

## Research Article

# ***Bacillus* Strains Most Closely Related to *Bacillus nealsonii* Are Not Effectively Circumscribed within the Taxonomic Species Definition**

**K. Kealy Peak,<sup>1</sup> Kathleen E. Duncan,<sup>2</sup> Vicki A. Luna,<sup>1</sup> Debra S. King,<sup>1</sup> Peter J. McCarthy,<sup>3</sup> and Andrew C. Cannons<sup>1</sup>**

<sup>1</sup>USF Center for Biological Defense, College of Public Health, University of South Florida, 3602 Spectrum Boulevard, Tampa, FL 33612-9401, USA

<sup>2</sup>Department of Botany and Microbiology, University of Oklahoma, 770 Van Vleet Oval, Norman, OK 73019-6131, USA

<sup>3</sup>Harbor Branch Oceanographic Institute at Florida Atlantic University, 5600 US No. 1 North, Fort Pierce, FL 34946, USA

Correspondence should be addressed to K. Kealy Peak, kpeak1@health.usf.edu

Received 10 May 2011; Revised 12 August 2011; Accepted 19 August 2011

Academic Editor: Luca Simone Cocolin

Copyright © 2011 K. Kealy Peak et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Bacillus* strains with >99.7% 16S rRNA gene sequence similarity were characterized with DNA:DNA hybridization, cellular fatty acid (CFA) analysis, and testing of 100 phenotypic traits. When paired with the most closely related type strain, percent DNA:DNA similarities (% S) for six *Bacillus* strains were all far below the recommended 70% threshold value for species circumscription with *Bacillus nealsonii*. An apparent genomic group of four *Bacillus* strain pairings with 94%–70% S was contradicted by the failure of the strains to cluster in CFA- and phenotype-based dendrograms as well as by their differentiation with 9–13 species level discriminators such as nitrate reduction, temperature range, and acid production from carbohydrates. The novel *Bacillus* strains were monophyletic and very closely related based on 16S rRNA gene sequence. Coherent genomic groups were not however supported by similarly organized phenotypic clusters. Therefore, the strains were not effectively circumscribed within the taxonomic species definition.

## **1. Introduction**

CBD 118 was one of the two first *Bacillus* strains not related to the *B. cereus* group reported to harbor the capsule genes carried on pXO2 by *Bacillus anthracis* (USF Center for Biological Defense (CBD)) [1, 2]. Luna et al. isolated and sequenced the capsule operon (*capA*, *capB*, *capC*, *capD*, and promoter), *repA*, *capR*, *acpA*, *IS1627*, *ORF43*, *ORF48*, and *ORF61* on a large plasmid in CBD 118 [1]. Its status as a carrier of *B. anthracis* capsule genes spurred research into determining its closest relatives, to aid in circumscribing the reservoir of genes essential for virulence in *B. anthracis*. When near full length 16S rRNA gene sequences were compared, the most similar type strains to strain CBD 118 were *Bacillus circulans* ATCC 4513<sup>T</sup> (98.9%) and *Bacillus nealsonii* DSM 15077<sup>T</sup> (99.3%). Strain CBD 118 differed from *B. circulans* ATCC 4513<sup>T</sup> and *B. nealsonii* DSM 15077<sup>T</sup> for 10 and 12 of

100 phenotypic traits evaluated, respectively. The percentages of DNA:DNA binding in two pairings each of strain CBD 118 to *B. circulans* ATCC 4513<sup>T</sup> and *B. nealsonii* DSM 15077<sup>T</sup> were 12.5 and 10.2% and 10.8 and 8.3%, respectively. Thus, strain CBD 118 is differentiated by phenotypic and genome-based methods from the only validly named species with greater than 98.7% 16S rRNA gene sequence similarity [3–5]. Strain CBD 118 was the sole exemplar of a novel species. Prior to the proposal of novel species, studies of ten or more strains are recommended in order to detail intra-species diversity and to foster appropriate type strain assignment [6–8]. To identify the requisite closely related strains, the V1–V3 hypervariable regions of the 16S rRNA gene [9] from strain CBD 118 were compared to sequences available in GenBank. Eight potential sibling strains were obtained for study. Although the eight strains tested negative for capsule production and for the pXO2 genetic marker by PCR, the

group retained taxonomic—if not biodefense—significance. This work presents the polyphasic taxonomic characterization of these eight strains with respect to CBD 118. Incongruent strain-strain associations within this polyphasic data set illustrate the difficulties in applying a pragmatic, taxonomic, bacterial species definition to groups of strains that do not fall into coherent clusters based on genetic and phenotypic analyses.

Bacterial species are currently defined by pragmatic criteria in a coordinated, polyphasic scheme of 16S rRNA sequence-based phylogeny, indirect whole genome comparisons by DNA:DNA hybridization and analysis of numerous covariant phenotypic characters [3, 5, 10, 11]. Key requisites of the taxonomic species definition can be condensed as follow: (i) a species should be a monophyletic group with a high degree of genetic similarity, (ii) the recommended thresholds of 70% DNA similarity and 5°C  $\Delta T_m$  are guidelines, not absolute limits for circumscribing new species, (iii) genomic boundaries for a separate species should be defined after analysis of the collective phenotype, (iv) phenotypic intra-group homo- or heterogeneity can only be understood after analysis of as many traits as possible among at least five and preferably more strains, (v) a bacterial species should not be classified unless it can be recognized by multiple independent methods and possesses a set of determinative phenotypic properties [3, 5, 11].

Underlying these guidelines are assumptions about the genetic and phenotypic characteristics of bacterial species that may not be equally applicable to all groups of bacteria [12–16]. That is, it is usually assumed that there are clusters of strains, for example, “sequence clusters” [17], “ecotypes” [18], and so forth, distinct from other clusters. Investigators have been encouraged to develop other genomic-based methods to supplement or even supplant DNA:DNA hybridization as the acknowledged standard for delineating genospecies clusters [3, 4, 6, 16, 19]. Various methods are increasingly used to define genetic and phenotypic similarity among strains—from multilocus sequence typing (MLST) [20] up to the analysis of whole genomes [13, 14]. Ever more precise and detailed descriptions of similarity among strains and between clusters can be obtained by advances in sequencing technology, its application to more isolates and by polyphasic phenotypic analysis of increased numbers of characters. But a more fundamental and less tractable problem is that of the species level circumscription of related bacteria that do not appear to fit readily into sequence clusters and hence within the current taxonomic species definition [14]. Taxonomic species definitions continue to be refined as new techniques become available and new strains are described [3, 4, 6, 16, 19]. Our study illustrates complexities that can be encountered as polyphasic methods are applied to greater numbers of strains forming a broader sample of the microbial world.

## 2. Materials and Methods

**2.1. Bacterial Strains.** Nine *Bacillus* strains in this study were deposited in the Agricultural Research Service Culture Collection (ARSCC), U.S. Department of Agriculture, Peoria,

Ill, USA. Accession numbers for the CBD collection and the ARSCC (NRRL) follow the original strain identifiers. *Bacillus* sp. CBD 118 = NRRL B-51264 was isolated from a powder initially suspected of harbouring *B. anthracis* [1]. *Bacillus* strains provided by colleagues: OSS 25 = CBD 1278 = NRRL B-59473 [21]; P307 = CBD 1284 = NRRL B-59474 and P308 = CBD 1285 = NRRL B-59475 (Harbor Branch Marine Microbial Culture Collection); C4T1F3B3 = CBD 1286 = NRRL B-59476 [22]; IAFILS6 = CBD 1287 = NRRL B-59477 [23]; AD5A = CBD 1288 = NRRL B-59478, U4A = CBD 1289 = NRRL B-59479 and ADP4II = CBD 1290 = NRRL B-59480 [24]. Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under accession numbers DQ374636, EU656111, EU683686, FJ554672, FJ943256–FJ943261. The type strains of *B. circulans*, ATCC 4513<sup>T</sup> and of *B. nealsonii*, DSM 15077<sup>T</sup> were acquired from the American Type Culture Collection (ATCC) and the Deutsche Sammlung von Microorganismen und Zellkulturen (GmbH) (DSMZ), respectively.

**2.2. Preservation and Authentication of *Bacillus* Strains.** Upon receipt, each strain was subcultured by streaking to tryptic soy agar (TSA) and TSA with 5% sheep red blood cells (TSA-BA) and grown at 30°C. After 24, 48, and 72 h incubation, plated strains were examined for purity based on the presence of colonies of a single morphotype. A single, well-isolated and representative colony was designated as the progenitor colony and streaked for pure culture reisolation on plates of TSA, TSA-BA, and TSA with 5 mg L<sup>-1</sup> MnSO<sub>4</sub>, incubated at 30°C for up to 72 h. Characteristic, well-isolated colonies on these plates served as first passage sources of inocula for initial phenotypic characterization as detailed below. Colony morphologies for each strain were observed at 24 and 48 h for consistency with the progenitor colony and were described for standard colony features including color, surface texture and degree of luster, relative transmittance of direct light through the colony, shape, margin configuration, elevation, diameter in mm and hemolysis reaction. Each of the *Bacillus* strains in this study including the type strains presented one or more differential colony features that were documented and subsequently monitored as evidence of purity and authenticity whenever strains were subcultured. Phenotypic tests and other procedures utilizing broths were routinely subcultured at the incubation end point to TSA-BA check plates. After 24 and 48 h incubation at 30°C, check plates were reviewed for the presence of colonies of a single, differential morphotype, characteristic of each strain.

