ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 18/05/2022

DOI: https://doi.org/10.3897/arphapreprints.e86638

Biochemical, physiological, and molecular characterisation of a large collection of aerobic endospore-forming bacteria isolated from Brazilian soils

Paulo Henrique Rosa Martins, Leon Rabinovitch, Juliana Orem, Waldeyr Silva, Felipe Mesquita, Maria Magalhaes, Danilo Cavalcante, Adriana Vivoni, Josiane Brito, Edmar Oliveira, Vera Lima, Marlene De-Souza

Biochemical, physiological, and molecular characterisation of a large collection of aerobic endospore-forming bacteria isolated from Brazilian soils

Paulo Henrique Rosa Martins[1], ph.biomedicina@gmail.com;
Leon Rabinovitch[2], leon@ioc.fiocruz.br;
Juliana Capela de Orem[1], jucorem@gmail.com;
Waldeyr Mendes C. Silva[3], waldeyr.mendes@ifg.edu.br;
Felipe de Araujo Mesquita[1], felipedearaujomesquita@gmail.com;
Maria Ines Andre de Magalhães[1], miamagalhaes@gmail.com;
Danilo de Andrade Cavalcante[1], danilo.ac9@gmail.com;
Adriana Marcos Vivoni[2], avivoni@ioc.fiocruz.br;
Edmar Justo de Oliveira2, edsea@hotmail.com.br;
Vera Cristina Pessoa de Lima[2], vera@ioc.fiocruz.br;
Josiane Teixeira Brito[2], josidebrito@hotmail.com;
and Marlene Teixeira De-Souza[1*], marlts@unb.br.

[1]Department of Cellular Biology, Institute of Biological Sciences, University of Brasília, DF, Brazil.
[2]Laboratório de Fisiologia Bacteriana, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brazil
[3]Federal Institute of Goias, Formosa, Brazil

*Corresponding author
Address: Universidade de Brasília; IB/Departamento de Biologia Celular;
Campus Universitário Darcy Ribeiro; 70.910-900 – Brasília – DF, Brazil.
Telephone: +55 61 3107-3044
e-mail: marlts@unb.br
https://orcid.org/0000-0003-1538-2657

Abstract

The aerobic endospore-forming bacteria (AEFB) comprise species of *Bacillus* and related genera, allocated in the phylum *Firmicutes*. Although *Bacillus* spp. are among the first bacteria to be characterised, the wide diversity renders appropriate categorisation and generalisations challenging tasks. To determine genetic diversity, analyses at the molecular level are the most accurate. However, gene expression, morphological, biochemical, and physiological aspects must also be considered. The metabolism of bacteria is adapted to their natural environment or host. Thus, metabolic outlines can be used for identifying AEFB, form the basis of the formal description of bacterial taxa, and are strongly recommended for taxonomic purposes. This work addressed the biochemical and physiological profiles of 312 environmental AEFB—designated as SDF (*Solo do Distrito Federal*)—by performing 30 tests. Out of it, 246 were classified by 16S rRNA gene sequences. We summarised the phenotypic test relationships among selected SDF strains using a Pearson correlation-based clustering represented in heatmaps. In practice, biochemical and physiological profiles are often less discriminatory than molecular data and may be unstable because of the loss of traits. Though these test reactions are not universally positive or negative within species, they may define biotypes and be efficient strain markers, enhancing the accuracy of unknown sample identification. It can be also helpful in selecting the best represent the phenotypes of samples. Along with the other phenotypic and genotypic data, the present results will be of great importance for the robust classification of the SDF strains within the scope of the polyphasic approach.

Keywords: *Bacillales*; *Bacillaceae*; endosporulation; *Firmicutes*; bacterial identification; bacterial metabolism; phenotyping; taxonomy.

Running head: Soil aerobic endospore-formers phenotypic and molecular profiles

Introduction

Aerobic endospore-forming bacteria (AEFB) encompass species from genus *Bacillus* and related genera and produce dormant and highly resistant cells called spores (Fritze 2004; Logan and Halket 2011; Setlow 2014; Driks and Eichenberger 2016). Spores can germinate within seconds when external conditions become favourable (Moir and Cooper 2014). Strains of AEFB are widely distributed in nature, and soil is recognised as the main reservoir (Fritze 2004; Logan et al. 2009; Mandic-Mulec and Prosser 2011; De Vos 2011). AEFB harbour species of significant importance in health, environment, and biotechnology (Fritze 2004; Logan et al. 2009; Ehling-Schulz and Messelhäusser 2013; Alina et al. 2015).

AEFB exhibit high levels of genetic, biochemical, and physiological diversity and appreciable resistance to adverse environmental (De Vos et al. 2009; Logan et al. 2009; Logan and Halket 2011; Galperin 2013; Setlow 2014; Driks and Eichenberger 2016). The high heterogeneity in the phenotypic and genotypic characteristics has been hampering the taxonomy of these species (Ash 1991; Fritze 2004; Logan et al. 2009; Galperin 2013).

The first identification and classification schemes of AEFB were based on the morphology of the colonies, vegetative cells, sporangia, spores, Gram-staining response, besides biochemical, physiological, and chemotaxonomic properties (Logan et al. 2009). Today's polyphasic taxonomy distinguishes and classifies strains based on these classical phenotypic data, supplemented with genotypic and other phenotypic results obtained at the molecular level (Colwell 1970; Fritze 2004; Prakash et al. 2007; Logan et al. 2009; Das et al. 2014). Combining classical and molecular data, notably 16S rRNA gene sequencing, has revolutionised our understanding of domain Bacteria (Bochner 2009) and led to a rapid increase in the number of descriptions of novel AEFB taxa, especially at genus and species levels (Fritze 2004; Logan et al. 2009; Maughan and Van der Auwera 2011).

AEFB are allocated in the phylum *Firmicutes*, within the class *Bacilli*, order *Bacillales*, where seven families harbour aerobic spore-forming genera: *Bacillaceae*, Alicyclo*bacillaceae*, *Paenibacillaceae*, *Planococcaceae*, *Pasteuriaceae*, *Sporolactobacillaceae*, and *Thermoactinomycetaceae* (De Vos et al. 2009; Logan and Halket 2011; Galperin 2013; Parte 2018).

Phylogenetic studies based on the 16S rRNA gene sequences suggest clusters of closed related AEFB species, designated groups (Ash et al. 1991; Stackebrandt and Swiderski 2002; Fritze 2004; De Vos et al. 2009; Logan 2009; Alina 2015). The early rRNA groups 1 to 5 of *Bacillus* species proposed by Ash et al. in 1991 were expanded to house alkaliphilic and alkalitolerant species, while other groups of species, such as those allocated in genera *Paenibacillus* (group 3, *Brevibacillus* (group 4), and other distinct taxa have been reclassified (Stackebrandt and Swiderski 2002).

Within genus *Bacillus*, members of *B. cereus* group or *sensu lato* (*sl*) and *B. subtilis* complex are composed of highly related members (>99% similarity), restricting species delimitation when considering only the 16S rRNA gene analyses. Differently from the other *Bacillus* groups described above that harbour high genome identity, *B. megaterium* and *B. aryabhattai* share 99.7% of identity in the 16S rRNA gene sequences. Nevertheless, the genomes are less than 70% identical (Shivaji et al.

2009). Therefore, the distinction of these two strains using only this technique is also challenging.

Since observable features from growth conditions and enzymatic reactions are related to the genome expression, the resulting profiles allow detecting phenotypic patterns for the species evaluated. Thus, investigating these intrinsic metabolic activities are still essential for the identification and classification of new AEFB isolates. These assays are highly recommended in the characterization of AEFB strains (Logan et. al 2009).

To help understand AEFB diversity and explore their biotechnological potential, we isolated 312 strains from soil samples collected at random areas of the Federal District, Midwest region of Brazil (Cavalcante et al. 2019; Orem et al. 2019; Martins et al. 2020). These strains, designated SDF0001-SDF0312 (*Solo do Distrito Federal* or SDF) are deposited at the *Coleção de Bactérias Aeróbias Formadoras de Endósporos* (AEFB Collection—AEFBC), hosted at the University of Brasilia. For taxonomic purposes, the SDF strains are being analysed by a polyphasic strategy.

In the present work, 30 biochemical and physiological tests were performed to investigate substrates utilisation and transformation, in addition to the growth conditions capabilities of 312 SDF strains. Among them, 246 were classified by 16S rRNA sequences. A Pearson correlation based on a clustering method (Gu et al. 2016) was used to construct heatmaps to summarise the relationships of selected SDF strains to these phenotypic tests.

Methods

Bacterial strains. The 312 SDF strains evaluated in this study were isolated as described in Cavalcante et al. (2019) and Orem et al. (2019). The reference strains used as positive and negative controls for the physiological and biochemical tests (Table 1) are deposited at *Coleção de Culturas do Gênero Bacillus e Gêneros Correlatos* (CCGB), of the Instituto Oswaldo Cruz (LFB-Fiocruz-RJ, Brazil).

Ethics statement. Specific permissions required to collect bacterial strains used in this study were endorsed by the Federal Brazilian Authority (CNPq; Authorization of Access and Sample of Genetic Patrimony nº 010439/2015-3). Sampling did not involve endangered or protected species.

Biochemical and physiological assays. Strains were grown in nutrient agar (33 °C, 24 h) under atmospheric aerobic conditions. Cells from a single colony were transferred to a tube containing nutrient broth and incubated at 33 °C, under constant stirring (200 rpm), for about 16 h. The 30 biochemical and physiological tests (Table 1) were performed according to Bergey's Manual of Systematic Bacteriology (Smith et al. 1952; Gordon et al. 1973; Claus and Berkeley 1986; Oliveira and Rabinovitch 1998; De Vos et al. 2009; Rabinovitch and Oliveira 2015). All tests were performed in duplicate in two independent experiments.

Taxonomic assignments of SDF strains. DNA preparation, PCR amplification, sequencing, and sequence analyses were performed as described in Orem et al. (2019). Briefly, the nearly full length of both strands of 16S rRNA genes was amplified using total DNA and primers 27F (5’ AGA GTT TGA TCM TGG CTC AG 3’) and 1492R (5’ GGY TAC CTT GTT ACG ACT T 3’). PCR products were bi-directionally sequenced employing the Sanger method. These sequences were filtered for Q≥20 in Phred

scores and taxonomically assigned using BLAST and Classifier as described in Orem et al. (2019).

