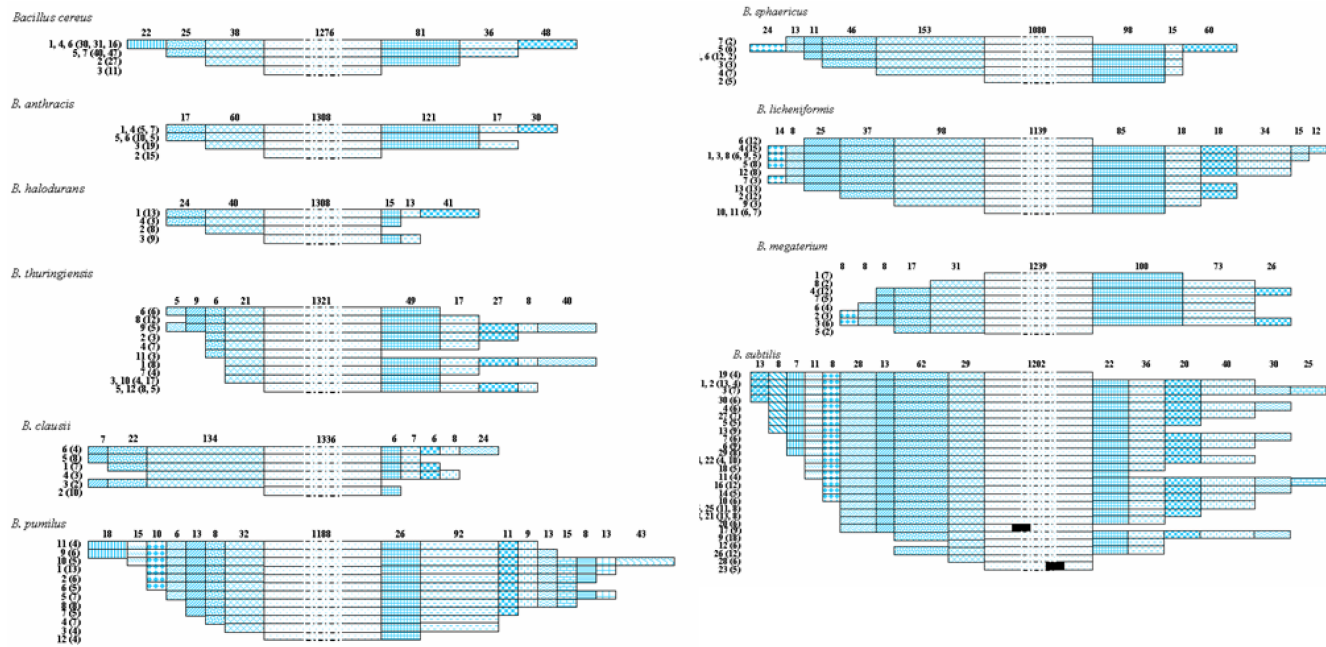
Multiple alignment of 16S gene sequences within each group showed that the conserved region varies from 1443 to 1548 nts in length. Further alignment of the representative sequences of each of the group revealed that the core conserved region is 1416 nts long. It represents 91% of the total length. The low variability in the 16S rDNA gene sequence is evident by the length of the flanking regions which varies from 65 to 70 nts. However, the 4 clusters were still maintained in spite of low genetic variability.

***Bacillus clausii*.** *Bacillus clausii* group is constituted of 39 strains. On the basis of the 16S rDNA gene phylogenetic tree, 6 clusters were observed for 34 of the strains (Figure S9). Five strains did not fall clearly into clusters. The 6 clusters BCI to BCVI were represented by 2 to 10 different strains. Multiple alignments of gene sequences of members of each of the cluster revealed the size of the conserved region, which varied in length from 1343 nts to 1551 nts. However, subsequently when the representative 16S rDNA sequences from each of the 6 clusters were considered, the conserved region was found to be 1337 nts long. It represented 86.2% of the total length of which 16S rDNA gene of *B. clausii* may extend. It reflects that there is quite a large genetic variability among the different clusters. The variable region extends only up to 164 nts upstream of the 16S rDNA core region and 52 nts downstream. In spite of a large homologous region among all the clusters of *B. clausii*, they were still distinguishable from each other particularly the clusters BCII, BCIII and BCVI.

***B. megaterium*.** The *B. megaterium* group was represented by 47 strains. A phylogenetic tree based on the 16S rDNA gene segregated these 47 strains in to 8 clusters: BMI to BMVIII (Figure S10). Each of the clusters was composed of 2 to 12 isolates, 6 strains could not be clustered concretely. Multiple alignments of gene sequences of members of these 8 clusters revealed the diversity of the core region of the 16S rDNA, which 1240 nts long. It represented 81.2% of the total length of the *B. megaterium* 16S rDNA gene. The 18.8% diversity in the region flanking the core region reflects the extent and range of genetic variability among the clusters BMI to BMVIII. Some of the clusters had high bootstrap value of 963 to 1000, while others had moderate bootstrap value of 562 to 661.

### Phylogeny of *Bacillus* Core Groups

The 1121 strains of *Bacillus cereus* group, *B. subtilis*, *B. licheniformis*, *B. pumilus*, *B. sphaericus*, *B. megaterium*, *B. clausii* and *B. halodurans* strains could be represented by 89 clusters/sequences (Figure 1). A phylogenetic tree drawn on the basis of the representative sequences of the 16S rDNA gene of the 89 clusters (Figure 2) could segregate 10 *Bacillus* species into different clusters (bootstrap value varied from 135 to 716). Members of *B. sphaericus*, *B. halodurans*, *B. pumilus* and *B. licheniformis*, *B. clausii* and *B. megaterium* were very clearly segregated. Since *B. subtilis* clusters were subdivided into two groups. *B. subtilis* Gr I was more similar to *B. sphaericus* (bootstrap value 312) whereas *B. subtilis* Gr II was more close to *B. pumilus* (bootstrap value 296). *B. subtilis* group of 271 strains was in fact observed to form 26 clusters on their 16S rDNA gene phylogenetic tree, which is indicative of greater diversity. *B. cereus* group members, *B. cereus*, *B. thuringiensis* and *B. anthracis* were placed next to each other on the 16S rDNA phylogenetic tree. Except one subclade consisting only of *B. thuringiensis* strains, the other clade was represented by all the three members. Members from each of the clusters were selected to define the range of genetic variability in each of the *Bacillus* species. From each clade, two members were selected for representing *B. sphaericus*, *B. subtilis* Gr I, *B. halodurans*, *B. cereus*. For *B. pumilus*, *B. anthracis*, *B. licheniformis* and *B. clausii* three members were chosen to represent them. Four members of *B. subtilis* Gr II and *B. megaterium*



**Figure 1. Variability in the terminal regions of the 16S rRNA gene sequences of *Bacillus* spp.:** (a). *B. cereus*; (b). *B. anthracis*; (c). *B. halodurans*; (d). *B. thuringiensis*; (e). *B. clausii*; (f). *B. pumilus*; (g). *B. sphaericus*; (h). *B. licheniformis*; (i). *B. megaterium*; (j). *B. subtilis*. Values represent group number within a species. Values within parentheses represent number of strains within the same group showing  $\geq 95\%$  similarity among them. Values on the top of horizontal bar represent the number of nucleotides showing similarity among groups. Dotted region in the middle is not to scale. Colored pattern corresponds to distinct fragments. Solid black box represents gap with in the sequences in *B. subtilis*. doi:10.1371/journal.pone.0004438.g001

and five members of *B. thuringiensis* were selected to represent these species. Thus 34 sequences were chosen to represent the 10 *Bacillus* species (Figure 2). This reference phylogenetic framework tree has been used to segregate those *Bacillus* isolates which have been presently classified only as *Bacillus* sp.

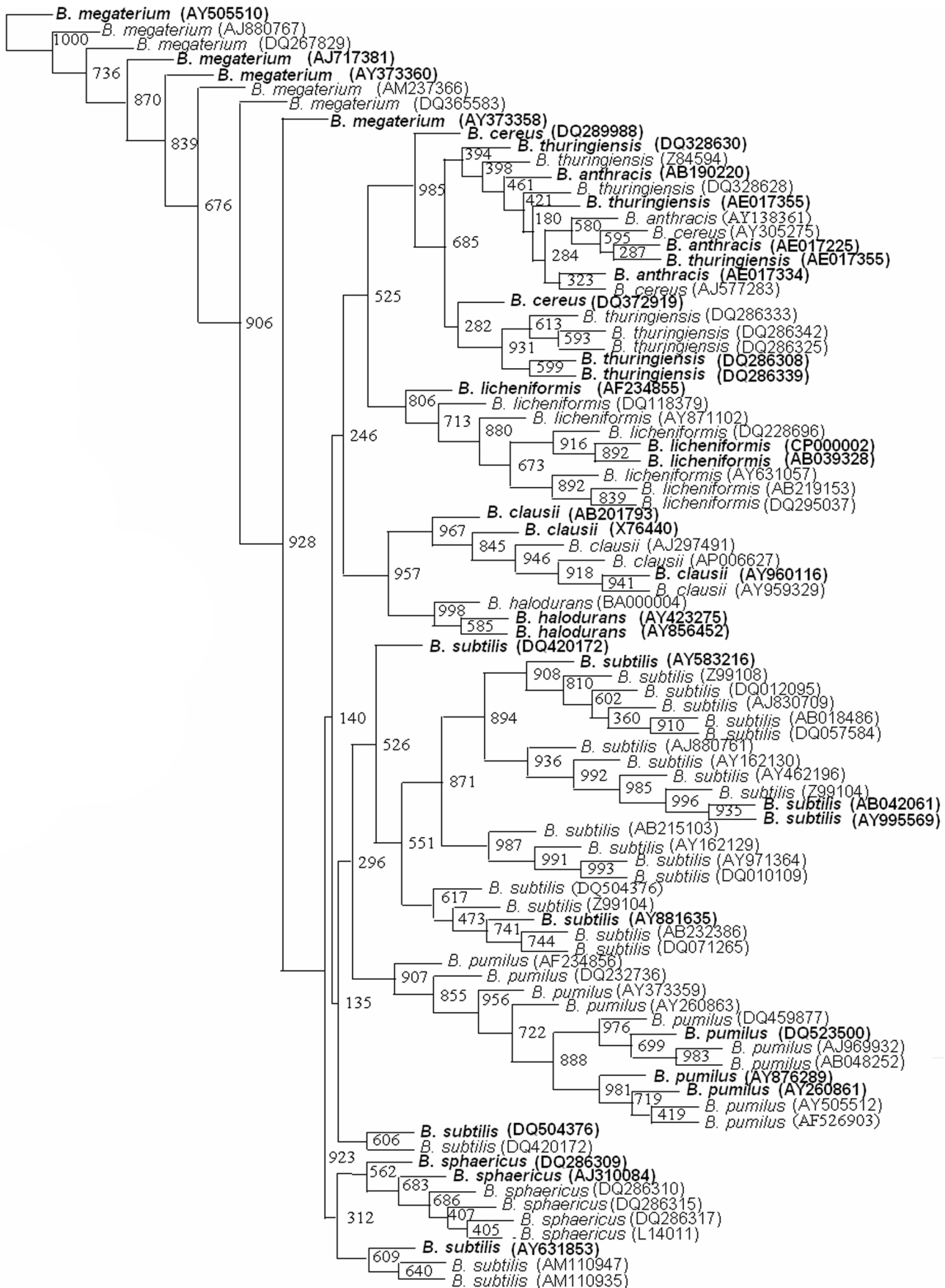
A total of 1025 sequences of 16S rDNA gene from *Bacillus* sp. were screened at the rate of 52 entries along with 34 reference phylogenetic framework sequences to generate 22 phylogenetic trees. From all the phylogenetic trees it was possible to classify a total of 305 isolates of the *Bacillus* sp: 75 isolates as *B. cereus*, 2 isolates as *B. thuringiensis*, 44 isolates as *B. subtilis*, 21 isolates as *B. licheniformis*, 32 isolates as *B. pumilus*, 23 isolates as *B. sphaericus*, 7 isolates as *B. halodurans*, 69 isolates *B. megaterium*, 31 isolates as *B. clausii* and 1 as *B. anthracis*. Final phylogenetic trees were drawn on the basis of this preliminary screening showed that 305 out of these 1025 show more similarity having 800 to 1000 bootstrap values with their respective species (Figures 3–7). The accession numbers of these *Bacillus* species are given in Table S1. At this rate about 29.75% of the unclassified *Bacillus* sp. could be identified up to species level. This phylogenetic framework based on 34, 16S rDNA sequences from 10 *Bacillus* species can be used to identify *Bacillus* strains up to species level.

## Signature Analysis

***Bacillus* species-specific signature.** The sequences of 10 data sets were submitted groupwise to MEME program (<http://meme.nbcr.net/meme/meme.html>). Ten signatures were identified for each species, which were 25–30 nts long (Table S2). Signature, which was not present in other *Bacillus* spp. and found as distinct was used to blast against the NCBI database. As seen from the Table 2, there were 1–5 regions in each of the *Bacillus* species, which had the potential to be used as signature except in case of *B. anthracis*, *B. licheniformis*, and *B. subtilis*. Signature

analyses of 16S rDNA sequences of known *Bacillus* spp. show unique signatures exclusive to them. *B. cereus* and *B. thuringiensis* had unique 16S rDNA signatures of 29 to 30 nts length. These occurred with a frequency of 151/211 and 43 to 63/108 sequences. By this approach, no signatures specific to *B. anthracis* alone could be detected. It may be remarked that among the 3 members of the *B. cereus* group, *B. cereus* shared 8/10 signatures with *B. anthracis* and 5/10 signatures with *B. thuringiensis*. Four out of 10 signatures were shared by all the three members. *B. clausii*, *B. megaterium* and *B. halodurans* can be identified with the aid of 2 to 3 signatures. For *B. licheniformis* and *B. subtilis*, no unique signatures could be identified. Some signatures such as 5'TGTGGTTTAAATTCGAAGCAACGCGAAGAA 3' were shared by all the *Bacillus* spp. except *B. subtilis*, *B. pumilus* and *B. sphaericus*, providing “evidences” of common origins. Similarly, certain signatures 5'TAAAGCTCTGTTGTTAGGGAAGAA-CAAGT 3' were shared only between *B. subtilis* and *B. pumilus*. Such similarities in signature sequences were also recorded among *B. licheniformis* and *B. megaterium* on one hand and *B. licheniformis* and *B. pumilus* on the other. For all these signatures, the closest match within the top 50 best hits (BLAST) with their respective species was in the range of 60 to 98%. Exceptionally the best matches were in the lower range of around 24% in the case of *B. cereus*.

**Cluster-specific signature.** In case of 11 clusters (Figure 8), on the basis of the ClustalW alignment, the size of the conserved region varied from 1226 to 1547 nts. These 11 clusters consisting of 18 to 50 isolates each did not appear to fall within any of the aligned species and may represent sub-species or novel lineages. The sequences of these 11 clusters were groupwise submitted to MEME program (<http://meme.nbcr.net/meme/meme.html>). Ten signatures were identified for each cluster which was 25–30 nts long (Table S3). As seen from the Table, there were 1–2 regions in each of the cluster, which had the potential to be used as



1000

**Figure 2. Phylogenetic tree based on 89, 16S rRNA gene sequences from 10 *Bacillus* species.** A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes, 1000 bootstrap replicates were run. Bold sequences (34) are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>). doi:10.1371/journal.pone.0004438.g002

signature except in case of Cluster 2–4 and Cluster 8 (Table 2). The signatures identified through MEME program revealed that there are certain similarities and even overlapping sequence stretches. The usage of signature and the conserved sequence regions against all other known *Bacillus* spp. indicate that they were not significantly similar and are not homologous. The signature sequence - 5' TTTAATTTCGAAGCAACGCGAAGAACCCTTA3' - shared by all the Clusters except 5 and 7 also indicates that these sequences may have a common origin similar to *Lactobacillus*, *Paenibacillus*, *Clostridium*, rumen bacterium, etc. [105]. Each of the clusters has certain unique signatures, which did not match with the 10 *Bacillus* spp. Typically, Clusters 2, 3, 4 and 8 did not show any unique signatures by this approach. However, in Cluster 2 – 5' AAATGATTGGGGTGAAGTCGTAACAAGGTA3' was the only signature which showed remarkable closeness to *B. clausii* in 31 out of 39 sequences. Similarly, only one out of 10 signatures in Clusters 3 and 4 showed closest matches with *B. licheniformis* and *B. megaterium* at a frequency of 54/131 and 13/47 sequences, respectively. As far as Cluster 8 is concerned, one of the signatures could be traced among 18/37 sequences of Cluster 7 and with negligible frequency in other clusters.