The *Bacillus* including type strains were inoculated from the progenitor colony to aerated tryptic soy broth (TSB), grown to late log phase, subcultured to a TSA check plate, aseptically harvested by centrifugation, resuspended in TSB with 10% glycerol, aliquoted to multiple cryovials, and subjected to a controlled freeze prior to storage at –85°C. One week after cryostocking, a cryovial of each strain was thawed, subcultured on TSA and TSA-BA plates, enumerated for viability and again evaluated for the single, differential colony morphotype. Prior to retesting of phenotypic characters and other analyses, strains were subcultured from the cryostocks and endospore production was induced on TSA with

5 mg L<sup>-1</sup> MnSO<sub>4</sub> plates. Serial transfer of strains was restricted by the use of single, characteristic endospore-producing colonies as inoculation sources for subsequent testing.

Preservation of strain authenticity was evaluated at the end of the study. Four strains that formed an apparent genomic cluster were subcultured from cryopreserved stock, retested for seven differential phenotypic traits including colony morphotype and resequenced for the 16S rRNA gene. The resultant sequences were compared to the original sequences deposited in GenBank.

**2.3. 16S rRNA Gene Sequence Analysis.** Amplification of 16S rRNA gene sequences from *Bacillus* strains, sequencing of the approximately 1500 bp long products, fragment assemblies and alignment of 16S rRNA gene sequences from type strains of selected *Bacillus* species were performed as previously reported [2]; an additional primer, 534R, was employed in some amplifications. Identification of phylogenetic neighbors was carried out by BLAST 2.2.20+ [25] and megaBLAST (discontinuous option) [26] searches of GenBank [27]. Calculation of pairwise sequence similarity to nearest neighbors used the EzTaxon global alignment algorithm [28]. Alignments of 16S rDNA sequences were also made using the Infernal secondary structure based aligner and SeqMatch scores (S.ab) calculated with RDP10 at the Ribosomal Database Project website [29]. Nucleotide (nt) positions in the hypervariable regions V1–V3 were identified in the 16S sequence of strain CBD 118 by alignment with conserved regions at nt positions 48–70, 346–366, and 490–511 in *rrnE* of *B. subtilis* subsp. *subtilis* strain 168<sup>T</sup> (NC\_000964.2, Locus tag: BSUr022, GeneID: 2914197, updated 3/2010 to NC\_000964.3, Locus tag: BSU\_rRNA\_30, GeneID: 8303085) [9]. In the *E. coli* numbering system, regions V1–V3 correspond to nt positions 69–99, 137–242 and 433–497, respectively [30]. Using GenBank bl2seq, pairwise alignments of 461 bp from strain CBD 118 (nt 26–461 corresponding to *rrnE* nt positions 48–511) were made to 16S rDNA sequences from closely related strains. Presumptive signature sequences (PSS) within the V1–V3 region were identified and compared to all GenBank sequences using BLASTN 2.2.20+ with parameters adjusted for short input sequences [25]. Dendrograms were constructed from approximately 1390 bp or 448 bp using neighbor-joining, maximum parsimony and maximum likelihood algorithms (PHYLIP v. 3.6.80) [31] with 1000 bootstrap replications performed to estimate support for each branch.

**2.4. DNA:DNA Hybridization Studies.** Strains were subcultured from cryopreserved stock, grown to late log phase in aerated TSB, harvested by centrifugation and provided to the Deutsche Sammlung von Microorganismen und Zellkulturen (GmbH) (DSMZ) as ≥3 g wet weight biomass preserved in 50:50 sterile dI H<sub>2</sub>O:2-propanol. Prior to harvesting, each broth culture was screened for a single characteristic morphology in wet mounts using phase contrast microscopy at x1000 under oil and subcultured on TSA-BA check plates grown at 30°C. After 24 and 48 h incubation, check plates corresponding to the preserved biomass for each strain were

reviewed for the presence of colonies of a single morphotype, consistent with that previously determined to be characteristic of and differential for the strain. Biomass was shipped only after no apparent evidence of contamination or mislabeling of strains was detected.

DNA:DNA hybridizations were performed by the Identification Service of the DSMZ. Cells of preserved biomass were disrupted in a French pressure cell and the DNA purified by chromatography on hydroxyapatite. DNA:DNA hybridization was carried out at 65°C using a model Cary 100 Bio UV/VIS spectrophotometer equipped with a Peltier-thermostatted 6 × 6 multicell changer and a temperature controller with in situ temperature probe (Varian) [32, 33].

**2.5. Cellular Fatty Acid Analysis.** Cellular fatty acid (CFA) composition of *Bacillus* strains was determined by Microbial ID, Inc. (MIDI) at both study start and end point. Each strain was subcultured from cryopreserved stock, inoculated from a single, characteristic colony to a TSA slant and grown at 30°C for 48 h to foster endospore formation. Prior to shipment, each slant was subcultured on a TSA-BA check plate grown at 30°C and observed after 24 and 48 h incubation. The check plate for each strain was reviewed for colonies of a single, differential morphotype and the slant shipped to MIDI only after no evidence of contamination or mislabeling of strains was discerned.

Strains were grown under standardized conditions on tryptic soy broth agar quadrant streak plates at 28°C for 24 h. To reduce disparities in the effective physiological age of the cells, biomass was harvested from colonies growing in the third streaked quadrant. Fatty acid methyl esters were extracted by a four-step procedure of saponification, methylation, extraction and sample clean-up. Fatty acid peaks were analyzed by gas chromatography and named by comparing retention times to those in a known mixture. A dendrogram program used a multivariate clustering algorithm to produce unweighted pair matching based on similar CFA content between strains and generated a tree scaled to Euclidian distance (ED).

**2.6. Phenotypic Characterization.** All tests were incubated at 30°C unless otherwise noted [34]; incubation periods are specified. Differential tests were performed at minimum twice or as specified; prior to re-testing, strains were subcultured from cryostocks held at –85°C in TSB, 10% glycerol. Control strains of *Bacillus* and *Paenibacillus* included *B. cereus* ATCC 14579<sup>T</sup>, *B. circulans* ATCC 4513<sup>T</sup>, *B. megaterium* ATCC 14581<sup>T</sup>, *B. nealsonii* DSM 15077<sup>T</sup>, *B. pumilus* ATCC 7061<sup>T</sup>, *B. thuringiensis* ATCC 10792<sup>T</sup>, and *P. polymyxa* ATCC 43865<sup>T</sup>. Sporulation was induced on TSA with 5 mg L<sup>-1</sup> MnSO<sub>4</sub>, grown for 40–48 h. Hemolysis reaction was determined on TSA with 5% sheep red blood cells (REMEL), grown for 48 h. Pigment production and mean colony diameter were evaluated on TSA, tryptone blood agar base and tryptone glucose yeast extract plates, 24, 48, 72 h and 1 week. Motility was determined by either stab inoculation of motility test medium (REMEL), observed at 24 and 48 h, or phase contrast observation of wet mounts made with aerated cells grown in TSB to log phase, 3 to 6 h. Cell morphology,



endospore characterization and swelling of the sporangium, presence of parasporal bodies and motility were observed in wet mounts using phase contrast microscopy at  $\times 1000$  under oil. Anaerobic growth was evaluated after 1 week in the Mit-su-bishi Pack-Anaero anaerobic gas generating system with the following pre-reduced media: fluid thioglycollate medium with dextrose and indicator (REMEL), tryptone glucose yeast extract agar plates, and anaerobic agar [34], inoculated in the molten state. Oxidase reaction was tested with Kovács' phenylenediamine redox dye reagent (Becton, Dickinson). Growth of cells at defined temperatures was tested in 3 mL of TSB in  $13 \times 100$  mm tubes for 48 h in water baths set to 30, 35, 40, 45, 50, 55, and  $60 \pm 1^\circ\text{C}$  and examined for turbidity at 24 h intervals. Growth of cells at defined pH was tested in the same manner in TSB adjusted to pH 4.6, 5.6, 6.1, 6.5, 6.8, 7.3, 7.8, 8.1, and 8.5. Salt tolerance was tested on nutrient agar plates supplemented with 0, 1, 3, 7 and 10% NaCl, incubated for 5 days. Physiological tests performed on cells grown in commercial media (REMEL) included casein and starch hydrolysis, incubated 14 days; growth on mannitol egg yolk polymyxin agar, incubated 48 h; growth in methyl red Voges-Proskauer (MRVP) broth for final pH and VP reaction, tested at 3, 5, and 7 days; nitrate reduction tested in nitrate broth at 3, 7, and 14 days and on nitrate agar slants, at 3 and 5 days; and growth on Sabouraud's 4% glucose agar, pH 5.6, incubated 72 h. Gelatin hydrolysis was tested in 12% nutrient gelatin (REMEL) for 2 weeks. Hydrolysis of Tween-80 was tested on plates of a peptone-based medium [35], incubated for 4 weeks. Acid production from 49 carbohydrates or carbohydrate derivatives was tested using the API 50 CH panel and API CHB/E medium with mineral oil overlay,  $\geq 4$  test panels per strain, in combination with eleven biochemical tests from the API 20 E kit,  $\geq 2$  test panels per strain, incubated for 48 and 24 h, respectively (bioMérieux). Acid production was read at 24 and 48 h in a semiquantitative way, where 0 was assigned to negative reactions of the same alkaline red as the no-carbohydrate control and 5 assigned to yellow indicator shifts of maximum intensity. Values of 1, 2, 3, or 4 were given to intermediate reactions with 3, 4, and 5 being considered positive. Differential phenotypic traits between paired strains were enumerated. Each differential character state was assigned a numerical value—that is, 1 = negative, 2 = variable, 3 = positive—and subjected to hierarchical cluster analysis (SPSS for Windows, Release 15.0.1.1, 2007). A dendrogram was generated using average linkage between groups, scaled in Euclidian distance units.