Heatmaps. The biochemical and physiological assays results were arranged in heatmaps (Gu et al. 2016) to enhance the potential of visually revealing patterns and correlations among them. We took the dichotomous values 0 (for Negative) and 1 (for Positive) as binary variables representing the association among the species' and its biochemical and physiological assay results. Using Pearson's correlation, the species were clustered taking similar biochemical and physiological results (Hummel et al. 2017). R scripts are available at https://github.com/waldeyr/bafes_figures.

Results and discussion

Due to metabolism importance for identification and classification of AEFB new isolates (Fritze 2002; Logan et al. 2009), we applied 30 biochemical and physiological tests (Table 1) to 312 AEFB strains isolated from Brazilian soils, designated SDF strains (Cavalcante et al. 2019; Orem et al. 2019; Martins et al. 2020). The profiles obtained from enzymatic reactions and growth conditions are described in Table S1, available in the online Supplementary Material. It is important to state that, all the 312 SDF strains studied are aerobic or facultative anaerobic endospore-formers, and Gram-positive or Gram-variable cells (Cavalcante et al. 2019; Orem et al. 2019; Martins et al. 2020). The latter characteristics are common to taxa found in the order *Bacillales* (Fritze 2004; De Vos et al. 2009; Logan et al. 2009; Galperin 2013), where these environmental AEFB strains are allocated.

Of these 312 SDF strains, the taxonomic assignments of 246 were addressed using the standard tool of taxonomists for bacterial identification, classification, and phylogenetic relatedness, the 16S rRNA gene sequences (Tringe and Hugenholtz 2008; De Vos et al. 2009; Hakorvita et al. 2016), as described in Orem et al. (2019). The lowest and highest inter-species pairwise 16S rRNA gene sequence similarities spanned from 90% to 100% (Table S1). Considering the similarity thresholds for genera 96%, and ≥97% for species (Stackebrandt and Goebel 1994), the classification obtained segregated 238 SDF strains into 6 genera, being 4 part of family *Bacillaceae* and 2 of *Paenibacillaceae* (Fig. 1A). Among the SDF strains described in the present work, *Bacillus* spp., belonging to the family *Bacillaceae,* are the most prevalent (207 strains; 84.14%), followed by species of genera *Paenibacillus* (14; 5.69%; family *Paenibacillaceae*), *Lysinibacillus* (7; 2.84%; *Bacillaceae*), *Brevibacillus* (6; 2.43%; *Paenibacillaceae*), *Terribacillus* (1; 0.40%; *Bacillaceae*), and *Rummeliibacillus* (1; 0.40%; *Bacillaceae*). These findings are not surprising since the selective procedure we used to isolate SDF strains intended to favour non-fastidious AEFB species, excluding strict anaerobic endospore-forming and Gram-negative cells (Cavalcante et al. 2019; Orem et al. 2019; Martins et al. 2020).

Included in the 224 SDF strains classified at the species level (Table S1), members of *B. pumilus* subgroup were predominant (83 strains; 37.05%), followed by *B. cereus* group species (48; 21.42%), *B. megaterium* group (35; 15.62%), other members of *B. subtilis* complex (12; 5.35%); *B. simplex* (7; 3.12%); *B. clausii* (3; 1.33%); *B. subterraneus* (2; 0.89%); besides 1 (0.44%) of each: *B. australimaris*; *B. arbutinovorans; B. circulans; B. kochii*; *B. luciferensis*; *B. oleronius*; *B. siamensis*, and *B. senegalensis*. Outside genus *Bacillus*, other species belonging to family *Bacillaceae* were *Lysinibacillus* sphaericus (3; 1.33%); *L. xylanilyticus* (2; 0.89%); *L. fusiformis* (2;

0.89%), and *Terribacillus goriensis* (1; 0.44%). *Paenibacillus* spp. (12 strains; 5.35%), and *Brevibacillus* spp. (5 strains; 2.23%) allocated in the family *Paenibacillaceae* complete the list of SDF strains classified at the species level (see below). The diversity of the SDF strains is represented in Fig. 1B.

It is mentionable that members of *B. cereus sl* and *B. subtilis* subgroups are composed of very related members (>99% similarity), restricting species delimitation when considering only the 16S rRNA gene analyses. Conversely, *B. megaterium* and *B. aryabhattai* share 99.7% of identity in the 16S rRNA gene sequences, even though the genomes are less than 70% identical (Shivaji et al. 2009). Therefore, the distinction of these two species using only this technique is also challenging.

Thus, our taxonomic assignments based on 16S rRNA gene sequences are a preliminary inference of genera or species. Accordingly, when 16S rRNA gene profiling placed these strains within these AEFB taxa, a sample analysed can belong to two or even more species alternatives within the same affiliation cluster. In these instances, this approach can find groups of bacteria, nevertheless cannot assign it accurately to a species according to its low discrimination ability. Since 10 SDF strains exhibited similarity rates spanning 90-95% (Table S1), the 16S rRNA gene-sequencing tool failed to classify these environmental strains even at the genus level. While this genetic marker was insufficient to set up genus or species, low gene-sequence similarity might suggest that novel species could have been isolated (Tindal et al. 2010). Nevertheless, the description of the new taxa is beyond the scope of this article.

Bacillus is the genus type of order *Bacillales*, and *Bacillus* spp. have been isolated from a wide range of environments (Tamames et al. 2010; Mandic-Mulec and Prosser 2011; Alina et al. 2015; Orem et al. 2019; Cavalcante 2019; Salgado et al. 2020). Soils, along with freshwaters, are one of the least restrictive for these species. It is worthy to note that certain species found in soils are inactive in these environments. It could be the case for some SDF strains isolated from Brazilian soils. The method of isolation based on heat shock allowed the dormant spores to germinate and grow in vitro.

B. cereus and *B. anthracis* are human pathogens causing food-borne illness and anthrax, respectively (Arnesen et al. 2008; Ehling-Schulz and Messelhäusser 2013). On the other hand, the metabolic breadth of *Bacillus* spp. has been explored by the industry for producing a vast range of antibiotics; plant growth promotion molecules; hydrolyses; toxins against plants, fungus, insect, and nematode, in addition to other bioproducts (de Maagd et al. 2003; Berkeley et al. 2008; Logan et al. 2009; Galperin 2013; Alina 2015). Therefore, despite the danger of few members, most species are beneficial.

Genus *Bacillus* stays the largest AEFB taxon, accommodating 614 species, as registered at the List of Prokaryotic Names with Standing in Nomenclature (LPSN: https://www.bacterio.net/*Bacillus*.html; accessed on 01 February 2022). Taxonomy within genus *Bacillus* is hampered by high heterogeneity at phenotypic and genotypic levels (Ash 1991; Fritze 2004; Logan and De Vos 2009; Logan et al. 2009). Further, these divergencies restrict the distinction between *Bacillus* spp. and those allocated in other genera inside *Bacillaceae*.

Typically, *Bacillus* spp. are considered aerobic, although at least 20 species are facultatively anaerobic (Logan and De Vos 2009). Furthermore, nitrate reduction is frequently observed in this genus. Members of *Bacillus* can be rods or cocci, motile or non-motile, organotrophic or lithotrophic (Fritze 2004; Logan and De Vos 2009; Logan et al. 2009). Cell size, varying from 0.4 to 1.8 µm in diameter and from 0.9 to 10.0 µm in length, can also be used to differentiate *Bacillus* spp. (Logan and De Vos 2009). Phylogenetically, most recognised species are arranged into subclusters or rRNA groups.

Due to the significant relevance in economy and health issues, the *B. cereus* group and *B. subtilis* complex have been received considerable attention (Fritze 2004; Maughan and Van der Auwera 2011). The *B. cereus* group hosts *B. cereus sensu stricto* (or *ss* or *B. cereus*), *B. anthracis*, *B. thuringiensis*, *B. mycoides*, *B. pseudomycoides*, *B. weihenstephanensis*, *B. toyonensis*, and *B. cytotoxicus* (Fritze 2004; Maughan and Van der Auwera 2011; Ehling-Schulz and Messelhäusser 2013). Organisms placed in this group belong to 16S rRNA/DNA group 1. Cells are typically wider than 1 µm, Gram-positive, and endospores are oval to cylindrical paracentral or subterminally localised in unswelling sporangia.

Traditionally, these bacteria have been differentiated based on phenotypic characteristics, especially pathogenic potential. Nonetheless, this group is a highly homogeneous subdivision inside genus *Bacillus* (Helgason et al. 2000; Chen and Tsen 2002). Furthermore, they are hardly distinguishable with standard biochemical and chemotaxonomic methods or phylogenetically relevant target genes (Bavykin et al. 2004; Arnesen et al. 2008). However, specific biochemical and physiological characteristics of the *B. cereus sl* are advantageous to differentiate these taxa from the other aerobic endospore-forming species.

Out of 224 SDF strains classified at the species level (Table S1), 48 (21.42%) were members of the *B. cereus* group. Using a Pearson correlation-based clustering method (Gu et al., 2016), we constructed a heatmap (Fig. 2) to summarise the relationships of these 48 environmental strains to the 30 biochemical and physiological tests performed (Table 1). Each column shows the metabolic pattern (bottom side) of individual SDF strain (rows at the right side), classified based on 16S rRNA sequences. The green and the red colours represent positive and negative responses, respectively. Fig. 2 shows an assembling of these essays based on the prevalence of the positive responses. It is doable to distinguish which strains respond similarly to the tests when they are in the same clade. For example, those strains in distant clades respond differently. It is also possible to discern which SDF strains respond similarly to each test, as well as discriminate them by correlating rows and columns. The clusters from the upper-side dendrogram (Fig. 2) stands for the similarity of the 30 tests responses. The left-most contains 17 columns, while the right-most contains 13, where the majority of these SDF strains responded positively and negatively, respectively.

Although many AEFB may not respond positively to the catalase test, most species rod-shaped, either Gram-positive or Gram-positive only in the initial stages of growth, are catalase-positive, especially members of the genus *Bacillus* (Logan and De Vos, 2009a). However, in most cases, respiratory metabolism occurs at low 0_2 levels. Here, all the 48 SDF members of the *B. cereus* group responded positively to this enzyme linked to respiration in the presence of atmospheric 0_2 (Fig. 2). This

positivity seems to be a characteristic of this group of sporulating procaryotes. As assessed in this work, it is worth noting that *B. cereus* ss can grow under certain anaerobiosis conditions (Logan and De Vos, 2009a).