The unique signatures of sequences present in Cluster 1 showed similarity to uncultured *Bacillus* sp. and uncultured *Macrococcus* sp. with a frequency of 28% among the top 50 hits (BLAST). Similarly, the signatures of Clusters 6, 9 and 11 shared similar sequences to *Virgibacillus*, *B. megaterium* and *Geobacillus*, respectively, at a frequency in the range of 16 to 74%. In spite of employing all approaches to assign the signatures and sequences to known organisms, the signatures of Clusters 5, 7 and 10 could not be properly categorized. In brief, all these 11 clusters consisting of 18 to 50 sequences each did not appear to fall within any of the aligned species and may represent sub-species or novel lineages. It is probable that the *Bacillus* community is more diverse than reported so far (<http://rdp.cme.msu.edu/>). This suggests that bacterial communities in a variety of soils, environmental habitats, etc., may be very similar when assessed by molecular methods for their 16S rDNA than for other metabolic genes.

### Restriction Enzyme Analysis

A total of 14 Type II restriction enzymes, which are independent of methylase and cleave at very specific sites within or close to the recognition sequence (4 to 6 mer) (Table 3), varied in their response to 632, 16S rDNA gene sequences: 344 belonging to 10 known species of *Bacillus* and 288 belonging to strains classified as *Bacillus* sp. Restriction enzymes (REs) sites for *SmaI*, *EcoRI*, *DpnII*, *RsaI*, *BfaI*, *HaeIII*, *Tru9I* and *AluI* in 16S rDNA gene sequences occurred with a frequency of 1 to 7 resulting in 2 to 9 fragments of varied nucleotide lengths. On the other hand, enzymes *BamHI*, *NoI*, *SacI*, *NruI*, *HindIII* and *PstI* proved to be “non”-cutters in most of the cases. Of these later 6 REs, *BamHI*, *NoI*, and *SacI* could cut only 0.3 to 2.3% of the 344 sequences belonging to 10 known species of *Bacillus*: *B. cereus*, *B. anthracis*, *B. thuringiensis*, *B. subtilis*, *B. licheniformis*, *B. pumilus*, *B. sphaericus*, *B. halodurans*, *B. megaterium*, *B. clausii*. In spite of low frequency of RE sites for *NruI*, *HindIII* and *PstI* in 16S rDNA genes sequences, these can be still exploited for distinguishing different *Bacillus* spp. such that i) *NruI* could cleave only *B. subtilis* and *B. sphaericus* at the rate of 12/30 and 41/42 sequences, respectively; ii) *HindIII* sites

appeared in 43/47 sequences of *B. megaterium*; iii) *PstI* sites occurred at high frequency of 35/42 in *B. sphaericus*; 28/30 in *B. pumilus*; 44/47 in *B. megaterium*; 29/30 in *B. licheniformis* and 5/30 in *B. subtilis* (Table S4). The sites for the two enzymes *EcoRI* and *SmaI* appeared in 96 to 97% of the sequences but due to the presence of only one site per sequence, these two enzymes could not serve any significant purpose at this stage.

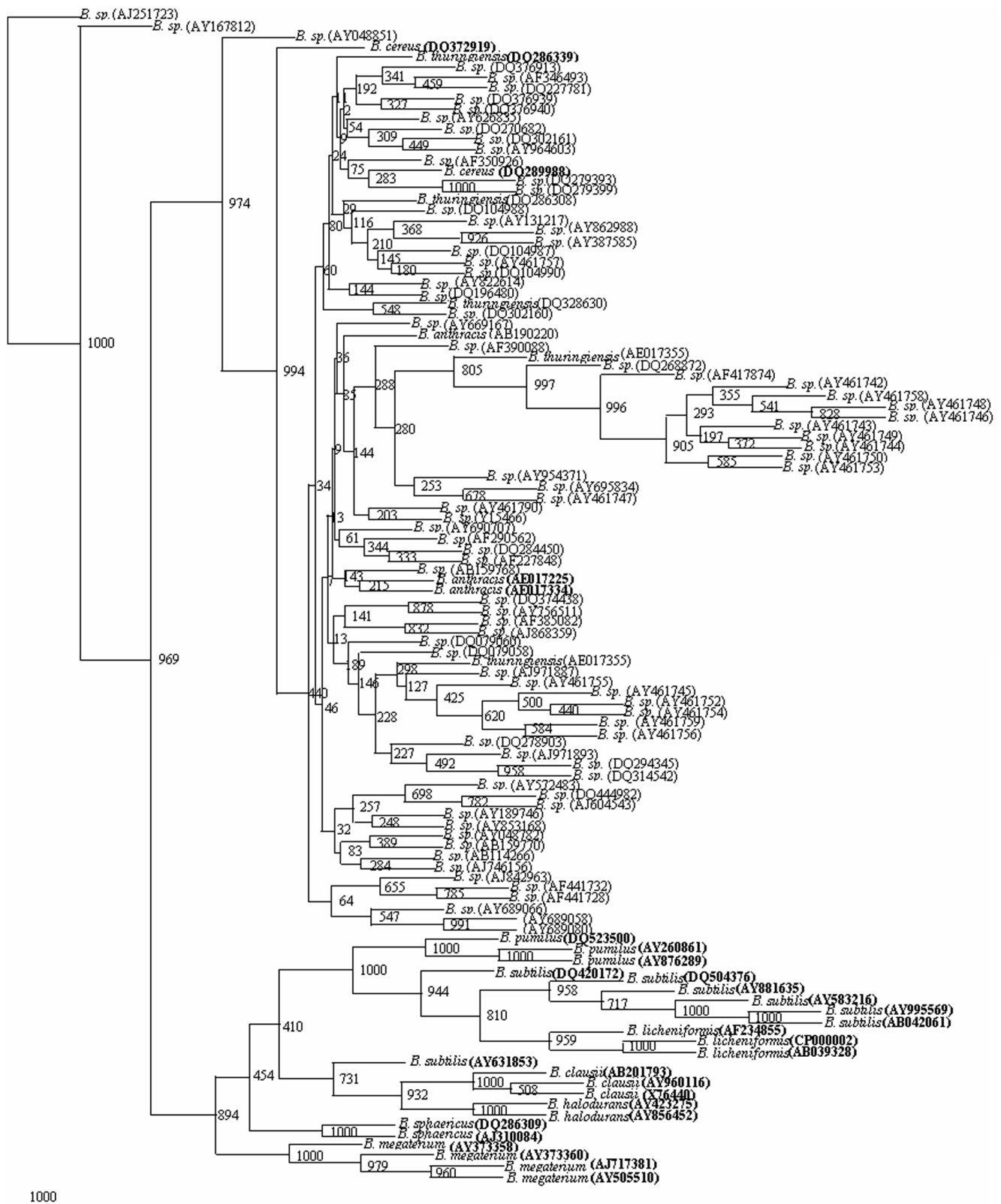
Six REs with more than 1 site for their action were compared for all the 344 sequences of 10 *Bacillus* species (Figure 9). The pattern of the number and lengths of the fragments served as reference standards for those strains which have been designated so far as *Bacillus* sp.

A comparative analysis of *in silico* cleavage of 344 sequences of 10 *Bacillus* spp. revealed very unique patterns. With *RsaI*, 3 fragments of 406, 355 and 146 nts (in this order from 5' end) were obtained for *B. cereus*, *B. anthracis*, *B. thuringiensis*. *B. megaterium* and *B. clausii* had two fragments of 16/19 and 11 nts in addition to the 3 recorded with the 3 members of *B. cereus* group. *B. subtilis*, *B. licheniformis* and *B. sphaericus* showed similar patterns of fragment size and order: 18/19 – 11 – 405/406/407 – 496/501/502. *B. pumilus* appeared to have lost two 5' RE sites for *RsaI* such that only two clear cut fragments could be recorded with high frequency. *B. halodurans* seems to have acquired an additional RE site and assumed a status intermediate to *B. cereus* group on one hand and the rest of the *Bacillus* spp. on the other.

A very distinct order and length of the fragments could be seen with the other 5 REs - *DpnII*, *BfaI*, *HaeIII*, *Tru9I* and *AluI* as well (Figure 9). A comparison of *in silico* digestions of the 16S rDNA gene sequences with all the 6 REs revealed two features. *B. subtilis* showed two distinct subgroups with *DpnII*, *BfaI* and *AluI*. Here, no similarities in size of the fragments and the positions of the RE sites were evident within each of the three sets of subgroups. With *DpnII*, *B. subtilis* revealed 4 fragments in one group: 119-196-87-51 (nts) and 6 fragments of 77-24-83-317-589-12 (nts) in the other. The three additional RE sites appeared on the 5' end of the 16S rDNA genes. Similarly with *AluI*, *B. subtilis* groups I and II differed not only in the number of RE sites but also in their positions. It may not be too premature to conclude that *B. subtilis* need to be subdivided in to at least two groups, since a similar situation was observed with respect to the signature analyses presented in the previous sections. Nakamura et al. [77] and Chun and Bae [10] divided *B. subtilis* in to two subspecies, namely *B. subtilis* subsp. *subtilis* and *B. subtilis* subsp. *spizizenii* on the basis of cell wall chemistry and DNA-DNA relatedness data.

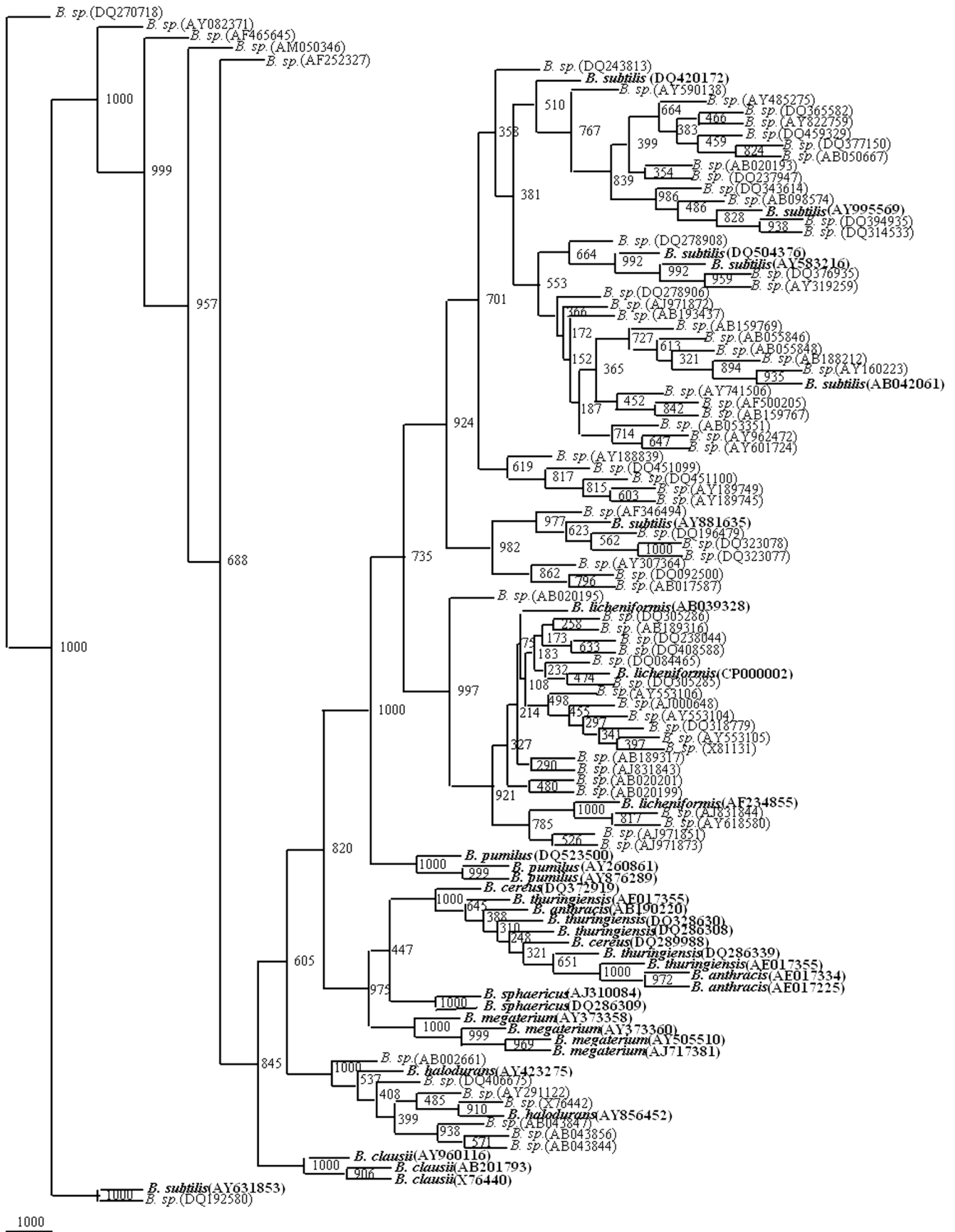
The second distinct feature which emerged is with regards to the classification of these 10 *Bacillus* spp. on the basis of the 6 REs is with respect to members of the *B. cereus* group - *B. cereus*, *B. anthracis*, *B. thuringiensis*, which were indistinguishable within the group. However, they were very clearly distinguishable from all other *Bacillus* spp. with respect to the action of *DpnII* and *AluI*. *B. halodurans* could be distinguished from all other species on the basis of a combination of actions of *TruI* and *HaeIII*. *B. clausii* could be segregated from the others by using *Tru9I* and *RsaI*. *B. megaterium* though quite close to *B. licheniformis* for *BfaI* and *Tru9I* and *B. clausii* for *RsaI*; was distinguishable by cutting its 16S rDNA gene sequences with a set of 4 REs: *BfaI*, *Tru9I*, *HaeIII* and *RsaI*. *B. megaterium* appears to be the originator of *B. halodurans*, *B. clausii*, *B. licheniformis*, *B. subtilis* and *B. pumilus*, with whom it shares different RE sites and fragments lengths (Figure 10).





**Figure 3. Phylogenetic tree of 34 framework sequences (bold values) and 78 *Bacillus* sp., which could be designated as *B. thuringiensis*, *B. anthracis* and *B. cereus* (Figures S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31 to S32 and Table S1).** A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes, 1000 bootstrap replicates were run. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

doi:10.1371/journal.pone.0004438.g003



**Figure 4. Phylogenetic tree of 34 framework sequences (bold values) and 72 *Bacillus* sp., which could be designated as *B. subtilis*, *B. licheniformis* and *B. halodurans* (Figures S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31 to S32 and Table S1).** A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes, 1000 bootstrap replicates were run. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>). doi:10.1371/journal.pone.0004438.g004

*B. sphaericus* shows similarity to *B. clausii* and *B. halodurans* with respect to *Tru9I*; to *B. licheniformis* and *B. subtilis* for *HaeIII* and *RsaI* and to *B. pumilus* for *HaeIII*. Hence a combination of 3 enzymes *Tru9I*, *HaeIII* and *RsaI* will be necessary to distinguish it from its closely resembling species. *B. subtilis* is quite similar to *B. licheniformis* but can be distinguished on the basis of *Tru9I*. On the other hand *B. subtilis* and *B. pumilus*, which show similar RE patterns for *BfaI* and *HaeIII*, can be separated on the basis of *RsaI* (Figure 10).

#### Pattern of restriction enzyme sites in *Bacillus* spp.

On the basis of sequence similarity (phylogenetic closeness), and signature patterns (previous sections), 288 *Bacillus* sp. and 47 sequences of *Jeotgalibacillus*, *Brevibacillus*, *Geobacillus*, *Marinibacillus*, *Paenibacillus*, *Pontibacillus* and *Virgibacillus*, etc. could be segregated into 11 clusters. RE patterns of the 344 16S rDNA sequences of 10 *Bacillus* spp. served as references for the segregation of strains designated as *Bacillus* sp. (Figure 9).

Out of the 14 TypeII REs employed for *in silico* digestions, *SmaI*, *EcoRI*, *DpnII*, *RsaI*, *BfaI*, *HaeIII*, *Tru9I* and *AluI* could cleave 16S rDNA gene sequences with a frequency of 1 to 7 as was observed with known *Bacillus* spp. RE sites for *BamHI*, *NotI*, *SacI* occurred with a very low frequency of 0.4 to 3.0% and thus proved to be “non”-cutters. On the other hand, the sites for enzymes *NruI*, *HindIII* and *PstI* were observed to occur with moderate frequency in some of the clusters and could not be detected in some others. *SmaI* and *EcoRI* sites were observed with high frequency but with only one site per sequence, strong conclusions were difficult to draw. Hence, once again, information based on 6 REs - *DpnII*, *RsaI*, *BfaI*, *HaeIII*, *Tru9I* and *AluI* - generating more than 2 fragments proved useful for reaching meaningful conclusions.

