Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Programa de Pós-Graduação em Genética e Biologia Molecular

Defining the Genus Boundaries for the Paenibacillaceae Family Using Comparative Genomic Analysis

Autor: Felipe Lhywinskh Guella

Orientadora: Dra. Luciane Maria Pereira Passaglia

Porto Alegre

Agosto de 2023

Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Programa de Pós-Graduação em Genética e Biologia Molecular

Defining the Genus Boundaries for the Paenibacillaceae Family Using Comparative Genomic Analysis

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular)

Fonte Financiadora

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

AGRADECIMENTOS

À minha orientadora, pela paciência e tolerância que foram necessárias para que esse trabalho se tornasse realidade.

Aos professores e funcionários do Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM), por construírem um ambiente acolhedor e intelectualmente estimulante.

Aos meus colegas de laboratório e de pós-graduação, por sempre oferecerem suporte nos momentos de necessidade.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), por conceder a bolsa de pós-graduação que possibilitou o desenvolvimento desse trabalho.

SUMMARY

ABBREVIATION LIST

AAI – average amino acid identity

ANI – average nucleotide identity

DDH – DNA-DNA Hybridization

GBDP – Genome Blast Distance Phylogeny

GGDC – genome-to-genome distance calculator

ICNP – International Code of Nomenclature of Prokaryotes

ICSP – International Committee on Systematics of Prokaryotes

LPSN – List of Prokaryotic names with Standing in Nomenclature

MLSA – multi-locus sequence analysis

NCBI – National Center for Biotechnology Information

OGRI – overall genome relatedness index

POCP – percentage of conserved protein

RESUMO

Os avanços na área de sequênciamento de genoma completo de procariotos provocaram um aumento significativo na quantidade de informações disponíveis em bancos de dados públicos para pesquisadores da área de sistemática bacteriana. Em consequência disso, esforços têm sido feitos para desenvolver metodologias, usualmente chamadas de OGRIs, que utilizam todo o genoma de um organismo para ajudar a definir sua posição taxonômica. Com isso em mente, este trabalho utilizou algumas dessas técnicas desenvolvidas, como ANI e AAI, para avaliar a situação taxonômica da família Paenibacillaceae utilizando sequências depositadas no banco de dados RefSeq do NCBI. A primeira parte deste trabalho utilizou genômica comparativa e análise filogenética para comprovar que as sequências de cepas definidas como de espécies distintas de *Paenibacillus* na verdade são subespécies uma da outra. Na segunda parte, sequências do gênero *Paenibacillus* que tiveram resultados inconclusivos para sua situação taxonômica dentro do RefSeq, foram comparadas com sequências de cepas tipo de outros gêneros da família. Foram encontradas diversas sequências que, ou pertenciam a outro gênero, ou não pertenciam a nenhum gênero analisado. Na terceira parte, o foco foi para melhor definir as fronteiras que separam os gêneros dentro da família e, ao mesmo tempo, encontrar grupos de sequências que potencialmente pertencem a um gênero ainda não descrito. Para isso, todas as sequências da família Paenibacillaceae disponíveis no RefSeq foram baixadas e o teste AAI foi utilizado para isolar grupos dentro da família e, concomitantemente, selecionar uma sequência genômica em cada grupo que pudesse servir como referência de comparação. Adicionalmente, também foi feita uma análise filogenética utilizando MLSA para corroborar os resultados do teste genômico e auxiliar na seleção de grupos. Com isso, foi possível identificar diversos grupos que aparentemente pertencem a gêneros ainda não descritos, e grupos com membros representando mais de um gênero. Contudo, muitos gêneros da família ainda não possuem sequências de suas espécies depositadas no RefSeq para verificar se esses grupos isolados pertençam a esses gêneros já descritos. Finalmente, este trabalho demonstrou os benefícios de se utilizar genômica comparativa para ajudar a definir as fronteiras entre gêneros dentro da família Paenibacillaceae através da seleção de uma sequência referência para comparação.

ABSTRACT

Advances in whole genome sequencing of prokaryotes resulted in a significant increase in the amount of information available in public databases for researchers on the bacterial systematics field. Consequently, efforts have been made to develop methodologies, usually called OGRIs, that use the whole genome of an organism to help define its taxonomic status. With that in mind, I used some of those methods developed, such as ANI and AAI, to evaluate the taxonomic status of the Paenibacillaceae family using sequences deposited in the NCBI RefSeq database. In the first part of this work, I used comparative genomic analysis and phylogeny to prove that the sequences of two strains defined as distinct species of the *Paenibacillus* genus are in fact subspecies of each other. In the second part, sequences of the Paenibacillus genus that showed inconclusive results for their taxonomic status on RefSeq, were compared against sequences of type strains from other genera of the family. Several sequences were found to either belong to another genus, or do not belong to any genus evaluated. On the third part, the focus was to better define the boundaries that separate the genera of the family and, at the same time, find sequence groups that potentially belong to a genus not yet described. To do that, all sequences of the Paenibacillaceae family available at RefSeq were downloaded and the AAI test was used to isolate groups within the family and, concomitantly, select a genome sequence that could serve as a referential for comparison. Additionally, a phylogenetic analysis, using MLSA, was also made to corroborate the genomic test results and assist in the group selection. With those results, it was possible to identify several groups that apparently belong to genera not yet described, as well as groups with members of more than one genus. However, several genera of the family do not have sequences of its species deposited in RefSeq yet to verify if those isolated groups belong to those already described genus. Finally, this work demonstrated the benefits of using comparative genomics to help define the boundaries between genera within the Paenibacillaceae family through the selection of a reference sequence for comparison.

Objectives

1. **Primary**

- Evaluate the taxonomic status of the Paenibacillaceae family using whole genomic analysis and phylogenomics.

2. **Secondary**

- Identify and propose a reference strain for each genus in the family to use whole genome comparison for genus boundary delineation.
- Propose new monophyletic genera within the family using whole genomic comparison.

