Peertechz

LIFE SCIENCES GROUP

OPEN JOURNAL OF Bacteriology

ISSN: 2640-8007 DOI: https://dx.doi.org/10.17352/ojb

Research Article

Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria

Paulo Henrique R Martins[1], Luciano Paulino da Silva[2], Juliana Capella de Orem[1], Maria Inês A de Magalhães[1], Danilo de Andrade Cavalcante[1] and Marlene Teixeira De-Souza[1]*

[1]Department of Cellular Biology, University of Brasilia, Darcy Ribeiro University Campus, Brasilia, DF, Brazil

[2]Embrapa Genetic Resources and Biotechnology, Caixa Postal 02372, Brasilia, DF, Brazil

Received: 11 January, 2020
Accepted: 11 March, 2020
Published: 12 March, 2020

***Corresponding author:** Marlene Teixeira De-Souza, Department of Cellular Biology, University of Brasilia, Darcy Ribeiro University Campus, 70.910-900 Brasilia, DF, Brazil, Tel: +55 61 3107-3044;
E-mail: marlts@unb.br

ORCiD: https://orcid.org/0000-0003-1538-2657

Keywords: *Bacillales*; Taxonomy; Phenotype; Genotype; MALDI-TOF-MS

https://www.peertechz.com

Abstract

Aerobic Endospore-Forming Bacteria (AEFB) are taxonomically and physiologically diverse, comprising species of genus *Bacillus* and related genera of industrial and medical importance. For taxonomic purpose, we applied the matrix-assisted laser desorption/ionization mass spectrometry with time-of-flight to identify 64 environmental AEFB (SDF for *Solo do Distrito Federal*) and compare the results with those obtained using 16S rRNA gene sequencing. Concordance between the two methods was observed for 93,75% samples at the genus level. Strains were clustered between 2 genera (family *Bacillaceae)*: *Bacillus*, the most prevalent, and Lysinibacillus. Two other genera, *Brevibacillus* and *Paenibacillus* (family *Paenibacillaceae*) were also distinguished. Gene similarity discriminated an additional genus (*Rummeliibacillus*). At the species level, the genotyping method achieved superior capacity identifying 93,75% strains. Among 31 strains identified at the species level by protein profiling, 61.29% coincided and both, protein and gene profiling, placed other 32.25% strains within groups of closely related species of *Bacillus* bearing two or even more species alternatives within the same affiliation cluster. These results suggested the applicability of the score and sequence similarity ranges in a complementary way for initial identification and clustering of closely related samples inside these 64 SDF strains. Our assignments are useful because they clearly identify the genera and restrict the identity of a strain to one or two possible species in the genera, thus clarifying their genetic interrelationships. This study also stresses that combining phenotypic and genotypic methods into polyphasic approaches is essential for a robust assignment of the remarkable genetic and ecological diversity of AEFB.

Introduction

Species of the genus *Bacillus* and related genera are collectively designated Aerobic Endospore-Forming Bacteria (AEFB). Inside the phylum *Firmicutes*, these species are allocated in the class *Bacilli*, order *Bacillales* which contains seven out of ten families harbouring endospore-formers: *Alicyclobacillaceae*, *Bacillaceae*, *Paenibacillaceae*, *Pasteuriaceae*, *Planococcaceae*, *Sporolactobacillaceae* and *Thermoactinomycetacea* [1,2]. AEFB are widely distributed in nature, including extreme environments, and the soil is considered their main repository [3]. *Bacillus anthracis* and *B. cereus* are known for infecting humans. To highlight the ecological and economic importance of some AEFB strains we can mention a wide range of properties, including nitrogen fixation; plant growth promotion; activity toward insects, nematodes, and fungi; soil phosphorus solubilisation; production of exopolysaccharides, high diversity of hydrolytic enzymes, antibiotics, cytokinins, among other bioproducts [1,2].

AEFB present a high level of genetic and physiological diversity which render the demarcation of genus and species borders very complex [1,2,4,5]. Currently, 16S rRNA gene sequences are used to assign taxa in a phylogenetic tree and

Citation: R Martins PH, da Silva LP, de Orem JC, de Magalhães MIA, De-Souza MT, et al. (2020) Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. Open J Bac 4(1): 001-007. DOI. https://dx.doi.org/10.17352/ojb.000012

draw the largest frontiers in the prokaryotic classification system [5]. However, phenotype can influence the depth of a hierarchical line consistency and is necessary to generate useful characterization [6,7].

Among the phenotype-based methods for the identification of microorganisms, the use of the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has dramatically increased [8,9]. Analyses by MALDI-TOF MS do not require lengthy biochemical reactions and are faster than other conventional phenotypic identification methods, presenting similar or even superior reliability [8]. Besides, the toleration of varying growth conditions and the high reproducibility of this technique resulted in the elaboration of standard protocols [10,11]. Indeed, clinical laboratories have been successfully using MALDI-TOF MS to identify microorganisms at the species level, allowing that most of the clinically relevant pathogens to be rapidly included in the spectra database [12-14]. The efficacy of method relies on the stability of mass spectral patterns generated, since some cell components, routinely used on the analyses, are ubiquitous, highly conserved, integral, and abundant in living cells [15,16]. Mass spectra resulting from whole cells, or protein extracts, are compared to reference spectra available in commercial databases, based, in particular, on clinical strains. The more similar the mass spectral patterns are, the closer to the phylogenetic relationships. Given the predominance of ribosomal and regulatory proteins, besides clinical diagnoses, these biomarkers are also useful for taxonomic studies of bacteria.