On the basis of RE patterns (Figure 9), Cluster 1 comprising of 46 sequences (22 *Bacillus* sp. and 24 sequences of *Jeotgalibacillus*, *Marinibacillus*, *Ureibacillus* and *Sporosarcina*) showed similarity to 4 different *Bacillus* spp.: *B. subtilis*, *B. licheniformis*, *B. sphaericus* and *B. halodurans* (Table 4). Clusters numbered 4 and 7, showed similarity to 7 and 8 *Bacillus* spp., respectively. Incidentally, except for *Tru9I*, pattern of fragment length (nts) and order showed that Clusters 1, 4 and 8 resemble *B. licheniformis*, *B. sphaericus*, *B. megaterium*. However, there was quite a bit of variation among these clusters for the other 5 REs, which implies that the organisms in these clusters might have a common origin but are presently well distinguishable from each other. Clusters numbered 2, 3, 5, 6, 8, 9, and 10 showed RE fragment length and order pattern to be quite similar to 1 to 5 of the known *Bacillus* spp. but for certain REs there was no resemblance to known species. Clusters numbered 3, 6, 9 and 10 were quite distinct and hence categorized as “unique” on the basis of their response to *AluI*, *DpnII* and *HaeIII*. The fragment lengths and order were unique to each of them. A comparison of the different Clusters numbered 3, 6, 9, 10 with respect to their similarity to known species revealed that no two clusters showed exact match. The only exception to this observation, were the clusters 6 and 7 with respect to the action of *HaeIII*, where the order and size of the fragments were quite similar. For *BfaI*, *RsaI* and *Tru9I*, the similarity to known *Bacillus* spp. varied from cluster to cluster in two respects: firstly, the species identification varied from RE to RE, and secondly, no two clusters showed similar results with any two REs (Table 4).

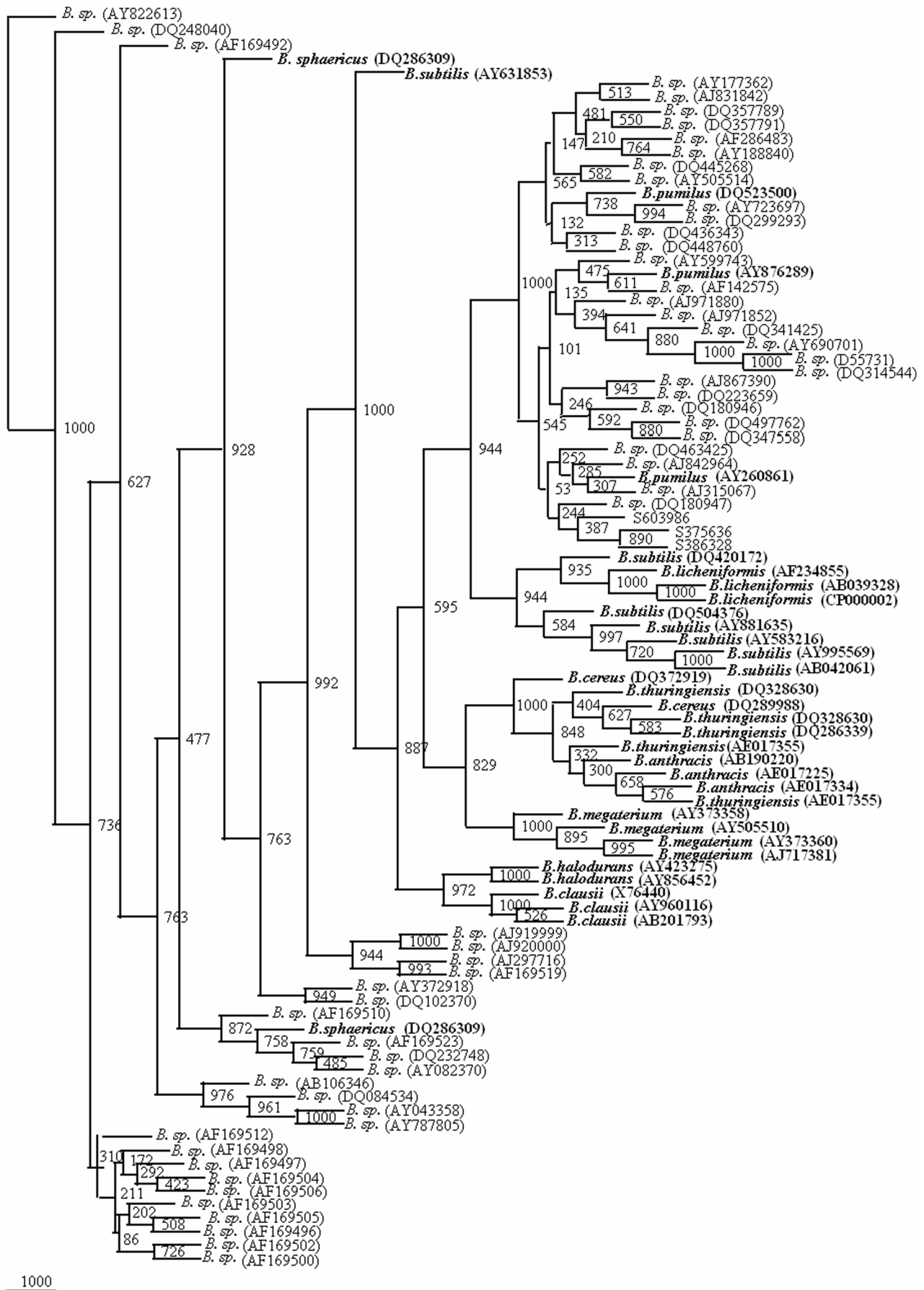
Clusters 1, 2, 6 and 11 were composed of *Bacillus* sp. which could be categorized into two groups: Group 1 consisted of *Bacillus* sp., which has now been reclassified as *Jeotgalibacillus*, *Marinibacillus*, *Ureibacillus*, *Sporosarcina* and Group 2 consisted of those which have been defined only up to genus level (*Bacillus* sp.). These served as controls for the observations made for *Bacillus* spp. with different REs. Cluster 1 showed closest matches with *Marinibacillus*, *B. subtilis* and *B. licheniformis* for *RsaI* (Figure 9). Such a dual relationship was also recorded for Cluster 1 with *BfaI*, *DpnII* and *Tru9I*. In these cases, the closest matches were *Sporosarcina*, *Jeotgalibacillus*, *Marinibacillus*, *B. licheniformis* and *B. sphaericus*. The variability in the closest matches among different REs reflects that Cluster 1 is a “unique” group intermediate to *Bacillus* and other closely related species such as *Jeotgalibacillus*, *Marinibacillus*, *Sporosarcina*, etc. Similarly, Clusters 2 and 6 showed uniqueness in their fragment order and length with 3 and 5 different REs, respectively. It implies that these groups of organisms need attention from taxonomists and may be classified as new organisms. Cluster 11 turned out to be primarily a set of 46 sequences of *Geobacillus* and 4 sequences of *Anoxybacillus*. These sequences largely resembled known *Geobacillus* sp. with respect to their RE activities. It thus served primarily as validation of the results recorded with *Bacillus* sp (Table 4).

#### Discussion

Among the group of aerobic endospore-forming bacteria of 25 genera and over 200 species, *Bacillus* is the largest and most prominent. *Bacillus* is comprised by heterogeneous assembly of gram-positive, rod shaped, spore forming bacteria, which may grow aerobically or as facultative anaerobes. Further characterization and identification has been traditionally based on biochemical tests and fatty acid methylester (FAME) analysis [106,107]. With further developments, API (Analytab Products, Inc) system of identification is quite reproducible and reliable [108]. Microbial genotype or DNA sequence based analytical approaches are highly reproducible and enable large number of samplings at a time [100]. 16S rDNA sequencing has proved to be one of the most powerful tools for the classification of microorganisms [109–111]. 16S rDNA sequencing technique has been used for the identification of *Bacillus* spp. such as *B. subtilis* [101]; *B. cereus* and *B. thuringiensis* [102]; *P. alvei* (formerly *B. alvei*) [103].

The genus *Bacillus* has undergone considerable taxonomic changes. The number of spp. within this genus was reduced from 146 in the 5<sup>th</sup> Edition of Bergey’s Manual of Determinative Bacteriology [112] to 22 [113]. In the Approved List of Bacterial Names [114], 31 of the 38 aerobic endospore formers were *Bacillus*. However, at present there are 175 *Bacillus* species (<http://rdp.cme.msu.edu/>). The two factors for the rapid increase are the application of more diverse and intelligent methods for enrichment and isolation and the development of new and ever more sophisticated methods of amplification and sequencing of genes [115].

The identification of *Bacillus* species on the basis of 16S rDNA gene sequence is done by blasting it against the available databases. The need for developing a tool for identifying *Bacillus* species arose due to two reasons: (i) more than 50% of the 16S rDNA sequences deposited in the databases, have been annotat-



1000

**Figure 5. Phylogenetic tree of 34 framework sequences (bold values) and 55 *Bacillus* sp., which could be designated as *B. sphaericus* and *B. pumilus* (Figures S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31 to S32 and Table S1).** A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes, 1000 bootstrap replicates were run. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>). doi:10.1371/journal.pone.0004438.g005

ed/identified only as *Bacillus* sp., ii) *B. subtilis* strains were seen to cluster in different clades and widely placed clusters on the phylogenetic tree (Data not shown here) raising doubts about the sequencing quality. It gives an impression that *B. subtilis* perhaps defies the phylogenetic pattern. It also suggests that the genera *Bacillus* may be further divided into sub-species, particularly as far as *B. subtilis* is concerned. Our data imply that the 16S rDNA molecule of the *B. subtilis* may exist in two different states much similar to *B. cereus*. This may lead to a different structure of the 30S subunit since binding of primary binding proteins affects binding of the secondary and tertiary binding proteins [116]. In fact, Nakamura et al. [77] divided *B. subtilis* in to two subspecies, namely *B. subtilis* subsp. *subtilis* and *B. subtilis* subsp. *spizizenii* on the basis of cell wall chemistry and DNA-DNA relatedness data. These were therefore regarded as genomovars. In view of this scenario, we thought of developing a tool which will enable identification of new *Bacillus* isolates at species level. In fact, a few *Bacillus* species, such as *B. kaustophilus*, *B. sterothomophilus*, *B. thermoglucosidarius* and *B. thermoleovorans* have been transferred to the newly created genus *Geobacillus* [65]. Certain others such as *Bacillus globisporus*, *B. pasteurii* and *B. psychrophilus* have been reclassified to the genus *Sporosarcina* [66]. In the same manner, *B. marinus* has been reclassified as *M. marinus* [67].

### Problems with *Bacillus* Species

The members of *B. cereus* group show 99.5 to 100% similarity for their 16S and 23S rDNA sequences [117,118]. The genetic diversity of *B. cereus* isolates is enhanced by extra chromosomal elements [92,119–122]. The need to develop approaches that rapidly identify the “near neighbours” of *B. cereus* group are of great interest for the study of *B. anthracis* virulence mechanisms as well to prevent the use of such strains for *B. anthracis* based bioweapon development [123]. In this study all the three members of the *B. cereus* group have very large conserved regions and appear indistinguishable on the basis of 16S rDNA gene sequence. *B. thuringiensis* and *B. mycoides* differ from *B. anthracis* and *B. cereus* by 0 to 9 nucleotides [47]. Even single strand conformation of polymorphism (SSCP) did not allow species discrimination within *B. cereus* group [124]. Variable region VI of 16S rDNAs of *B. cereus* and *B. thuringiensis* are useful for differentiation between these species [125]. However, our approaches have been able to distinguish *B. cereus* and *B. thuringiensis* from *B. anthracis*. With the first approach of using the representative sequences for each species, these three though distinguishable were always placed next to each other. On the other hand, with specific signature sequence, it was possible to distinguish *B. cereus* from other two. It was difficult to identify species specific signatures for *B. anthracis* and *B. thuringiensis*. A strategy has been proposed by Daffonchio et al. [123] for the identification of near neighbours of *B. anthracis* based on single nucleotide polymorphism (SNP) in the 16S–23S rDNA ITS containing tRNA genes, characteristic of *B. anthracis*. Two *B. cereus* strains and one *B. thuringiensis* strains showed RSI-PCR profiles identical to that of *B. anthracis*. The strict relationship with *B. anthracis* was confirmed by MLST of four independent loci: the 16S–23S rDNA long ITS, the SG-749 fragment that included a region homologous to *B. subtilis ypnA* gene; the AC-390 fragment that is homologous to the *B. subtilis ywfK* gene, encoding a hypothetical transcriptional regulator belonging to the LysR

family; the *pleR* gene encoding for a pleiotropic regulator previously identified as one of the principal regulators of *B. cereus* virulence gene and the *cerA* gene that encodes the cereolysin A phospholipase. In yet another population genetic study among a strain collection of *B. cereus* group species, it was found by MLEE and MLST that the strains could be divided into two main groups [123]. The difficulty in distinguishing *B. cereus* group members on 16S rDNA based diagnosis however, correlated well with gyrase B (*gyrB*) as a molecular diagnostic marker [126].

Rep-PCR has been shown to be a useful technique in the subtyping of *Bacillus* species [37,38]. However, protein coding genes such as *gyrA* and *rpoB* exhibit much higher genetic variation. These genes have been thus used for the classification of closely related taxa within the *B. subtilis* group [10,39]. In spite of such difficulties our frame work is proving an efficient tool to handle such problems. With our approach two *B. subtilis* groups could be easily segregated on the basis of 16S rDNA gene sequence itself. These reference sequences in fact could segregate the two subspecies proposed by Nakamura et al. [77] namely *B. subtilis* subsp. *subtilis* and *B. subtilis* subsp. *spizizenii* on the basis of cell wall chemistry and DNA-DNA relatedness data. Of the two strains chosen within the *B. subtilis* Gr1, one of them was highly similar to *B. licheniformis* (Accession No. DQ504376), whereas the other strain was quite distinct (*B. subtilis* Accession No. AY631853).

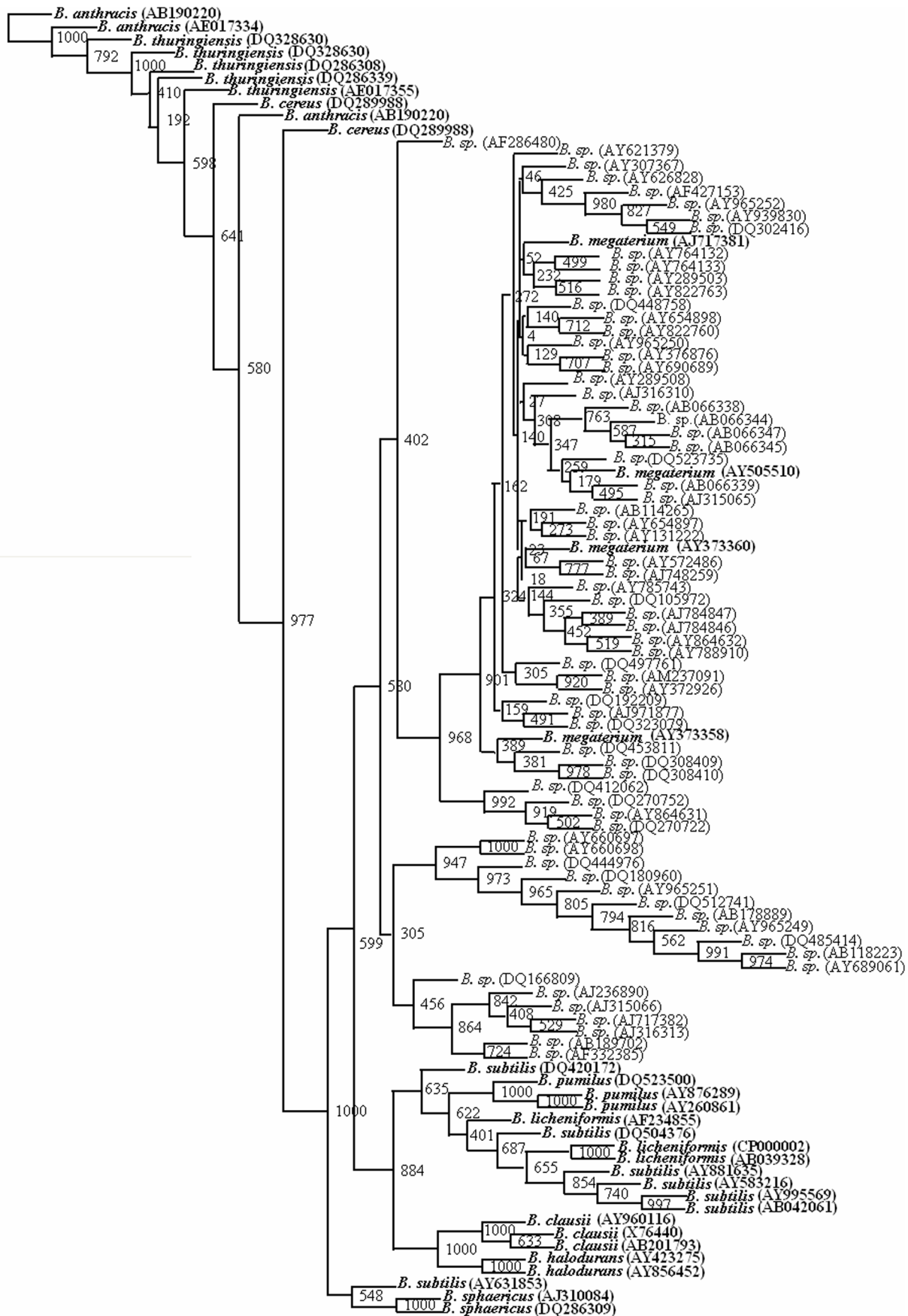
The large degree of variation in the individual group fingerprints suggests that a substantial intra-species genetic diversity may exist and highlights the very high resolution of 16S rDNA gene sequencing. The data presented here is to our knowledge the first time, where molecular technique has been exploited and applied in this manner for assigning *Bacillus* isolates to different species. It would be difficult to extrapolate to define the potential limits of genetic variability within the *Bacillus*. However, such a strategy can be extended to other genera as well. It could form the basis for developing reference phylogenetic tree for various genera for which large number of 16S rDNA sequencing data is available.