Chapter 1 – First Paper

Genome-based reclassification of *Paenibacillus panacisoli* DSM 21345^T as *Paenibacillus massiliensis* subsp. *panacisoli* subsp. nov. and description of *Paenibacillus massiliensis* subsp. *massiliensis* subsp. nov.

Fabiana Tonial[1]†, Felipe Guella[2]†, Luciane Maria Pereira Passaglia[2] and Fernando Hayashi Sant'Anna[2,*]

Author affiliations: [1]Departamento de Patologia, Instituto de Ciências Biológicas, Universidade de Passo Fundo, Passo Fundo, Brazil; [2]Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

*Correspondence: Fernando Hayashi Sant'Anna, santanna.fh@gmail.com

†These authors contributed equally to this work

Published at *International Journal of Systematic and Evolutionary Microbiology (IJSEM)*

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC NOTE

Tonial *et al.*, *Int. J. Syst. Evol. Microbiol.* 2020;70:4838–4842

DOI 10.1099/ijsem.0.004328

MICROBIOLOGY SOCIETY

Genome-based reclassification of *Paenibacillus panacisoli* DSM 21345^T as *Paenibacillus massiliensis* subsp. *panacisoli* subsp. nov. and description of *Paenibacillus massiliensis* subsp. *massiliensis* subsp. nov.

Fabiana Tonial[1]†, Felipe Guella[2]†, Luciane Maria Pereira Passaglia[2] and Fernando Hayashi Sant'Anna[2,*]

Abstract

Bacteria of the genus *Paenibacillus* are relevant to humans, animals and plants. The species *Paenibacillus massiliensis* and *Paenibacillus panacisoli* are Gram-stain-positive and endospore-forming bacilli isolated from a blood culture of a leukemia patient and from soil of a ginseng field, respectively. Comparative analyses of their 16S rRNA genes revealed that the two *Paenibacillus* species could be synonyms (99.3% sequence identity). In the present study we performed different genomic analyses in order to evaluate the phylogenetic relationship of these micro-organisms. *Paenibacillus massiliensis* DSM 16942^T and *P. panacisoli* DSM 21345^T presented a difference in their G+C content lower than 1 mol%, overall genome relatedness index values higher than the species circumscription thresholds (average nucleotide identity, 95.57%; genome-wide ANI, =96.51%; and orthologous ANI, 96.25%), and a monophyletic grouping pattern in the phylogenies of the 16S rRNA gene and the proteome core. Considering that these strains present differential biochemical capabilities and that their computed digital DNA–DNA hybridization value is lower than the cut-off for bacterial subspecies circumscription, we suggest that each of them form different subspecies of *P. massiliensis*, *Paenibacillus massiliensis* subsp. *panacisoli* subsp. nov. (type strain DSM 21345^T) and *Paenibacillus massiliensis* subsp. *massiliensis* subsp. nov. (type strain DSM 16942^T).

Paenibacillus is the type genus of the family *Paenibacillaceae* and is the most studied group of the family due to its diversity and its ecological and economic relevance [1]. Bacteria from this genus are able to survive in a broad spectrum of habitats, from extreme environments such as volcanic soils [2] and the Antarctic continent [3], to the earthworm gut [4]. Nonetheless, many of them are commonly found associated with plants [1], and some of them present plant growth-promoting abilities [5–10]. Species of this genus have heterogeneous characteristics, but most are Gram-stain–positive, endospore-forming and facultative anaerobic bacilli [11]. *Paenibacillus* strains share the same basal characteristics as *Bacillus* [1], in which the composition and structure of the major cell-wall polysaccharide varies depending on the species being examined, a fact that influences the characterization of the micro-organism when subjected to Gram staining. The species of this genus present variations in the S layer standing over the peptidoglycan, which may be thinner in some species [12].

Despite the methods for bacterial species identification being well-established and recorded in successive editions of Bergey's manuals, the advent of high-throughput sequencing has influenced the bacterial taxonomy field. Currently, it has been proposed that overall genome relatedness indexes (OGRI) are included in order to identity novel bacterial species [13], such as average nucleotide identity (ANI) [14]. The portability and wealth of information generated by genome sequencing are leading to the identification of many taxonomic misclassifications on all levels [15].

In the genus *Paenibacillus*, some species have already been reclassified using established genome metrics [16–19], since 16S rRNA gene analyses lack enough resolution to correctly identify species from this genus [16–18]. However, it was

Author affiliations: [1]Departamento de Patologia, Instituto de Ciências Biológicas, Universidade de Passo Fundo, Passo Fundo, Brazil; [2]Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.
***Correspondence:** Fernando Hayashi Sant'Anna, santanna.fh@gmail.com
Keywords: 16S rRNA phylogeny; Proteoma-core phylogeny; Genomic metrics; OGRI; overall genome relatedness index; genome metrics; taxonomy; reclassification.
Abbreviations: ANIb, average nucleotide identity based on BLAST+; dDDH, digital DNA–DNA hybridization; gANI, genome-wide average nucleotide identity; OGRI, overall genome relatedness index; OrthoANI, orthologous average nucleotide identity.
†These authors contributed equally to this work

004328 © 2020 The Authors

only in 2011 that *Paenibacillus massiliensis* DSM 16942^T and *Paenibacillus panacisoli* DSM 21345^T were compared [19] and it was proved that they presented 16S rRNA gene sequence similarity above 99%. In order to clarify the taxonomic status of both strains, this study used phylogenetic and genomic analyses.

For these analyses, closely related species to both strains were selected: *Paenibacillus illinoisensis* NBRC 15959^T, *Paenibacillus pabuli* NBRC 13638^T, *Paenibacillus amylolyticus* NBRC 15957^T, *Paenibacillus barcinonensis* CECT 7022^T and *Paenibacillus polymyxa* IAM 13411^T; *P. polymyxa* was utilized as outgroup in the phylogenetic analyses. In the genomic analyses, all strains were compared to *P. panacisoli* DSM 21345^T.