### 3. Results and Discussion

**3.1. Bacterial Strains Obtained for Taxonomic Study.** Eight *Bacillus* strains having  $\geq 99.7\%$  16S rRNA gene sequence similarity [19, 36] and  $\geq 0.980$  SeqMatch S<sub>ab</sub> scores to 1495 bp of strain CBD 118 (DQ374636) were provided to the CBD for taxonomic comparisons: strain OSS 25 (EU683686), isolated from a metallurgic waste site in Italy [21]; strains P307 (FJ943260) and P308 (FJ554672), isolated from deep-water marine sponge (*Discodermia* sp.), Bahamas; strain

C4T1F3B3 (FJ943258), isolated from cultured flounder, USA [22]; strain IAFILS6 (FJ943259) isolated from a consortium degrading polyaromatic hydrocarbons, Canada [23]; strains AD5A (FJ943256), U4A (FJ943261), and ADP4II (FJ943257), isolated from plant thorns, Israel [24]. The most closely related type strains were *Bacillus circulans* ATCC 4513<sup>T</sup> (AY724690) (98.9%; S<sub>ab</sub> not ranked) and *B. nealsonii* DSM 15077<sup>T</sup> (EU656111) (99.3%; S<sub>ab</sub> 0.961).

**3.2. Presumptive Signature Sequences.** Hypervariable regions V1–V3 in the *Bacillus* 16S rRNA gene sequence had been reported to be discriminatory for most *Bacillus* species [9]. The eight *Bacillus* strains were identified as potential sibling strains when 461 bp spanning V1–V3 hypervariable regions of the 16S rRNA gene sequence of strain CBD 118 were compared to GenBank database sequences. Presumptive signature sequences (PSS) were identified in V1 at nt positions 71–92 (PSS<sub>1A</sub> and PSS<sub>1B</sub>) and in V2 at nt positions 183–223 (PSS<sub>2</sub>) (Table 1). The eight *Bacillus* strains, strain CBD 118 and *B. nealsonii* DSM 15077<sup>T</sup> differed from *B. circulans* ATCC 4513<sup>T</sup> by two nucleotide changes in PSS<sub>1</sub> and at five positions in PSS<sub>2</sub>. *B. nealsonii* DSM 15077<sup>T</sup> also differed in PSS<sub>2</sub> at four nucleotide positions and by insertion of two thymines. Since not all 16S rRNA sequences from *B. circulans* and *B. nealsonii* strains have been examined for these signatures, the PSS are termed presumptive. Strains CBD 118 and OSS 25 were identical for 461 bp spanning V1–V3 and differed from the other seven *Bacillus* strains by one bp in PSS<sub>2a</sub> (Table 1). In nucleotide blast searches of the GenBank database, eight sequences—including that of strain CBD 118—had 100% coverage and 100% identity to both PSS<sub>1B</sub> and PSS<sub>2a</sub>. When 1495 bp from *Bacillus* sp. CBD 118 (DQ374636) was used as the reference sequence, four near full length sequences from cultured strains (EU683686, *Bacillus* sp. OSS 25; DQ333291, *Bacillus benzoovorans* LLG; EU660368, *Bacillus nealsonii* CT18; GU471201, *Bacillus* sp. Q2CJ3) each had  $\geq 99.8\%$  similarity, suggesting all four were *Bacillus* strains more closely related to CBD 118 than to *B. benzoovorans* or *B. nealsonii* type strains. Searches coupling PSS<sub>1B</sub> with PSS<sub>2b</sub> or PSS<sub>2c</sub> identified other isolates with  $>99\%$  sequence similarity to strains C4T1F3B3, P307 and P308 and to IAFILS6, AD5A, U4A, and ADP4II, respectively.

**3.3. Phylogenetic Analysis.** In a neighbor-joining (N-J) tree (Figure 1) based on approximately 1390 bp of 16S rRNA sequence, strains most closely related to *B. circulans* and *B. nealsonii* were divided into two well-supported sister clades. Strain CBD 118 and *Bacillus* strains with  $\geq 99.7\%$  sequence similarity to CBD 118 formed a complex clade with *B. nealsonii* DSM 15077<sup>T</sup>. In the subtree (Figure 1), the *Bacillus* strains grouped according to PSS<sub>2</sub> type, but without strong bootstrap support. The *B. circulans* clade included *B. circulans* ATCC 4513<sup>T</sup>, one strain identified as *B. circulans* and two *Bacillus* spp. with  $\geq 99.5\%$  similarity to the type strain. The attribution of species-level identity based solely on 16S rRNA gene sequence similarity is known to be unreliable [6, 7] especially among *Bacillus* [36]. Keswani and Whitman [36] studied the relationship of 16S rRNA sequence similarity (S) to DNA:DNA hybridization (D). Among 40 *Bacillus*

TABLE 1: Presumptive signature sequences (PSS) conserved in 16S rRNA gene at nt positions 71–92<sup>a,b</sup> (PSS<sub>1</sub>) and 183–223<sup>a</sup> (PSS<sub>2</sub>) in respective V1 and V2 hypervariable regions of *Bacillus nealsonii*-related strains. <sup>a</sup>Numbering based on *rrnE* of *Bacillus subtilis* subsp. *subtilis* strain 168<sup>T</sup> (NC\_000964.3); <sup>b</sup>gap at nt position 80 in PSS<sub>1</sub> sequence alignments to *rrnE* is reflected in the numbering.

PSS type	<i>Bacillus</i> Strain no.	GenBank Accession no.	Nucleotides of PSS
PSS <sub>1A</sub>	ATCC 4513 <sup>T</sup>	AY724690	G A C T T T A A A <sup>b</sup> A G C T T G C T T T T A
	<i>B. circulans</i>		
PSS <sub>1B</sub>	N3 <i>B. circulans</i>	AB215100	G A C T T T A A A A G C T T G C T T T T A
	DSM 15077 <sup>T</sup>	EU656111	G A C T T A A A A A G C T T G C T T T T T
	<i>B. nealsonii</i>		
	CBD 118	DQ374636	G A C T T A A A A A G C T T G C T T T T T
	OSS 25	EU683686	G A C T T A A A A A G C T T G C T T T T T
	P307	FJ943260	G A C T T A A A A A G C T T G C T T T T T
	P308	FJ554672	G A C T T A A A A A G C T T G C T T T T T
	C4T1F3B3	FJ943258	G A C T T A A A A A G C T T G C T T T T T
	IAFILS6	FJ943259	G A C T T A A A A A G C T T G C T T T T T
	AD5A	FJ943256	G A C T T A A A A A G C T T G C T T T T T
	U4A	FJ943261	G A C T T A A A A A G C T T G C T T T T T
	ADP4II	FJ943257	G A C T T A A A A A G C T T G C T T T T T
PSS <sub>2a</sub>	CBD 118	DQ374636	CCTTTTCTACTCATGTAGGAAAG—CTGAAAGACGGTTTA—CGC
	OSS 25	EU683686	CCTTTTCTACTCATGTAGGAAAG—CTGAAAGACGGTTTA—CGC
	CT18	EU660368	CCTTTTCTACTCATGTAGGAAAG—CTGAAAGACGGTTTA—CGC
	Q2CJ3	GU471201	CCTTTTCTACTCATGTAGGAAAG—CTGAAAGACGGTTTA—CGC
PSS <sub>2b</sub>	P307	FJ943260	CCTTTCCTACTCATGTAGGAAAG—CTGAAAGACGGTTTA—CGC
	P308	FJ554672	CCTTTCCTACTCATGTAGGAAAG—CTGAAAGACGGTTTA—CGC
	C4T1F3B3	FJ943258	CCTTTCCTACTCATGTAGGAAAG—CTGAAAGACGGTTTA—CGC
PSS <sub>2c</sub>	IAFILS6	FJ943259	CCTTTTCTACTCATGTAGAAAAG—CTGAAAGACGGTTTA—CGC
	AD5A	FJ943256	CCTTTTCTACTCATGTAGAAAAG—CTGAAAGACGGTTTA—CGC
	U4A	FJ943261	CCTTTTCTACTCATGTAGAAAAG—CTGAAAGACGGTTTA—CGC
	ADP4II	FJ943257	CCTTTTCTACTCATGTAGAAAAG—CTGAAAGACGGTTTA—CGC
PSS <sub>2</sub> <i>B. nealsonii</i>	DSM 15077 <sup>T</sup>	EU656111	CCTTTTCTACTCATGTAGAGAAGTCTGAAAGACGGCATCTCGC
PSS <sub>2</sub> <i>B. circulans</i>	ATCC 4513 <sup>T</sup>	AY724690	CCTTTTCCTCTCATGAGGAAAAG—CTGAAAGACGGTTTA—CGC

spp., for an *S* of 0.997, 0.993 or 0.991, *D* could be expected to be <0.70 about 50, 95 or 99% of the time, respectively. Strains OSS 25 and WZ12 in the *B. nealsonii* clade were identified in GenBank as *B. circulans*, however each had <99% sequence similarity to *B. circulans* ATCC 4513<sup>T</sup>, indicating the strains do not belong to that species [36]. *Bacillus benzoovorans* DSM 5391<sup>T</sup> (D78311) was placed outside the *B. circulans* cluster. Strains C4T1F3B3, AD5A and U4A in the *B. nealsonii* clade were identified in GenBank as *B. benzoovorans*; however, each had 96% sequence similarity to *B. benzoovorans* DSM 5391<sup>T</sup>, significantly below the threshold for species level relatedness [19, 36]. Sequences of strains OSS 25, C4T1F3B3, AD5A and U4A, when compared to the EzTaxon database of curated type strain sequences, were most closely related to *B. nealsonii* DSM 15077<sup>T</sup> (99.3%, 99.3%, 99.4%, and 99.5%, resp.). Phylogenies of N-J, maximum parsimony, and maximum likelihood trees [31] (not shown) were consistent in placement of strains into the same clades described above, whether based on 1390 bp or 448 bp encompassing the hypervariable regions. Based on 16S rRNA gene sequence-based phylogeny, the strains gathered for this study were considered to be *Bacillus* spp., most closely related to *B. nealsonii*.