Using MALDI-TOF MS we generated spectra from 64 environmental AEFB samples isolated from Brazilian soils and quoted as SDF (*Solo do Distrito Federal*) strains [17]. The predictive molecular relationship of protein profiling obtained for these environmental AEFB was further compared with classification based on the reference-method for taxonomic assignment of prokaryotes, the 16S rRNA gene sequencing.

Material and methods

SDF strains. Soil sampling and SDF strains isolation were described in Cavalcante, *et al.*, [17]. The 64 SDF strains used in this work (Tables 1-5) were randomly selected among SDF0001 to SDF0154, deposited at the *Coleção de Bactérias aeróbias formadoras de endósporos* (CBafes, or AEFB Collection), hosted at the University of Brasilia, Brazil.

Ethics statement. Specific permissions required for

Table 1: Genus-level concordant identifications between MALDI-TOF MS and 16S rRNA gene sequencing analyses when challenged against 64 SDF strains against 64 SDF strains.

SDF strain	Bruker Biotyper MALDI TOF MS		16S rRNA sequencing	
	Species (best match)	Score value	Species (% of similarity)	Acc. N⁰*
0016	*Bacillus* sp.	1.868	*Bacillus simplex* (98)	MH356301
0043	*Bacillus pumilus*	1.911	*Bacillus safensis* (98)	MH356317
0051	*Bacillus megaterium*	1.982	*Bacillus megaterium* (99)	MH356323
0053	*Bacillus cereus*	1.915	*Bacillus toyonensis* (98)	MH356325
0061	*Bacillus cereus*	1.913	*Bacillus cereus* (99)	MH356331
0068	*Bacillus simplex*	1.960	*Bacillus simplex* (100)	MH356337
0080	*Paenibacillus alvei*	1.857	*Paenibacillus alvei* (99)	MH356346
0099	*Bacillus megaterium*	1.944	*Bacillus megaterium* (97)	MH356360
0103	*Bacillus pumilus*	1.886	*Bacillus safensis* (98)	MH356362
0112	*Bacillus pumilus*	1.729	*Bacillus pumilus* (100)	MH356367
0113	*Bacillus pumilus*	1.830	*Bacillus pumilus* (100)	MH569356
0114	*Bacillus altitudinis*	1.762	*Bacillus pumilus* (98)	MH569357
0115	*Bacillus pumilus*	1.945	*Bacillus pumilus* (99)	MH356368
0118	*Bacillus altitudinis*	1.845	*Bacillus megaterium* (99)	MH356370
0120	*Bacillus altitudinis*	1.822	*Bacillus pumilus* (100)	MH356372
0121	*Bacillus safensis*	1.796	*Bacillus pumilus* (100)	MH356373
0123	*Bacillus pumilus*	1.744	*Bacillus pumilus* (100)	MH356375
0126	*Bacillus pumilus*	1.927	*Bacillus pumilus* (100)	MH356377
0127	*Bacillus amyloliquefaciens*	1.719	*Bacillus siamensis* (98)	MH356378
0130	*Bacillus safensis*	1.836	*Bacillus pumilus* (100)	MH356380
0132	*Bacillus altitudinis*	1.742	*Bacillus pumilus* (99)	MH356382
0135	*Bacillus pumilus*	1.815	*Bacillus pumilus* (100)	MH356385
0136	*Bacillus pumilus*	1.771	*Bacillus pumilus* (99)	MH356386
0139	*Bacillus pumilus*	1.765	*Bacillus pumilus* (94)	MH569360
0145	*Bacillus pumilus*	1.799	*Bacillus pumilus* (99)	MH356393
0146	*Bacillus pumilus*	1.713	*Bacillus pumilus* (98)	MH356394
0147	*Bacillus altitudinis*	1.766	*Bacillus safensis* (98)	MH356395
0152	*Bacillus pumilus*	1.779	*Bacillus pumilus* (99)	MH356400
0153	*Bacillus pumilus*	1.993	*Bacillus altitudinis* (99)	MH569362
0154	*Bacillus pumilus*	1.851	*Bacillus pumilus* (99)	MH569363

Score value is the identification log score given by the standard Bruker interpretative criteria (unreliable identification: 0.000–1.699; genus level identification: 1.700–1.999; species identification: ≥2.000). Similarity of 95-96% and ≥97% were considered as the threshold values for identification at the genus and species levels, respectively. Strain SDF0139 (highlighted in grey) was considered unidentified by sequencing similarity. ACC. N⁰ : GenBank accession number.

Citation: R Martins PH, da Silva LP, de Orem JC, de Magalhães MIA, De-Souza MT, et al. (2020) Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. Open J Bac 4(1): 001-007. DOI. https://dx.doi.org/10.17352/ojb.000012

collection of 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.