The 16S rDNA sequences therefore give consistent separation of the strains into 9 major, non-overlapping clusters. The significance of this formal cluster assignment is made clear by the inclusion in our data set of sequences from independently reported strains. The signature nucleotide offers the opportunity for designing species-specific probes as primers for a rapid identification/segregation of new isolates.

Since pathogenic capacities of *Bacillus* species are often plasmid linked i.e. the pXO2 encoded capsule gene cluster, it may not necessarily be linked to internal genotypic grouping of taxa. The fact that plasmids can easily be transferred or lost makes these criteria unacceptable for typing purposes. Our reference phylogenetic tree may partially replace the use of schemes such as MLST.

### Signatures

16S rDNA sequences vary due to substitutions and not due to insertion or deletion of bases [127,128]. Based on this they could identify repeating elements that are highly conserved across different species of *Pseudomonas* [128]. There are programs available for designing polymerase chain reaction (PCR) primer pairs as a means of rapid detection [129,130]. These programs select primer pairs based on the user defined parameters.



1000

**Figure 6. Phylogenetic tree of 34 framework sequences (bold values) and 69 *Bacillus* sp., which could be designated as *B. megaterium* (Figures S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31 to S32 and Table S1).** A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes, 1000 bootstrap replicates were run. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>). doi:10.1371/journal.pone.0004438.g006

However, it becomes difficult to select the best primers since they do not provide information regarding the specificity of the oligonucleotides or patterns [131–134]. *B. cereus* has time and again been reported to be highly homogenous however, signature sequences of the 16S rDNA could distinguish the psychrotolerant and mesophilic strains. Single base pair substitutions were randomly distributed over the gene. The most obvious differences was one signature located at bp 180 to 192 (*E. coli* nomenclature) or bp 180 to 201 (*B. cereus* nomenclature) [127].

### Restriction Enzyme Analysis

The phenomenon of host specific restriction and modification of bacterial viruses stemmed from endonucleases within the cells that destroy foreign DNA molecules. REs cleave DNA at specific sites, generating discrete and gene-size fragments and have proved to be a remarkable tool for investigating gene organization, function and expression [135–137]. REs occur in combination with 1 or 2 modification enzymes (DNA methyl transferases) that protect the cell's own DNA from cleavage by the RE. Since ME methylates at the same site where RE cuts, R-M system perhaps ensures that 16S rDNA remains conserved in spite of the fact that a large number of sites for each RE are present. In other genes, presence of RE sites increases diversity by promoting recombination [138,139]. PCR Restriction analysis (PRA) of 16S rDNA has been shown to contribute to rapidly and reliably identify newly isolated strains belonging to recognized species [140]. However as a result of analyzing a large set of data encompassing many species, we may extend the statement that this method can be applied for recognizing so far unrecognized species as well. In their study, they have applied four REs: *HaeIII*, *HinfI*, *TaqI* and *RsaI*. However, *HinfI* and *RsaI* showed no distinctive patterns for the strains tested. In fact, *TaqI* proved instrumental in clearly distinguishing one of the groups. A wide range of REs such as *HaeIII*, *DpnII*, *RsaI*, *BfaI* and *Tru9I* were used for defining the genus *Virgibacillus* [141,142]. *B. licheniformis* 16S rDNA sequence was spliced with *AluI* into five bands –70; 140; 180; 200 and 800 nts, whereas *RsaI* resulted in four bands –110; 400; 450; 500 nts [105].

An innovative approach for revealing intraspecific genomic variability of *B. cereus* and *B. licheniformis* was the PCR fingerprinting of the spaces between the 16S and 23S rRNA genes and of intergenic tRNA genes regions [60]. Although RAPD showed remarkable diversity among *B. cereus* strains however, it was realized that the genetic diversity can arise from plasmid wide variability in the plasmid profiles. *B. licheniformis* formed 2 groups with all the methods. Based on single strand conformation of polymorphism analysis after RE (*AluI* and *RsaI*) digestions of 16S rDNA, two different evolutionary schemes for the two *Bacillus* species, *B. cereus* and *B. licheniformis* were proposed [60]. In our analysis, only 6 out of 14 REs (*DpnII*, *RsaI*, *BfaI*, *HaeIII*, *Tru9I* and *AluI*) - proved beneficial in easy distinction. The rest 8 REs (*BamHI*, *EcoRI*, *HindIII*, *NoI*, *NruI*, *SacI*, *SmaI*, and *PstI*) could not be exploited to significant extents to be useful for this purpose. However, in a different gene such as *gyrB*, digestion with *EcoRI* (and *ClaI*) could distinguish the four members of the *B. cereus* group but *HindIII* did not [38]. So the same set of REs may not hold good for different gene sequences. In fact *HindIII* gave poor results even while screening *Corynebacterium* spp., hence was discontinued

and multiple RE usage was recommended [143]. *In silico* digestion with RE *AluI* was found to be most discriminative [144] and generated 3 to 13 fragments depending on the *Mycoplasma* species. Although 73 *Mycoplasma* species could be differentiated using *AluI*, other species gave undistinguishable patterns. For these, an additional restriction digestion typically with *BfaI* (or hpyF10VI) was needed for a final identification [145]. This was confirmed by application of ARDRA on 27 species and subspecies. We also validated our findings by applying RE to species of *Virgibacillus*, *Gracibacillus* and *Geobacillus* and can be exploited for describing new species [146].

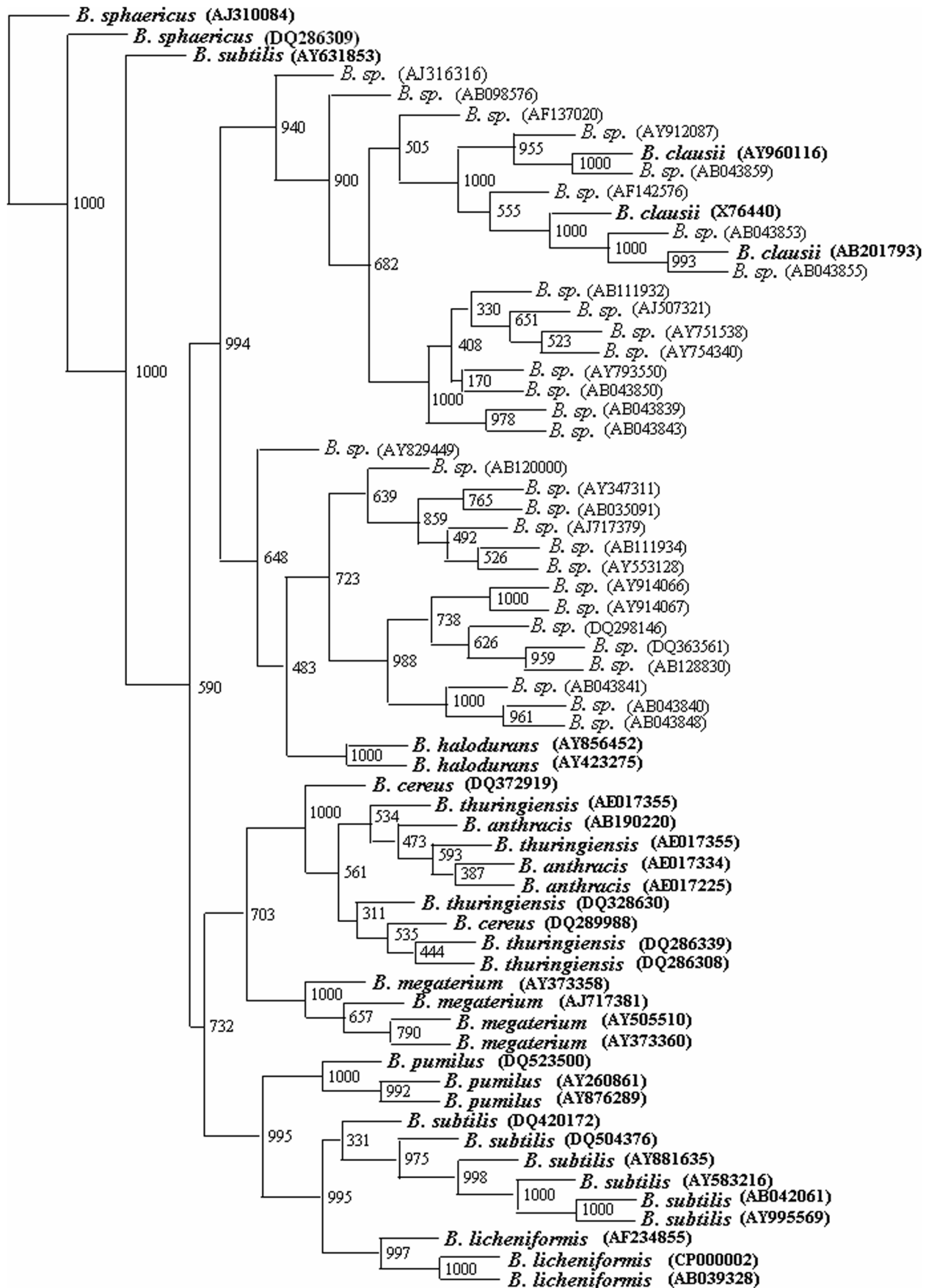
### Novel Lineages

Among the innovative strategies applied by various researchers, phylogenetic relationships between *Bacillus* species and related genera were inferred from comparison of 3' end 16S rDNA and 5' end 16S–23S ITS nucleotide sequences [82]. Among the 40 Bacillaceae species, *Bacillus circulans* remained ungrouped. Out of the ten groups, Group VI constituted of *B. licheniformis*, *B. subtilis*, *B. sphaericus* along with *B. amyloliquefaciens*, *B. atrophaeus*, *B. mojavensis*, *B. macroides* and *B. fusiformis*. Group X was placed independent of other 6 *Bacillus* groups and was comprised of *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. mycoides* and *B. lentus*. It indicates that *B. cereus* group is quite different from other *Bacillus* species. Separation of *Bacillus* species by *Paenibacillus*, *Brevibacillus*, *Geobacillus*, *Marinibacillus* and *Virgibacillus* species, indicates that in some cases, further divisions or conversely further grouping might be warranted. Our work has provided the tools for defining the thresholds of each species and enables us to pose the questions such as: Should current classifications be re-examined?

Lower levels of similarity were found with other alkalitolerant *Bacillus* strains particularly DSM8714 and DSM877, which still lack taxonomic standing. The results obtained confirm that the four *Eterogermina* strains belong to a unique genospecies which can be unequivocally identified as *B. clausii*. The finding is of intrinsic value, since some bacterial strains described as *B. clausii* strains have been reported to exhibit levels of DNA hybridization with the reference type strain of less than 61% [147]. Thus emphasizing the great genomic heterogeneity of the strains placed in the species, *B. clausii* [32].

As Ash et al., [42] predicted, their phylogenetic groups have been redefined as separate genera, and their outlying species have served as the basis for novel taxa as well. At the present pace of refined discoveries, we can expect new *Bacillus* - like genera to be redefined in the near future. In fact, Nazina et al., through physiological and genetic analysis submitted the validly described genus name of *Geobacillus* [65].

Heyndrickx et al., [142] undertook a polyphasic study, which revealed the presence within *Virgibacillus* of an as yet undescribed new species for which the name *Virgibacillus proomii* was proposed (*V. proomii* was distinguished from *V. pantothenicus* and members of *Bacillus sensu stricto* and from members of *Paenibacillus* and other aerobic endospore-forming bacteria by routine phenotypic tests). Comparisons of the 16S rDNA sequences of type strains of *Bacillus* and *Sporosarcina* species indicated that *Bacillus pantothenicus* lies at the periphery of rDNA gr1 (*Bacillus sensu stricto*) [42]. *Virgibacillus* was proposed to accommodate *B. pantothenicus* and two related organisms, which appeared to belong to an as yet (as on 1999)



1000



**Figure 7. Phylogenetic tree of 34 framework sequences (bold values) and 31 *Bacillus* sp., which could be designated as *B. clausii* (Figures S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31 to S32 and Table S1).** A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes, 1000 bootstrap replicates were run. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>). doi:10.1371/journal.pone.0004438.g007

undescribed new species. It appears that the relationship between *Bacillus laevolecticus* and *V. pantothenicus* (Group III) and between *Bacillusadius* and *M. marinus* (Group IV) could still be open to debate. The robustness of this classification tool will be assessed by comparison with the current Bacillaceae classifications.

The various parameters like signatures (generated by MEME), restriction enzyme (RE) sites, nucleotide stretches “generated” by RE and the phylogenetic framework together can enable to generate a battery of markers. These are likely to define the variability between the species of a specific genus and the specificity of the genus. The use of these parameters is a simple, rapid approach, suitable to larger screening programs and easily accessible to most laboratories.

## Materials and Methods

### Sequence data

A total of 2146, 16S rDNA sequences belonging to the genus *Bacillus* (from RDP/NCBI sites: <http://rdp.cme.msu.edu/>; <http://www.ncbi.nlm.nih.gov/>) were analysed in the present study. These included 271 sequences belonging to isolates of *B. subtilis*, 211 to isolates of *B. cereus*, 153 to isolates of *B. anthracis*, 131 to isolates of *B. licheniformis*, 108 to isolates of *B. thuringiensis*, 83 to isolates of *B. pumilus*, 47 to isolates of *B. megaterium*, 42 to isolates of *B. sphaericus*, 39 to isolates of *B. clausii*, 36 to isolates of *B. halodurans* and 1025 to isolates of *Bacillus* species (Table 1). The first ten sets of *Bacillus* species were used as the master species set for this analysis for generating a phylogenetic framework while the 1025 *Bacillus* species were used as a data set for segregating these unclassified *Bacillus* species. The other *Bacillus* species, which were represented by relatively minor numbers, were not considered for the development of the identification tool (Table S5).

### Phylogenetic Analyses

For phylogenetic analyses of each of these 10 species data sets, the sequences from each of them were assembled and aligned using the multiple alignment program CLUSTALW version 1.82 [133]. To estimate evolutionary distance, pairwise distances between all taxa were calculated with the DNADIST of the PHYLIP 3.6 package. The resultant distance matrix was then used to draw a neighbor joining tree with the program NEIGHBOR. The program SEQBOOT [148] was used for statistical testing of the trees by resampling the dataset 1000 times. The trees were viewed through HyperTree Version 1.0.0 [133] and TreeView Version 1.6.6 [149] (Figures S1, S2, S3, S4, S5, S6, S7, S8, S9, S10).

For each of these 10 data sets, sequences which fell in the same clade were grouped together. Candidate sequences of these individual groups were aligned and a consensus was obtained by removing ambiguous parts using JALVIEW sequence editor. Consensus from each group was chosen as a representative for the particular group. A phylogenetic tree was drawn on the basis of these representative sequences of the 16S rDNA gene. Members from each of the clusters of the tree were selected to define the range of each of the *Bacillus* species. Thus a reference set of 34 sequences was selected that contained two to five from each represented topology (Table 5) and regarded these as likely candidates that could give information about the organismal phylogeny.