The cytochrome C oxidase is especially useful to discriminate Gram-negative pathogens Vibrio spp. (oxidase positive) from the oxidase-negative enteric bacteria (Vila et al. 1992). This enzyme catalyses the oxidation of cytochrome C while reducing oxygen to form water. The oxidation test in vitro employs colourless artificial acceptors like dimethyl or tetramethyl p-phenylenediamine resulting in purple colour when positive. This essay also distinguishes *Neisseria* and *Moraxella* (both oxidase positive) from *Acinetobacter* spp. (oxidase negative) (Henriksen 1976; Powell and Marcon 2012).

From the 48 SDF strains allocated in the *B. cereus* group, 25 (52.08%) were oxidase-positive. Logan and De Vos (2009) point out this variability for this genus and related genera, demonstrating apparent inactivity of this enzyme, or even that the traditional method failed to detect the oxidase activity in almost half of these samples.

Anaerobiosis assays, performed in tubes containing aldehyde-reduced agar medium inoculated with a needle, revealed growth a few centimetres below the interface of the culture medium with atmospheric air to 24 (50%) of the SDF strains belonging to the *B. cereus sl*. This effect present in-depth denotes anaerobic growth, a property conserved among AEFB.

Some Bacillus species do not appear to utilise carbohydrates whatsoever (Logan and De Vos 2009). Yet the acid production profiles from monosaccharides and disaccharides are of great value in the characterisation and identification of these species. Most SDF strains allocated into *B. cereus* group used D-glucose, L-arabinose, D-xylose, and other fermentable carbohydrates as sole sources of carbon and energy (Table S1; Fig. 2). Probably they have the genetic information to conduct the pathway of Embden-Meyerhof-Parnas, coupled with the Krebs cycle, verified by acid production (Logan and De Vos 2009a).

Regarding glucose consumption, five SDF strains, one classified as *B. thuringiensis* (SDF0225), and four as *B.cereus ss* (SDF0124; SDF0229; SDF0237, and SDF248), responded negatively to the use of this monosaccharide, which is rare among rods AEFB, as they usually assimilate and degrade D-glucose. Though the formation of acid from D-mannitol is frequently negative for members of the *B. cereus sl* and positive for strains of other groups (Fritze 2002), three SDF strains classified as *B. cereus* ss (SDF0219; SDF0124, and SDF0022) and *B. anthracis* SDF0199 were able to ferment this sugar (Fig. 2). Interestingly, these strains were gathered in the uppermost and downmost rows of the strains' list (right side). Indeed, clustering heat maps can group samples based on the similarity of their phenotypic patterns, thus identifying atypical responses (Zhao et al. 2014).

The Voges-Proskauer test presented some species such as two *B. thuringiensis* strains (SDF0161 and SDF0178); three *B. anthracis* (SDF181; SDF0186, and SDF0199), besides seven *B. cereus* ss (SDF0155; SDF0159; SDF0182; SDF0184; SDF0239; SDF0270, and SDF0272) responding negatively to the acetyl-methylcarbinol production assay, which allows us to suspect that these strains may not

produce enzymes that decarboxylate lactic acid from the glycolytic pathway, or do not have an enzyme capable of bonding two molecules originating from the production of acetate ions.

Oliveira and Rabinovitch (1998) established a standardised protocol for detection of gelatin hydrolysis by *Lysinibacillus sphaericus*—former *B. sphaericus* (Seldin et al. 1984; Ash et al. 1994)—showing that 93.3% of strains belonging to this species hydrolyses this incomplete protein after four days of incubation. Here bulk 48 SDF strains accommodated in the B. cereus group could use gelatin. Providing the relatively high number of strains submitted to this type of biochemical test, we considered a very valid verification.

The development in the presence of lysozyme is another characteristic of the *B. cereus* group and hardly occur in the other species of other groups (Fritze 2002). However, *B. cereus* SDF0124 and *B. thuringiensis* SDF085 did not grow in this condition, indicating that the cell wall of these two strains can be hydrolysed by this enzyme.

The production of haemolysin, as well as cell morphology in a few strains of *B. cereus sl*, are also phenotypes with relevance for taxonomic studies (Fritze 2002; Fritze 2004; De Vos and Logan 2009; Logan et al. 2009). *B. cereus* ss, in general, mobile, is heavily haemolytic, but does not produce rhizoid growth pattern, a characteristic that can be used to differentiate from colonies of B. mycoides strains (Fritze 2002). Most *B. anthracis* strains are neither mobile nor haemolytic (Fritze 2004; Maughan and Van der Auwera 2011). However, non-mobile *B. cereus* strains, as well as hemolytic *B. anthracis*, may hinder the differentiation between these two species. In addition, the latter species can be differentiated by parasporal crystal formation typically described to *B. thuringiensis* (Fritze 2002).

Out of 48 strains, SDF allocated in the *B. cereus* group, three samples classified as *B. cereus* (SDF0159, SDF0237, and SDF0270); one *B. anthracis* SDF0181, and three *B. thuringiensis* (SDF0161; SDF0085, and SDF0030) presented no haemolysin activity. It is of great significance to mention that *B. thuringiensis* SDF0030 produces a typical parasporal crystal (Cavalcante et al. 2014), a classical feature distinguishing *B. thuringiensis* strains from *B. cereus* ss. (Dagmar 2014). Conversely, three strains classified as *B. anthracis* (SDF0199, SDF089, and SDF0186) were positive for haemolysin activity. The main phenotypical properties that are frequently used to distinguish *B. cereus*, *B. thuringiensis*, and *B. anthracis* are related to the presence or absence of large plasmids, where the replicons are localized (Maughan and Van der Auwera 2011). Future investigation on the extrachromosomal profiles of these SDF strains will help to understand the evolutionary relatedness of these species.

B. subtilis ss, the genus *Bacillus* type-species, is prominent in microbial history, and play a distinct role as a model for Gram-positive bacteria and in the understanding of stress-resistance of bacterial spores (Fritze 2004; Maughan and Van der Auwera 2011; Galperin 2013; Driks and Eichenberger 2016). Besides being recognised as a model, this species, along with other highly related accommodated in the *B. subtilis* complex, is extensively employed in industry and agriculture (Fan et al. 2017).

B. subtilis strains are aerobics, although some strict anaerobic growth may be observed in complex media with glucose or (less effectively) nitrate (Logan and De Vos 2009a). This organism is catalase-positive, oxidase variable, and can reduce nitrate to nitrite. The motile rod-cells of 0.7–0.8 x 2.0–3.0 µm are Gram-positive and can be frequently observed singly, in pairs, and, occasionally, in chains. This species forms ellipsoidal to cylindrical endospores at the central, paracentral, or subterminal position in unswollen sporangia.

Although optimal growth ranges from 28-30 °C, *B. subtilis* can tolerate temperatures from 5–20 °C and 45–55 °C (Logan and De Vos 2009a). Growth can occur from pH 5.5 to 8.5, with no limits recorded. The vegetative cells have a significant role in the early steps of organic matter decomposition. Growth in minimal medium containing glucose and ammonium salt—as sole sources of carbon— and nitrogen is also observed. Most strains can use citrate, as the sole carbon source, and growth occurs in the presence of up to 7% NaCl, and certain tolerate 10% NaCl. *B. subtilis* can hydrolyse casein, esculin, gelatin, and starch but not phenylalanine and urea. Extracellular dextran and levan are produced from sucrose. Voges–Proskauer test is positive, and the production of acid without gas can be detected from glucose, besides additional carbohydrates.

As a taxonomic unit above the species level, the *B. subtilis* species complex can be split into four clades (Fan et al. 2017). These recognizable monophyletic groups comprise clade I, consisting of three subspecies of *B. subtilis* (*subtilis*, *spizenii*, and *inaquosorum*), besides *B. tequilensis*, *B. vallismortis*, B. *mojavensis*, and B. atrophaeus; clade II containing species *B. amyloliquefaciens*, *B. siamensis*, and a conspecific complex embracing *B. methylotrophicus*, *B. velezensis*, and *B. amyloliquefaciens* subsp. *plantarum*; clade III encompassing *B. licheniformis*, *B. sonorensis*, and related species, and clade IV made of *B. pumilus* and *B. safensis*, *B. xiamenensis*, and a conspecific group involving the type strains of *B. altitudinis*, *B. stratosphericus*, and *B. aerophilus*. Like strains from the *B. cereus* group, these taxa are placed in 16S rRNA/DNA group 1 and are phylogenetic and physiologically remarkably similar (Fritze 2004). Strains from this complex are usually mesophiles and neutrophiles, but often tolerant to high pH values (Fritze 2004).

Employing 16S rRNA gene sequences, from the 224 SDF strains classified at the species level 95 (42.41%) were allocated in the *B. subtilis* complex (Table S1). Among them, the *B. pumilus* subgroup represented 83 (37.05%), most of it or 61 (27.23%) classified as *B. pumilus*; 16 (7.14%) as *B. safensis*, and 6 (2.67%) as *B. altitudinis*. Seven strains belonged to *B. amyloliquefaciens* subgroup or 3.12%, being 4 (1.78%) *B. amyloliquefaciens* strains and 3 (1.33%) *B. velezensis*. The remaining 4 (1.78%) SDF strains were accommodated in the *B. subtilis* subgroup, being 3 (1.33%) *B. subtilis* ss and 1 (0.44%) *B. tequilensis*.

The relationships of these 95 strains to the 30 biochemical and physiological tests described in Table 1 were also analysed. The resulting heat map (Gu et al. 2016) shown in Fig. 3 revealed two clusters (upper-side dendrogram) encompassing 13 columns at the left-most cluster, while the right-most contained 17, where these SDF strains responded positively and negatively, respectively.

Species belonging to the so-called *B. pumilus* subgroup are almost identical in the 16S rRNA gene sequences, sharing above 99.5% similarity (Alina et al. 2015). *B. pumilus* ss is aerobic, catalase-positive, and enabled to reduce nitrate (Logan and De Vos 2009a). Gram-positive or Gram-variable small rods (0.6–0.7 by 2.0–3.0 µm) cells can be observed as singly or in pairs and are motile. Cylindrical to ellipsoidal endospores can be central, paracentral, and subterminally localised in unswollen sporangia. Though optimal growth occurs at pH 6.0 and 9.5, some strains can reproduce at pH 4.5. This species can tolerate up to 10% NaCl, hydrolyse casein, esculin, and gelatin but cannot break down starch. Phenylalanine is not deaminated, and citrate is utilised as the sole carbon source, but propionate is not. Acid without gas is produced from glucose and many other carbohydrates, and Voges–Proskauer test is positive.