MALDI-TOF MS. Using a 10µL plastic loop, cells from 4 single colonies per SDF strain, cultured in solid Luria-Bertani (28 °C/24-48h), were transferred to 4 microtubes containing 300µL of ultra-pure water (Mille-Q™), resulting in 4 different extractions and 4 different measurements for each strain. After vortex stirring, 900 µL of 100% ethanol was added, the suspension stirred again, and centrifuged at 12,000×*g* for 2min. Air-dried pellets were resuspended in 30µL of 70% formic acid and acetonitrile in the ratio of 1:1 (*v/v*). The final mixture was stirred and centrifuged at 12,000×*g* for 2min, and 1µL of the supernatants was transferred to a spot onto a 96-well stainless steel MALDI target plate. The matrix, prepared in an organic solvent mixture to a final concentration of 10mg mL.$^{-1}$ in a 50:40:10 acetonitrile:water:3% Trifluoroacetic Acid (TFA) ration solution, was overlaid and allowed to dry. Each sample was spotted 4times. The mass spectra for the SDF strains were acquired (MicroFlex mass spectrometer; Bruker Daltonics, Bremen, Germany) at Embrapa Mass Spectrometry Laboratory (Brasilia, DF, Brazil). The spectra were recorded in the linear positive mode at a laser frequency of 60Hz within a mass range from *m/z* 2,000 to 20,000. For each spectrum, 240 laser shots in 40-shot steps from different positions of the target spot were collected and analysed. Spectra were externally calibrated employing *Escherichia coli* rProteins (Bruker Daltonics, Bremen, Germany). SDF strains spectra were loaded with the MALDI Biotyper software (Bruker Daltonics, Bremen, Germany) and analysed using the standard pattern-match algorithm, which compared the spectrum acquired to all inputs present in the manufacturer library. The results of the pattern-matching process were expressed as log values ranging from 0 to 3.000, according to the manufacturer instructions. Scores of <1.700 are interpreted as unreliable identification and of ≥1.700-1.999 and ≥2.000-2.99 indicate identification at the genus and species levels, respectively.

Taxonomic assignments of SDF strains. DNA preparation, PCR amplification, sequencing, and sequence analyses were performed as described in Orem *et al.* (2019) [18]. Briefly, 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 bidirectionally sequenced by Sanger method and Phred scores of ≥20 used to assess quality of sequences. Taxonomic assignments of the sequences were performed using BLAST and Classifier. Both forward and reverse chromatograms of the sequenced 16S rDNA fragments were analysed by Chromas software (Technelysium Pty Ltd) to determine best quality regions. Consensus sequences (550-600 nucleotides) were created using BioEdit 7.2.6 software and deposited at NCBI (Tables 1-5 for accession numbers). Similarity of 95%-96% and ≥97% were considered as the threshold values for identification at the genus and species levels, respectively.

Results

Fresh cells from 4 single colonies per SDF strain were used to obtain 4 different protein extractions. Each of these biological replicates was spotted and analysed 4times. The spectra were acquired by MALDI-TOF mass spectrometer (Bruker Daltonics: MicroFlex model) and recorded in the linear positive mode at a laser frequency of 60Hz within a mass range from m/z 2,000 to 20,000. For each spectrum, 240 laser shots in 40-shot steps from different positions of the target spot were collected and analysed. External calibration employed *E. coli* rProteins (Bruker Daltonics, Bremen, Germany). The spectra of 64 SDF strains analysed with the FlexAnalysis 3.3 and MALDI-Biotyper 3.0 programs (Bruker Daltonics) were used to identify and classify these AEFB according to resulting mass spectra.

This analysis revealed that for 33(51.56%) strains the score ranged from 1.700 to 1.999, thus identifying these SDF strains at genus level (Tables 1,2). The remaining 31(48.43%) presented log score values >2.000, which indicates species-level identification (Tables 3-5). Overall, genus *Bacillus* was predominant comprising 60(93.75%) strains, while the other 4(6.25%) strains were distributed among 3 genera: *Lysinibacillus* (1), *Brevibacillus* (2), and *Paenibacillus* (1).

Amidst the 33(51.56%) strains identified at genus level by the MALDI Biotyper database (Tables 1,2), 16S rRNA sequence similarity-based analysis coincided in terms of identifying 30(90.90%) samples (Table 1). Considering the best match suggested by protein profiling, 15(50.00%) out of these 30 coincide with the *Bacillus* spp. discrimination obtained by the genotype method at species level, and 13(43.33%) belonged to the same close-related AEFB groups (Table 1).

Regarding the remaining 3(9.09%) strains, both identification methods yielded different genera (Table 2). MALDI-TOF MS-based analysis allocated SDF0063 and SDF0133 at the genera *Brevibacillus* and *Bacillus*, respectively. However, these two strains were classified at species level by 16S rRNA gene sequencing, with both presenting 99% of similarity with *Lysinibacillus xylanilyticus* and *Paenibacillus alvei*, respectively. The third strain (SDF0066) was assigned to 2 different genera: *Bacillus* (score 1.915) and *Rummeliibacillus* (96% similarity) by MALDI-TOF MS and 16S rRNA gene sequencing, respectively.