### *Bacillus* Species-Specific Signature

MEME (Multiple EM for Motif Elicitation) is used for searching for novel motifs or signatures in sets of biological sequences. MEME works by searching for repeated, ungapped sequence patterns that occur in the DNA or protein sequences [150,151]. MEME searches can be performed via the web server (<http://meme.nbcr.net>) and its mirror sites [151]. The same web server also allows access to motif alignment and search tool to search sequence databases for matches to motifs. To successfully discover motifs with MEME, it is necessary to choose and prepare the input sequences carefully. Ideally, the sequences should be <1000 base pairs long [152]. In our analysis, sequences of 10 data sets in FASTA format were submitted group wise in MEME program Version 4.0.0 (<http://meme.nbcr.net/meme4/cgi-bin/meme.cgi>). In order to obtain maximum number of motifs in our sequences, we modified default settings from 3 motifs to 10 motifs. MEME search stops when this number of motifs has been found, or when none can be found with E-value less than 10000 (<http://meme.nbcr.net/meme4/meme-input.html#width>). We used default setting zero or one motif per sequence to get the occurrence of single motif which is distributed among the sequences. The default value of motif widths, set between 6 (minimum) and 50 (maximum) were modified and re-set between 25 and 30, respectively. Each of the 10 signatures (25 to 30 nucleotides long) (Table S2) was checked for its frequency of occurrence among all the sequences of a particular *Bacillus* sp. and the ones with highest frequency and did not appear in other *Bacillus* spp. were considered as unique to this species. These unique motifs were used as query sequence to BLAST against the sequenced microbial genomes available in NCBI database (<http://www.ncbi.nlm.nih.gov/>), to validate the results.

### Restriction Enzyme Analysis

A total of 14 Type II Restriction enzymes (Table 3) were considered for these analyses. The criteria for selecting type II RE are independence of methylase and occurrence of cleavage(s) at very specific sites that are within or close to the recognition sequence.

All the 10 *Bacillus* species considered were checked for all the 14 Type II RE using the online software: Restriction Mapper Version 3 (<http://restrictionmapper.org/index.htm>). Sequences (one at a time) were entered in the restriction mapper site, results obtained were analyzed and consensus pattern was determined for each species depending upon its frequency of occurrence in the sequences. Ten known *Bacillus* spp. and 11 clusters belonging to *Bacillus* sp. were used as data sets. For *B. megaterium*, *B. sphaericus*, *B. clausii*, *B. halodurans* all the sequences were analysed and for *B. cereus*, *B. anthracis*, *B. licheniformis*, *B. thuringiensis*, *B. pumilus* 30 sequences were taken into consideration but for *B. subtilis* since no conclusive pattern could be made using 30 sequences, so 150 sequences were taken. The patterns developed for each *Bacillus* species were considered as a representative for that specific species.

### Identifying *Bacillus* Species

The reference set of 34 sequences has been used to segregate unclassified *Bacillus* species. A total of 1025 16S rDNA sequences from *Bacillus* sp. were screened at the rate of 52 *Bacillus* species

**Table 2.** Characteristics of unique nucleotide signatures for 16S rDNA sequences of different *Bacillus* spp. and clusters of *Bacillus* sp.

<i>Bacillus</i> spp./Cluster <sup>a</sup>	Signature	Length (nts) <sup>b</sup>	Frequency <sup>c</sup>
<i>Bacillus cereus</i>	AAAGTGGAAATTCATGTGTAGCGGTGAAAT	30	151/211
<i>B. thuringiensis</i>	ATAACATTTTGAAGTGCATGGTTCGAAAT CTTTAGTCTGAAGTTAACGCATTAAGCA	30 29	43/108 63/108
<i>B. anthracis</i>	No unique signature was detectable		0/153
<i>B. clausii</i>	AATCCATAAAGCCATTCTCAGTTCGGATT AAATGATTGGGGTGAAGTCGTAACAAGGTA AAACCGGAGCTAATACCGGATAATCCCTTT GCATTAGCTAGTTGGTAAGGTAACGGCTTA GTAGTGCCGAAGTTAACACATTAAGCACT	30 30 30 30 29	14/39 31/39 14/39 15/39 23/39
<i>B. halodurans</i>	ATAATAAAAAGAACTGCATGGTCTTTTTT ACCAAAGGGAGCTTGCTCCTAGAGGTTAGC	30 30	21/36 01/36
<i>B. pumilus</i>	AAGGTTTAGCCAATCCCACAAATCTGTTCT AAGGTTTAGCCAATCCCATAAATCTGTTCT ATAGTTCCTGAACCGCATGGTTCAGGAT	30 30 30	14/83 36/83 42/83
<i>B. megaterium</i>	ATGATTGAAAGATGGTTTCGGCTATCACTT AATCCATAAAAACCATTCAGTTCGGATT AACTGATTAGAAGCTTGCTTCTATGACGTT TCTTGACATCCTCTGACAACCTAGAGATA TGGGATAACTTCGGGAACCGAAGCTAATA	30 30 30 30 30	36/47 19/47 28/47 33/47 34/47
<i>B. sphaericus</i>	TAAAACCTGTTGTAAGGGGAAGAACAGTA ATAGTGGAAATCCAAGTGTAGCGGTGAAAT TAATCCGATAAAGTCTGTTCTCAGTTCGGAT AGTAACACGTGGGCAACCTACCTTATAGTT TAACTGGCTGTACCTTGACGGTACCTTATT	30 30 30 30 30	23/42 14/42 23/42 10/42 27/42
<i>B. subtilis</i>	No unique signature was detectable		0/211
<i>B. licheniformis</i>	No unique signature was detectable		0/131
<i>Bacillus</i> sp. Cluster 1	TAAAACCTGTTGTAAGGGGAAGAACAGTA AATCCATAAAAACCGTTCAGTTCGGAT	30 29	19/46 05/46
<i>Bacillus</i> sp. Cluster 5	AATCCATAAATCTATTCTCAGTTCGGATT	30	05/32
<i>Bacillus</i> sp. Cluster 6	AAGCAAATCCCATAAAAACCATTCAGT TCAAGCAAATCCCATAAAAACCATTCAGT	28 30	09/23 07/23
<i>Bacillus</i> sp. Cluster 7	ATAACTCATTCTCGCATGAGGAAATGTT	30	10/37
<i>Bacillus</i> sp. Cluster 9	AATCCATAAAAACCACTCTCAGTTCGGATT	30	01/25
<i>Bacillus</i> sp. Cluster 10	AATCCACAAAACCGTTCAGTTCGGATT	30	03/25
<i>Bacillus</i> sp. Cluster 11	TAAACGATGAGTGCTAAGTGTAGAGGGGT GTCGTAAGCTCTGTTGTGAGGGACGAAGG	30 30	23/50 25/50

<sup>a</sup>Cluster represent isolates defined only up to genus level *Bacillus* sp.

<sup>b</sup>nucleotides.

<sup>c</sup>Frequency of occurrence of the signature out of the total sequences screened.

Out of 11 Clusters, no unique signatures were detectable in Clusters 2, 3, 4 and 8.

doi:10.1371/journal.pone.0004438.t002

with 34 reference sequences by generating 22 phylogenetic trees (Figures S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32). In each of the phylogenetic tree, bootstrap percentage values that showed a stability of at least 80% were considered. These *Bacillus* species were then taken to reconstruct final phylogenetic tree with their respective species. These strains were also checked for the *Bacillus* species-specific signatures (similarly as done for 10 data sets) and restriction enzyme analysis as benchmark for concluding that the strain belonged to the specific *Bacillus* species.

### Clusters/Potential Novel Lineages

In addition to the identification of some *Bacillus* spp. up to species level, [366 sequences were aligned {according to clusters made in 22 trees of *Bacillus* sp.} and 52 representatives were chosen and a phylogenetic tree was made] 335 isolates were found to cluster in to 11 groups {44 representatives out of 52} (Figure 8). These 11 clusters consisting of 18 to 50 isolates each did not

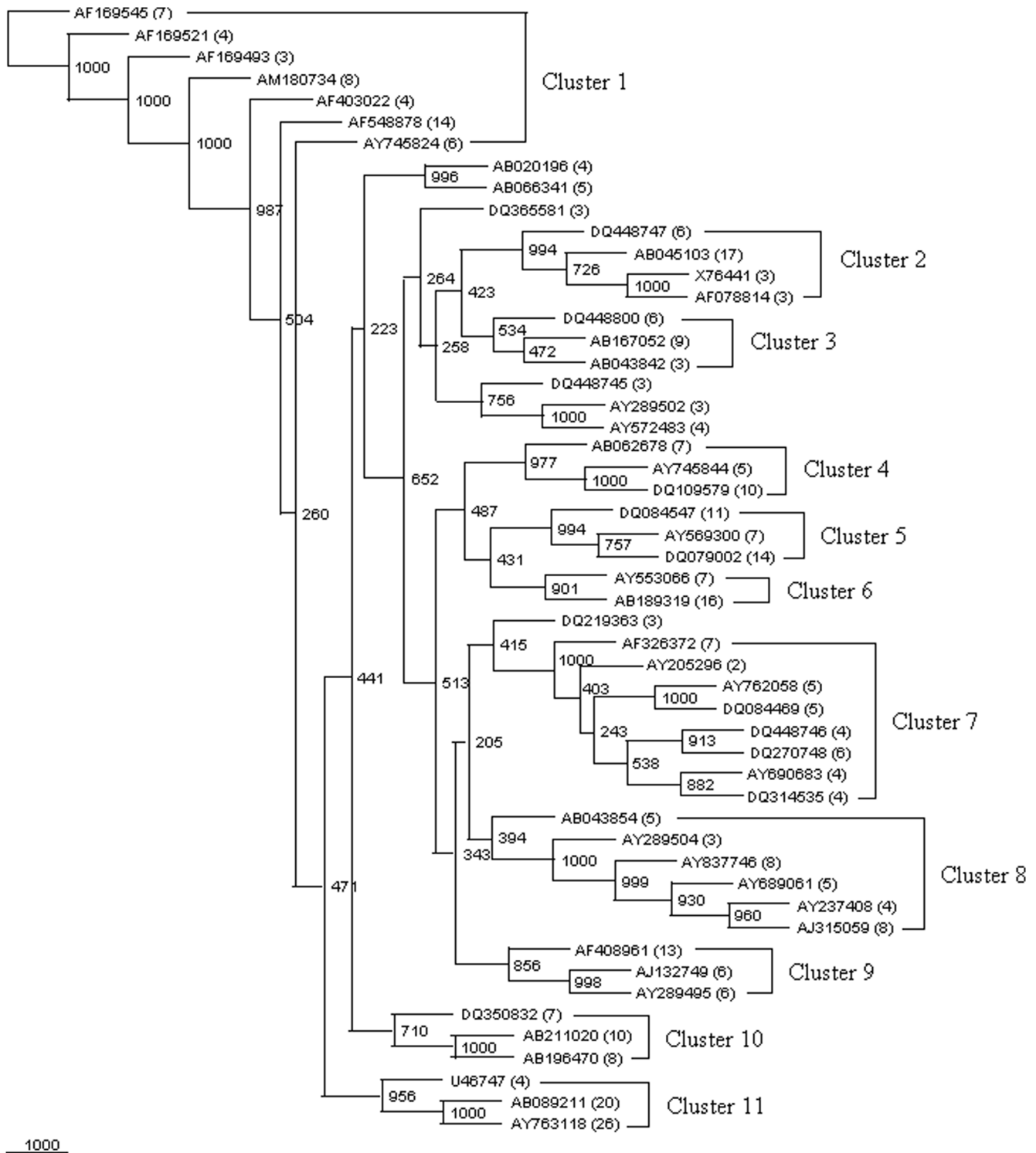
appear to fall within any of the aligned species and so may represent sub-species or novel lineages. These were also checked for signatures identified through MEME program to reveal that if there are certain similarities. Each of the 11 clusters was also checked for the 14 type II restriction enzyme digestion. The pattern so obtained was checked against those of the representative *Bacillus* species for a match so that a conclusion could be drawn that these might belong to some *Bacillus* species.

### Supporting Information

**Table S1** Accession numbers of 16S rDNA sequences of *Bacillus* sp. identified up to species

Found at: doi:10.1371/journal.pone.0004438.s001 (0.05 MB DOC)

**Table S2** Signatures of *Bacillus* species obtained through MEME software (<http://meme.sdsc.edu/meme/meme.html>)



**Figure 8. Phylogenetic tree of 52 representative sequences (appeared as 52 clusters {equivalent to 366 sequences} in Figures S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31 to S32).** These 52 representative sequences were observed to segregate in 11 clusters: cluster 1 to cluster 11. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes, 1000 bootstrap replicates were run. Accession numbers of the representative sequences are given and values in parentheses against these are equal to the number of sequences in that group within each cluster (<http://rdp.cme.msu.edu/>).

doi:10.1371/journal.pone.0004438.g008

**Table 3.** Restriction enzyme (Type II) used in the study with their cut sites (rebase.neb.com/rebase/rebase.html).

S. No.	Restriction Enzyme	Cut site
1.	<i>AluI</i>	AG ↓ CT
2.	<i>BamHI</i>	G ↓ GATCC
3.	<i>Bfal</i>	C ↓ TAG
4.	<i>DpnII</i>	↓ GATC
5.	<i>EcoRI</i>	G ↓ AATTC
6.	<i>HaeIII</i>	GG ↓ CC
7.	<i>HindIII</i>	A ↓ AGCTT
8.	<i>NotI</i>	GC ↓ GGCCGC
9.	<i>NruI</i>	TCG ↓ CGA
10.	<i>RsaI</i>	GT ↓ AC
11.	<i>SacI</i>	GAGCT ↓ C
12.	<i>SmaI</i>	CCC ↓ GGG
13.	<i>Tru9I</i>	T ↓ TAA
14.	<i>PstI</i>	CTGCA ↓ G

doi:10.1371/journal.pone.0004438.t003

Found at: doi:10.1371/journal.pone.0004438.s002 (0.10 MB DOC)

**Table S3** Characteristics of nucleotide signatures for 16S rDNA gene of clusters of *Bacillus* sp.

Found at: doi:10.1371/journal.pone.0004438.s003 (0.05 MB DOC)

**Table S4** Occurrence of restriction endonuclease digestion sites in 16S rDNA sequence(s) of *Bacillus* spp. and clusters of *Bacillus* sp. with low frequency or limited RE sites.

Found at: doi:10.1371/journal.pone.0004438.s004 (0.03 MB DOC)

**Table S5** List of *Bacillus* species available at <http://rdp.cme.msu.edu/>.

Found at: doi:10.1371/journal.pone.0004438.s005 (0.19 MB DOC)

**Figure S1** Phylogenetic tree based on 153, 16S rRNA gene sequences of *Bacillus anthracis*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s006 (7.75 MB TIF)

**Figure S2** Phylogenetic tree based on 211, 16S rRNA gene sequences of *Bacillus cereus*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s007 (2.16 MB TIF)

**Figure S3** Phylogenetic tree based on 108, 16S rRNA gene sequences of *Bacillus thuringiensis*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the

study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s008 (1.52 MB TIF)

**Figure S4** Phylogenetic tree based on 271, 16S rRNA gene sequences of *Bacillus subtilis*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s009 (10.12 MB TIF)

**Figure S5** Phylogenetic tree based on 131, 16S rRNA gene sequences of *Bacillus licheniformis*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s010 (1.63 MB TIF)

**Figure S6** Phylogenetic tree based on 83, 16S rRNA gene sequences of *Bacillus pumilus*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s011 (7.14 MB TIF)

**Figure S7** Phylogenetic tree based on 42, 16S rRNA gene sequences of *Bacillus sphaericus*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s012 (0.75 MB TIF)

**Figure S8** Phylogenetic tree based on 36, 16S rRNA gene sequences of *Bacillus halodurans*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s013 (0.87 MB TIF)

**Figure S9** Phylogenetic tree based on 39, 16S rRNA gene sequences of *Bacillus clausii*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s014 (0.89 MB TIF)

**Figure S10** Phylogenetic tree based on 47, 16S rRNA gene sequences of *Bacillus megaterium*. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Bold sequences are the ones considered as framework in the study. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s015 (0.88 MB TIF)

**Figure S11** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136

*Bacillus* spp. – Reference

RsaI		HaeIII		AhaI	
<i>Bacillus anthracis</i>	406 • 355 • 146	<i>Bacillus anthracis</i>	22 • 565 • 34 • 456	<i>Bacillus anthracis</i>	174 • 224 • 599 • 185 • 21 • 58
<i>B. thuringiensis</i>	406 • 355 • 146	<i>B. thuringiensis</i>	• 22 • 568 • 34 • 457	<i>B. thuringiensis</i>	174 • 224 • 599 • 185 • 21 • 58
<i>B. cereus</i>	406 • 355 • 146	<i>B. cereus</i>	• 22 • 565 • 34 • 457	<i>B. cereus</i>	174 • 224 • 599 • 185 • 21 • 58
<i>B. halodurans</i>	171 • 235 • 357 • 146	<i>B. halodurans</i>	70 • 22 • 591 • 459	<i>B. halodurans</i>	173 • 224
<i>B. megaterium</i>	16 • 11 • 406 • 356 • 146	<i>B. megaterium</i>	70 • 22 • 571 • 459	<i>B. megaterium</i>	86 • 18 • 612 • 209 • 265 • 123
<i>B. clausii</i>	• 19 • 11 • 406 • 356 • 146	<i>B. clausii</i>	• 70 • 22 • 55 • 216 • 264 • 4	<i>B. clausii</i>	27 • 22 • 104 • 211 • 417 • 4
<i>B. licheniformis</i>	• 19 • 11 • 406 • 351 • 146	<i>B. licheniformis</i>	• 22 • 591 • 457 • 4	<i>B. licheniformis</i>	138 • 33 • 322 • 363
<i>B. sphaericus</i>	• 18 • 11 • 405 • 406	<i>B. sphaericus</i>	• 70 • 22 • 596 • 452 • 4	<i>B. sphaericus</i>	173 • 613 • 209 • 265 • 4
<i>B. subtilis</i>	• 19 • 11 • 407 • 302 • 146	<i>B. subtilis</i>	• 22 • 600 • 457	<i>B. subtilis</i>	• 173 • 186 • 431 • 207 • 265 • 4
<i>B. pumilus</i>	• 409 • 302	<i>B. pumilus</i>	• 22 • 599 • 458 • 4	<i>B. pumilus</i>	• 164 • 42 • 779 • 374
				<i>B. pumilus</i>	83 • 18 • 183 • 430 • 207 • 265 • 4