3.4. % DNA:DNA Similarity (% *S*). DNA:DNA pairings were conducted between *B. nealsonii* DSM 15077<sup>T</sup> and two strains representing each PSS<sub>2</sub> type—CBD 118, OSS 25 (PSS<sub>2a</sub>); P308, C4T1F3B3 (PSS<sub>2b</sub>); IAFILS6, AD5A (PSS<sub>2c</sub>) (Table 2). DNA:DNA% similarity (% *S*) data reveals clusters of closely-related strains termed genomic groups or genomic species and remains the “gold standard” for species circumscription [3, 5, 19, 37, 38]. More distantly related groups are separated by discontinuities, commonly in the range of 50–70% *S*; usually only a few intermediate strains are found [37, 38]. The % *S* for each of the six *Bacillus* strains to *B. nealsonii* DSM 15077<sup>T</sup> was less than 40%, well below the recommended 70% threshold value to be circumscribed in that species [3, 5]. The % *S* between PSS<sub>2a</sub> strains CBD 118 and OSS 25 were 49.9% and 55.5%. In a previous set of pairings, % *S* of 63.8 and 61.6 suggest differences in DNA quality between the two testing events. The % *S* for the four CBD 118 and OSS 25 pairings fall within the 70–50% transitional range for species circumscription [6, 7, 37]. Pairings of P308 (PSS<sub>2b</sub>), C4T1F3B3 (PSS<sub>2b</sub>) and IAFILS6 (PSS<sub>2c</sub>) with CBD 118 (PSS<sub>2a</sub>) resulted in at least one % *S* per strain at or very close to the 70% threshold for species delineation with CBD 118.

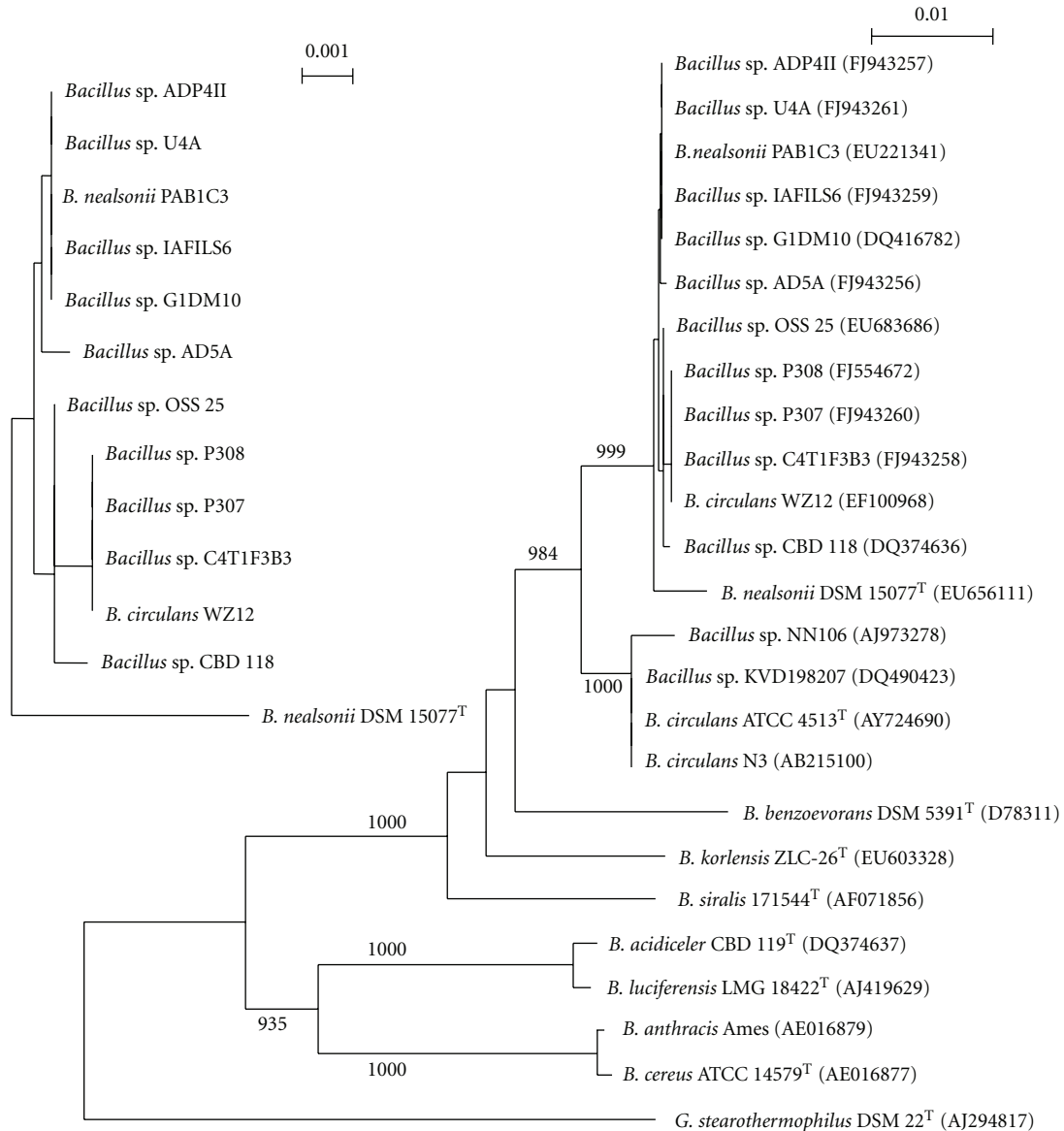


FIGURE 1: Phylogenetic relationship of strain CBD 118 with type strains of other *Bacillus* spp. and with strains from GenBank with highest similarities (accession numbers in parentheses). The Neighbor-Joining tree and subtree are based on approximately 1390 bp of the 16S rRNA gene. Numbers show the level of bootstrap support from 1000 repetitions with only those >750 shown. Bars, nucleotide substitutions per nucleotide.

TABLE 2: % DNA:DNA similarity between seven *Bacillus* species including *B. nealsonii* DSM 15077<sup>T</sup>, with reference to presumptive signature sequence 2 (PSS<sub>2</sub>) types. <sup>a</sup>Duplicate measurements from DNA:DNA hybridizations performed using spectrophotometric determinations.

Strain ID	CBD 118	DSM 15077 <sup>T</sup>	OSS 25	P308	C4T1F3B3	IAFILS6	AD5A
PSS <sub>2</sub> type	PSS <sub>2a</sub>	PSS <sub>2Bn</sub>	PSS <sub>2a</sub>	PSS <sub>2b</sub>	PSS <sub>2b</sub>	PSS <sub>2c</sub>	PSS <sub>2c</sub>
Strain ID							
CBD 118							
DSM 15077 <sup>T</sup>	17.8 (11.1) <sup>a</sup>						
OSS 25	49.9 (55.5)	13.9 (19.3)					
P308	63.6 (67.8)	38.6 (29.7)	76.7 (79.7)				
C4T1F3B3	70.8 (66.3)	15.6 (21.8)	85.6 (88.5)	94.0 (93.0)			
IAFILS6	67.9 (68.1)	30.6 (25.4)	72.2 (68.3)	86.2 (80.6)	67.3 (72.7)		
AD5A	49.2 (41.9)	8.1 (6.9)	35.6 (28.2)	19.7 (23.4)	15.6 (25.2)	19.3 (21.1)	

TABLE 3: Comparison of percent cellular fatty acid content between strain CBD 118, eight closely related strains, and most closely related type strains. 1: CBD 118; 2: OSS 25; 3: P307; 4: P308; 5: C4T1F3B3; 6: IAFILS6; 7: AD5A; 8: U4A; 9: ADP4II; 10: *Bacillus nealsonii* DSM 15077<sup>T</sup>; 11: *Bacillus circulans* ATCC 4513<sup>T</sup>. <sup>†</sup>C17:1 anteiso B/Iso I. Other summed features with composite less than 1% for all strains are not presented. CBD 118 and OSS 25 have PSS<sub>2a</sub>; P307, P308, and C4T1F3B3 have PSS<sub>2b</sub>; IAFILS6, AD5A, U4A, and ADP4II have PSS<sub>2c</sub>.