Table 2: Genus-level discrepant identifications between MALDI-TOF MS and 16S rRNA gene sequencing analyses when challenged against 64 SDF strains.

SDF strain	Bruker Biotyper MALDI TOF MS		16S rRNA sequencing	
	Species (best match)	Score value	Species (% of similarity)	Acc. Nº*
0063	*Brevibacillus formosus*	1.871	*Lysinibacillus xylanilyticus* (99)	MH356333
0066	*Bacillus simplex*	1.915	*Rummeliibacillus pycnus* (96)	MH356336
0133	*Bacillus pumilus*	1.760	*Paenibacillus alvei* (99)	MH356383

Score value is the identification log score given by the standard Bruker interpretative criteria (unreliable identification: 0.000–1.699; genus level identification: 1.700–1.999; species identification: ≥2.000). Similarity of 95-96% and ≥97% were considered as the threshold values for identification at the genus and species levels, respectively. ACC. Nº: GenBank accession number.

Citation: R Martins PH, da Silva LP, de Orem JC, de Magalhães MIA, De-Souza MT, et al. (2020) Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. Open J Bac 4(1): 001-007. DOI. https://dx.doi.org/10.17352/ojb.000012

Table 3: Species-level concordant identifications between MALDI-TOF MS and 16S rRNA gene sequencing analyses when challenged against 64 SDF strains.

SDF strain	Bruker Biotyper MALDI TOF MS		16S rRNA sequencing	
	Species (best match)	Score value	Species (% of similarity)	Acc. N°*
0005	*Lysinibacillus fusiformis*	2.079	*Lysinibacillus fusiformis* (98)	MH356291
0006	*Bacillus cereus*	2.203	*Bacillus cereus* (99)	MH356292
0015	*Bacillus oleronius*	2.155	*Bacillus oleronius* (99)	MH356300
0050	*Bacillus megaterium*	2.258	*Bacillus megaterium* (98)	MH356322
0056	*Bacillus megaterium*	2.335	*Bacillus megaterium* (99)	MH356328
0057	*Bacillus megaterium*	2.198	*Bacillus megaterium* (98)	MH356329
0058	*Bacillus megaterium*	2.036	*Bacillus megaterium (98)*	MH569349
0064	*Bacillus megaterium*	2.273	*Bacillus megaterium* (99)	MH356334
0069	*Bacillus megaterium*	2.234	*Bacillus megaterium* (99)	MH356338
0072	*Bacillus megaterium*	2.240	*Bacillus megaterium* (99)	MH569350
0074	*Bacillus megaterium*	2.159	*Bacillus megaterium* (99)	MH356341
0075	*Bacillus simplex*	2.004	*Bacillus simplex* (99)	MH356342
0076	*Bacillus megaterium*	2.064	*Bacillus megaterium* (99)	MH356343
0078	*Bacillus megaterium*	2.047	*Bacillus megaterium* (99)	MH356344
0094	*Bacillus megaterium*	2.091	*Bacillus megaterium* (98)	MH356355
0095	*Bacillus megaterium*	2.227	*Bacillus megaterium* (100)	MH356356
0096	*Bacillus megaterium*	2.243	*Bacillus megaterium* (100)	MH356357
0097	*Bacillus megaterium*	2.266	*Bacillus megaterium* (98)	MH356358
0119	*Bacillus cereus*	2.152	*Bacillus cereus* (100)	MH356371

Score value is the identification log score given by the standard Bruker interpretative criteria (unreliable identification: 0.000–1.699; genus level identification: 1.700–1.999; species identification: ≥2.000). Similarity of 95-96% and ≥97% were considered as the threshold values for identification at the genus and species levels, respectively. ACC. N°: GenBank accession number.

Table 4: Species-level discrepant identifications between MALDI-TOF MS and 16S rRNA gene sequencing analyses when challenged against 64 SDF strains.

SDF strain	Bruker Biotyper MALDI TOF MS			16S rRNA sequencing
	Species (best match)	Score value	Species (% of similarity)	Acc. N°*
0029	*Bacillus megaterium*	2.252	*Bacillus aryabhattai* (99)	MH356310
0030	*Bacillus cereus*	2.028	*Bacillus thuringiensis* (99)	MH356311
0055	*Bacillus megaterium*	2.218	*Bacillus aryabhattai* (98)	MH356327
0065	*Bacillus megaterium*	2.163	*Bacillus aryabhattai (98)*	MH356335
0082	*Bacillus megaterium*	2.221	*Bacillus aryabhattai* (99)	MH356347
0086	*Bacillus megaterium*	2.183	*Bacillus aryabhattai* (98)	MH356349
0089	*Bacillus cereus*	2.100	*Bacillus anthracis* (100)	MH356351
0098	*Bacillus megaterium*	2.053	*Bacillus safensis* (98)	MH356359
0110	*Bacillus pumilus*	2.043	*Bacillus amyloliquefaciens* (100)	MH356366
0117	*Bacillus pumilus*	2.018	*Bacillus altitudinis* (99)	MH356369

Score value is the identification log score given by the standard Bruker interpretative criteria (unreliable identification: 0.000–1.699; genus level identification: 1.700–1.999; species identification: ≥2.000). Similarity of 95-96% and ≥97% were considered as the threshold values for identification at the genus and species levels, respectively. ACC. N°: GenBank accession number.