The 16S rRNA gene sequences from all strains were obtained from the Genbank database, aligned with the SINA aligner (version 1.2.11) [20] and sequence gaps were excluded with trimAl on the NGPhylogeny.fr platform [21, 22]. Phylogeny was reconstructed using the maximum-likelihood method, with the Hasegawa–Kishino–Yano substitution model [23] and Gamma distribution with five categories, with 1000 bootstrap replications, using MEGA-X software [24]. There were a total of 1533 positions in the final dataset. The parameters used for the analysis were those suggested by MEGA-X software, based on the aligned sequences. The resulting tree was rooted on the outgroup strain branch, *P. polymyxa* IAM 13411^T. The phylogenetic reconstruction of the 16S rRNA genes indicated that *P. panacisoli* DSM 21345^T and *P. massiliensis* DSM 16942^T form a monophyletic group with high bootstrap values (Fig. 1). The 16S rRNA gene identity values between both strains and their closely related species were computed in BioEdit version 7.0.5.3 [25]. *P. panacisoli* DSM 21345^T and *P. massiliensis* DSM 16942^T presented identity values for 16S rRNA gene higher than the species delimitation threshold of 98.7% [26] as shown on Table 1.

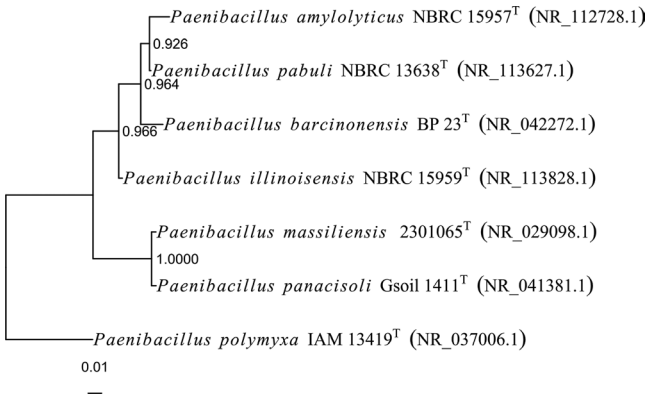

Fig. 1. Phylogeny of the 16S rRNA genes of *Paenibacillus massiliensis* 2301065^T, *Paenibacillus panacisoli* DSM 21345^T and their closest neighbours. The tree was built using the maximum-likelihood method. Bootstrap values greater than 0.9 are shown next to the branches. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. *Paenibacillus polymyxa* IAM 13419^T is the outgroup. Accession numbers of 16S rRNA gene sequences are next to the species name.

Table 1. Identity values of 16S rRNA genes of *Paenibacillus* type strains in relation to their counterpart from *Paenibacillus panacisoli* DSM 21345^T

Organism	Identity value (%)
Paenibacillus massiliensis* 2301065^T	**99.3**
Paenibacillus pabuli NBRC 13638^T	94.5
Paenibacillus amylolyticus NBRC 15957^T	93.9
Paenibacillus barcinonensis CECT 7022^T	93.7
Paenibacillus illinoisensis NBRC 15959^T	95.1
Paenibacillus polymyxa IAM 13411^T	93.1

*Bold, value above 98.7 of sequence identity. Accession numbers are provided in Fig. 1.

For genomic metrics and core-proteome analysis, all genomes were downloaded from the NCBI RefSeq Database [27] (Table 2). Core proteome analysis was conducted using GET_HOMOLOGUES version 03012019 [28], and the orthologous proteins were clustered using OrthoMCL version 1.4 [29] included in the GET_HOMOLOGUES package. The raw data from the OrthoMCL analysis are available on Figshare and can be accessed at https://figshare.com/articles/Core_proteome_files/11994258/1. The 2255 protein sequences of the core proteome were aligned using MUSCLE version 3.8.1551 [30], concatenated with the MEGA-X software [24] and sequence gaps were removed with trimAl on the NGPhylogeny.fr platform [21, 22], resulting in 726 049 amino acid positions. The core-proteome phylogeny was reconstructed with MEGA-X software [24], using the neighbour-joining method [31] with the Jukes–Cantor substitution model and 1000 bootstrap replicates in MEGA-X software [24]. The resulting tree was rooted on the *P. polymyxa* IAM 13411^T

Table 2. Genome sequences utilized

Strain	Accession no.	Length	G+C content (mol%)
Paenibacillus massiliensis 2301065^T	GCF_000377505	6385800	48.5
Paenibacillus panacisoli DSM 21345^T	GCF_000426545	6326414	48.27
Paenibacillus amylolyticus NBRC 15957^T	GCF_004001025	7110896	45.6
Paenibacillus barcinonensis CECT 7022^T	GCF_003217495	6261136	46.9
Paenibacillus illinoisensis NBRC 15959^T	GCF_004000925	6623886	47.1
Paenibacillus pabuli NBRC 13638^T	GCF_001514495	7329063	46.5
Paenibacillus polymyxa IAM 13411^T	GCF_900454525	5984949	45.1

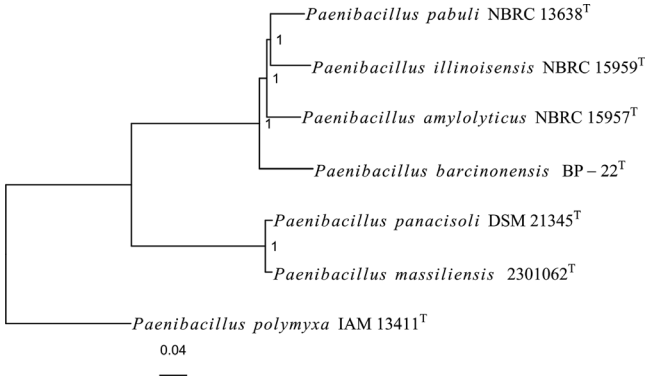

Fig. 2. Phylogeny of concatenated the core proteome of *Paenibacillus massiliensis* 2301065^T, *Paenibacillus panacisoli* DSM 21345^T and their closest *Paenibacillus* type strains. The tree was built using the neighbour-joining method. Bootstrap values greater than 0.9 are shown next to the branches. There were a total of 726049 positions in the final dataset.

branch. The core-proteome phylogenetic tree corroborates the 16S gene tree, in which *P. panacisoli* DSM 21345^T and *P. massiliensis* DSM 16942^T form a monophyletic clade with high bootstrap values (Fig. 2).