Strains	1	2	3	4	5	6	7	8	9	10	11
C <sub>13:0</sub> iso	1.55	1.69	0.91	0.94	1.15	0.72	0.48	0.72	0.76	3.41	0.67
C <sub>13:0</sub> anteiso	0.24	0.53	0.21	—	0.22	0.24	0.18	—	—	0.79	0.17
C <sub>14:0</sub> iso	2.52	2.95	1.48	1.73	1.82	1.28	1.81	2.96	1.81	2.39	4.78
C <sub>14:0</sub>	9.93	3.97	6.37	7.90	7.04	4.25	4.71	3.82	3.58	13.67	11.66
C <sub>15:0</sub> iso	25.95	20.57	20.17	21.53	23.24	16.59	14.68	21.42	21.36	29.28	20.25
C <sub>15:0</sub> anteiso	42.78	60.69	47.94	43.91	44.73	46.86	51.51	45.18	44.28	35.09	40.64
C <sub>16:1</sub> $\omega$ 7c alcohol	0.23	—	0.20	—	0.21	0.18	0.73	0.60	0.50	0.59	1.02
C <sub>16:0</sub> iso	1.07	1.32	0.92	1.04	1.08	1.22	1.09	2.54	1.67	0.74	4.36
C <sub>16:1</sub> $\omega$ 11c	2.66	0.33	2.82	2.02	2.23	2.06	4.37	2.69	2.79	4.69	2.64
C <sub>16:0</sub>	8.84	4.22	7.59	12.55	9.58	5.68	3.47	3.73	3.02	4.60	5.58
C <sub>15:0</sub> iso 3OH	—	0.33	—	—	—	—	—	0.96	1.31	—	—
C <sub>15:0</sub> 2OH	—	0.61	—	—	—	—	—	0.83	1.42	—	—
C <sub>17:1</sub> iso $\omega$ 10c	—	—	—	—	—	0.28	0.44	0.28	0.26	0.25	0.23
Sum in feature 4 <sup>†</sup>	—	—	0.34	—	—	0.50	2.17	1.30	1.20	0.36	0.40
C <sub>17:0</sub> iso	0.80	0.36	1.22	1.25	1.21	3.09	1.21	1.61	1.54	0.87	1.27
C <sub>17:0</sub> anteiso	3.41	2.42	9.35	6.90	7.49	16.11	12.15	8.84	10.23	3.27	6.07
C <sub>16:0</sub> iso 3OH	—	—	—	—	—	—	—	0.42	0.34	—	—
C <sub>16:0</sub> 2OH	—	—	—	—	—	—	—	0.21	0.54	—	—
C <sub>17:0</sub> iso 3OH	—	—	—	—	—	—	—	—	0.30	—	—
C <sub>18:0</sub>	—	—	0.26	0.23	—	0.28	0.19	—	—	—	—
C <sub>17:0</sub> 2OH	—	—	—	—	—	—	0.16	0.59	1.46	—	—
C <sub>19:0</sub> anteiso	—	—	0.22	—	—	0.66	0.63	0.34	0.66	—	—

Given the estimated 10% reproducibility of % S values (DSMZ), DNA:DNA pairings that have  $\geq 80\%$  S should meet or exceed the recommended 70% threshold to delineate taxa at the species level. Three of 21 strain pairings—P308 with C4T1F3B3; OSS 25 with C4T1F3B3; P308 with IAFILS6—tested at  $\geq 80\%$  S; therefore, these four strains appeared to represent a coherent genomic cluster. A diagram (Figure 2) in which the two measurements per pairing were averaged, illustrates varying degrees of genomic coherency among all 7 strains. Averaged % S of 93.5% strongly supports DNA relatedness between strains P308 (PSS<sub>2b</sub>) and C4T1F3B3 (PSS<sub>2b</sub>); % S of 83.4% also supports relatedness between P308 (PSS<sub>2b</sub>) and IAFILS6 (PSS<sub>2c</sub>). But relatedness between C4T1F3B3 (PSS<sub>2b</sub>) and IAFILS6 (PSS<sub>2c</sub>) is not similarly well supported at 70% S, thus the degree of relatedness of each strain to P308 was not reproduced in relation to each other. Also, within this apparent genomic cluster, averaged % S was 87.1% between OSS 25 (PSS<sub>2a</sub>) and C4T1F3B3 (PSS<sub>2b</sub>), 78.2% between OSS 25 (PSS<sub>2a</sub>) and P308 (PSS<sub>2b</sub>), but was 70.25% between OSS 25 (PSS<sub>2a</sub>) and IAFILS6 (PSS<sub>2c</sub>). A genomic cluster based on these four strains incorporates an  $\sim 24\%$  range for % S and values for two pairings that may—given  $\sim 10\%$  reproducibility—lie in the transitional range for species circumscription. Some strains of a species may show less than 70% S with the type strain or other strains of the same species, thus internal heterogeneity within genomic groupings and species is permitted [6, 7, 11, 16]. However,

studies of the average nucleotide identity (ANI) of all conserved genes between any two genomes [13, 14, 39] support adoption of a higher rather than a relaxed threshold for species circumscription. The 70% threshold for species delineation based on DNA:DNA pairings corresponds to 95% ANI and 85% or 79% conserved protein coding genes between a pair of strains [39], thus substantial phenotypic differences were possible among two or more of these four strains.

**3.5. Cellular Fatty Acid Analysis (CFA).** Cellular fatty acid compositions of the nine *Bacillus* strains, *B. circulans* ATCC 4513<sup>T</sup> and *B. nealsonii* DSM 15077<sup>T</sup> are compared in Table 3. Consistent with *Bacillus* [34], the major cellular fatty acids (CFA) measured in the strains were C<sub>14:0</sub>, C<sub>15:0</sub>-anteiso, C<sub>15:0</sub>-iso and C<sub>16:0</sub>. Profiles from a second CFA analysis performed at the end of study (not shown) were consistent with those in Table 3. The second data set deviated in the absence of 1–7 very low % CFAs (most  $< 0.5\%$ ; three  $< 1.5\%$ ) from the profiles of each of the strains, suggesting differences between the two testing events in the effective physiological age of the strains. The ability to reproduce profiles for a single strain is dependent on standardized conditions for growth medium, incubation time and temperature, and effective physiological age of the cells ([34], MIDI technical literature). In the second analysis, slight changes in values of major and other CFAs for  $\geq 9$  strains followed a parallel pattern of elevation (C<sub>15:0</sub>-iso, C<sub>17:0</sub>-iso) or reduction (C<sub>14:0</sub>, C<sub>16:1</sub>  $\omega$ 11c, C<sub>16:0</sub>),



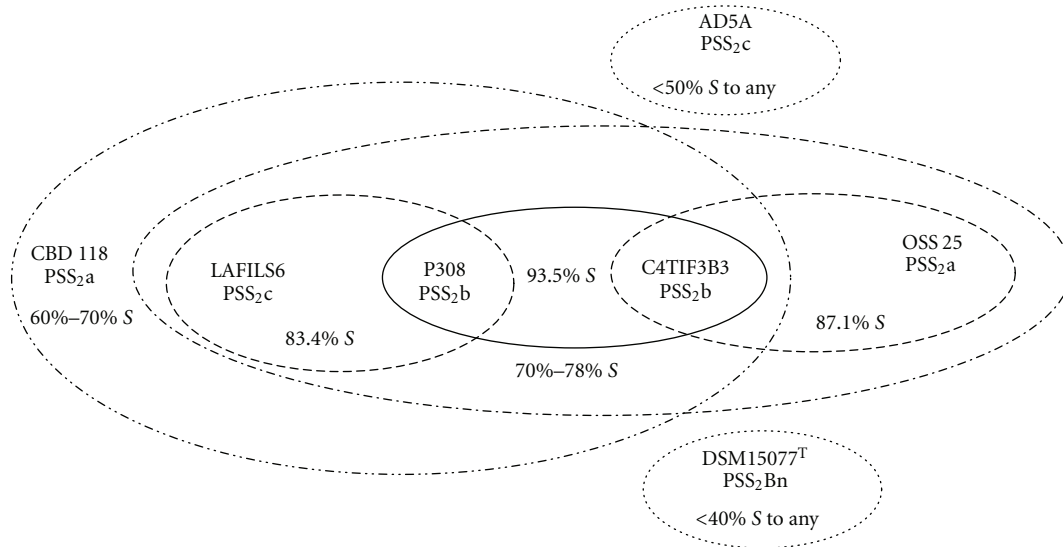


FIGURE 2: Diagram of averaged % DNA:DNA similarities (% S) between *Bacillus* strains DSM 15077<sup>T</sup>, CBD 118, OSS 25, P308, C4T1F3B3, IAFILS6, and AD5A with reference to the presumptive signature sequence 2 type (PSS<sub>2</sub>).

also suggesting differences in effective physiological age for those strains relative to the first testing event. However, salient differential CFAs were reproduced in the profiles of both data sets, supporting the authenticity of the strain set at both time points in our study. *C*<sub>15:0</sub>-anteiso was 60% in both profiles for strain OSS 25. Strain P307 was twice distinguished by the summed feature *C*<sub>17:1</sub>-anteiso B/Iso I and *C*<sub>19:0</sub>-anteiso. In both data sets, PSS<sub>2c</sub> strains IAFILS6, AD5A, U4A, and ADP4II were differentiated from all other strains by *C*<sub>17:1</sub> iso  $\omega$ 10c, the summed feature *C*<sub>17:1</sub> anteiso B/Iso I and *C*<sub>19:0</sub> anteiso.