In general, the SDF strains belonging to this group corroborates the traits described in Bergeys' *Firmicutes* (Logan and De Vos 2009a). Furthermore, according to Logan and Forsyth, unpublished observations cited in this manual, *B. pumilus* strains isolated from Antarctic soils and penguin rookeries present phenotypic peculiarities, such as producing a diffusible yellow pigment.

Outside *Bacillaceae*, 17/224 (7.58%) SDF strains were allocated in two genera of the family *Paenibacillaceae*. *Paenibacillus* spp. accounted for 12 (5.35%) strains being *7* (3.12%) of *P. alvei* and 1 (0.44%) of each: *P. chibensis; P. ginsengagri*; *P. lautus; P. susongensis,* and *P. terrígena* (Table S1). Five (2.23%) strains of the genus *Brevibacillus* (quoted here as Br.): *Br. laterospor*us (4 or 1.78%), and 1 (0.44%) of *Br. agrii* completed the SDF strains allocated into the family *Paenibacillaceae* (Table S1). The mutual connection between these 18 strains and the 30 biochemical and physiological tests (Table 1) is represented in Fig. 4. The two clusters (upper-side dendrogram) distinguishable by this heat map (Gu et al. 2016) comprehend 11 (left-most) and 19 (right-most) columns, embracing most of these SDF strains responding positively and negatively, respectively.

The genus *Paenibacillus* was created to reallocate species previous*ly* accommodated in the RNA group 3 of genus *Bacillus* (Priest 2009). The family *Paenibacillaceae* was subsequently proposed to house genus *Paenibacillus* and close relatives' genera (Ash et al. 1993; Shida et al. 1997). This family encloses two monophyletic clusters, the first consisting of genera *Paenibacillus*, *Brevibacillus*, *Cohnella*, and *Thermobacillus*, the second of genera *Aneurinibacillus*, *Ammoniphilus*, and *Oxalophagus* (De Vos et al. 2009). The type-genus is *Paenibacillus*.

Members of this family may be strictly aerobic, microaerophilic, facultative aerobic, or obligate anaerobic, being catalase-positive or -negative (De Vos et al. 2009). Cells are straight to curved rods of 0.5–1.0 x 2–6 µm, Gram-positive but may stain Gram-negative or variable. Oval or ellipsoidal endospores are frequently formed inside a swelling sporangium. Peritrichous flagella may be observed, but some species are nonmotile. Although they can utilise oxalic acid as the sole carbon and energy source, these cells are organoheterotrophs and grow in complex media, using carbohydrates and amino acids. They can be mesophilic or thermophilic, neutrophilic or alkaliphilic and have been isolated from soil, roots, faeces, blood, and other substrates.

After *Bacillus*, the genus *Paenibacillus* accommodates the second largest number of AEFB species known (342), as registered at the LPSN (https://www.bacterio.net/; accessed on 01 February 2022). *Paenibacillus* harbours species aerobic or facultative rod-shaped cells (Priest 2009; Galperin 2013; Parte 2018) and bears a typical Gram-positive cell-wall structure (Shida et al. 1996). Nevertheless, even young cells react weakly or even negatively to Gram staining. It should be noted that the 12 SDF strains classified as *Paenibacillus* spp. in this work stained weakly, or yet, Gram-negative (not shown).

Brevibacillus species are aerobic, though some strains are microaerophilic and facultatively anaerobic (Logan and De Vos, 2009b). Most species are catalase-positive. Oxidase reaction and nitrate reduction can differ among strains. Rod-shaped cells are 0.7–1.0 µm x 3.0–6.0 µm, occur singly, in pairs, in chains, are motile through peritrichous flagella, and are Gram-positive or Gram-variable. The ellipsoidal endospores swell the sporangia.

This genus includes a high diversity of thermophilic, psychrophilic, acidophilic, alkalophilic, and halophilic strains that use a variety of carbon sources for either heterotrophic or autotrophic growth (Panda et al. 2014). Carbohydrates may be assimilated, but acid is produced weakly, if at all, by most species. Some amino acids and organic acids may be used as carbon and energy sources (Logan and De Vos 2009b). Casein, gelatin, and starch hydrolysis vary among species. Optimum growth occurs at pH 7.0 and can be inhibited by 5% NaCl. The type-species is *Br. brevis* (Shida et al. 1996), former *Bacillus brevis*.

Brevibacillus spp. are used as a factory for the expression of biotechnologically-important enzymes (e.g., alpha-amylase, sphingomyelinase, xylanase, CGTase, and chitosanase), as well as heterologous proteins including cytokines (EGF, IL-2, NGF, IFN-c, TNF-a, and GM-CSF), antigens, and adjuvants (Mizukami et al. 2010). Besides, *Brevibacillus* spp. are considered a valuable tool for structural and functional biology studies (Panda et al. 2014).

Br. brevis, *Br. choshinensis*, and *Br. laterosporus* have attracted considerable interest owing to the production or transformation of valuable compounds and the biocontrol proprieties (De Vos et al. 2009b). The broad entomopathogenic activity includes species from orders *Coleoptera*, *Lepidoptera*, and *Diptera* and from phyla *Nematoda* and *Mollusca* (Ruiu et al. 2013).

The recent improvements in the tools have been helping in uncovering the vast physiological and genetic diversity within the AEFB, resulting in more appropriate taxonomic arrangements (Fritze 2004; Maughan and Van der Auwera 2011; Galperin 2013). As a result, many new descriptions of genera and species, and reclassifications have occurred.

Molecular methods, especially 16S rRNA gene sequencing, have become the prevailing technique in procaryotic identification, but significant restrictions in our ability to identify environmental bacteria to the genus and species levels remain (Fritze 2004; Maughan and Van der Auwera 2011; Galperin 2013).

Here, the performance of the 16S rRNA sequence analysis was excellent. This tool resolved 238 (96.74%) out of 246 SDF strains at the genus level, unrevealing four and two genera within *Bacillaceae* and *Paenibacillaceae*, respectively. Among the 246 samples, 224 SDF samples (91.05%) were classified at the species level. As mentioned above, using this technique, closely related strains such as those belonging to the *B. cereus* group, *B. subtilis* complex, and other AEFB taxa cannot be resolved at the species level. Still, our classifications are suitable since they clearly show the genera and restrict the identity of part of these SDF strains to one or a few species in the genera described. The positions of the SDF strains in this initial clustering and identification of closely related species may be more accurately determined by incorporating additional data obtained at both genotypic and phenotypic analyses.

Furthermore, our SDF strain classifications revealed well-known AEFB species, together with others that are scarcely described in the literature. Identifying multiple species and strains from different genera may help resolve the order *Bacillales* at the family, genus, and species levels.

Conclusion

In the present study, 30 biochemical and physiological tests provided profiles of all the 312 SDF strains deposited at AEFBC. From the genetic point of view, a large number of samples such as those originating from the environment, as the SDF strains' collection, will hardly display 100% equal answers for all tests, as seen in taxonomic studies of strains isolated from non-clinical substrates (Logan and De Vos 2009; Logan and Halket 2011). In such cases, there are always taxonomically diverging strains. The ubiquitous species *B. pumilus*, isolated from Antarctic soils and penguin rookeries, corroborate this statement as a phenotypic distinction from other lineages can be observed (Logan and Forsyth, unpublished observations, apud Logan and De Vos, 2009a). The divergent samples need to have a separate and improved taxonomic study.

Biochemical and physiological profiles are utile for identifying these microorganisms. These essays are also part of the minimum standards proposed by Logan et al. (2009) for characterising new species of these taxa. However, the value of these tests to accurately identify large numbers of environmental species is limited (Fritze 2004). Therefore, phenotypic similarities cannot be taken with certainty to indicate close evolutionary relatedness.

However, along with the other phenotypic and genotypic data (Cavalcante et al. 2019; Orem et al. 2019; Martins et al. 2020), including complete genome sequences in progress, the profiles described in Table S1 will be significant for robust identification, consequently, classification and differentiation of these environmental strains. The biochemical and physiological profiles can also help optimise the culture conditions for further characterisation and the production of bioactive metabolites by the SDF strains.

Hence, the classification of AEFB at the species levels is not straightforward. And to classify and differentiate closely related SDF strains, these essays should be coupled to other classical and molecular methods involving phenotypic and genotypic types (Cavalcante et al. 2019; Orem et al. 2019; Martins et al. 2020) in a polyphasic

approach (Colwell 1970; Fritze 2004; Prakash et al. 2007; Logan et al. 2009; Das et al. 2014).

This strategy will facilitate the establishment of accurate classification of the SDF strains. It will also allow responsible exploitation of the extraordinary AEFB biotechnological potential, the reliable use as insect control agents, and the handling of animal pathogens.

Funding

The authors have no funding to report.

Competing interests

The authors have declared that no competing interests exist.

Acknowledge

We thank University of Brasilia, and the Brazilian research funding agencies *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (Capes) and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq). We are in debt with Arthur S. Araujo, and Liliam de Oliveira F. Marceneiro for excellent technical assistance.