Furthermore, OGRI methods were conducted in order to assess their genomic similarities and solve this taxonomic issue. ANI values were computed using the JSpeciesWS Online Service [32]. Orthologous ANI (OrthoANI) values were calculated using OAT software version 0.93 [33]. MiSI method was utilized as described in Varghese *et al.* [34]. gANI values were computed using ANI calculator version 1.0 [34]. dDDH values were estimated using the GGDC 2.1 web service (Formula 2) [35]. All strains were compared to *P. panacisoli* DSM 21345^T. OGRI results and G+C content difference (Table 3) indicate that *P. panacisoli* DSM 21345^T and *P. massiliensis* DSM 16942^T belong to the same species with ANIb, OrthoANI and gANI values being above the species threshold – ANIb ≥95%; gANI ≥96.5% – and G+C percentage content differing less than 1% between them.

Lastly, the phenotypic and chemotaxonomic profiles of *P. panacisoli* DSM 21345^T and *P. massiliensis* DSM 16942^T were compared using their original published profiles [36, 37]. From 25 available phenotypic traits, they differ in 11 (Table 4). These differences could be consequence of intra-specific variation, and the low reproducibility of phenotypic and chemotaxonomic tests should be considered [36].

Results of phylogenetic and genomic analyses suggests that *P. panacisoli* DSM 21345^T and *P. massiliensis* DSM 16942^T are in fact members of the same species. Since the name *P. massiliensis* was validly published before *P. panacisoli* [36], the latter should be considered a later heterotypic synonym of the former.

Considering that these strains present a dDDH value lower than the cut-off for bacterial subspecies differentiation (79–80%) [38] and that they have distinct biochemical capabilities, we suggest that they form different subspecies of *P. massiliensis*. Therefore, we propose the reclassification of *Paenibacillus panacisoli* DSM 21345^T as *Paenibacillus massiliensis* subsp. *panacisoli* subsp. nov.

According to Rule 40d of the International Code of Nomenclature of Bacteria [39], the description of a novel subspecies which excludes the type strain of the species *P. massiliensis* automatically creates the subspecies *Paenibacillus massiliensis* subsp. *massiliensis* subsp. nov.

EMENDED DESCRIPTION OF *PAENIBACILLUS MASSILIENSIS* ROUX AND RAOULT 2004

Paenibacillus massiliensis (mas.si.li.en′sis. L. masc. adj. *massiliensis* of Massilia, the old Greek and Roman name for Marseille, where the type strain was isolated).

The description is based on Roux and Raoult [37], Ten *et al.* [36] and this study.

Cells are Gram-stain-positive, facultatively anaerobic rods (0.5 µm wide, 2.0–4.0 µm long). Ellipsoidal endospores are formed in swollen sporangia. The organism grows on routine media and forms translucent, beige-coloured, flat

Table 3. Genomic metrics of *Paenibacillus* type strains in relation to *Paenibacillus panacisoli* DSM 21345^T

Bold text indicates values above the species threshold. ANIb and OrthoANI threshold ≥95%, gANI threshold ≥96.5%. Brackets, confidence interval.

Strain	ANIb (%)	gANI (%)	OrthoANI (%)	G+C content difference (mol%)	dDDH (%)
Paenibacillus massiliensis 2301065^T	**95.57***	**96.51†**	**96.25**	**0.23**	66.9 [64–69.8]
Paenibacillus pabuli NBRC 13638^T	87.22	72.36	69.09	1.77	19.8 [17.6–22.2]
Paenibacillus amylolyticus NBRC 15957^T	74.15	72.19	68.86	2.67	19.4 [17.2–21.8]
Paenibacillus barcinonensis CECT 7022^T	72.47	72.42	70.44	1.37	20.3 [18.1–22.7]
Paenibacillus illinoisensis NBRC 15959^T	75.93	72.35	70.13	1.17	20.3 [18.1–22.8]
Paenibacillus polymyxa IAM 13411^T	84.32	72.97	70.06	3.17	19.5 [17.3–21.9]

*Aligned nucleotides=0.83.
†Alignment fraction=0.86.

Table 4. Phenotypic characteristics of the type strains of *Paenibacillus massiliensis* and *Paenibacillus panacisoli* recorded from the original descriptions in [37] and [36]

Strains: 1, *Paenibacillus massiliensis* 2301065^T; 2, *Paenibacillus panacisoli* DSM 21345^T. Highlighted (bold) text indicates divergent phenotypic characteristics between the compared *Paenibacillus* species and acid higher concentration for both bacterial isolates. –, Negative; +, positive; ND, not detected.