CFA profiles are known to vary widely in many named *Bacillus* spp., thus circumscribing species based on CFA content is usually possible only in cases of genomically-homogeneous strains [34]. Therefore, we considered only the linkage of nearest neighbors without attribution of taxonomic level. In a dendrogram scaled to Euclidian distance (ED) and based on the initial data set (Figure 3), the three strains sharing PSS<sub>2b</sub>—P307, P308, and C4T1F3B3—clustered together at near 6 ED. Strain OSS 25 (PSS<sub>2a</sub>) was distantly linked at near 20 ED to the other *Bacillus* spp. including CBD118 (PSS<sub>2a</sub>). Among PSS<sub>2c</sub> strains, ADP4II linked to U4A at  $\leq 3$  ED and IAFILS6 linked to AD5A at 7.5 ED but the linkage between the two pairs was at  $\leq 13$  ED. In the dendrogram (not shown) based on the end of study data set, small cumulative differences in individual fatty acid percentages relative those in the initial data set resulted in changes in the level of ED linkage among strains. However, the profile similarities between strains P308 and C4T1F3B3, the 60% of *C*<sub>15:0</sub>-anteiso that distinguished strain OSS 25, and the differentiation of IAFILS6 by *C*<sub>17:1</sub> iso  $\omega$ 10c, summed feature *C*<sub>17:1</sub> anteiso B/Iso I and *C*<sub>19:0</sub> anteiso were all reproduced in both data sets. In both dendrograms, strains P308 and C4T1F3B3 were clustered at  $\leq 6$  ED, IAFILS6 clustered with AD5A at  $\geq 7.5$  ED, and OSS 25 was isolated at  $\geq 16$  ED. The genome-based cluster of four strains—P308, C4T1F3B3, OSS 25, and

IAFILS6—with DNA:DNA pairings of 94–70% S was not reproduced in the CFA-based dendrogram from either data set.

**3.6. Differential Phenotypic Characterization.** Twenty-five of 100 phenotypic traits differentiate among the nine *Bacillus* and the two most closely related type strains (Table 4). Type strains *B. circulans* ATCC 4513<sup>T</sup> and *B. nealsonii* DSM 15077<sup>T</sup> are distinguished from the other strains by lack of acetoin production and by acid from 2-ketogluconate. The numbers of characters that separate each pair of *Bacillus* strains were compiled in a matrix (Table 5) and presented in a dendrogram (Figure 4). The only consistency between the CFA-based (Figure 3) and phenotype-based dendrogram was the close linkage of two PSS<sub>2c</sub> strains, U4A and ADP4II—one of only two instances in which strains of the same PSS<sub>2</sub> type were directly linked in the phenotype-based dendrogram. Only 3 of 36 phenotypic pairings resulted in  $\leq 5$  character differences while 19 strain pairs had  $\geq 10$  differences (highest number = 13). As a group, strains IAFILS6, AD5A, U4A and ADP4II sharing PSS<sub>2c</sub> formed the most coherent group with between 3–8 character differences. But in the phenotype dendrogram, closely paired strains U4A and ADP4II clustered with P307 (PSS<sub>2b</sub>), while AD5A links with OSS 25 (PSS<sub>2a</sub>) and C4T1F3B3 (PSS<sub>2b</sub>). Strains U4A and ADP4II differed by one CFA (Table 3) and 3 phenotypic characters, and may represent strain variants of a novel species. Only 6 or 5 characters differentiated P307 (PSS<sub>2b</sub>) from U4A and ADP4II, respectively. DNA:DNA pairing data on these three strains is not, however, available for comparison. The four strains—P308, C4T1F3B3, OSS 25, and IAFILS6—that comprised the genomic-based cluster were differentiated by 9–13 characters, including nitrate reduction, temperature range, and acid production from carbohydrates.

**3.7. Incongruence of Character Sets and Application of a Bacterial Species Definition.** The taxonomic species definition



TABLE 4: Comparison of differential phenotypic traits for strain CBD 118, eight closely related strains, and most closely related type strains. Strains: 1, *Bacillus circulans* ATCC 4513<sup>T</sup>; 2, *Bacillus nealsonii* DSM 15077<sup>T</sup>; 3, CBD 118; 4, OSS 25; 5, P307; 6, P308; 7, C4T1F3B3; 8, IAFILS6; 9, AD5A; 10, U4A; 11, ADP4II. <sup>a</sup>Approximate mean diameter of 5 well-isolated colonies grown  $24 \pm 2$  h, 30°C on tryptic soy agar plates. <sup>b</sup>Larger, more opaque zone of hydrolysis for these strains. Subtle, <50% increase in diameter of sporangium for preponderance of cells field<sup>-1</sup>; Overt,  $\geq 50\%$  increase in diameter of sporangium for  $\geq 25\%$  of cells field<sup>-1</sup>; +: positive; -: negative; Scant: optical density just above threshold for unaided detection; v: variable, indicating both + and - reactions for strain in  $\geq 4$  tests. Scant and v were differentiated from both + and - reactions in pairwise tabulation of differential traits.

Characteristic	1	2	3	4	5	6	7	8	9	10	11
Swelling of sporangium	Subtle	—	Subtle	Subtle	Overt	Subtle	Overt	Overt	Overt	Overt	Overt
Colony diameter (mm) <sup>a</sup>	2	2	2	0.75	2	2	2	4	1.5	4	3
Hemolysis reaction	—	—	—	alpha	—	—	—	—	—	—	—
Growth at 45°C	+	Scant	Scant	Scant	+	Scant	+	—	—	+	+
50°C	+	—	—	—	—	—	—	—	—	—	—
Growth with 7% NaCl	+	—	+	+	+	—	+	+	+	+	+
Tween-80 hydrolysis	+	—	—	—	Scant	—	—	+	+ <sup>b</sup>	+ <sup>b</sup>	+ <sup>b</sup>
Acetoin production	—	—	+	+	+	+	+	+	+	+	+
Nitrate reduction	—	—	+	—	+	+	—	+	—	+	+
Acid from L-Arabinose	+	+	—	—	—	v	—	—	—	—	—
D-Arabitol	—	v	+	+	+	+	+	+	—	+	+
Dulcitol	—	—	—	—	—	—	+	—	—	—	—
$\alpha$ -methyl-D-Glucoside	+	+	+	—	+	+	v	v	+	+	+
Glycerol	+	+	+	—	v	+	—	v	—	—	+
Glycogen	+	—	v	—	v	+	—	—	—	—	—
Inulin	+	—	—	—	—	—	—	—	—	—	—
2-Ketogluconate	+	+	—	—	—	—	—	—	—	—	—
5-Ketogluconate	—	v	—	+	+	+	+	+	—	—	—
D-Lyxose	—	+	—	—	—	—	—	—	—	—	—
$\alpha$ -methyl-D-Mannoside	+	v	v	v	—	—	—	—	—	—	—
Sorbitol	+	+	+	+	+	+	—	—	—	+	+
D-Tagatose	—	+	—	—	—	—	—	—	—	—	—
Xylitol	+	—	v	—	—	—	v	—	—	—	—
D-Xylose	+	+	—	—	+	+	—	v	v	v	+
$\beta$ -methyl-D-Xyloside	+	v	—	v	+	+	v	—	+	+	+

TABLE 5: Number of differential phenotypic traits between paired *Bacillus* strains among 100 traits tested with reference to the presumptive signature sequence 2 type (PSS<sub>2</sub>).

Strain ID	CBD 118	OSS 25	P307	P308	C4T1F3B3	IAFILS6	AD5A	U4A	ADP4II
PSS <sub>2</sub> type	PSS <sub>2a</sub>	PSS <sub>2a</sub>	PSS <sub>2b</sub>	PSS <sub>2b</sub>	PSS <sub>2b</sub>	PSS <sub>2c</sub>	PSS <sub>2c</sub>	PSS <sub>2c</sub>	PSS <sub>2c</sub>
CBD 118									
OSS 25	9								
P307	8	12							
P308	8	11	7						
C4T1F3B3	11	9	10	13					
IAFILS6	12	12	8	12	9				
AD5A	13	12	10	13	10	7			
U4A	10	11	6	10	10	6	5		
ADP4II	9	12	5	8	11	8	7	3	

mandates that a species be a monophyletic group with a high degree of genomic similarity that also shares a high order of similarity in many independent phenotypic features [3, 5, 10, 11]. The eight strains collected for comparison to strain CBD 118 are monophyletic (Figure 1) when considering 16S

rRNA gene similarity. While recognizing that 16S rRNA sequence lacks resolving power at the level of bacterial species [6, 7, 11, 36], we hypothesized that the PSS types might yet function as exclusionary thresholds, for example, species that shared a PSS<sub>2</sub> type might or might not be the same

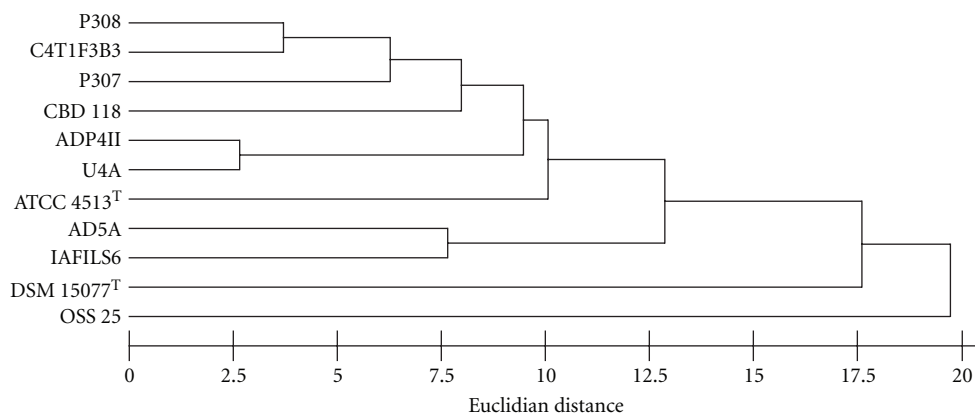


FIGURE 3: Dendrogram based on cluster analysis of cellular fatty acid composition for 11 *Bacillus* strains including *B. circulans* ATCC 4513<sup>T</sup> and *B. nealsonii* DSM 15077<sup>T</sup>, in Euclidian Distance units.