Cluster 3, 4, 5, 7-10

RfaI		HaeIII		DpaII	
Cluster 3 (18)	• 579 • 183 • 335	<i>Bacillus anthracis</i>	527 • 270 • 5 • 86 • 134 • 44	<i>Bacillus anthracis</i>	298 • 814 • 237 • 175
Closest Match	<i>B. halodurans</i>	<i>B. thuringiensis</i>	527 • 270 • 5 • 86 • 134 • 44	<i>B. thuringiensis</i>	297 • 814 • 238 • 175
Cluster 4 (22)	• 763 • 335 • 4	<i>B. cereus</i>	• 527 • 270 • 5 • 85 • 134 • 44	<i>B. cereus</i>	• 814 • 238 • 174
Closest Match	<i>B. megaterium</i>	<i>B. halodurans</i>	270 • 5 • 86 • 136	<i>B. halodurans</i>	283 • 119 • 934 • 174
Cluster 5 (32)	• 763 • 335 • 4	<i>B. megaterium</i>	280 • 86 • 137 • 4	<i>B. megaterium</i>	• 132 • 815 • 238 • 4
Closest Match	<i>B. megaterium</i>	<i>B. clausii</i>	270 • 5 • 86 • 134 • 4	<i>B. clausii</i>	296 • 814 • 238 • 4
Cluster 7 (37)	• 763 • 335 • 4	<i>B. licheniformis</i>	• 271 • 86 • 134 • 4	<i>B. licheniformis</i>	• 119 • 695 • 187 • 51 • 4
Closest Match	<i>B. subtilis</i> , <i>B. megaterium</i>	<i>B. sphaericus</i>	• 271 • 86 • 136	<i>B. sphaericus</i>	• 806 • 238 • 4
Cluster 8 (28)	763 • 334 • 4	<i>B. subtilis</i>	• 271 • 89 • 134 • 181 • 4	<i>B. subtilis</i>	• 119 • 695 • 187 • 51 • 4
Closest Match	<i>B. licheniformis</i> , <i>B. pumilus</i> , <i>B. megaterium</i> , <i>B. subtilis</i>	<i>B. pumilus</i>	331 • 86 • 134 • 4	<i>B. pumilus</i>	• 74 • 24 • 83 • 317 • 589 • 12 • 4
Cluster 9 (25)	764 • 335 • 4				
Closest Match	<i>B. licheniformis</i> , <i>B. pumilus</i> , <i>B. megaterium</i>				
Cluster 10 (25)	1099				
Closest Match	<i>B. sphaericus</i>				

RsaI		AhaI		Tru9I	
Cluster 3 (18)	• 19 • 11 • 404 • 357 • 146	Cluster 3 (18)	15 • 52 • 33 • 185 • 637	Cluster 3 (18)	268 • 8 • 86 • 136
Closest Match	<i>B. megaterium</i>	Cluster 4 (22)	• 11 • 615 • 209 • 265 • 4	Closest Match	<i>B. halodurans</i> / <i>B. clausii</i>
Cluster 4 (22)	• 404 • 357 • 146	Cluster 5 (32)	36 • 32 • 42 • 615 • 209 • 265 • 123	Cluster 4, 5, 7, 8, 9, 10	• 278 • 86 • 136
Closest Match	<i>B. anthracis</i> , <i>B. cereus</i> , <i>B. thuringiensis</i>	Cluster 6 (7)	36 • 32 • 50 • 124 • 4	Closest Match	<i>B. licheniformis</i> , <i>B. sphaericus</i> , <i>B. megaterium</i>
Cluster 5 (32)	• 18 • 11 • 404 • 303	Cluster 7 (37)	143 • 33 • 184 • 438 • 209 • 265 • 4		
Closest Match	<i>B. sphaericus</i> , <i>B. licheniformis</i>	Cluster 8 (28)	35 • 32 • 615 • 209 • 4		
• 357 • 146		Cluster 9 (25)	83 • 32 • 31 • 822 • 4		
Closest Match	unique	Cluster 10 (25)	73 • 614 • 4		
Cluster 7 (37)	• 403 • 357 • 146	Cluster 11 (25)	173 • 188 • 438 • 209 • 4		
Closest Match	<i>B. anthracis</i> , <i>B. cereus</i> , <i>B. thuringiensis</i>				
Cluster 8 (28)	• 18 • 11 • 404 • 357 • 146				
Closest Match	<i>B. megaterium</i>				
Cluster 9 (25)	• 18 • 11 • 404 • 356 • 146				
Closest Match	<i>B. megaterium</i> , <i>B. clausii</i>				
Cluster 10 (25)	• 404 • 357 • 146				
Closest Match	unique				
Cluster 11 (25)	• 11 • 404 • 4				
Closest Match	unique				

Cluster 2

RsaI		AhaI		HaeIII	
<i>Brevibacillus</i> (1)	18 • 11 • 404 • 357 • 146	<i>Brevibacillus</i> (1)	158 • 33 • 403 • 212 • 161 • 46	<i>Brevibacillus</i> (1)	34 • 64 • 22 • 85 • 313 • 437
<i>Paenibacillus</i> (16)	• 404 • 341 • 15 • 147	<i>Paenibacillus</i> (4)	85 • 89 • 186 • 637	<i>Paenibacillus</i> (4)	70 • 22 • 591 • 908
Cluster 2 (12)	18 • 11 • 404 • 351 • 146	Cluster 2 (12)	85 • 90 • 186 • 228 • 373 • 47	Cluster 2 (12)	70 • 22 • 85 • 4
Closest Match	<i>Brevibacillus</i> , <i>B. megaterium</i>	Closest Match	Unique	Closest Match	Unique

Cluster 6

RsaI		HaeIII		DpaII	
<i>Fornbacillus</i> (1)	18 • 11 • 170 • 231 • 355	<i>Fornbacillus</i> (1)	11 • 70 • 22 • 209 • 240 • 34 • 335	<i>Fornbacillus</i> (1)	24 • 47 • 172 • 119
<i>Oreanobacillus</i> (2)	406 • 501	<i>Oreanobacillus</i> (2)	70 • 22 • 565 • 34 • 457	<i>Oreanobacillus</i> (2)	246 • 119 • 695 • 238
<i>Pirgibacillus</i> (13)	• 171 • 235 • 405	<i>Pirgibacillus</i> (13)	70 • 22 • 565 • 34 • 457	<i>Pirgibacillus</i> (13)	• 246 • 119 • 695 • 238 • 4
<i>Pirgibacillus</i> Ref (25)	• 171 • 235 • 501 • 146	<i>Pirgibacillus</i> Ref (25)	• 70 • 22 • 566 • 34 • 457	<i>Pirgibacillus</i> Ref (25)	• 243 • 119 • 695 • 238 • 4
Cluster 6 (7)	18 • 11 • 406 • 354 • 146	Cluster 6 (7)	34 • 44 • 22 • 85 • 478 • 34 • 457	Cluster 6 (7)	1050 • 119
Closest Match	<i>B. megaterium</i> , <i>B. clausii</i>	Closest Match	unique	Closest Match	unique

Cluster 11

RsaI		HaeIII		DpaII	
<i>Anoxybacillus</i> (4)	• 406 • 359 • 146	<i>Anoxybacillus</i> (4)	78 • 22 • 85 • 513 • 461 • 4	<i>Anoxybacillus</i> (4)	278 • 86 • 138
<i>Geobacillus</i> (46)	• 406 • 359 • 146	<i>Geobacillus</i> (46)	• 25 • 61 • 17 • 22 • 85 • 326 • 154 • 34 • 461	<i>Geobacillus</i> (46)	• 364 • 137 • 4
<i>Geobacillus</i> Ref (25)	• 406 • 359 • 146	<i>Geobacillus</i> Ref (25)	• 25 • 61 • 17 • 22 • 85 • 326 • 153 • 32 • 455 • 4	low freq	• 278 • 86 • 4
				<i>Geobacillus</i> Ref (25)	360 • 133 • 4

RfaI		Tru9I		AhaI	
<i>Anoxybacillus</i> (4)	• 711 • 344 • 4	<i>Anoxybacillus</i> (4)	1100 • 4	<i>Anoxybacillus</i> (4)	135 • 33 • 186 • 429 • 211
<i>Geobacillus</i> (46)	• 1296 • 175	<i>Geobacillus</i> (46)	• 878 • 27 • 195 • 4	<i>Geobacillus</i> (46)	• 86 • 88 • 186 • 429 • 211 • 350
low freq	417 • 937	<i>Geobacillus</i> Ref (25)	• 870 • 24 • 195 • 4	<i>Geobacillus</i> Ref (25)	86 • 89 • 186 • 429 • 203 • 350
<i>Geobacillus</i> Ref (25)	• 1296 • 175				



**Figure 9. Nucleotide fragments generated as a result of *in silico* restriction enzyme (a). *RsaI* (b). *HaellI* (c). *AluI* (d). *BfaI* (e). *Tru9I* (f). *DpnII*, action on 16S rDNA gene sequences in 10 *Bacillus* spp. and 11 clusters (Figure 8). A solid black box represents the restriction enzyme cut site. At the terminal points, solid black box are shown only if highly variable fragments are observed beyond this cut site.**  
doi:10.1371/journal.pone.0004438.g009

<i>Bacillus</i> spp.	<i>BfaI</i>	<i>Tru9I</i>	<i>HaellI</i>	<i>RsaI</i>
<i>B. clausii</i>				
<i>B. licheniformis</i>				
<i>B. megaterium</i>				
<i>B. subtilis</i>				
<i>B. pumilus</i>				
<i>B. halodurans</i>				

**Figure 10. *Bacillus* species showing distinct patterns with different restriction enzymes.**  
doi:10.1371/journal.pone.0004438.g010

{including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s016 (3.62 MB TIF)

**Figure S12** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s017 (0.74 MB TIF)

**Figure S13** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s018 (0.37 MB TIF)

**Figure S14** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136

**Table 4. Closest matches of Clusters of *Bacillus* sp. with known *Bacillus* sp. on the basis of restriction enzyme pattern analysis.**

<i>Bacillus</i> sp.	Restriction enzyme						No. of Best matches		
	<i>BfaI</i>	<i>DpnII</i>	<i>RsaI</i>	<i>Tru9I</i>	<i>AluI</i>	<i>HaellI</i>	<i>Bacillus</i> spp.	Unique <sup>a</sup>	
<b>Cluster 1 (22)<sup>b</sup></b>	<i>B. sphaericus</i>	unique	<i>B. licheniformis</i> <i>B. subtilis</i>	<i>B. licheniformis</i>	unique	<i>B. sphaericus</i> <i>B. halodurans</i>	4	2	
<b>Cluster 2 (12)</b>	<i>B. halodurans</i>	<i>B. clausii</i>	<i>B. megaterium</i>	unique	unique	Unique	3	3	
<b>Cluster 3 (18)</b>	<i>B. halodurans</i>	unique	<i>B. megaterium</i>	<i>B. halodurans</i> <i>B. clausii</i>	unique	Unique	3	3	
<b>Cluster 4 (22)</b>	<i>B. megaterium</i>	unique	<i>B. anthracis</i> <i>B. cereus</i> <i>B. thuringiensis</i>	<i>B. licheniformis</i> <i>B. megaterium</i> <i>B. licheniformis</i>	unique	<i>B. sphaericus</i> <i>B. halodurans</i> <i>B. megaterium</i>	7	2	
<b>Cluster 5 (32)</b>	<i>B. megaterium</i>	unique	<i>B. sphaericus</i> <i>B. licheniformis</i>	<i>B. licheniformis</i> <i>B. megaterium</i> <i>B. sphaericus</i>	<i>B. megaterium</i> unique	Unique	3	3	
<b>Cluster 6 (7)</b>	Unique	unique	<i>B. megaterium</i> <i>B. clausii</i>	unique	unique	Unique	2	5	
<b>Cluster 7 (37)</b>	<i>B. subtilis</i> <i>B. megaterium</i>	unique	<i>B. anthracis</i> <i>B. cereus</i> <i>B. thuringiensis</i>	<i>B. licheniformis</i> <i>B. megaterium</i> <i>B. sphaericus</i>	unique	<i>B. licheniformis</i> <i>B. subtilis</i> <i>B. pumilus</i>	8	2	
<b>Cluster 8 (28)</b>	<i>B. pumilus</i> <i>B. megaterium</i> <i>B. licheniformis</i> <i>B. subtilis</i>	<i>B. sphaericus</i>	<i>B. megaterium</i>	<i>B. licheniformis</i> <i>B. megaterium</i> <i>B. sphaericus</i>	unique	Unique	5	2	
<b>Cluster 9 (25)</b>	<i>B. licheniformis</i> <i>B. pumilus</i> <i>B. megaterium</i>	<i>B. sphaericus</i>	<i>B. megaterium</i> <i>B. clausii</i>	<i>B. licheniformis</i> <i>B. megaterium</i> <i>B. sphaericus</i>	unique	unique	5	2	
<b>Cluster 10 (25)</b>	<i>B. sphaericus</i>	unique	unique	<i>B. licheniformis</i> <i>B. sphaericus</i> <i>B. megaterium</i>	unique	unique	3	4	
<b>Cluster 11</b>	All were designated <sup>c</sup>								

<sup>a</sup>Did not match with any of the known RE patterns of the *Bacillus* sp. used in this study.

<sup>b</sup>Number of 16S rDNA sequences of *Bacillus* sp.

<sup>c</sup>Cluster 11 constituted of members of known species such as: *Anoxybacillus* and *Geobacillus*.

doi:10.1371/journal.pone.0004438.t004

**Table 5.** Accession numbers of 16S rDNA sequences of *Bacillus* species used for generating phylogenetic framework (<http://rdp.cme.msu.edu/>).

Organism	Reference sequence(s)
<i>Bacillus thuringiensis</i>	DQ286308(T) <sup>a</sup> , DQ286339, DQ328630, AE017355, DQ286329
<i>B. anthracis</i> <sup>b</sup>	AB190218, AE017334, AE017225
<i>B. cereus</i> <sup>b</sup>	DQ372919, DQ289988
<i>B. subtilis</i>	AB042061(T), DQ420172, AY995569, DQ504376, AY583216, AY881635, AY631853
<i>B. licheniformis</i>	AB039328(T), CP000002, AF234855
<i>B. pumilus</i>	AY260861(T), AY876289, DQ523500
<i>B. megaterium</i>	AJ717381(T), AY373358, AY505510, AY373360
<i>B. sphaericus</i>	AJ310084(T), DQ286309
<i>B. clausii</i>	X76440(T), AB201793, AY960116
<i>B. halodurans</i>	AY423275(T), AY856452
<b>Total</b>	<b>34 strains</b>

<sup>a</sup>Type strain.

<sup>b</sup>For *B. cereus* group only one type strain was used.  
doi:10.1371/journal.pone.0004438.t005

{including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s019 (0.26 MB TIF)

**Figure S15** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s020 (0.28 MB TIF)

**Figure S16** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s021 (0.29 MB TIF)

**Figure S17** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s022 (0.22 MB TIF)

**Figure S18** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with

Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s023 (0.25 MB TIF)

**Figure S19** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s024 (0.67 MB TIF)

**Figure S20** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s025 (0.26 MB TIF)

**Figure S21** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).  
Found at: doi:10.1371/journal.pone.0004438.s026 (0.25 MB TIF)

**Figure S22** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s027 (0.24 MB TIF)

**Figure S23** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s028 (0.24 MB TIF)

**Figure S24** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s029 (0.25 MB TIF)

**Figure S25** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s030 (0.25 MB TIF)

**Figure S26** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s031 (0.26 MB TIF)

**Figure S27** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s032 (0.21 MB TIF)

**Figure S28** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add

authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s033 (0.21 MB TIF)

**Figure S29** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s034 (0.63 MB TIF)

**Figure S30** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s035 (0.26 MB TIF)

**Figure S31** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s036 (0.22 MB TIF)

**Figure S32** Phylogenetic tree of 34 framework sequences (bold values) and *Bacillus* sp. at the rate of 52 sequences (1–1136 {including 1025 *Bacillus* sp. and rest other species to add authenticity to the results}) which could be designated as 10 known *Bacillus* spp. used in the study. A neighbor-joining analysis with Jukes-Cantor correction and bootstrap support was performed on the gene sequences. Bootstrap values are given at nodes. Values in parentheses are accession numbers (<http://rdp.cme.msu.edu/>).