Considering similarity ≥97% and 95-96% as the threshold values for identification at the species and genus levels, respectively [19], the 16S rRNA gene sequence analysis classified 60(93.75%) SDF strains at species (Tables 1-4) and 2(3.12%) at genus levels, respectively (Tables 2,5). The prevalent genus was *Bacillus* harbouring 56(87.50%) strains, followed by 4 additional genera: 2(3.12%) of *Lysinibacillus* and of *Paenibacillus*, besides 1(1.56%) of *Brevibacillus* and of *Rummeliibacillus*. The sequence similarity-based approach failed to identify strains SDF0108 and SDF0139 at genus level, since the similarity in both cases was 94% (Tables 5,1, respectively; highlighted in grey).

When comparing performance of both methods, concerning the classification of SDF strains at species level, it was observed that 19 (61.29%) out of 31 SDF strains identified by MALDI-TOF MS presented 16S rRNA gene sequence similarity ≥97%, therefore, being also characterized at species level (Table 3). For strains identified at the species level by MALDI-TOF MS, strains SDF0014 and SDF0066, or 6.45%, presented similarity of 96%, consequently, being identified at the genus level by gene sequence similarity. In addition, other 10 out of 64(15.62%) strains identified at the species level by both methods were classified as different species in each case (Table 4).

Table 5: Identification at species level by MALDI-TOF MS and genus level or above by 16S rRNA gene sequencing when challenged against 64 SDF strains.

SDF strain	Bruker Biotyper MALDI TOF MS		16S rRNA sequencing	
	Species (best match)	Score value	Species (% of similarity)	Acc. N°*
0014	*Brevibacillus borstelensis*	2.209	*Brevibacillus borstelensis* (96)	MH356299
0108	*Bacillus pumilus*	2.023	*Bacillus pumilus* (94)	MH569355

Score value is the identification log score given by the standard Bruker interpretative criteria (unreliable identification: 0.000–1.699; genus level identification: 1.700–1.999; species identification: ≥2.000). Similarity of 95-96% and ≥97% were considered as the threshold values for identification at the genus and species levels, respectively. Strain SDF0108 (highlighted in grey) was considered unidentified by sequencing similarity. ACC. N°: GenBank accession number.

Citation: R Martins PH, da Silva LP, de Orem JC, de Magalhães MIA, De-Souza MT, et al. (2020) Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. Open J Bac 4(1): 001-007. DOI. https://dx.doi.org/10.17352/ojb.000012

Discussion

Despite their phenotypic diversity, many species of AEFB share high genetic homogeneity. In 1991, Ash, Dorsch, & Stackebrandt sequenced the 16S rRNA gene of 51 standard strains, at that point defined as *Bacillus* spp., and showed that they can be segregated into several distinct phylogenetic groups. Two of these sequenced helped in the proposition of the novel genera *Paenibacillus* and *Brevibacillus* [20]. Along with other genera, these two taxa are now recognized to comprise a separate *Bacillales* family, designated *Paenibacillaceae*. Likewise, based on clear-cut differences in discriminative taxonomic markers and the distant placement, *B. pycnus* is reclassified into a separate genus [21]. According to current 16S rRNA gene sequence-based relatedness, the latter and strains from other species of this clade are presently allocated into the genus *Rummeliibacillus*. It is noteworthy that inside order *Bacillales*, genera *Bacillus*, *Lysinibacillus*, and the novel genus *Rummeliibacillus* all belong to the family *Bacillaceae*.

Members of *B. cereus* group share 99.5 to 100% of similarity for their 16S rRNA gene sequences [22,23]. The subgroup of *B. pumilus* belong to the *B. subtilis* complex and the species *B. pumilus sensu stricto* share 99%-100% of similarity with the species *B. safensis, B. altitudinis, and B. amyloliquefaciens* [24]. Correspondingly, *Bacillus megaterium/aryabhattai* are also among many pairs of distinct taxa of AEFB that bear extreme close evolutionary relationship sharing 99.7% similarity of 16S rRNA sequences [25].

Currently, MALDI-TOF MS is well-established as a fast and reliable technique in clinical laboratories to identify microorganism species [13].—However, application of this technique in other fields of microbiology, whose reference databases cover only a small portion of the vast range of microbial diversity, has been limited [26,27]. Even though, protein profiling has been found to be useful in discriminating many closed-related Bacillus *sp.* [28-33].

In this work, we compared MALDI-TOF MS analysis of 64 environmental AEFB to the standard 16S rRNA gene sequencing method for the identification and classification of AEFB isolated from Brazilian soils, designated SDF strains [17].