Phenotypic characteristic	**1**	**2**
Catalase	+	+
Oxidase	–	+
Growth at 50 °C	+	–
Growth at 44 °C	+	+
Anaerobic growth	+	+
Spore presence	+	+
Growth in presence of 5% NaCl	+	+
Nitrate reduction	+	+
Voges–Proskauer test	–	–
Gelatin liquefaction	–	+
Acid production from:		
Glycerol	+	+
D-Arabinose	–	+
Ribose	+	+
D-Xylose	–	+
Glucose	+	–
Rhamnose	–	–
Mannitol	+	+
N-Acetylglucosamine	–	–
Sucrose	+	–
Inulin	–	+
Raffinose	+	+
Starch	+	–
Glycogen	+	–
L-Fucose	–	–
Gluconate	–	+
Fatty acid composition:		
$C_{14:0}$	2.9	ND
$C_{15:0}$	1.6	2.8
$C_{16:0}$	14.1	9.4
iso-$C_{14:0}$	8.0	6.9
iso-$C_{15:0}$	5.4	5.0
iso-$C_{16:0}$	13.6	20.2
iso-$C_{17:0}$	1.0	2.1
anteiso-$C_{15:0}$	**42.8**	**44.8**
anteiso-$C_{17:0}$	2.9	5.6

colonies after incubation for 24 h at 30 °C. Bacteria are motile by means of peritrichous flagella. Catalase-positive and oxidase-variable. Optimal growth occurs at 30–37 °C, but variable growth at 50 °C. Growth occurs in the presence of 5% (w/v) NaCl. Nitrate is reduced, but gelatin liquefaction is variable. Acid is produced from glycerol, ribose, galactose, fructose, mannose, mannitol, amygdalin, arbutin, aesculin, salicin, cellobiose, maltose, lactose, melibiose, trehalose, raffinose and gentiobiose. Acid production from D-arabinose, D-xylose, inulin, gluconate, glucose, sucrose, starch and glycogen are variable. Acid is not produced from erythritol, L-xylose, adonitol, methyl β-D-xyloside, sorbose, rhamnose, dulcitol, inositol, sorbitol, methyl α-D-mannoside, methyl α-D-glucoside, *N*-acetylglucosamine, melezitose, xylitol, turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, 2-ketogluconate or 5-ketogluconate. The major fatty acids are anteiso-$C_{15:0}$, iso-$C_{16:0}$ and $C_{16:0}$.

The type strain, which was isolated from blood culture, is strain 2301065^T (=CIP 107939^T=CCUG 48215^T=DSM 16942^T). The DNA G+C content of the type strain is 48.5mol% with a genome size of 6.38 Mpb. The name *Paenibacillus panacisoli* (Ten *et al.* [36]) is a later heterotypic synonym.

DESCRIPTION OF *PAENIBACILLUS MASSILIENSIS* SUBSP. *MASSILIENSIS* SUBSP. NOV.

Description is as that given for *Paenibacillus massiliensis* by Roux and Raoult [37]. The type strain is 2301065^T (=CIP 107939^T=CCUG 48215^T=DSM 16942^T).

DESCRIPTION OF *PAENIBACILLUS MASSILIENSIS* SUBSP. *PANACISOLI* SUBSP. NOV.

Description as that given for *Paenibacillus panacisoli* by Ten *et al.* [36]. The type strain is Gsoil 1411 (=KCTC 13020^T=LMG 23405^T=DSM 21345^T).

Funding information
This work was funded by the CNPq/INCT-FBN (Conselho Nacional de Desenvolvimento Científico e Tecnológico/Instituto Nacional de Ciência e Tecnologia da Fixação Biológica de Nitrogênio, Brazil). F.G. and F.H.S. received scholarships from CAPES and CNPq, respectively.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Grady EN, MacDonald J, Liu L, Richman A, Yuan Z-C. Current knowledge and perspectives of *Paenibacillus*: a review. *Microb Cell Fact* 2016;15:203.
2. Uetanabaro AP, Wahrenburg C, Hunger W, Pukall R, Spröer C *et al*. *Paenibacillus agarexedens* sp. nov., nom. rev., and *Paenibacillus agaridevorans* sp. nov. *Int J Syst Evol Microbiol* 2003;53:1051–1057.
3. Montes MJ, Mercadé E, Bozal N, Guinea J. *Paenibacillus antarcticus* sp. nov., a novel psychrotolerant organism from the Antarctic environment. *Int J Syst Evol Microbiol* 2004;54:1521–1526.