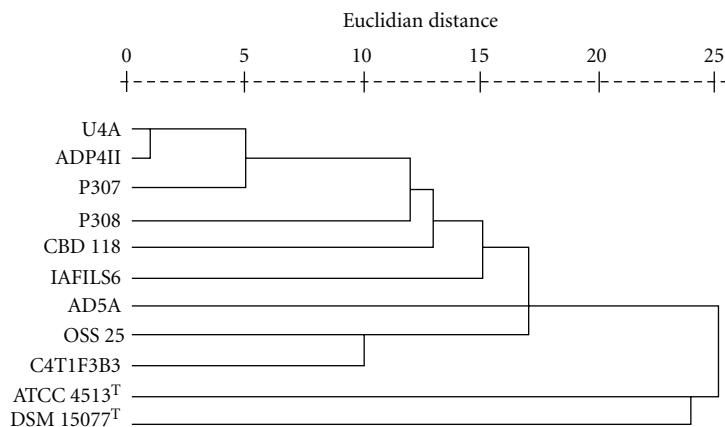


FIGURE 4: Dendrogram based on cluster analysis of differential phenotypic characters among 11 *Bacillus* strains including *B. circulans* ATCC 4513<sup>T</sup> and *B. nealsonii* DSM 15077<sup>T</sup>, in Euclidian Distance units.

species, but strains with different PSS<sub>2</sub> types would not be the same species. While the highest degree of DNA relatedness (93.5% *S*) was between PSS<sub>2</sub>b strains P308 and C4T1F3B3 (Figure 2), the hypothesis of an exclusionary threshold was contradicted by  $\geq 70\%$  *S* between strains of different PSS<sub>2</sub> types. However, we suggest that the PSS<sub>2</sub> types remain effective tools to search 16S rRNA sequence databases for more strains of  $>99.7\%$  similarity.

No strain tested at greater than 70%–50% *S* in pairings to CBD 118 (Table 2) and the three strains most closely-related to CBD 118 (PSS<sub>2</sub>a) based on % *S*—P308 (PSS<sub>2</sub>b), C4T1F3B3 (PSS<sub>2</sub>b), and IAFILS6 (PSS<sub>2</sub>c)—can be differentiated by 8, 11 and 12 characters, respectively (Table 5). Strain OSS 25, most closely related to CBD 118 based on 16S rRNA sequence similarity and PSS<sub>2</sub>a, differs by 9 phenotypic characters as well as having only transitional range % *S* to CBD 118 and distant linkage based on CFA. It is recommended for *Bacillus* and related genera [6] that the 70% *S* threshold for species delineation not stand alone in delimiting species but should be supported by other characteristics that differentiate strains of the proposed species from other species. In the application of the

taxonomic species definition, phenotype continues to have a salient role in the determination of break-points in genomic data for species circumscription and no single parameter—genomic properties or phenotypic traits—should be given undue prominence [3, 6, 37]. The classification that results from application of the taxonomic species definition should be predictive, establishing determinative properties and therefore cannot be based only on genomic characters [3, 7, 11]. The 70% threshold could be interpreted flexibly [6, 7, 11, 16, 37] and a more relaxed boundary used to circumscribe a genomic grouping of these four strains with CBD 118. The resultant grouping would, however, lack sufficient phenotypic cohesion to be of predictive value and therefore does not justify circumscription as a taxonomic species.

In polyphasic taxonomic studies when the strains and phenotypic characters tested were both sufficiently numerous, the resultant clustering pattern has generally reproduced the genomic grouping [7]. In this instance, the four strains with highest % *S* to support species circumscription are differentiated by multiple phenotypic, species level discriminators (Tables 2, 3, 5). Strain OSS 25 (PSS<sub>2</sub>a) paired with

P308 (PSS<sub>2</sub>b) and with C4T1F3B3 (PSS<sub>2</sub>b) at 78% and 87% *S* (Figure 2), but were differentiated by 11 and 9 characters respectively, as well as significant differences in CFA profiles. In the phenotype dendrogram (Figure 4), OSS 25 was linked most closely with C4T1F3B3 but not with P308. Strains P308 (PSS<sub>2</sub>b) and IAFILS6 (PSS<sub>2</sub>c) share 83% *S* but are differentiated by CFA profiles and 12 traits. Strains P308 (PSS<sub>2</sub>b) and C4T1F3B3 (PSS<sub>2</sub>b) share the strongest DNA relatedness with 94% *S* and were closely linked based on CFA profiles but can be differentiated by 13 phenotypic characters and failure to be linked in the phenotype dendrogram. At the end of the study, these four strains were subcultured from cryopreserved stock, retested for six differential phenotypic traits and resequenced for the 16S rRNA gene. The resultant sequence for each strain was subjected to BLAST analysis and in each case resulted in a 100% match to the region of overlap with the ~1500 bp previously accessioned into GenBank for the strain. For each strain, the re-evaluation of six phenotypic characters—degree of endospore-driven swelling, colony diameter, hemolysis reaction, growth at 45°C, growth with 7% NaCl, and nitrate reduction—reproduced the results of previous testing shown in Table 4. These results indicate that the authenticity of these strains was maintained though the course of the study. The internal diversity of strains P308, C4T1F3B3, OSS 25 and IAFILS6 confounds delineation in a phenotypically coherent unit and their circumscription as one species accommodating multiple biovars or ecovars does not, in our minds, support a predictive taxonomy. No common ecological or disease state can be cited to justify the nomination of a pragmatic species epithet for these strains. The designation of genomovars, as originally proposed by Ursing et al. [38], applies to two or more genomic strain clusters within a phenotypically coherent named species that cannot be phenotypically delimited from other strains of the nomenospecies. With these four strains, the converse is the case—one apparent genomic group of strains with four differential phenotypes. It is possible that each of these four strains is the sole exemplar of a novel species and that cohesive phenotypic clusters await the isolation and robust polyphasic characterization of more sibling strains. On the other hand, MLSA [20] on these and more sibling isolates could support the description of one or more species with a high degree of intraspecies diversity—thereupon, a species description could be justified. At this point, rather than being reinforced by coherent phenotypic clustering, potentially coherent genomic clusters among strains are contradicted by interstrain variability and are not therefore effectively circumscribed within the taxonomic species definition.

Difficulties in applying the taxonomic species definition are not new—see the taxonomic histories of *Pseudomonas stutzeri* [38] and *Acinetobacter* [40] to cite just two—whereas these nine *Bacillus* strains are demonstrably novel and their degrees of relatedness appear to confound the taxonomic species definition. Polyphasic data did not clarify relationships and illuminate coherent clusters among these strains—instead, potentially “transitional” forms were revealed. While acknowledging the current insufficiency of our data set, these strains are reminiscent of Model 9 of Istock et al. [12], “Highly variable partially recombining nonspecies”, in which

clusters of strains may be discerned, but transitional strains erase any clear demarcation between clusters. Likewise, these strains may be an example of the “continuum of diversity” suggested to characterize groups in which forces promoting coherence dominate those promoting divergence of populations [13]. More data is required to clarify relationships among these strains—particularly sampling more strains in order to determine the range of variation and whether or not discrete phenotypic clusters exist. Indeed, it is hoped that researchers holding closely related strains recognizable by 100% identity to the PSS<sub>2</sub> types will join in collaborating with labs having expertise in recommended methods of *Bacillus* identification [6] to characterize an expanded number of strains. To this end, the nine *Bacillus* strains in this study have been deposited in a publicly accessible culture collection.

## Authors' Contribution

K. K. Peak and K. E. Duncan contributed equally to this work.

## Acknowledgments

The authors appreciate strains graciously provided by Chiara Alisi, Environmental Microbiology, UTPRA-GEOC, Ente per le Nuove tecnologie, l'Energia e l'Ambiente (ENEA-Casaccia), Rome, Italy (OSS 25); Marta Gomez-Chiarri, Dept. of Fisheries, Animal and Veterinary Sciences, University of Rhode Island (C4T1F3B3); Richard Villemur, INRS-Institut Armand-Frappier, Laval, QC, Canada (IAFILS6); Malka Halpern, Dept. of Biology, University of Haifa, Oranim, Tivon, Israel (AD5A, U4A, ADP4II) as well as invaluable technical assistance provided by William Veguilla. This HBOI contribution number 1838. This work was supported by the Department of the Army, U.S. Army Research, Development and Engineering Command, contract DAAD13-01-C-0043, and Joint Program Executive Office for Chemical and Biological Defense, JPM Guardian, Contract no. W911SR-07-C-0084.