MALDI-TOF MS results were evaluated using cut-off scores ≥1.7000 to <1.999, and ≥2.000 for acceptable identification at genus and species levels, respectively, as suggested by the manufacturer. At genus level, the overall concordance between the two methods was 60(93,75%) SDF strains. Biotyper identified 33(51.56%) and 31(48.84%) strains at genus (Tables 1,2) and species (Tables 3-5) levels, respectively. Conversely, 16S rRNA gene sequencing approach identified all strains at species level, expect for 2(3.12%) that were identified only at genus level (Tables 2,5) and 2 others that could not be identified even at genus level (Tables 1,5). The genus *Bacillus* prevails comprising 60(93.75%) and 56(87.50%) strains classified by protein profiling and gene sequence similarity, respectively. MALDI-TOF MS-based analysis distributed the remaining 4(6.25%) strains among 3(4.68%) other genera: 2(3.12%) to *Brevibacillus* and 1(1.56%) to either *Paenibacillus* and or *Lysinibacillus*. In contrast, the genotype method assigned 6 strains to 4 additional genera: 2(3.12%) to either *Lysinibacillus* or *Paenibacillus*, besides 1(1.56%) to either *Brevibacillus* or *Rummeliibacillus.*

Decreasing the cut-off point for identification to a score of 1.700 had little effect on the overall classification, as the inclusion of SDF strains with a MALDI Biotyper score of ≥1.700 and <2.0 did not significantly affect the results obtained using the recommended score of ≥2.0 (Table 1). From the 30 out of 33 strains identified at genus level by MALDI-TOF MS (Table 1), the best match suggested coincided with 15(50.00%) strains discriminated at species level by the genotype method, and the 13(43.33%) discrepant belonged to the same close-related groups of the genus *Bacillus*, mostly belonging to the *B. subtilis* complex.

Although further studies would be required to accurately discriminate these species, this is a relevant guide towards the many different genetic clusters found inside genus bearing hundreds of species, as in the case of *Bacillus* consisting of almost 400 species (List of Prokaryotic names with Standing in Nomenclature: http://www.bacterio.net/index.html; retrieved 11 October 2019).

In this study, protein profiling and 16S rRNA gene sequencing were discordant in only 3 classifications at genus level (Table 2). Strains SDF0063 and SDF0133 were assigned to genera *Brevibacillus* and *Bacillus*, respectively, by the first technique. On the other hand, these strains were classified at species level by 16S rRNA gene sequencing, both presenting 99% of similarity with *Lysinibacillus xylanilyticus* and *Paenibacillus alvei*, respectively. The strain SDF0066 was assigned to genus *Bacillus* (score 1.915) and to *Rummeliibacillus* (96% similarity) by MALDI-TOF MS and 16S rRNA gene sequencing, respectively. As for the results discussed above, these discrepancies are most likely due to the insufficient coverage of bacterial species in the databases. Indeed, at the time these analyses were performed, most environmental species studied here were underrepresented with one or few spectra in the reference library.

With respect to 31 out of 64 SDF strains which MALDI BioTyper identifications reached scores of >2.000 (species identification), 19(61.29%) were concordant (Table 3) and SDF0014 and SDF0108 (6.45%) were identified only at the genus level by 16S rRNA gene sequencing (Table 5). Nevertheless, the remaining 10(32.25%) strains were also identified at the species level by both methods (Table 4), although classified as different species of genus *Bacillus*. Interestingly, in this case, the results obtained by both techniques also pointed out to a pair of alternative species. Five strains (SDF0029, SDF0055, SDF0065, SDF0082, and SDF0086) were classified as *B. megaterium* and *B. aryabhattai* by MALDI-TOF MS and 16S rRNA gene sequences, respectively. Considering the high genetic similarity between these 2 species and that there was no representative strain of species *B. aryabhattai* in the BioTyper 3.0 library (Bruker Daltonics), these results were not surprising. Therefore, the availability of higher number of *B. megaterium* spectra and library entries belonging to *B. aryabhattai* might improve MALDI-TOF

Citation: R Martins PH, da Silva LP, de Orem JC, de Magalhães MIA, De-Souza MT, et al. (2020) Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. Open J Bac 4(1): 001-007. DOI. https://dx.doi.org/10.17352/ojb.000012

MS accuracy. Likewise, the other 10 strains (Table 4), allocated by both methods as either *B. cereus* (5) and *B. pumilus* (5) group strains are also scarcely represented in the library. Thus, at the species level, considering the various groups of closed-related AEFB, the overall concordance between the two methods was 29(45,31%) SDF strains.

Our classification based on 16S rRNA gene sequences is a preliminary determination of genera or species. Thus, when 16S rRNA gene profiling placed these strains within these *Bacillus sp.* groups, the sample analysed can belong to two or even more species alternatives within the same affiliation cluster. Therefore, in these instances the 16S rRNA gene sequencing can only identify these sets of bacteria but cannot assign it accurately to a certain species according to its low discrimination capacity. Even so, our assignments are useful because they clearly identify the genera and restrict the identity of SDF strains to one or two possible species in the genera described.

Since strains SDF0108 and SDF0139 presented similarity of 94%, the 16S rRNA gene-sequencing tool failed to classify both even at genus level (Tables 5,1, respectively; highlighted in grey). Though sometimes the use of these sequences as a single marker is not enough to delineate species, low gene sequence similarity may grant the first indication that a novel species could have been isolated [34]. However, description of new species is beyond the scope of this study.