4. Validov S, Kamilova F, Qi S, Stephan D, Wang JJ *et al*. Selection of bacteria able to control *Fusarium oxysporum* f. sp. radicis-lycopersici in stonewool substrate. *J Appl Microbiol* 2007;102:461–471.
5. de Souza R, Meyer J, Schoenfeld R, da Costa PB, Passaglia LMP. Characterization of plant growth-promoting bacteria associated with rice cropped in iron-stressed soils. *Ann Microbiol* 2015;65:951–964.
6. Fürnkranz M, Adam E, Müller H, Grube M, Huss H *et al*. Promotion of growth, health and stress tolerance of Styrian oil pumpkins by bacterial endophytes. *Eur J Plant Pathol* 2012;134:509–519.
7. Ker K, Seguin P, Driscoll BT, Fyles JW, Smith DL. Switchgrass establishment and seeding year production can be improved by inoculation with rhizosphere endophytes. *Biomass and Bioenergy* 2012;47:295–301.
8. Liu D, Yang Q, Ge K, Hu X, Qi G *et al*. Promotion of iron nutrition and growth on peanut by *Paenibacillus illinoisensis* and *Bacillus* sp. strains in calcareous soil. *Braz J Microbiol* 2017;48:656–670.
9. Ferchichi N, Toukabri W, Boularess M, Smaoui A, Mhamdi R *et al*. Isolation, identification and plant growth promotion ability of endophytic bacteria associated with lupine root nodule grown in Tunisian soil. *Arch Microbiol* 2019;201:1333–1349.
10. Abdallah Y, Yang M, Zhang M, Masum MMI, Ogunyemi SO *et al*. Plant growth promotion and suppression of bacterial leaf blight in rice by *Paenibacillus polymyxa* Sx3. *Lett Appl Microbiol* 2019;68:423–429.
11. Ash C, Priest FG, Collins MD. Molecular identification of rRNA group 3 bacilli (ash, Farrow, Wallbanks and Collins) using a PCR probe test. proposal for the creation of a new genus *Paenibacillus*. *Antonie van Leeuwenhoek* 1993;64:253–260.
12. Beveridge TJ. Mechanism of gram variability in select bacteria. *J Bacteriol* 1990;172:1609–1620.
13. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009;106:19126–19131.
14. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* 2005;102:2567–2572.
15. Sangal V, Goodfellow M, Jones AL, Schwalbe EC, Blom J *et al*. Next-Generation systematics: an innovative approach to resolve the structure of complex prokaryotic taxa. *Sci Rep* 2016;6:38392.
16. Sant'Anna FH, Ambrosini A, de Souza R, de Carvalho Fernandes G, Bach E *et al*. Reclassification of *Paenibacillus riograndensis* as a genomovar of *Paenibacillus sonchi*: genome-based metrics improve bacterial taxonomic classification. *Front Microbiol* 2017;8:8.
17. Sant'Anna FH, Ambrosini A, Guella FL, Porto RZ, Passaglia LMP. Genome-based reclassification of *Paenibacillus dauci* as a later heterotypic synonym of *Paenibacillus shenyangensis*. *Int J Syst Evol Microbiol* 2019;69:177-182.
18. Guella F, Porto RZ, Sant'Anna FH, Ambrosini A, Passaglia LMP. Genomic metrics analyses indicate that *Paenibacillus azotofixans* is not a later synonym of *Paenibacillus durus*. *Int J Syst Evol Microbiol* 2019;69:2870–2876.
19. Kim KK, Lee KC, Lee J-S. Reclassification of *Paenibacillus ginsengisoli* as a later heterotypic synonym of *Paenibacillus anaericanus*. *Int J Syst Evol Microbiol* 2011;61:2101–2106.
20. Pruesse E, Peplies J, Glöckner FO. Sina: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 2012;28:1823–1829.
21. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
22. Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F *et al*. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* 2019;47:W260–W265.
23. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 1985;22:160–174.
24. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
25. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT.. *Nucl Acids Symp Ser*1999.
26. Erko S, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006;8:6–9.
27. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D *et al*. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;44:D733–745.
28. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 2013;79:7696–7701.
29. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;13:2178–2189.
30. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
31. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425.
32. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 2016;32:929–931.
33. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 2016;66:1100–1103.
34. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K *et al*. Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 2015;43:6761–6771.
35. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:14.
36. Ten LN, Baek S-H, Im W-T, Lee M, Oh HW *et al*. *Paenibacillus panacisoli* sp. nov., a xylanolytic bacterium isolated from soil in a ginseng field in South Korea. *Int J Syst Evol Microbiol* 2006;56:2677–2681.
37. Roux V, Raoult D. *Paenibacillus massiliensis* sp. nov., *Paenibacillus sanguinis* sp. nov. and *Paenibacillus timonensis* sp. nov., isolated from blood cultures. *Int J Syst Evol Microbiol* 2004;54:1049–1054.
38. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V *et al*. Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci* 2014;9:2.
39. Parker CT, Tindall BJ, Garrity GM. International Code of Nomenclature of bacteria (2008 revision). *Int J Syst Evol* 2019;69:S1–DS111.

Final remarks

The number of currently known and described bacterial species is only a fraction of the actual quantity of species that exist on Earth. There are currently 23,566 species with names validly published under the ICNP, with this number falling to 19,409 when excluding synonyms (Parte et al. 2020). A recent study using 16S rRNA gene sequence estimated that there are between 2.2 and 4.3 million full-length prokaryotic operational taxonomic units (OTUs) worldwide (Louca et al. 2019). One of the reasons for this large disparity is because a significant number of microorganisms are hard to isolate or simply are "unculturable" (Rappé and Giovannoni 2003; Keller and Zengler 2004; Achtman and Wagner 2008).

Recent advancements in metagenomics provided the opportunity for researchers to sequence an entire community within a microbiome and identify how many potential species are present using computational analysis, with several studies managing to successfully identify novel bacterial species using computational analysis on metagenomic datasets (Wang et al. 2012; Tu et al. 2014; Pust and Tümmler 2021). Furthermore, recent studies using long-read genome assembly methods, such as nanopore (Jain et al. 2016), could sequence the complete genome of all bacterial organisms present in a metagenomic community with four or fewer contigs for each genome sequenced (Moss et al. 2020; Cuscó et al. 2021), although with reduced nucleotide accuracy in comparison to short-read assembly methods.

Currently, a new bacterial species is only recognized by ICNP when there is proof of deposit of viable cultures in at least two public culture collections from different countries with no restriction on availability (Trujillo and Oren 2018), which significantly hinders the identification and naming of uncultivated or unculturable microorganisms (Konstantinidis et al. 2017). Additionally, some countries like Brazil have proprietary laws that impose several restrictions when trying to deposit biological cultures of organisms isolated from their environments in international collections. Considering the mandatory deposit of cell cultures to be too constraining, a group of renowned scientists in the prokaryotic systematics field launched SeqCode (Hedlund et al. 2022), a code of nomenclature "based upon isolate genome, metagenome-assembled genome or single-amplified genome sequences" (Hedlund et al. 2022). Consequently, the improvements in metagenome

sequencing, allied with this new code of nomenclature based on genomic data will significantly increase the rate of newly identified species within the bacterial kingdom, which, as a result, will increase our reliance on in silico computational analysis for prokaryotic taxonomy.

In the first chapter of this thesis, we used comparative genomic analysis to prove that *Paenibacillus panacisoli* DSM 21345^T and *Paenibacillus massiliensis* DSM 16942^T are in fact subspecies of *P. massiliensis*, showing the potential of using OGRIs to improve the taxonomic assignment of bacterial species. Meanwhile, on the *Paenibacillus* genus alone, there are 278 strains deposited on the NCBI RefSeq database without species assignment that could be identified as a novel species via OGRIs if the SeqCode rules of nomenclature were to be adopted.