Found at: doi:10.1371/journal.pone.0004438.s037 (0.73 MB TIF)

## Author Contributions

Conceived and designed the experiments: VCK. Performed the experiments: SP SL SC. Analyzed the data: SP SL SC VCK. Wrote the paper: VCK.

## References

- Baldauf SL (2003) Phylogeny for the faint of heart: a tutorial. *Trends Genet* 19: 345–351.
- Goldenfeld N, Woese C (2007) Biology's next revolution. *Nature* 445: 369.
- Lee H, Hee-Jung P, Sang-Nae C, Gill-Han B, Sang-Jae K (2000) Species identification of *Mycobacteria* by PCR-restriction fragment length polymorphism of the *rpoB* gene. *J Clin Microbiol* 38: 2966–2971.
- Lal S, Cheema S, Kalia VC (2008) Phylogeny vs genome reshuffling: Horizontal gene transfer. *Ind J Microbiol* 48: 228–242.
- Kim BJ, Lee SH, Lyu MA, Kim SJ, Bai GH, et al. (1999) Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). *J Clin Microbiol* 37: 1714–1720.
- Yamamoto S, Harayama S (1998) Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int J Syst Bacteriol* 48: 813–819.
- Kasai H, Ezaki T, Harayama S (2000) Differentiation of phylogenetically related slowly growing mycobacteria by their *gyrB* sequences. *J Clin Microbiol* 38: 301–308.
- Yamamoto S, Bouvet PJ, Harayama S (1999) Phylogenetic structures of the genus *Acinetobacter* based on *gyrB* sequences: comparison with the grouping by DNA-DNA hybridization. *Int J Syst Bacteriol* 49: 87–95.
- Venkateswaran K, Moser DP, Dollhopf ME, Lies DP, Saffarini DA, et al. (1999) Polyphasic taxonomy of the genus *Shevanelia* and description of *Shevanelia oneidensis* sp. nov. *Int J Syst Bacteriol* 49: 705–724.



10. Chun J, Bae KS (2000) Phylogenetic analysis of *Bacillus subtilis* and related taxa based on partial *gyrA* gene sequences. *Antonie van Leeuwenhoek* 78: 123–127.
11. Claus D, Berkeley RCW (1986) Genus *Bacillus*. In: Sneath PHA, Mair NS, Sharpe ME, Holt JG, eds. *Bergey's manual of systematic bacteriology*, vol. 2. Baltimore: The Williams and Wilkins. pp 1105–1140.
12. Priest FG (1993) In: Sonenshein AL, Hoch JA, Losick R eds. *Bacillus subtilis* and other gram-positive bacteria. Washington, D.C.: American Society for Microbiology, Systematics and Ecology of *Bacillus*. pp 3–16.
13. Hofte H, Whiteley HR (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol Rev* 53: 242–255.
14. Kalia VC, Jain SR, Kumar A, Joshi AP (1994) Fermentation of biowaste to H<sub>2</sub> by *Bacillus licheniformis*. *World J Microbiol Biotechnol* 10: 224–227.
15. Porwal S, Kumar T, Lal S, Rani A, Kumar S, et al. (2008) Hydrogen and polyhydroxybutyrate producing abilities of microbes from diverse habitats by dark fermentative process. *Bioresour Technol* 99: 5444–5451.
16. Sonakya V, Raizada N, Kalia VC (2001) Microbial and enzymatic improvement of anaerobic digestion of waste biomass. *Biotechnol Lett* 23: 1463–1466.
17. Drobniowski FA (1993) *Bacillus cereus* and related species. *Clin Microbiol Rev* 6: 324–338.
18. Jackson SG, Goodbrand RB, Ahmed R, Kasatiya S (1995) *Bacillus cereus* and *Bacillus thuringiensis* isolated in a gastroenteritis outbreak investigation. *Lett Appl Microbiol* 21: 103–105.
19. Pérez C, Muñoz-Garay C, Portugal LC, Sánchez J, Gill SS, et al. (2007) *Bacillus thuringiensis* ssp. *israelensis* CytI<sub>Aa</sub> enhances activity of CryII<sub>Aa</sub> toxin by facilitating the formation of a pre-pore oligomeric structure. *Cell Microbiol* 9: 2931–2937.
20. Al-Ajlani MM, Sheikh MA, Ahmad Z, Hasnain S (2007) *Bacillus subtilis* strains produce a broad spectrum of bioactive peptides with great potential for biotechnological and biopharmaceutical applications. *Microb Cell Fact* 6: 17.
21. Ishihara H, Sasaki T, Shimura K (1968) Biosynthesis of bacitracin. II. Incorporation of <sup>14</sup>C-labeled amino acids into bacitracin by a cell-free preparation from *Bacillus licheniformis*. *Biochem Biophys* 166: 496–504.
22. Ishihara H, Takoh M, Nishibayashi R, Sato A (2002) Distribution and variation of bacitracin synthetase gene sequences in laboratory stock strains of *Bacillus licheniformis*. *Curr Microbiol* 45: 18–23.
23. Martirani L, Varcamonti M, Naclerio G, De Felice M (2002) Purification and partial characterization of bacillicin 490, a novel bacteriocin produced by a thermophilic strain of *Bacillus licheniformis*. *Microbiol Cell Fact* 1: 1.
24. De Clerck E, De Vos P (2002) Study of the bacterial load in a gelatine production process, focussed on *Bacillus* and related endospore forming genera. *Syst Appl Microbiol* 25: 611–618.
25. Heyndrickx M, Scheldeman P (2002) *Bacilli* associated with spoilage in dairy products and other food. In: Berkeley R, Heyndrickx M, Logan NA, De Vos P, eds. *Applications and Systematics of Bacillus and Relatives*. Oxford: Blackwell.
26. Suihko ML, Stackebrandt E (2003) Identification of aerobic mesophilic *Bacilli* isolated from board and paper products containing recycled fibres. *J Appl Microbiol* 94: 25–34.
27. Brown KL (2000) Control of bacterial spores. *Br Med Bull* 56: 158–171.
28. Sorokulova IB, Reva ON, Smirnov VV, Pinchuk IV, Lapa SV, et al. (2003) Genetic diversity and involvement in bread spoilage of *Bacillus* strains from flour and rony bread. *Lett Appl Microbiol* 37: 169–173.
29. Green DH, Wakeley PR, Page A, Barnes A, Baccigalupi L, et al. (1999) Characterization of two *Bacillus* probiotics. *Appl Environ Microbiol* 65: 4288–4291.
30. Hoa NT, Baccigalupi L, Huxham A, Smertenko A, Van PH, et al. (2000) Characterization of *Bacillus* species used for oral bacteriotherapy and bacteriophagephylaxis of gastrointestinal disorders. *Appl Environ Microbiol* 66: 5241–5247.
31. Barbosa TM, Serra CR, La Ragione RM, Woodward MJ, Henriques AO (2005) Screening for *Bacillus* isolates in the broiler gastrointestinal tract. *Appl Environ Microbiol* 71: 968–978.
32. Senesi S, Celandroni F, Tavanti A, Ghelardi E (2001) Molecular characterization and identification of *Bacillus clausii* strains marketed for use in oral bacteriotherapy. *Appl Environ Microbiol* 67: 834–839.
33. Fox GE, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, et al. (1980) The phylogeny of prokaryotes. *Science* 209: 457–463.
34. Rasko DA, Altherr MR, Han CS, Ravel J (2005) Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol Rev* 29: 303–329.
35. Harrell LJ, Andersen GL, Wilson KH (1995) Genetic variability of *Bacillus anthracis* and related species. *J Clin Microbiol* 33: 1847–1850.
36. Priest FG (1981) DNA homology in the genus *Bacillus*. In: Berkeley RCW, Goodfellow M, eds. *The Aerobic Endospore-forming Bacteria*. London: Academic Press. pp 35–57.
37. Da Silva KRA, Rabinovitch L, Seldin L (1999) Phenotypic and genetic diversity among *Bacillus sphaericus* strains isolated in Brazil, potentially useful as biological control agents against mosquito larvae. *Res Microbiol* 150: 153–160.
38. Herman L, Heyndrickx M (2000) The presence of intragenically located REP-like elements in *Bacillus sporothermodurans* is sufficient for REP-PCR typing. *Res Microbiol* 151: 255–261.
39. Palmisano MM, Nakamura LK, Duncan KE, Istock CA, Cohan FM (2001) *Bacillus sonoriensis* sp. nov., a close relative of *Bacillus licheniformis*, isolated from soil in the Sonoran desert, Arizona. *Int J Syst Evol Microbiol* 51: 1671–1679.
40. De Clerck E, De Vos P (2004) Genotypic diversity among *Bacillus licheniformis* strains from various sources. *FEMS Microbiol Lett* 231: 91–98.
41. Stackebrandt E, Swiderski J (2002) From phylogeny to systematics: the dissection of the genus *Bacillus*. In: Berkeley R, Heyndrickx M, Logan NA, De Vos P, eds. *Applications and Systematics of Bacillus and Relatives*. Oxford: Blackwell.
42. Ash C, Farrow JAE, Wallbanks S, Collins MD (1991) Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-rRNA sequences. *Lett Appl Microbiol* 13: 202–206.
43. Ash C, Collins MD (1992) Comparative analysis of 23S rRNA gene sequences of *Bacillus anthracis* and emetic *Bacillus cereus* determined by PCR-direct sequencing. *FEMS Microbiol Lett* 94: 75–80.
44. Bavykin SG, Lysov YP, Zakhariyev V, Kelly JJ, Jackman J, et al. (2004) Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *J Clin Microbiol* 42: 3711–3730.
45. Gurtler V (1999) The role of recombination and mutation in 16S–23S rDNA spacer rearrangement. *Gene* 238: 241–252.
46. Gurtler V, Stanisich VA (1996) New approaches to typing and identification of bacteria using the 16S–23S rDNA spacer region. *Microbiology* 142: 3–16.
47. Ash C, Farrow JAE, Dorsch M, Stackebrandt E, Collins MD (1991) Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S ribosomal RNA. *Int J Syst Bacteriol* 41: 343–346.
48. Daffonchio D, Cherif A, Borin S (2000) Homoduplex and heteroduplex polymorphisms of the amplified ribosomal 16S–23S internal transcribed spacers describe genetic relationships in the “*Bacillus cereus* group”. *Appl Environ Microbiol* 66: 5460–5468.
49. Bouvet A, Grimont F, Grimont PAD (1991) Intraspecies variations in nutritionally variant streptococci: rRNA gene restriction patterns of *Streptococcus defectivus* and *Streptococcus adjacens*. *Int J Syst Bacteriol* 41: 483–486.
50. Blanc DS, Seigrist HH, Sahli R, Francioli P (1993) Ribotyping of *Pseudomonas aeruginosa*: discriminatory power and usefulness as a tool for epidemiological studies. *J Clin Microbiol* 31: 71–77.
51. Fukunaga M, Sohnaka M, Yanagihara Y (1993) Analysis of *Borrelia* species associated with Lyme disease by rRNA gene restriction fragment length polymorphism. *J Gen Microbiol* 139: 1141–1146.
52. Ning W, Mackie RI, Gaskins RH (1997) Biotype and ribotype diversity among *Lactobacillus* isolates from the mouse ileum. *Syst Appl Microbiol* 20: 423–431.
53. Jalava K, Hiemi S, Hirvi U, Hanninen ML (1999) Evaluation of a molecular identification scheme based on 23S rRNA gene polymorphisms for differentiating canine and feline gastric *Helicobacter* spp. *Lett Appl Microbiol* 28: 269–274.
54. Joung KB, Cote JC (2002) A single phylogenetic analysis of *Bacillus thuringiensis* strains and *bacilli* species inferred from 16S rRNA gene restriction fragment length polymorphism is congruent with two independent phylogenetic analyses. *J Appl Microbiol* 93: 1075–1082.
55. Vandamme P, Pot B, Gillis M, De Vos P, Kersters K, et al. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 60: 407–438.
56. Vold BS (1985) Structure and organization of genes for transfer ribonucleic acid in *Bacillus subtilis*. *Microbiol Rev* 49: 71–80.
57. Maes N, De Gheldre Y, De Ryck R, Vaneechoutte M, Meugnier H, et al. (1997) Rapid and accurate identification of *Staphylococcus* species by tRNA intergenic spacer length polymorphism analysis. *J Clin Microbiol* 35: 2477–2481.
58. De Gheldre Y, Vandamme P, Goossens H, Strulens MJ (1999) Identification of clinically relevant viridans streptococci by analysis of transfer DNA intergenic spacer length polymorphism. *Int J Syst Bacteriol* 49: 1591–1598.
59. Borin S, Daffonchio D, Sorlini S (1997) Single strand conformation polymorphism analysis of PCR-tDNA fingerprinting to address the identification of *Bacillus* species. *FEMS Microbiol Lett* 157: 87–93.
60. Daffonchio D, Borin S, Frova G, Manachini PL, Sorlini C (1998) PCR fingerprinting of whole genomes: the spacer between the 16S and 23S rRNA genes and intergenic tRNA gene regions reveal a different intraspecific genomic variability of *Bacillus cereus* and *Bacillus licheniformis*. *Int J Syst Bacteriol* 48: 107–116.
61. Ehrenstein B, Bernards AT, Dijkshoorn L, Gerner-Smidt P, Towner KJ, et al. (1996) *Acinetobacter* species identification by using tRNA spacer fingerprinting. *J Clin Microbiol* 34: 2414–2420.
62. Guinebretiere MH, Berge O, Normand P, Morris C, Carlin F, et al. (2001) Identification of bacteria in pasteurized zucchini purees stored at different temperatures and comparison with those found in other pasteurized vegetable purees. *Appl Environ Microbiol* 67: 4520–4530.
63. Rowan NJ, Deans K, Anderson JG, Gemmill CG, Hunter IS, et al. (2001) Putative virulence factor expression by clinical and food isolates of *Bacillus* spp. after growth in reconstituted infant milk formulae. *Appl Environ Microbiol* 67: 3873–3881.
64. Zeigler DR (2001) *Bacillus* genetic stock center catalog of strains, 7th edition. Volume 3: The Genus *Geobacillus*. Ohio: The Ohio State University.
65. Nazina TN, Tourova TP, Poltarau AB, Novikova EV, Grigoryan AA, et al. (2001) Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp nov and *Geobacillus uzonensis* sp nov from petroleum reservoirs and transfer of *Bacillus stearothersophilus*, *Bacillus thermo-*

- catenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *Int J Syst Evol Microbiol* 51: 433–446.
66. Yoon JH, Lee KC, Weiss N, Kho YH, Kang KH, et al. (2001) *Sporosarcina aquimarina* sp. nov., a bacterium isolated from seawater in Korea, and transfer of *Bacillus globisporus* (Larkin and Stokes 1967), *Bacillus psychrophilus* (Nakamura 1984) and *Bacillus pasteurii* (Chester 1898) to the genus *Sporosarcina* as *Sporosarcina globispora* comb. nov., *Sporosarcina psychrophila* comb. nov. and *Sporosarcina pasteurii* comb. nov. and emended description of the genus *Sporosarcina*. *Int J Syst Evol Microbiol* 51: 1079–1086.
  67. Yoon JH, Weiss N, Lee KC, Lee IS, Kang KH, et al. (2001) *Jeotgalibacillus alimentarius* gen. nov., sp. nov., a novel bacterium isolated from jeotgal with *L-lysine* in the cell wall, and reclassification of *Bacillus marinus* Rueger 1983 as *Marinibacillus marinus* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 51: 2087–2093.
  68. Vilas-Bóas GT, Peruca APS, Arantes OMN (2007) Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis*. *Canadian J Microbiol* 53: 673–687.
  69. Ticknor LO, Kolsto AB, Hill KK, Keim P, Laker MT, et al. (2001) Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. *Appl Environ Microbiol* 67: 4863–4873.
  70. Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, et al. (2004) Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc Natl Acad Sci USA* 101: 8449–8454.
  71. Carlson CR, Caugant DA, Kolsto AB (1994) Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. *Appl Environ Microbiol* 60: 1719–1725.
  72. Helgason E, Caugant DA, Olsen I, Kolsto AB (2000) Genetic structure of population of *Bacillus cereus* and *B. thuringiensis* isolates associated with periodontitis and other human infections. *J Clin Microbiol* 38: 1615–1622.
  73. Chen ML, Tsen HY (2002) Discrimination of *Bacillus cereus* and *Bacillus thuringiensis* with 16S rRNA and *gyrB* gene based PCR primers and sequencing of their annealing sites. *J Appl Microbiol* 92: 912–919.
  74. Helgason E, Caugant DA, Lecadet MM, Chen YH, Mahillon J, et al. (1998) Genetic diversity of *Bacillus cereus* and *B. thuringiensis* isolates from natural sources. *Curr Microbiol* 37: 80–87.
  75. Hill KK, Ticknor LO, Okinaka RT, Asay M, Blair H, et al. (2004) Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. *Appl Environ Microbiol* 70: 1068–1080.
  76. Helgason E, Tourasse NJ, Meisler R, Caugant DA, Kolsto AB (2004) Multilocus sequencing scheme for bacteria of the *Bacillus cereus* group. *Appl Environ Microbiol* 70: 191–201.
  77. Nakamura LK, Roberts MS, Cohan FM (1999) Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: a proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov. and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. *Int J Syst Bacteriol* 49: 1211–1215.
  78. Duncan KE, Ferguson N, Kimura K, Zhou X, Istock CA (1994) Fine-scale genetic and phenotypic structure in natural populations of *Bacillus subtilis* and *Bacillus licheniformis*: implications for bacterial evolution and speciation. *Evolution* 48: 2002–2025.
  79. Manachini PL, Fortina MG, Levati L, Parini C (1998) Contribution to phenotypic and genotypic characterization of *Bacillus licheniformis* and description of new genomovars. *Syst Appl Microbiol* 21: 520–529.
  80. Rey MW, Ramaiya P, Nelson BA, Brody-Karpin SD, Zaretsky EJ, et al. (2004) Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biol* 5: R77.
  81. Lapidus A, Galleron N, Andersen JT, Jørgensen PL, Ehrlich SD, et al. (2002) Co-linear scaffold of the *Bacillus licheniformis* and *Bacillus subtilis* genomes and its use to compare their competence genes. *FEMS Microbiol Lett* 209: 23–30.
  82. Xu D, Cote JC (2003) Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S–23S ITS nucleotide sequences. *Int J Syst Evol Microbiol* 53: 695–704.
  83. Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, et al. (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res* 28: 4317–4331.
  84. Takami H, Horikoshi K (1999) Reidentification of facultatively alkaliphilic *Bacillus* sp. C-125 to *Bacillus halodurans*. *Biosci Biotechnol Biochem* 63: 943–945.
  85. Tena D, Martinez-Torres JA, Perez-Pomata MT, Saez-Nieto JA, Rubio V, et al. (2007) Cutaneous infection due to *Bacillus pumilus*: Report of 3 Cases. *Clin Infectious Diseases* 44: 40–42.
  86. From C, Pukall R, Schumann P, Hormazabal V, Granum PE (2005) Toxin producing ability among *Bacillus* spp outside the *Bacillus cereus* group. *Appl Environ Microbiol* 71: 1178–1183.
  87. Vary PS, Biedendieck R, Fuerch T, Meinhardt F, Rohde M, et al. (2007) *Bacillus megaterium*—from simple soil bacterium to industrial protein production host. *Appl Microbiol Biotechnol* 76: 957–967.
  88. Han B, Liu H, Hu X, Cai Y, Zheng D, et al. (2007) Molecular characterization of a glucokinase with broad hexose specificity from *Bacillus sphaericus* strain C3-41. *Appl Environ Microbiol* 73: 3581–3586.
  89. Davidson EW (1995) Biochemistry and mode of action of the *Bacillus sphaericus* toxins. *Mem Inst Oswaldo Cruz Rio J* 90: 81–86.
  90. Schenkel RGM, Dias SC, Lopes JB (1993) Serological characterization of 23 isolates of *Bacillus sphaericus* pathogenic to mosquito larvae screened from Brazilian soil. *Rev Microbiol* 24: 166–167.
  91. Krych VK, Johnson JL, Yousten AA (1980) Deoxyribonucleic acid homologies among strains of *Bacillus sphaericus*. *Int J Syst Bacteriol* 30: 476–484.
  92. Rippert KE, Johnson JL, Yousten AA (1997) DNA similarities among mosquito-pathogenic and non pathogenic strains of *Bacillus sphaericus*. *Int J Syst Bacteriol* 47: 214–216.
  93. Yousten AA, DeBarjac H, Hendrich J, Dumanoir VC, Myers P (1980) Comparison between bacteriophage typing and serotyping for the differentiation of *Bacillus sphaericus* strains. *Ann Microbiol* 131B: 297–308.
  94. DeBarjac H, Veron M, Dumanoir VC (1980) Caractérisation biochimie et serologique de souches de *Bacillus sphaericus* pathogènes ou non pour les moustiques. *Ann Microbiol* 131B: 191–201.
  95. DeBarjac H, Thiery I, Dumanoir VC, Ripouteau H (1985) Serological classification of *Bacillus sphaericus* strains on the basis of toxicity to mosquito larvae. *Appl Microbiol Biotechnol* 21: 85–90.
  96. Frachon E, Hamon S, Nicholas L, DeBarjac H (1991) Cellular fatty acid analysis as a potential tool for predicting mosquitocidal activity of *Bacillus sphaericus*. *Appl Environ Microbiol* 57: 3394–3398.
  97. Zahner V, Rabinovitch L, Cavados CFG, Momen H (1994) Multilocus enzyme electrophoresis on agarose gels as an aid to the identification of entomopathogenic *Bacillus sphaericus* strains. *J Appl Bacteriol* 76: 327–335.
  98. Alexander B, Priest FG (1990) Numerical classification and identification of *Bacillus sphaericus* including some strains pathogenic for mosquito larvae. *J Gen Microbiol* 136: 367–376.
  99. Guerinéau M, Alexander B, Priest FG (1991) Isolation and identification of *Bacillus sphaericus* strains pathogenic for mosquito larvae. *J Invert Pathol* 57: 325–333.
  100. Olive DM, Bean P (1999) Principles and applications of methods for DNA-based typing of microbial organisms. *J Clin Microbiol* 37: 1661–1669.
  101. Wattiau P, Renard ME, Ledent P, Debois V, Blackman G, et al. (2001) A PCR test to identify *Bacillus subtilis* and closely related species and its application to the monitoring of wastewater biotreatment. *Appl Microbiol Biotechnol* 56: 816–819.
  102. Hansen BM, Leser TD, Hendriksen NB (2001) Polymerase chain reaction assay for the detection of *Bacillus cereus* group cells. *FEMS Microbiol Lett* 202: 209–213.
  103. Djordjevic SP, Smith WA, Hornitzky MA (2000) Genetic and biochemical diversity among isolates of *Paenibacillus alvei* cultured from Australian honeybee (*Apis mellifera*) colonies. *Appl Environ Microbiol* 66: 1098–1106.
  104. Wu XY, Walker MJ, Hornitzky M, Chin J (2006) Development of a group-specific PCR combined with ARDRA for the identification of *Bacillus* species of environmental significance. *J Microbiol Methods* 64: 107–119.
  105. Arcuri EF, Wiedmann M, Boor KJ (2000) Phylogeny and functional conservation of  $\sigma E$  in endospore-forming bacteria. *Microbiology* 146: 1593–1603.
  106. Bobbie RJ, White DC (1980) Characterization of benthic microbial community structure by high resolution gas chromatography of fatty acid methyl esters. *Appl Environ Microbiol* 39: 1212–1222.
  107. Vaerewijck MJM, De Vos P, Lebbe L, Scheldeman P, Hoste B, et al. (2001) Occurrence of *Bacillus sporothermodurans* and other aerobic spore-forming species in feed concentrate for dairy cattle. *J Appl Microbiol* 91: 1074–1084.
  108. Logan NA, Berkeley RC (1984) Identification of *Bacillus* strains using the API system. *J Gen Microbiol* 130: 1871–1882.
  109. Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51: 221–271.
  110. Wang X, Heazlewood SP, Krause DO, Florin THJ (2003) Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J Appl Microbiol* 95: 508–520.
  111. Yamada Y, Hoshino K, Ishikawa T (1997) The phylogeny of acetic acid bacteria based on the partial sequences of 16S ribosomal RNA: the elevation of the subgenus *gluconacetobacter* to the generic level. *Biosci Biotechnol Biochem* 61: 1244–1251.
  112. Bergey DH, Breed RS, Murray EGD, Hitchens AP, eds (1939) *Bergey's Manual of Determinative Bacteriology*. 5th ed. Baltimore: Williams and Wilkins.
  113. Buchanan RE, Gibbons NE, eds (1974) *Bergey's Manual of Determinative Bacteriology*. 8th ed. Baltimore: Williams and Wilkins.
  114. Skerman VBD, McGowan V, Sneath PHA, eds (1980) *Approved lists of bacterial names*. Washington, DC: American Society for Microbiology.
  115. Fritze D (2004) *Taxonomy of the Genus Bacillus and Related Genera: The Aerobic Endospore-Forming Bacteria. The Nature and Application of Biocontrol Microbes: Bacillus spp.* The American Phytopathological Society 94: 1245.
  116. Stern S, Powers T, Changchien LM, Noller HF (1989) RNA protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science* 244: 783–790.
  117. Keim P, Kalif A, Schupp J, Hill K, Travis SE, et al. (1997) Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. *J Bacteriol* 179: 818–824.
  118. Petersen DJ, Shishido M, Holl BF, Chanway CP (1995) Use of species- and strain-specific PCR primers for identification of conifer root-associated *Bacillus* spp. *FEMS Microbiol Lett* 133: 71–76.

119. Carlson CR, Kolsto AB (1994) A small (2–4 Mb) *Bacillus cereus* chromosome corresponds to one conserved region of a larger (5–3 Mb) *Bacillus cereus* chromosome. *Mol Microbiol* 13: 161–169.
120. Carlson CR, Kolsto AB (1993) A complete physical map of a *Bacillus thuringiensis* chromosome. *J Bacteriol* 175: 1053–1060.
121. Carlson CR, Johansen T, Kolsto AB (1996) The chromosome map of *Bacillus thuringiensis* subsp. *canadensis* HD224 is highly similar to that of the *Bacillus cereus* type strain ATCC 14579. *FEMS Microbiol Lett* 141: 163–167.
122. Jensen GB, Andrup L, Wilcks A, Smidt L, Poulsen OM (1996) The aggregation-mediated conjugation system of *Bacillus thuringiensis* subsp. *israelensis*: host range and kinetics of transfer. *Curr Microbiol* 33: 228–236.
123. Daffonchio D, Raddadi N, Merabishvili M, Cherif A, Carmagnola L, et al. (2006) Strategy for identification of *Bacillus cereus* and *Bacillus thuringiensis* strains closely related to *Bacillus anthracis*. *Appl Environ Microbiol* 72: 1295–1301.
124. Bourque SN, Valero JR, Lavoie MC, Levesque RC (1995) Comparative analysis of the 16S to 23S ribosomal intergenic spacer sequences of *Bacillus thuringiensis* strains and subspecies and of closely related species. *Appl Environ Microbiol* 61: 1623–1626.
125. Te Giffel MC, Beumer RR, Klijn N, Wagendorp A, Rombouts FM (1997) Discrimination between *Bacillus cereus* and *Bacillus thuringiensis* using specific DNA probes based on variable regions of 16S rRNA. *FEMS Microbiol Lett* 146: 47–51.
126. Yamada S, Ohashi E, Agata N, Venkateswaran K (1999) Cloning and Nucleotide Sequence Analysis of *gyrB* of *Bacillus cereus*, *B. thuringiensis*, *B. mycoides*, and *B. anthracis* and their application to the detection of *B. cereus* in rice. *Appl Environ Microbiol* 65: 1483–1490.
127. Prub BM, Francis KP, Stetten FV, Scherer S (1999) Correlation of 16S Ribosomal DNA signature sequences with temperature-dependent growth rates of mesophilic and psychrotolerant strains of the *Bacillus cereus* group. *J Bacteriol* 181: 2624–2630.
128. Purohit HJ, Raju DV, Kapley A (2003) Identification of signature and primers specific to genus *Pseudomonas* using mismatched patterns of 16S rDNA sequences. *BMC Bioinfo* 4: 19.
129. Lowe T, Sherkin J, Yang SQ, Dieffenbach CW (1990) A computer program for selection of oligonucleotide primers for polymerase chain reaction. *Nucleic Acids Res* 18: 1757–1761.
130. Mitsuhashi M, Cooper A, Ogura M, Shinagawa T, Yano K, et al. (1994) Oligonucleotide probe design – a new approach. *Nature* 367: 759–761.
131. Feng DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J Mol Evol* 25: 351–360.
132. Lipman DJ, Altschul SF, Kececioglu JD (1989) A tool for multiple sequence alignment. *Proc Natl Acad Sci USA* 86: 4412–4415.
133. Thompson D, Higgins DG, Gibson TJ (1994) Clustal W, improving the sensitivity of progressive multiple sequences alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
134. Thompson JD, Plewniak F, Poch O (1999) A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res* 27: 2682–2690.
135. Bujnicki JM (2003) Crystallographic and bioinformatic studies on restriction endonucleases: inference of evolutionary relationships in the ‘midnight zone’ of homology. *Curr Protein Pept Sci* 4: 327–337.
136. Nikolajewa S, Beyer A, Friedel M, Hollunder J, Wilhelm T (2005) Common patterns in type II restriction enzyme binding sites. *Nucleic Acids Res* 33: 2726–2733.
137. New England Biolabs (2007) restriction endonucleases. Over view. [http://www.neb.uk.com/technicalresources/restriction\\_enzymes/overview.asp](http://www.neb.uk.com/technicalresources/restriction_enzymes/overview.asp).
138. Arber W (1979) Promotion and limitation of genetic exchange. *Science* 205: 361–365.
139. Price C, Bickle TA (1986) A possible role for DNA restriction in bacterial evolution. *Microbiol Sci* 3: 296–299.
140. Blanc M, Marilley L, Belfa T, Aragno M (1997) Rapid identification of heterotrophic, thermophilic, spore-forming bacteria isolated from hot composts. *Int J Syst Bacteriol* 47: 1246–1248.
141. Heyndrickx M, Vauterin L, Vandamme P, Kersters K, De Vos P (1996) Applicability of combined amplified ribosomal DNA restriction analysis (ARDRA) patterns in bacterial phylogeny and taxonomy. *J Microbiol Methods* 26: 247–259.
142. Heyndrickx M, Lebbe L, Kersters K, Hoste B, De Wachter R, et al. (1999) Proposal of *Virgibacillus proomii* sp. nov. and emended description of *Virgibacillus pantothenicus* (Proom and Knight 1950) Heyndrickx et al. 1998. *Int J Syst Bacteriol* 49: 1083–1090.
143. Björkroth JK, Korkeala HJ, Funke G (1999) rRNA gene RFLP as an identification tool for *Corynebacterium* species. *Int J Syst Bacteriol* 49: 983–989.
144. Persson A, Pettersson B, Bolske G, Johansson KE (1999) Diagnosis of contagious bovine pleuropneumonia by PCR-laser-induced fluorescence and PCR-restriction endonuclease analysis based on 16S rRNA genes of *Mycoplasma mycoides* subsp. *mycoides* SC. *J Clin Microbiol* 37: 3815–3821.
145. Stakenborg T, Vicia J, Butaye P, Maes D, De Baere T, et al. (2005) Evaluation of amplified rDNA restriction analysis (ARDRA) for the identification of *Mycoplasma* species. *BMC Infectious Diseases* 5: 46.
146. Bradbury JM (2001) International committee on systematic bacteriology, subcommittee on the taxonomy of Mollicutes. *Int J Syst Evol Microbiol* 51: 2227–2230.
147. Nielsen P, Fritze D, Priest FG (1995) Phenetic diversity of alkaliphilic *Bacillus* strains: proposal for nine new species. *Microbiology* 141: 1745–1761.
148. Felsenstein J (1993) Phylip (Phylogeny Inference Package) version 3.57c. Department of Genetics, University of Washington, Seattle. Distribution: <http://evolution.genetics.washington.edu/phylip.html>.
149. Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computer. *Comput Appl Biosci* 12: 357–358.
150. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Altman RB, Brutlag DL, Karp PD, Lathrop RH, Searls DB, eds. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, CA. pp 28–36.
151. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res Web Server issue*. pp W369–W373.
152. Pevzner PA, Sze SH (2000) Combinatorial approaches to finding subtle signals in DNA sequences. In: Bourne PE, Gribskov M, Altman RB, Jensen N, Hope D, Lengauer T, Mitchell JC, Scheeff ED, Smith C, Strande S, Weissig H, eds. In Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, CA. pp 269–278.