References

Adimpong DB, Sørensen KI, Nielsen DS, Thorsen L, Rasmussen TB, Derkx PMF, Jespersen L (2013) Draft whole-genome sequence of *Bacillus* sonorensis strain L12, a source of nonribosomal lipopeptides. Genome Announcements 1(2): e0097-13. https://doi.org/10.1128/genomeA.00097-13

Alina SO, Constantinscu F, Petruţa CC (2015) Biodiversity of *Bacillus* subtilis group and beneficial traits of *Bacillus* species useful in plant protection. Romanian Biotechnological Letters 20(5):10737-10750. https://www.cabdirect.org/cabdirect/abstract/20153414554

Arnesen LPS, Fagerlund A, Granum PE (2008) From soil to gut: *Bacillus* cereus and its food poisoning toxins. FEMS Microbiol Rev 32(4): 579-606. https://doi.org/10.1111/j.1574-6976.2008.00112.x

Ash C, Farrow JAE, Wallbanks S, Collins MD (1991) Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. Letters in Applied Microbiology 13(6): 202-206. https://doi.org/10.1111/j.1472-765X.1991.tb00608.x

Ash C, Priest FG, Collins MD (1993) Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. Antonie Leeuwenhoek 64: 253–260. https://doi.org/10.1007/BF00873085

Ash C, Priest FG, Collins MD (1994) In Validation of the publication of new names and new combinations previous*l*y effectively published outside the IJSB. List no. 51. Int. J. Syst. Bacteriol 44: 852. https://doi.org/10.1099/00207713-44-4-852

Ash C, Priest FG, Collins MD (1995) In Validation of the publication of new names and new combinations previous*l*y effectively published outside the IJSB. List no. 52. Int. J. Syst. Bacteriol 45: 197–198. https://doi.org/10.1099/00207713-45-1-197

Bavykin SG, Lysov YP, Zakhariev V, Kelly JJ, Jackman J, Stahl DA, Cherni A (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(8): 3711-3730. https://doi.org/10.1128/JCM.42.8.3711-3730.2004

Berkeley R, Heyndrickx M, Logan N, De Vos P (2008) Applications and systematics of *Bacillus* and relatives. Blackwell Publishing, Maldon-Oxford-Carlton-Berlim, 336 pp.

Bochner BR (2009) Global phenotypic characterization of bacteria. FEMS Microbiol Rev 33 (1): 191-205. https://doi.org/10.1111/j.1574-6976.2008.00149.x

Cavalcante DA, Orem JC, De-Souza MT (2014) Diversity of Bacterial Spores from Brazilian Cerrado's Soil Strains by Transmission Electron Microscopy. In: Polychroniadis E, Oral A, Ozer M (eds) International Multidisciplinary Microscopy Congress. Springer Proceedings in Physics, Antalya (Turkey), October 2013. Springer, Cham, Berlim, 227 - 231. https://doi.org/10.1007/978-3-319-04639-6_32

Cavalcante DA, De-Souza MT, Orem JC, Magalhães MIA, Martins PH, Boone TJ, Castillo JA, Driks A (2019) Ultrastructural analysis of spores from diverse *Bacillales* species isolated from Brazilian soil. Environ Microbiol Rep 11(2):155-164. https://doi.org/10.1111/1758-2229.12713

Chen ML, Tsen HY. 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(5): 912-919. https://doi.org/10.1046/j.1365-2672.2002.01606.x

Claus D, and Berkeley RCW (1986) Genus *Bacillus* Cohn, 1872. In: Sneath, P.H.A., Mair, N.S., Sharpe, M.E. and Holt. J.G., Eds., Bergey's Manual of Systematic Bacteriology, The Williams & Wilkins Co., Baltimore, 2, 1105-1139.

Colwell RR (1970) Polyphasic taxonomy of the genus Vibrio: numerical taxonomy of Vibrio cholerae, Vibrio parahaemolyticus and related Vibrio species. J Bacteriol, 104 (1): 410–433. https://doi.org/10.1128/jb.104.1.410-433.1970

Das S, Dash HR, Mangwani N, Chakraborty J, Kumari S (2014) Understanding molecular identification and polyphasic taxonomic approaches for genetic relatedness and phylogenetic relationships of microorganisms. J Microbiol Methods 103: 80-100. https://doi.org/10.1016/j.mimet.2014.05.013

De Maagd, RA, Bravo A, Berry C, Crickmore N, Schnepf HE (2003) Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annual Review of Genetics 37: 409–433. https://doi.org/10.1146/annurev.genet.37.110801.143042

De Vos P, Ludwig W, Schleifer KH, Whitman W (2009) Family IV. *Paenibacillaceae* fam. nov. In: De Vos P, Garrity G.M, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman, WB (Eds) Bergey's Manual of Systematic Bacteriology. Springer-Verlag, New York, 269–327.

De Vos P (2011) Studying the Bacterial Diversity of the Soil by Culture-Independent Approaches. In: Logan NA, De Vos P (Eds) Endospore-forming Soil Bacteria. Springer-Verlag, Berlin-Heidelberg, 61-72.

Douterelo I, Sharpe R, Boxall J (2014) Bacterial community dynamics during the early stages of biofilm formation in a chlorinated experimental drinking water distribution system: implications for drinking water discolouration. J Appl Microbiol 117(1): 286-301. https://doi.org/10.1111/jam.12516

Driks A, Eichenberger P (2016) The spore coat. In: Driks A, Eichenberger P (Eds) The Bacterial Spore: From molecules to systems. ASM Press, Washington, 179– 200.

Dromigny E, Vincent P, Jouve JL (1994) *Bacillus* cereus. In: Bourgeois CM, Mescle JF, Zucca J (Eds) Microbiología alimentaria. Zaragoza, Acribia, 107-111.

Ehling-Schulz M, Messelhäusser U (2013) *Bacillus* "next generation" diagnostics: moving from detection toward subtyping and risk-related strain profiling. Frontiers in Microbiology 4 (32): 1-8. https://doi.org/10.3389/fmicb.2013.00032

Fan B, Blom J, Klenk H-P, Borriss R (2017) *Bacillus* amyloliquefaciens, *Bacillus* velezensis, and *Bacillus* siamensis Form an "Operational Group B. amyloliquefaciens within the *B. subtilis* Species Complex. Frontiers in Microbiology 8 (22): 1-15. https://doi.org/10.3389/fmicb.2017.00022

Fritze D (2002) *Bacillus* Identification-Traditional Approaches. In: Berkeley R, Heyndrickx M, Logan N, De Vos P (Eds) Applications and Systematics of *Bacillus* and Relatives. Blackwell Publishing, Cambridge, 100-123.

Fritze D (2004) Taxonomy of the genus *Bacillus* and related genera: The aerobic endospore-forming bacteria. Phytopathology 94(11): 1245-1248. https://doi.org/10.1094/PHYTO.2004.94.11.1245

Galperin MY (2013) Genome diversity of Spore-Forming *Firmicutes*. Microbiology Spectrum 1(2): TBS 0015-2012 https://doi.org/10.1128/microbiolspectrum.TBS-0015-2012

Gordon RE, Haynes WC, Pang CHN (1973) The genus *Bacillus*. U.S. Department of Agriculture, Washington DC, 283 pp.

Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32(18): 2847-2849. https://doi.org/10.1093/bioinformatics/btw313

Hakorvita JR, Prezioso S, Hodge D, Pillai SP, Weigel LM (2016) Identification and analysis of informative single nucleotide polymorphisms in 16S rRNA gene sequences of the *Bacillus* cereus. J Clin Microbiology 54(11): 2749-2756. https://doi.org/10.1128/JCM.01267-16

Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolstø AB (2000) *Bacillus* anthracis, *Bacillus* cereus, and *Bacillus* thuringiensis – One species on the basis of genetic evidence. Appl Environ Microbiology 66(6):2627-2630. https://doi.org/10.1128/AEM.66.6.2627-2630.2000

Henriksen SD (1976) Moraxella, Neisseria, Branhamella, and Acinetobacter. Annual Review of Microbiology 30: 63–83. https://doi.org/10.1146/annurev.mi.30.100176.000431

Hummel M, Edelmann D, Kopp-Schneider A (2017) Clustering of samples and variables with mixed-type data. PLo S One 12(11): e0188274. https://doi.org/10.1371/journal.pone.0188274

Hussain T, Roohi A, Munir S, Ahmed I, Khan J, Edel-Hermann V, Kim KY, Anees M (2013) Biochemical characterization and identification of bacterial strains isolated from drinking water sources of Kohat, Pakistan. African Journal of Microbiology Research 7(16):1579-1590. https://dx.doi.org/10.5897/AJMR12.2204

Jeyaram K, Romi W, Singh TA, Adewumi GA, Basanti K, Oguntoyinbo FA (2011) Distinct differentiation of closely related species of *Bacillus* subtilis group with industrial

780 importance. J Microbiol Methods 87(2): 161-164.
781 https://doi.org/10.1016/j.mimet.2011.08.011
782
783 Liu Y, Lai Q, Dong C, Sun F, Wang L, Li G, Shao Z (2013) Phylogenetic diversity of
784 the *Bacillus* pumilus group and the marine ecotype revealed by Multilocus Sequence
785 Analysis. PLo S One 8(11): e80097. https://doi.org/10.1371/journal.pone.0080097
786
787 Logan NA, Berkeley RCW (1984) Identification of *Bacillus* strains using the API
788 system. J Gen Microbiol 130(7): 1871-82. https://doi.org/10.1099/00221287-130-7-
789 1871
790
791 Logan NA, Berge O, Bishop AH, Busse HJ, De Vos P, Fritze D, Heyndrickx M,
792 K€ampfer P, Salkinoja-Salonen MS, Seldin L, Rabinovitch L, Ventosa A (2009)
793 Proposed minimal standards for describing new taxa of aerobic, endospore-forming
794 bacteria. Int J Syst Evol Microbiology 59: 2114 – 2121.
795 https://doi.org/10.1099/ijs.0.013649-0
796
797 Logan NA, De Vos, P (2009a) Genus I. *Bacillus*, Cohn 1872. In: De Vos P, Garrity
798 GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman, WB (Eds)
799 Bergey's Manual of Systematic Bacteriology. Springer-Verlag, New York, 21–127.
800
801 Logan NA, De Vos P (2009b) Genus IV. *Brevibacillus*. In: De Vos P, Garrity GM,
802 Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman, WB (Eds) Bergey's
803 Manual of Systematic Bacteriology. Springer-Verlag, New York, 306–316.
804
805 Logan NA, Halket G (2011) Developments in the taxonomy of aerobic, endospore
806 forming Bacteria. In: Logan NA, De Vos P (Eds) Endospore-forming Soil Bacteria.
807 Springer-Verlag, Berlin-Heidelberg, 1-30.
808
809 Mandic-Mulec I, Prosser JI (2011) Diversity of Endospore-forming Bacteria in Soil:
810 Characterization and Driving Mechanisms. In: Logan NA, De Vos P (Eds) Endospore-
811 forming Soil Bacteria. Springer-Verlag, Berlin-Heidelberg, 31-59.
812
813 Martins PHR, da Silva LP, Orem JC, de Magalhães MIA, Cavalcante DA, De-Souza
814 MT (2020) Protein profiling as a tool for identifying environmental aerobic endospore-
815 forming bacteria. Open J Bac 4(1): 001-007. https://dx.doi.org/10.17352/ojb.000012
816
817 Maughan H, Van der Auwera G (2011) *Bacillus* taxonomy in the genomic era finds
818 phenotypes to be essential though often mis*l*eading. Infect Genet Evol 11(5): 789 -797.
819 https://doi.org/10.1016/j.meegid.2011.02.001
820
821 Mizukami M, Hanagata H, Miyauchi A (2010) *Brevibacillus* expression system: host-
822 vector system for efficient production of secretory proteins. Curr Pharm Biotechnology
823 11(3):251-258. http://dx.doi.org/10.2174/138920110791112031
824
825 Moir A, Cooper G (2014) Spore germination. In: Driks A, Eichenberger P (Eds) The
826 Bacterial Spore: From molecules to systems. ASM Press, Washington, 217– 236.
827
828 Muyzer G, Stams AJM (2008) The ecology and biotechnology of sulphate-reducing
829 bacteria. Nat Rev Microbiol 6(6):441-454. https://doi.org/10.1038/nrmicro1892