## References

- [1] V. A. Luna, D. S. King, K. K. Peak et al., “*Bacillus anthracis* virulent plasmid pX02 genes found in large plasmids of two other *Bacillus* species,” *Journal of Clinical Microbiology*, vol. 44, no. 7, pp. 2367–2377, 2006.
- [2] K. K. Peak, K. E. Duncan, W. Veguilla et al., “*Bacillus acidiceler* sp. nov., isolated from a forensic specimen, containing *Bacillus anthracis* pX02 genes,” *International Journal of Systematic and Evolutionary Microbiology*, vol. 57, no. 9, pp. 2031–2036, 2007.
- [3] E. Stackebrandt, W. Frederiksen, G. M. Garrity et al., “Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology,” *International Journal of Systematic and Evolutionary Microbiology*, vol. 52, no. 3, pp. 1043–1047, 2002.
- [4] E. Stackebrandt and B. M. Goebel, “Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology,” *International*

- Journal of Systematic Bacteriology*, vol. 44, no. 4, pp. 846–849, 1994.
- [5] L. G. Wayne, D. J. Brenner, R. R. Colwell et al., "Report of the *ad hoc* committee on the recommendation of approaches to bacterial systematics," *International Journal of Systematic Bacteriology*, vol. 37, no. 4, pp. 463–464, 1987.
  - [6] N. A. Logan, O. Berge, A. H. Bishop et al., "Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria," *International Journal of Systematic and Evolutionary Microbiology*, vol. 59, no. 8, pp. 2114–2121, 2009.
  - [7] R. Rosselló-Móra and R. Amann, "The species concept for prokaryotes," *FEMS Microbiology Reviews*, vol. 25, no. 1, pp. 39–67, 2001.
  - [8] B. J. Tindall, R. Rosselló-Móra, B.-J. Huse, W. Ludwig, and P. Kämpfer, "Notes on the characterization of prokaryote strains for taxonomic purposes," *International Journal of Systematic and Evolutionary Microbiology*, vol. 60, no. 1, pp. 249–266, 2010.
  - [9] K. Goto, T. Omura, Y. Hara, and Y. Sadaie, "Application of the partial 16S rDNA sequence as an index for rapid identification of species in the genus *Bacillus*," *Journal of General and Applied Microbiology*, vol. 46, no. 1, pp. 1–8, 2000.
  - [10] M. Buckley and R. J. Roberts, *Reconciling Microbial Systematics & Genomics, a Report Based on a Colloquium, Sponsored by the American Academy of Microbiology*, ASM Press, Washington, DC, USA, 2006.
  - [11] R. Rossello-Mora and P. Kämpfer, "Defining microbial diversity—the species concept for prokaryotic and eukaryotic microorganisms," in *Microbial Diversity and Bioprospecting*, A. T. Bull, Ed., pp. 29–39, ASM Press, Washington, DC, USA, 2004.
  - [12] C. A. Istock, J. A. Bell, N. Ferguson, and N. L. Istock, "Bacterial species and evolution: theoretical and practical perspectives," *Journal of Industrial Microbiology and Biotechnology*, vol. 17, no. 3–4, pp. 137–150, 1996.
  - [13] K. T. Konstantinidis and J. M. Tiedje, "Genomic insights that advance the species definition for prokaryotes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 7, pp. 2567–2572, 2005.
  - [14] K. T. Konstantinidis, A. Ramette, and J. M. Tiedje, "The bacterial species definition in the genomic era," *Philosophical Transactions of the Royal Society B*, vol. 361, no. 1475, pp. 1929–1940, 2006.
  - [15] J. Maynard Smith, N. H. Smith, M. O'Rourke, and B. G. Spratt, "How clonal are bacteria?" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 10, pp. 4384–4388, 1993.
  - [16] P. Vandamme, B. Pot, M. Gillis, P. De Vos, K. Kersters, and J. Swings, "Polyphasic taxonomy, a consensus approach to bacterial systematics," *Microbiological Reviews*, vol. 60, no. 2, pp. 407–438, 1996.
  - [17] W. P. Hanage, C. Fraser, and B. G. Spratt, "Sequences, sequence clusters and bacterial species," *Philosophical Transactions of the Royal Society B*, vol. 361, no. 1475, pp. 1917–1927, 2006.
  - [18] F. M. Cohan, "What are bacterial species?" *Annual Review of Microbiology*, vol. 56, pp. 457–487, 2002.
  - [19] E. Stackebrandt and J. Ebers, "Taxonomic parameters revisited: tarnished gold standards," *Microbiology Today*, vol. 33, no. 6, pp. 152–155, 2006.
  - [20] M. C. J. Maiden, J. A. Bygraves, E. Feil et al., "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 6, pp. 3140–3145, 1998.
  - [21] A. R. Sprocati, C. Alisi, F. Tasso, L. Segre, and C. Cremisini, "Comparison of microbial communities native to three differently polluted ecological niches in the industrial site of Bagnoli (Naples, Italy)," in *Modern Multidisciplinary Applied Microbiology, Exploiting Microbes and Their Interactions*, A. Meddez-Vilas, Ed., pp. 488–493, Wiley-VCH, Weinheim, Germany, 2006.
  - [22] E. Gauger, R. Smolowitz, K. Uhlinger, J. Casey, and M. Gómez-Chiarri, "*Vibrio harveyi* and other bacterial pathogens in cultured summer flounder, *Paralichthys dentatus*," *Aquaculture*, vol. 260, no. 1–4, pp. 10–20, 2006.
  - [23] I. Lafortune, P. Juteau, E. Déziel, F. Lépine, R. Beaudet, and R. Villemur, "Bacterial diversity of a consortium degrading high-molecular-weight polycyclic aromatic hydrocarbons in a two-liquid phase biosystem," *Microbial Ecology*, vol. 57, no. 3, pp. 455–468, 2009.
  - [24] M. Halpern, D. Raats, and S. Lev-Yadun, "Plant biological warfare: thorns inject pathogenic bacteria into herbivores," *Environmental Microbiology*, vol. 9, no. 3, pp. 584–592, 2007.
  - [25] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
  - [26] Z. Zhang, S. Schwartz, L. Wagner, and W. Miller, "A greedy algorithm for aligning DNA sequences," *Journal of Computational Biology*, vol. 7, no. 1–2, pp. 203–214, 2000.
  - [27] D. A. Benson, M. S. Boguski, D. J. Lipman et al., "GenBank," *Nucleic Acids Research*, vol. 27, no. 1, pp. 12–17, 1999.
  - [28] J. Chun, J. H. Lee, Y. Jung et al., "EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences," *International Journal of Systematic and Evolutionary Microbiology*, vol. 57, no. 10, pp. 2259–2261, 2007.
  - [29] J. R. Cole, Q. Wang, E. Cardenas et al., "The ribosomal database project: improved alignments and new tools for rRNA analysis," *Nucleic Acids Research*, vol. 37, no. 1, pp. D141–D145, 2009.
  - [30] S. Chakravorty, D. Helb, M. Burday, N. Connell, and D. Alland, "A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria," *Journal of Microbiological Methods*, vol. 69, no. 2, pp. 330–339, 2007.
  - [31] J. Felsenstein, "PHYLIP, phylogeny inference package, version 3.68," 2008, <http://evolution.gs.washington.edu/phylip.html>.
  - [32] J. De Ley, H. Cattoir, and A. Reynaerts, "The quantitative measurement of DNA hybridization from renaturation rates," *European Journal of Biochemistry*, vol. 12, no. 1, pp. 133–142, 1970.
  - [33] V. A. R. Huss, H. Festl, and K. H. Schleifer, "Studies on the spectrophotometric determination of DNA hybridization from renaturation rates," *Systematic and Applied Microbiology*, vol. 4, no. 2, pp. 184–192, 1983.
  - [34] N. Logan and P. De Vos, "Genus I. *Bacillus* Cohn 1872, 174<sup>AL</sup>," in *The Firmicutes, Bergey's Manual of Systematic Bacteriology*, vol. 3, pp. 21–128, Springer, New York, NY, USA, 2009.
  - [35] R. M. Atlas, "Tween-80 Hydrolysis Medium," in *Handbook of Microbiological Media*, L. C. Parks, Ed., p. 957, CRC Press, Boca Raton, Fla, USA, 1993.
  - [36] J. Keswani and W. B. Whitman, "Relationship of 16S rRNA sequence similarity to DNA hybridization in prokaryotes," *International Journal of Systematic and Evolutionary Microbiology*, vol. 51, no. 2, pp. 667–678, 2001.
  - [37] R. Rosselló-Móra, "DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation," in *Molecular Identification, Systematics, and Population Structure*



of *Prokaryotes*, E. Stackebrandt, Ed., pp. 23–50, Springer, Berlin, Germany, 2006.

- [38] J. B. Ursing, R. A. Rosselló-Móra, E. Garcia-Valdes, and J. Lalucat, “Taxonomic note: a pragmatic approach to the nomenclature of phenotypically similar genomic groups,” *International Journal of Systematic Bacteriology*, vol. 45, no. 3, p. 604, 1995.
- [39] J. Goris, K. T. Konstantinidis, J. A. Klappenbach, T. Coenye, P. Vandamme, and J. M. Tiedje, “DNA-DNA hybridization values and their relationship to whole-genome sequence similarities,” *International Journal of Systematic and Evolutionary Microbiology*, vol. 57, no. 1, pp. 81–91, 2007.
- [40] L. Dijkshoorn and A. Nemec, “The diversity of the genus *Acinetobacter*,” in *Acinetobacter, Molecular Biology*, pp. 1–34, Caister Academic Press, Norfolk, UK, 2008.