Finally, in the second and third chapters, we shift our focus to using comparative analysis to evaluate genus delineation within the Paenibacillaceae family. Even though the paper from the second chapter is mostly focused on the *Paenibacillus* genus, it serves as an introduction to what would be the main subject of the paper of the third chapter. In the second chapter, using only *Paenibacillus* strain sequences that showed inconclusive results by the NCBI taxonomic check, we were able to identify several strains assigned as *Paenibacillus* in the RefSeq database that do not show sufficient genomic similarity with the strain sequence of the type species of the genus, *Paenibacillus polymyxa*, which was corroborated by phylogenetic analysis. Beyond that, at least two monophyletic clades isolated from the main *Paenibacillus* clade formed exclusively of *Paenibacillus* strains were observed, indicating the possibility of two new genera. In the third and final chapter, we expanded our genomic analysis to all available sequences from the Paenibacillaceae family and proposed a method that selects a reference sequence for each current, and potentially new, genus and uses it to represent them at the genomic level, serving as a referential for comparison when using genomic analysis to identify the genus of a new bacterial organism.

While the initial intention of the work in the third chapter was to propose a taxonomic revision of the Paenibacillaceae family, the initial phylogenetic and genomic analysis results proved insufficient to define a clear genus boundary for the family. While

phylogenetic analysis included strain sequences from more than one genus on the same clade, the comparative genomic analysis showed AAI results above the cutoff value for genus delineation between sequences of different genera while at the same time having results below the cutoff point within sequences of the same genus. Even though those results clearly showed several genera misidentifications within the family, there were no clear boundaries that could properly isolate each group to effectively propose a taxonomic revision. The selection of a reference sequence for each group intended to add a second threshold layer for genus delineation when using OGRI and at the same time intended to develop a pipeline that uses AAI or POCP for genus delineation in a way that resembles the standard process of species delineation using ANI or GGDC. With the genus boundaries better defined within the Paenibacillaceae family, the next step is to propose the creation of all newly identified genera, while reclassifying any strain sequence that had its genus misidentified.

The use of whole genomic comparison for genus delineation is relatively recent in the scientific community but has shown great potential for taxonomic analysis in both new and old genera (Lopes-Santos et al. 2017; Chan et al. 2019; Xu et al. 2019; Yamano et al. 2022; Sreya et al. 2023). In case the SeqCode rules of nomenclature are generally adopted by taxonomists, the rate of inclusion of new organisms within the prokaryotic empire will increase significantly at every taxon level. Concomitantly, faster, and more reliable OGRI methods will need to be developed to be capable to evaluate continually growing datasets efficiently. In conclusion, in this thesis, by proposing the use of a reference strain, only one AAI test must be run to identify if a sequence does not belong to a determined genus, significantly reducing the time needed in the taxonomic assignment process. While the solution found in this thesis was extremely time-consuming at first, any further attempt to evaluate the Paenibacillaceae family will benefit greatly from the results found in this study. Furthermore, a method called FastAAI, which is about five orders of magnitude faster than the standard AAI, was proposed by Konstantinidis and collaborators and is currently available as a preprint (Konstantinidis et al. 2022), which could significantly reduce the process of reference strain selection.

REFERENCES

Achtman M and Wagner M (2008) Microbial diversity and the genetic nature of microbial species. Nature Reviews Microbiology 2008 6:6 6:431–440.

Amin AKMR, Tanaka M, Al-saari N, Feng G, Mino S, Ogura Y, Hayashi T, Meirelles PM, Thompson FL, Gomez-Gil B et al. (2017) *Thaumasiovibrio occultus* gen. nov. sp. nov. and *Thaumasiovibrio subtropicus* sp. nov. within the family Vibrionaceae, isolated from coral reef seawater off Ishigaki Island, Japan. Syst Appl Microbiol 40:290–296.

Ash C, Farrow JAE, Wallbanks S and Collins MD (1991) Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. Lett Appl Microbiol 13:202–206.

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

Auch AF, von Jan M, Klenk HP and Göker M (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci. doi: 10.4056/sigs.531120

Chan KG, See-Too WS, Chua KO, Peix Á, Goh KM, Hong KW, Yin WF and Lee LS (2019) *Aquella oligotrophica* gen. nov. sp. nov.: A new member of the family Neisseriaceae isolated from laboratory tap water. Microbiology Open. doi: 10.1002/MBO3.793

Chhe C, Uke A, Baramee S, Tachaapaikoon C, Pason P, Waeonukul R, Ratanakhanokchai K and Kosugi A (2023) *Insulambacter thermoxylanivorax* sp. nov., a thermophilic xylanolytic bacterium isolated from compost. Int J Syst Evol Microbiol 73:005724.

Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S et al. (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 68:461–466.

Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A and DiCuccio M (2018) Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. Int J Syst Evol Microbiol 68:2386–2392.

Collins MD, Lawson PA, Willems A, Cordoba JJ, Fernandez-Garayzabal J, Garcia P, Cai J, Hippe H and Farrow JAE (1994) The phylogeny of the genus Clostridium: Proposal of five new genera and eleven new species combinations. Int J Syst Bacteriol 44:812–826.

Cuscó A, Pérez D, Viñes J, Fàbregas N and Francino O (2021) Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. BMC Genomics 22:330.

Dsouza M, Taylor MW, Turner SJ and Aislabie J (2014) Genome-based comparative analyses of Antarctic and temperate species of *Paenibacillus*. PLoS One. doi: 10.1371/JOURNAL.PONE.0108009

Fang XM, Su J, Wang H, Zhang T, Zhao LL, Liu HY, Ma BP, Zhang YQ and Yu LY (2017) *Paenibacillus eucommiae* sp. nov., isolated from a traditional chinese medicinal herbal plant, *Eucommia ulmoides* oliver. Int J Syst Evol Microbiol 67:993–997.