Orem JC, Silva WMC, Raiol T, Magalhaes MIA, Martins PHR, Cavalcante DA, Kruger RH, Brigido MM, De-Souza MT (2019) Phylogenetic diversity of aerobic Spore-Forming Bacillalles isolated from Brazilian soils. International Microbiology 22: 511 – 520. https://doi.org/10.1007/s10123-019-00080-6

Oliveira EJ, Silva SEA, Rabinovitch L (1998) A standardized Protocol for the Rapid Detection of Gelatin Hydrolysis by *Bacillus* sphaericus. Israel Journal of Entomology 23:141-146. http://www.entomology.org.il/sites/default/files/pdfs/IJE-1998a.Oliveira.pdf

Palmisano MM, Nakamura LK, Duncan KE, Istock CA, Cohan FM (2001) *Bacillus* sonorensis sp. nov., a close relative of *Bacillus* licheniformis, isolated from soil in the Sonoran Desert, Arizona. Int J Syst Evol Microbiol 51(5):1671-1679. https://doi.org/10.1099/00207713-51-5-1671

Panda AK, Bisht SS, DeMondal SN, Kumar S, Gurusubramanian G, Panigrahi AK (2014) *Brevibacillus* as a biological tool: a short review. Antonie Van Leeuwenhoek 105(4):623-639. https://doi.org/10.1007/s10482-013-0099-7

Parte AC (2018) LPSN – List of Prokaryotic names with standing in nomenclature (bacterio.net), 20 years on. International Journal of Systematic and Evolutionary Microbiology 68(6): 1825-1829. https://doi.org/10.1099/ijsem.0.002786

Powell DA, Marcon MJ (2012) Acinetobacter Species. In: Long SS, Pickering LK, Prober CG (Eds) Principles and Practice of Pediatric Infectious Diseases. Elsevier, Amsterdam, 828 – 830.

Prakash O, Verma M, Sharma P, Kumar M, Kumari K, Singh A, Kumari H, Jit S, Gupta, SK, Khanna M, Lal R (2007) Polyphasic approach of bacterial classifification – An overview of recent advances. Indian J Microbiol 47(2):98–108. https://doi.org/10.1007/s12088-007-0022-x

Priest FG. Genus I. *Paenibacillus* Ash, Priest and Collins (1994) (Effective publication: Ash, Priest and Collins 1993, 259) emend. Shida, Takagi, Kadowaki, Nakamura and Komagata 1997a, 297. In De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey KH, Schleifer FA, & Williams BL (Eds). Bergey's Manual of Systematic Bacteriology, New York: Springer, 269-295.

Rabinovitch L, Oliveira EJ (2015) Coletânea de procedimentos técnicos e metodologias empregadas para o estudo de *Bacillus* e gêneros esporulados aeróbios correlatos. Montenegro Comunicação, Rio de Janeiro, 160 pp.

Ruiu L, Floris I, Satta A (2013) Emerging entomopathogenic bacteria for insect pest management. Bulletin of Insectology 66(2):181-186. https://www.cabdirect.org/cabdirect/abstract/20133416548

Salgado JRS, Rabinovitch L, Gomes MFS, Allil RCSB, Werneck MM, Rodrigues RB, Picão RC, Oliveira Luiz FB, Vivoni AM (2020) Detection of *Bacillus anthracis* and

Bacillus anthracis-like spores in soil from state of Rio de Janeiro, Brazil. Memórias do Instituto Oswaldo Cruz 115:1-10. https://doi.org/10.1590/0074-02760200370

Seldin L, Elsas JD, Penido EGC (1984) *Bacillus azotofixans* sp. nov. a Nitrogen-Fixing Species from Brazilian Soils and Grass Roots. Int J Syst Bacteriol, 34:451-456. https://doi.org/10.1007/BF02374784

Shivaji S, Chaturvedi P, Begum Z, Pindi PK, Manorama R, Padmanaban DA, Shouche YS, Pawar S, Vaishampayan P, Dutt CBS, Datta GN, Manchanda RK, Rao UR, Bhargava PM, Narlikar JV. (2009) Janibacter hoylei sp. nov., *Bacillus* isronensis sp. nov. and *Bacillus* aryabhattai sp. nov., isolated from cryotubes used for collecting air from the upper atmosphere. Int J Syst Evol Microbiol 59(12): 2977-2986. https://doi.org/10.1099/ijs.0.002527-0

Setlow P (2014) Spore Resistance Properties. Microbiology Spectrum 2(5): 2 – 5. https://doi.org/10.1128/microbiolspec.TBS-0003-2012

Shida O, Takagi H, Kadowaki K, Komagata K (1996) Proposal for two new genera, *Brevibacillus* gen. nov. and Aneurini*Bacillus* gen. nov. Int J Syst Bacteriol 46(4):939-946. https://doi.org/10.1099/00207713-46-4-939

Shida O, Takagi H, Kadowaki K, Nakamura LK, Komagata K (1997) Transfer of *Bacillus alginolyticus*, *Bacillus* chondroitinus, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus* kobensis, and *Bacillus thiaminolyticus* to the genus *Paenibacillus* and emended description of the genus *Paenibacillus*. Int J Syst Bacteriol 47(2): 289-298. https://doi.org/10.1099/00207713-47-2-289

Smith NR, Gordon RE, Clark FE (1952) Aerobic Sporeforming Bacteria. U.S. Department of Agriculture, Washington DC, 148 pp.

Stackebrandt E, Goebel BM (1994) Taxonomic Note: A Place for DNA-DNA Reassociation and 16s rRNA Sequence Analysis in the Present Species Definition in Bacteriology. International Union of Microbiological Societies 44(4):846-849. https://doi.org/10.1099/00207713-44-4-846

Stackebrandt E, Swiderski, J (2002) Chapter 2: From Phylogeny to Systematics: the dissection of the genus *Bacillus*. In: Berkley R, Heyndrickx M, Logan N, De Vos P (Eds) Applications and systems of *Bacillus* and relatives. Blackwell Publishing, Maldon-Oxford-Carlton-Berlim, 8–22.

Tamames J, Abellán JJ, Pignatelli M, Camacho A, Moya A (2010) Environmental distribution of prokaryotic taxa. BMC Microbiology 10(1): 1-14. http://www.biomedcentral.com/1471-2180/10/85

Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 60(1):249-266. https://doi.org/10.1099/ijs.0.016949-0

Tringe SG, Hugenholtz P (2008) A renaissance for the pioneering 16S rRNA gene. Curr Opin Microbiol 11(5):442-446. https://doi.org/10.1016/j.mib.2008.09.011

Vila J, Abdalla S, Gonzalez J, Garcia C, Bombi JA, Jimenez de Anta MT (1992) A one-minute oxidase test to detect Vibrio strains isolated from cultures on thiosulphate-citrate-bile salts-sucrose (TCBS) medium. J Appl Bacteriology 72(6):490-492. https://doi.org/10.1111/j.1365-2672.1992.tb01864.x

Wang LT, Lee FL, Tai CJ, Kuo HP (2008) *Bacillus* velezensis is a later heterotypic synonym of *Bacillus* amyloliquefaciens. Int J Syst Evol Microbiol 58(3):671-675. https://doi.org/10.1099/ijs.0.65191-0

Xiao Z, Ma C, Xu P, Lu JR (2009) Acetoin catabolism and acetylbutanediol formation by *Bacillus* pumilus in a chemically defined medium. PLo S One 4(5): e5627. https://doi.org/10.1371/journal.pone.0005627

Yoon J, Seo W, Shin YK, Kho YH, Kang KH, Park Y (2002) *Paenibacillus* chinjuensis sp. nov., a novel exopolysaccharide-producing bacterium. Int J Syst Evol Microbiol 52(2):415-421. https://doi.org/10.1099/00207713-52-2-415

Zhao S, Guo Y, Sheng Q, Shyr Y (2014) Advanced Heat Map and Clustering Analysis Using Heatmap3. BioMed Research International 2014:1-6. https://doi.org/10.1155/2014/986048

Tables and figures captions

Table 1. Biochemical and physiological tests used in this work and the respective controls

Table S1. Molecular, biochemical, and physiological profiles of SDF strains belonging to the AEFBC

Figure 1. Overall repartition of SDF strains according 16S rRNA gene sequencing classification. **(A)** Distribution of 238 SDF strains among six genera belonging to families *Bacillaceae* (*Bacillus*, *Lysinibacillus*, *Terribacillus*, and *Rummeliibacillus*) and *Paenibacillaceae* (*Paenibacillus* and *Brevibacillus*). (**B)** Species assignments of 224 SDF strains.

Figure 2. Correlation between SDF strains belonging to *B. cereus* group and growth conditions or enzymes activities. A Person correlation-based clustering method was employed to construct a heat map associating 48 SDF strains allocated in *B. cereus* group (right) and 30 phenotypical features (bottom) that contribute to AEFB identification and classification. The top dendrogram clustered the SDF strains into two parts based on the prevalence of positive responses (blue) to 30 growth conditions and enzyme reactions described at the bottom of the graphic. Negative responses are shown in red.