The results obtained here demonstrated that both techniques used for the identification of SDF strains had good resolution at the genus level. However, 16S rRNA gene sequences achieved superior capacity in identifying these environmental AEFB at the species level when compared with MALDI-TOF MS method. Both tools showed a lack of efficiency to discriminate closely related species. Nevertheless, this initial outline clarified the genetic interrelationships of these environmental strains. Hence, sequence similarity values and score ranges were complementary to each other and can help if comprehensive high-quality reference datasets are available.

Considering that in the present study less than 50% of the SDF strains were identified at the species level using MALDI-TOF MS, our results showed the importance of expanding the available spectrum libraries, since most spectra deposited in databases are of clinical source, thus presenting little information of other origins. Spectrum libraries of non-clinical samples may require special considerations concerning the clinical counterpart, because of the extent of stresses can be much more variable in the environment [35,36]. The stress-related proteins may lead to the misidentification of new isolates, since they may differ significantly from type strains in the proteotypic properties. Beyond this technical issue, the absence of public repositories for mass spectra may limit the use of MALDI-TOF MS, since existing libraries remain private and expensive to access.

This study also supports the need of using phenotypic along with genotypic methods into polyphasic approaches for taxonomic purposes of the diversity of AEFB.

Acknowledgement

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).

References

1. Logan NA, Berge O, Bishop AH, Busse HJ, De Vos P, et al. (2009) Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria. Int J Syst Evol Microbiol 59: 2114-2121. **Link:** https://bit.ly/38FzFyF

2. Galperin MY (2013) Genome Diversity of Spore-Forming Firmicutes. Microbiol Spectr 1: 1-15. **Link:** https://bit.ly/2TFGEmJ

3. Logan NA, Vos P (2011) Endospore-forming Soil Bacteria. 27. **Link:** https://bit.ly/2W3b7Nq

4. Fritze D (2004) Taxonomy of the Genus Bacillus and Related Genera: The Aerobic Endospore Forming Bacteria. Phytopathology 94: 1245-1248. **Link:** https://bit.ly/39KK79e

5. Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, et al. (2009) Bergey's Manual of Systematic Bacteriology. New York, Springer 3: 269. **Link:** https://bit.ly/2vco7Fj

6. 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. **Link:** https://bit.ly/2IA6Uc0

7. Maughan H, Van der Auwera G (2011) Bacillus taxonomy in the genomic era finds phenotypes to be essential though often misleading. Infect Genet Evol11: 789-797. **Link:** https://bit.ly/2VYSnyy

8. Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, et al. (2010) Comparison of Two Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Methods with Conventional Phenotypic Identification for Routine Identification of Bacteria to the Species Level. J Clin Microbiol 48: 1169-1175. **Link:** https://bit.ly/2xtJseg

9. AlMasoud N, Xu Y, Nicolaou N, Goodacre R (2014) Optimization of matrix assisted desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) for the characterization of Bacillus and Brevibacillus species. Anal Chim Acta 840: 49-57. **Link:** https://bit.ly/2xu0wAS

10. Seibold E, Maier T, Kostrzewa M, Zeman E, Splettstoesser W (2010) Identification of Francisella tularensis by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels. J Clin Microbiol 48:1061-1069. **Link:** https://bit.ly/3cMXurv

11. Šedo O, Sedláček I, Zdráhal Z (2011) Sample preparation methods for MALDI-MS profiling of bacteria. Mass Spectrom Rev 30: 417-434. **Link:** https://bit.ly/38I6KKn

12. Lasch P, Beyer W, Nattermann H, Stämmler M, Siegbrecht E, et al. (2009) Identification ofBacillusanthracisby using matrix-assisted laser desorption/ ionization-timeof flight mass spectrometry and artificial neural networks.Appl Environ Microbiol 75: 7229-7242. **Link:** https://bit.ly/2xuMhvB

13. Wieser A, Schneider L, Jung J, Schubert S (2012) MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). Appl Microbiol Biotechnol 93: 965-974. **Link:** https://bit.ly/3cZ3i1k

14. Singhal N, Kumar M, Kanaujia PK, Virdi JS (2015) MALDI TOF mass spectrometry: na emerging technology for microbial identification and diagnosis. Front Microbiol 6: 791. **Link:** https://bit.ly/2IDe4MG

15. Roberts E, Sethi A, Montoya J, Woese CR, Luthey-Schulten Z, (2008) Molecular signatures of ribosomal evolution. Proc Natl Acad Sci 105:13953-13958. **Link:** https://bit.ly/336FdRC

Citation: R Martins PH, da Silva LP, de Orem JC, de Magalhães MIA, De-Souza MT, et al. (2020) Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. Open J Bac 4(1): 001-007. DOI. https://dx.doi.org/10.17352/ojb.000012

16. Suarez S, Ferroni A, Lotz A, Jolley KA, Guérin P, et al. (2013) Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. J Microbiol Methods 94: 390-396. Link: https://bit.ly/2IDsUTj