Figure 3. Correlation between SDF strains belonging to *B. subtilis* complex and growth conditions or enzymes activities. A Person correlation-based clustering method was employed to construct a heat map associating 95 SDF strains allocated in *B. subtilis* complex (right) and 30 phenotypical features (bottom) that contribute to AEFB identification and classification. The top dendrogram clustered the SDF strains into two parts based on the prevalence of positive responses (green) to 30 growth conditions and enzyme reactions described at the bottom of the graphic. Negative responses are shown in red.

Figure 4. Correlation between SDF strains belonging to *family Paenibacillaceae* and growth conditions or enzymes activities. A Person correlation-based clustering method was employed to construct a heat map associating 18 SDF strains allocated in *B. subtilis* complex (right) and 30 phenotypical features (bottom) that contribute to AEFB identification and classification. The top dendrogram clustered the SDF strains into two parts based on the prevalence of positive responses (orange) to 30 growth conditions and enzyme reactions described at the bottom of the graphic. Negative responses are shown in red.

Table 1. Biochemical and physiological profiles analysed in this work and the respective controls

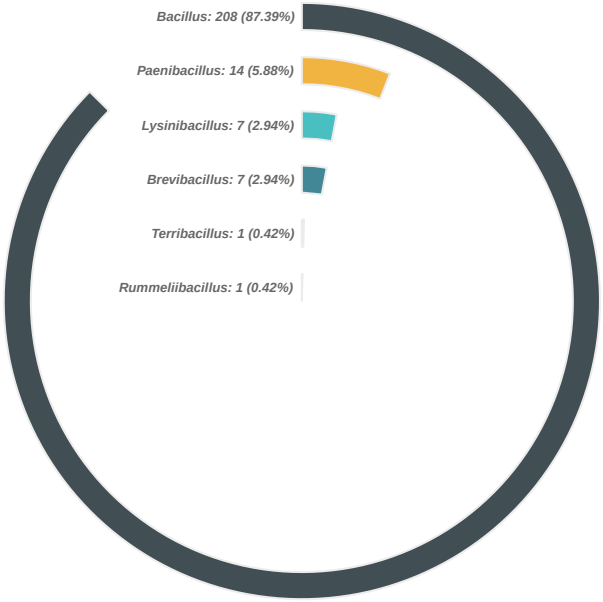
Test		Control	
		Positive	**Negative**
Growth condition	Citrate utilization	*Bacillus cereus CCGB406*	*Paenibacillus macerans CCGB126*
	Propionate utilization	*Bacillus licheniformis CCGB407*	*Bacillus subtilis CCGB1249*
	7% NaCl	*Bacillus amyloliquefaciens CCGB452*	*Paenibacillus macerans CCGB126*
	10% NaCl	*Bacillus amyloliquefaciens CCGB452*	*Paenibacillus macerans CCGB126*
	0.001% lysozyme	*Bacillus cereus CCGB406*	*Bacillus pumilus CCGB124*
	45 °C	*Geobacillus stearothermophilus CCGB412*	*ND**
	65 °C	*Geobacillus stearothermophilus CCGB412*	*Bacillus thuringiensis CCGB1163*
	pH 5.7	*Bacillus cereus CCGB406*	Paenibacillus alvei CCGB414
	Anaerobiosis	*Bacillus cereus CCG406*	*Bacillus megaterium CCGB408*
Enzyme	Catalase	*Bacillus cereus CCGB406*	*ND**
	Oxidase	*Lysinibacillus sphaericus CCGB745*	*Bacillus cereus CCGB406*
	Hemolysin	*Bacillus thuringiensis CCGB1163*	*Lysinibacillus sphaericus CCGB745*
	Nitrate reductatase	*Bacillus cereus CCGB406*	*Bacillus megaterium CCGB408*
Hydrolysis	Casein	*Bacillus megaterium CCGB408*	*Paenibacillus macerans CCGB126*
	Gelatin	*Bacillus cereus CCGB406*	*Geobacillus stearothermophilus CCGB412*
	Esculin	*Bacillus subtilis CCGB1249*	*Lysinibacillus fusiformis CCGB743*
	Starch	*Bacillus cereus CCGB406*	*Lysinibacillus sphaericus CCGB745*
Amino acid decomposition	Phenylalanine degradation	*Bacillus megaterium CCGB408*	*Bacillus cereus CCGB406*
	Tyrosine degradation	*Bacillus cereus CCGB406*	*L. sphaericus CCGB745*
	Arginine dihydrolase	*Bacillus licheniformis CCGB407*	*Bacillus megaterium CCGB408*
	Lysine decarboxylase	*Bacillus thuringiensis CCGB1163*	*Bacillus megaterium CCGB408*
	Ornithine decarboxylase	*Bacillus thuringiensis CCGB1163*	*Bacillus megaterium CCGB408*
Indole production		*Paenibacillus alvei CCGB414*	*Bacillus cereus CCGB406*
Production of acid from	D-Glucose	*Bacillus megaterium CCGB408*	*Lysinibacillus fusiformis CCGB743*
	L-Arabinose	*Bacillus megaterium CCGB408*	*Brevibacillus brevis CCGB052*
	Lactose	*Bacillus megaterium CCGB408*	*Lysinibacillus fusiformis CCGB743*
	Mannitol	*Bacillus megaterium CCGB408*	*Lysinibacillus fusiformis CCGB743*
	Sucrose	*Bacillus amyloliquefaciens CCGB452*	*Lysinibacillus sphaericus CCGB745*
	D-Xylose	*Bacillus megaterium CCGB408*	*Brevibacillus brevis CCGB052*
Voges-Proskauer test		*Bacillus cereus CCGB406*	*Bacillus megaterium CCGB408*

*not determined**.** **CCGB**: *Coleção de Culturas do Gênero Bacillus e Gêneros Correlatos*. CCGB is an integrant of the World Federation for Culture Collec6ons WFCC (#574).

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 18/05/2022* DOI: https://doi.org/10.3897/arphapreprints.e86638

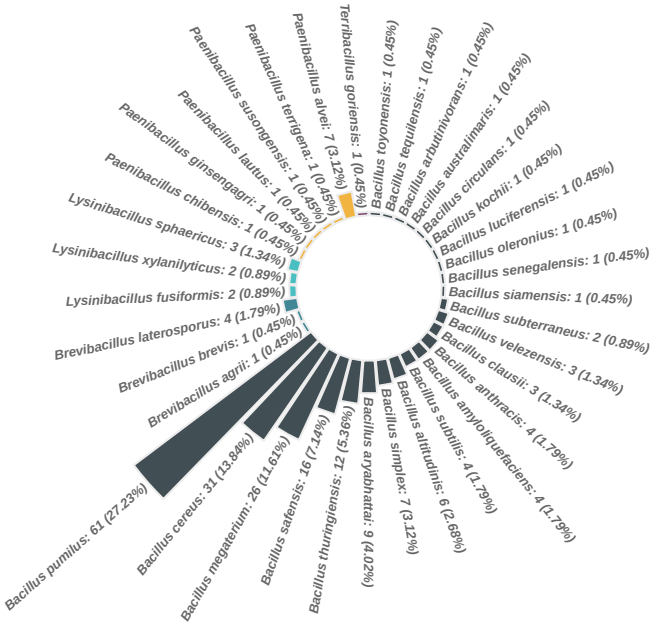
A

Genera level (238 SDF strains)

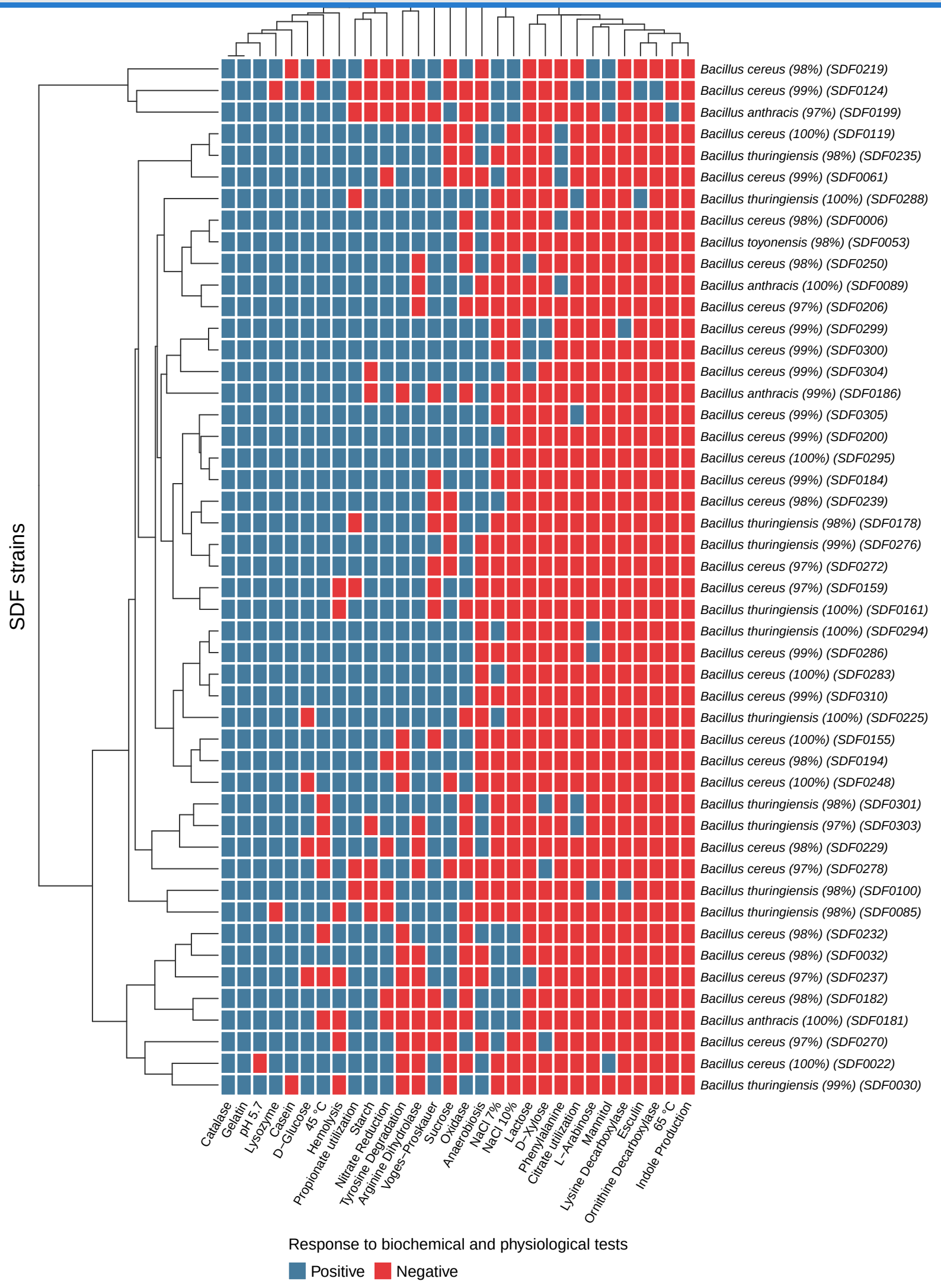


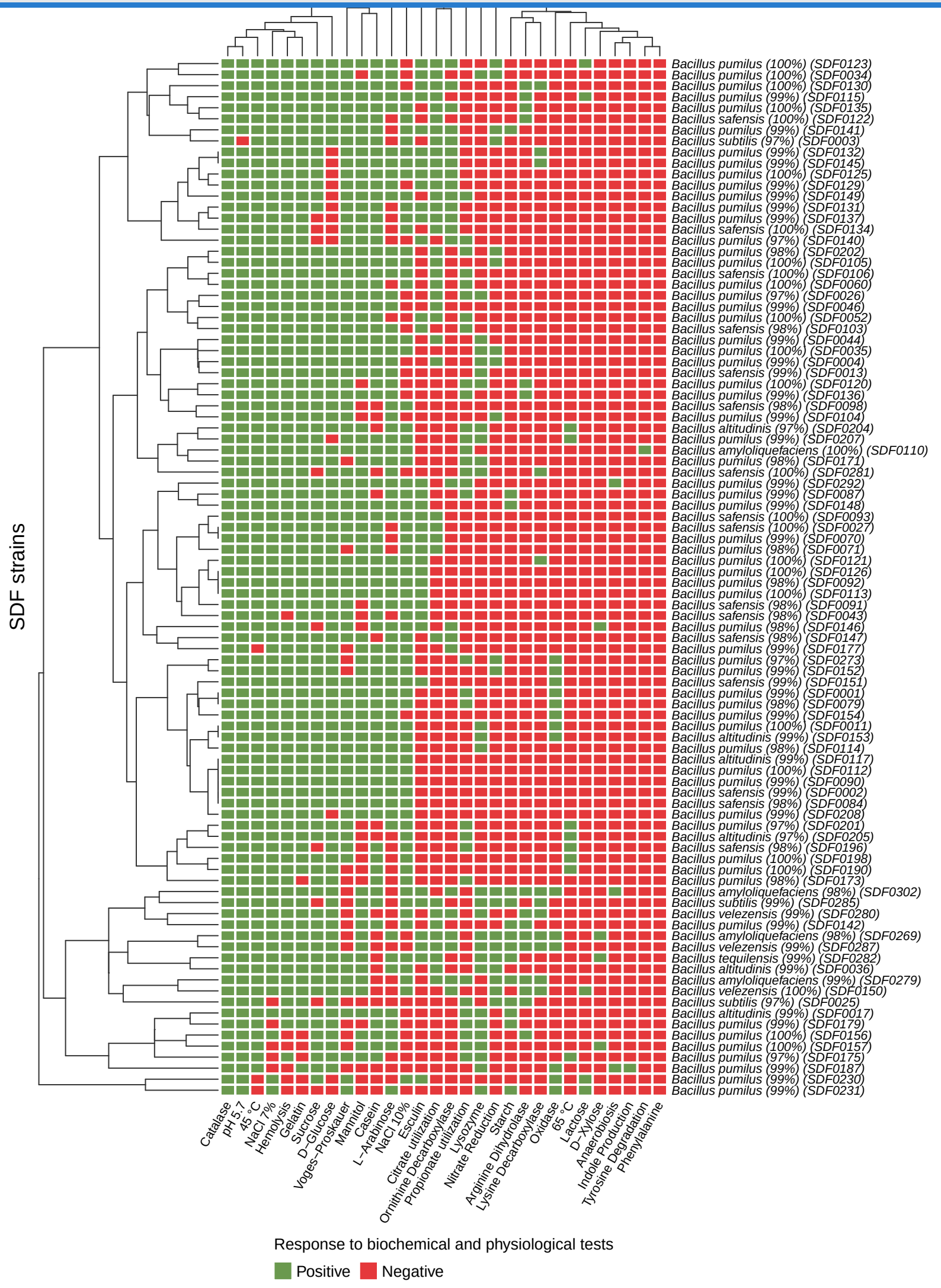
B

Species level (224 SDF strains)



ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 18/05/2022* DOI: https://doi.org/10.3897/arphapreprints.e86638




SDF strains
Bacillus pumilus (100%) (SDF0123)
Bacillus pumilus (100%) (SDF0034)
Bacillus pumilus (100%) (SDF0130)
Bacillus pumilus (99%) (SDF0115)
Bacillus pumilus (100%) (SDF0135)
Bacillus safensis (100%) (SDF0122)
Bacillus pumilus (99%) (SDF0141)
Bacillus subtilis (97%) (SDF0003)
Bacillus pumilus (99%) (SDF0132)
Bacillus pumilus (99%) (SDF0145)
Bacillus pumilus (100%) (SDF0125)
Bacillus pumilus (99%) (SDF0129)
Bacillus pumilus (99%) (SDF0149)
Bacillus pumilus (99%) (SDF0131)
Bacillus pumilus (99%) (SDF0137)
Bacillus safensis (100%) (SDF0134)
Bacillus pumilus (97%) (SDF0140)
Bacillus pumilus (98%) (SDF0202)
Bacillus pumilus (100%) (SDF0105)
Bacillus safensis (100%) (SDF0106)
Bacillus pumilus (100%) (SDF0060)
Bacillus pumilus (97%) (SDF0026)
Bacillus pumilus (99%) (SDF0046)
Bacillus pumilus (100%) (SDF0052)
Bacillus safensis (98%) (SDF0103)
Bacillus pumilus (99%) (SDF0044)
Bacillus pumilus (100%) (SDF0035)
Bacillus pumilus (99%) (SDF0004)
Bacillus safensis (99%) (SDF0013)
Bacillus pumilus (100%) (SDF0120)
Bacillus pumilus (99%) (SDF0136)
Bacillus safensis (98%) (SDF0098)
Bacillus pumilus (99%) (SDF0104)
Bacillus altitudinis (97%) (SDF0204)
Bacillus pumilus (99%) (SDF0207)
Bacillus amyloliquefaciens (100%) (SDF0110)
Bacillus pumilus (98%) (SDF0171)
Bacillus safensis (100%) (SDF0281)
Bacillus pumilus (99%) (SDF0292)
Bacillus pumilus (99%) (SDF0087)
Bacillus pumilus (99%) (SDF0148)
Bacillus safensis (100%) (SDF0093)
Bacillus safensis (100%) (SDF0027)
Bacillus pumilus (99%) (SDF0070)
Bacillus pumilus (98%) (SDF0071)
Bacillus pumilus (100%) (SDF0121)
Bacillus pumilus (100%) (SDF0126)
Bacillus pumilus (98%) (SDF0092)
Bacillus pumilus (100%) (SDF0113)
Bacillus safensis (98%) (SDF0091)
Bacillus safensis (98%) (SDF0043)
Bacillus pumilus (98%) (SDF0146)
Bacillus safensis (98%) (SDF0147)
Bacillus pumilus (99%) (SDF0177)
Bacillus pumilus (97%) (SDF0273)
Bacillus pumilus (99%) (SDF0152)
Bacillus safensis (99%) (SDF0151)
Bacillus pumilus (99%) (SDF0001)
Bacillus pumilus (98%) (SDF0079)
Bacillus pumilus (99%) (SDF0154)
Bacillus pumilus (100%) (SDF0011)
Bacillus altitudinis (99%) (SDF0153)
Bacillus pumilus (98%) (SDF0114)
Bacillus altitudinis (99%) (SDF0117)
Bacillus pumilus (100%) (SDF0112)
Bacillus pumilus (99%) (SDF0090)
Bacillus safensis (99%) (SDF0002)
Bacillus safensis (98%) (SDF0084)
Bacillus pumilus (99%) (SDF0208)
Bacillus pumilus (97%) (SDF0201)
Bacillus altitudinis (97%) (SDF0205)
Bacillus safensis (98%) (SDF0196)
Bacillus pumilus (100%) (SDF0198)
Bacillus pumilus (100%) (SDF0190)
Bacillus pumilus (98%) (SDF0173)
Bacillus amyloliquefaciens (98%) (SDF0302)
Bacillus subtilis (99%) (SDF0285)
Bacillus velezensis (99%) (SDF0280)
Bacillus pumilus (99%) (SDF0142)
Bacillus amyloliquefaciens (98%) (SDF0269)
Bacillus velezensis (99%) (SDF0287)
Bacillus tequilensis (99%) (SDF0282)
Bacillus altitudinis (99%) (SDF0036)
Bacillus amyloliquefaciens (99%) (SDF0279)
Bacillus velezensis (100%) (SDF0150)
Bacillus subtilis (97%) (SDF0025)
Bacillus altitudinis (99%) (SDF0017)
Bacillus pumilus (99%) (SDF0179)
Bacillus pumilus (100%) (SDF0156)
Bacillus pumilus (100%) (SDF0157)
Bacillus pumilus (97%) (SDF0175)
Bacillus pumilus (99%) (SDF0187)
Bacillus pumilus (99%) (SDF0230)
Bacillus pumilus (99%) (SDF0231)
Catalase
pH 5.7
45 °C
NaCl 7%
Hemolysis
Gelatin
Sucrose
D–Glucose
Voges–Proskauer
Mannitol
Casein
L–Arabinose
NaCl 10%
Esculin
Citrate utilization
Ornithine Decarboxylase
Propionate utilization
Lysozyme
Nitrate Reduction
Starch
Arginine Dihydrolase
Lysine Decarboxylase
Oxidase
65 °C
Lactose
D–Xylose
Anaerobiosis
Indole Production
Tyrosine Degradation
Phenylalanine
Response to biochemical and physiological tests
Positive
Negative

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 18/05/2022* DOI: https://doi.org/10.3897/arphapreprints.e86638