17. de Andrade Cavalcante D, De-Souza MT, de Orem JC, de Magalhães MIA, Martins PH, et al. (2019) Ultrastructural analysis of spores from diverse Bacillales species isolated from Brazilian soil. Environ Microbiol Rep 11: 155-164. Link: https://bit.ly/2W2qxkV

18. Orem JC, Silva WMC, Raiol T, Magalhães MI, Martins PH, et al. (2019) Phylogenetic diversity of aerobic spore-forming Bacillalles isolated from Brazilian soils. Int Microbiol 22: 511-520. Link: https://bit.ly/33a6Jhd

19. Schloss PD, Handelsman J (2005) Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. Appl Environ Microbiol 71: 1501-1506. Link: https://bit.ly/3aOQzMP

20. Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, et al. (2009b) Family IV. «Paenibacillaceae». Bergey's Manual of Systematic Bacteriology. New York, Springer 3: 269.

21. Vaishampayan P, Miyashita M, Ohnishi A, Satomi M, Rooney A, et al. (2009) Description of Rummeliibacillus stabekisii gen. nov., sp. nov. and reclassification of Bacillus pycnusNakamura et al. 2002 as Rummeliibacillus pycnus comb. nov. Int J Systematics Evolutionary Microbiol 59: 1094-1099. Link: https://bit.ly/39FABnS

22. Guinebretière MH, Auger S, Galleron N, Contzen M, De Sarrau B, et al. (2013) Bacillus cytotoxicus sp. nov. is a novel thermotolerant species of the Bacillus cereus group occasionally associated with food poisoning. Int J Syst Evol Microbiol 63: 31-40. Link: https://bit.ly/39I1hoa

23. Ash C, Farrow JA, 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 rRNA. Int J Syst Bacteriol 41: 343-346. Link: https://bit.ly/2v8U29A

24. Alina SO, Constantinscu F, Petruţa CC (2015) Biodiversity of Bacillus subtilis group and beneficial traits of Bacillus species useful in plant protection. Rom Biotechnol Lett 20: 10737–10750. Link: https://bit.ly/337mF3u

25. Shivaji S, Chaturvedi P, Begum Z, Pindi PK, Manorama R, et al. (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: 2977-2986. Link: https://bit.ly/3cSUc6i

26. Lovecka P, Pacovska I, Stursa P, Vrchotova B, Kochankova L, et al. (2015) Organochlorinated pesticide degrading microorganisms isolated from contaminated soil. N Biotechnol 32: 26-31. Link: https://bit.ly/38FAXJV

27. Emami K, Askari V, Ullrich M, Mohinudeen K, Anil AC, et al. (2012) Characterization of bacteria in Ballast water using MALDI-TOF mass spectrometry. PLoS One 7: e38515. Link: https://bit.ly/2Q3kmte

28. Ryzhov V, Hathout Y, Fenselau C (2000) Rapid characterization of spores of Bacillus cereus group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Appl Environ Microbiol 66: 3828-3834. Link: https://bit.ly/2VZmENL

29. Dickinson DN, La Duc MT, Satomi M, Winefordner JD, Powell DH, et al. (2004) MALDI-TOFMS compared with other polyphasic taxonomy approaches for the identification and classification of Bacillus pumilus spores. J Microbiol Methods 58: 1-12. Link: https://bit.ly/2IEm7Ja

30. Castanha ER, Fox A, Fox KF (2006) Rapid discrimination of Bacillus anthracis from other members of the B. cereus group by mass and sequence of "intact" small acid soluble proteins (SASPs) using mass spectrometry. J Microbiol Methods 67: 230-240. Link: https://bit.ly/2W1PtJd

31. Fernández-No IC, Böhme K, Díaz-Bao M, Cepeda A, Barros-Velázquez J, et al. (2013) Characterisation and profiling of Bacillus subtilis, Bacillus cereus and Bacillus licheniformis by MALDI-TOF mass fingerprinting. Food Microbiol 33: 235-242. Link: https://bit.ly/3cJdXNv

32. Branquinho R, Meirinhos-Soares L, Carriço JA, Pintado M, Peixe LV (2014) Phylogenetic and clonality analysis of bacillus pumilus isolates uncovered a highly heterogeneous population of different closely related species and clones. FEMS Microbiol Ecol 90: 689-698. Link: https://bit.ly/39HT9Uv

33. Branquinho R, Sousa C, Lopes J, Pintado ME, Peixe LV, et al. (2014) Differentiation of Bacillus pumilus and Bacillus safensis Using MALDI-TOF-MS. PLoS ONE 9: e116426. Link: https://bit.ly/2wIxbSN

34. Santos IC, Hildenbrand ZL, Schug KA (2016) Applications of MALDI-TOF MS in environmental microbiology. Analyst 141: 2827-2837. Link: https://bit.ly/2wQAjfe

35. 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: 249–266. Link: https://bit.ly/2TGxfeP

36. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73: 5261–5267. Link: https://bit.ly/2THCYkH

Citation: R Martins PH, da Silva LP, de Orem JC, de Magalhães MIA, De-Souza MT, et al. (2020) Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. Open J Bac 4(1): 001-007. DOI. https://dx.doi.org/10.17352/ojb.000012