G. E. Murray R, J. Brenner D, Colwell R, De Vos P, Goodfellow M, Grimont P, Pfennig N, Stackebrandt E and A. Zavarzin G (1990) Report of the Ad Hoc Committee on Approaches to Taxonomy within the Proteobacteria. Int J Syst Bacteriol. doi: 10.1099/00207713-40-2-213

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P and Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91.

Grady EN, MacDonald J, Liu L, Richman A and Yuan ZC (2016) Current knowledge and perspectives of *Paenibacillus*: a review. Microb Cell Fact. doi: 10.1186/S12934-016-0603-7

Hayashi Sant'Anna F, Bach E, Porto RZ, Guella F, Hayashi Sant'Anna E and Passaglia LMP (2019) Genomic metrics made easy: what to do and where to go in the new era of bacterial taxonomy. Crit Rev Microbiol 45:182–200.

Hedlund BP, Chuvochina M, Hugenholtz P, Konstantinidis KT, Murray AE, Palmer M, Parks DH, Probst AJ, Reysenbach A-L, Rodriguez-R LM et al. (2022) SeqCode: a nomenclatural code for prokaryotes described from sequence data. Nat Microbiol 7:1702–1708.

Henz SR, Huson DH, Auch AF, Nieselt-Struwe K and Schuster SC (2005) Whole-genome prokaryotic phylogeny. Bioinformatics. doi: 10.1093/bioinformatics/bth324

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT and Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications 2018 9:1 9:1–8.

Jain M, Olsen HE, Paten B and Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:239.

Kämpfer P, Rosselló-Mora R, Falsen E, Busse HJ and Tindall BJ (2006) *Cohnella thermotolerans* gen. nov., sp. nov., and classification of '*Paenibacillus hongkongensis*' as *Cohnella hongkongensis* sp. nov. Int J Syst Evol Microbiol 56:781–786.

Keller M and Zengler K (2004) Tapping into microbial diversity. Nature Reviews Microbiology 2004 2:2 2:141–150.

Konstantinidis K, Ruiz-Perez C, Gerhardt K, Rodriguez-R L, Jain C, Tiedje J and Cole J (2022) FastAAI: Efficient Estimation of Genome Average Amino Acid Identity and Phylum-level relationships using Tetramers of Universal Proteins. doi: 10.21203/rs.3.rs-1459378/v1

Konstantinidis KT, Rosselló-Móra R and Amann R (2017) Uncultivated microbes in need of their own taxonomy. The ISME Journal 2017 11:11 11:2399–2406.

Konstantinidis KT and Tiedje JM (2005a) Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences 102:2567–2572.

Konstantinidis KT and Tiedje JM (2005b) Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264.

Lee I, Kim YO, Park SC and Chun J (2016) OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103.

Lopes-Santos L, Castro DBA, Ferreira-Tonin M, Corrêa DBA, Weir BS, Park D, Ottoboni LMM, Neto JR and Destéfano SAL (2017) Reassessment of the taxonomic position of *Burkholderia andropogonis* and description of *Robbsia andropogonis* gen. nov., comb. nov. Antonie Van Leeuwenhoek 110:727–736.

Louca S, Mazel F, Doebeli M and Parfrey LW (2019) A census-based estimate of Earth's bacterial and archaeal diversity. PLoS Biol 17:e3000106.

Meier-Kolthoff JP, Auch AF, Klenk H-PP and Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:1–14.

Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL and Göker M (2022) TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res 50:D801–D807.

Moss EL, Maghini DG and Bhatt AS (2020) Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol 38:701–707.

Palmer M, Steenkamp ET, Blom J, Hedlund BP and Venter SN (2020) All anis are not created equal: Implications for prokaryotic species boundaries and integration of anis into polyphasic taxonomy. Int J Syst Evol Microbiol 70:2937–2948.

Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC and Göker M (2020) List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol 70:5607–5612.

Pust M-M and Tümmler B (2021) Identification of core and rare species in metagenome samples based on shotgun metagenomic sequencing, Fourier transforms and spectral comparisons. ISME Communications 2021 1:1 1:1–4.

Qin QL, Xie B Bin, Zhang XY, Chen XL, Zhou BC, Zhou J, Oren A and Zhang YZ (2014) A Proposed Genus Boundary for the Prokaryotes Based on Genomic Insights. J Bacteriol 196:2210.

Rappé MS and Giovannoni SJ (2003) The Uncultured Microbial Majority. https://doi.org/101146/annurev.micro57030502090759 57:369–394.

Richter M and Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences 106:19126–19131.

Rodriguez-R LM and Konstantinidis KT (2014) Bypassing cultivation to identify bacterial species. Microbe 9:111–118.

Rosselló-Móra R, Urdiain M and López-López A (2011) DNA–DNA Hybridization. Methods in Microbiology 38:325–347.

Schildkraut CL, Marmur J and Doty P (1961) The formation of hybrid DNA molecules and their use in studies of DNA homologies. J Mol Biol 3:595-IN16.

Schleifer KH (2009) Classification of Bacteria and Archaea: Past, present and future. Syst Appl Microbiol 32:533–542.

Shida O, Takagi H, Kadowaki K and Komagata K (1996) Proposal for two new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov. Int J Syst Bacteriol 46:939–946.

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

Sogin SJ, Sogin ML and Woese CR (1972) Phylogenetic measurement in procaryotes by primary structural characterization. J Mol Evol 1:173–184.

Soutar CD and Stavrinides J (2022) Phylogenomic analysis of the Erwiniaceae supports reclassification of *Kalamiella piersonii* to *Pantoea piersonii* comb. nov. and *Erwinia gerundensis* to the new genus *Duffyella* gen. nov. as *Duffyella gerundensis* comb. nov. Mol Genet Genomics 297:213–225.

Sreya P, Suresh G, Rai A, Ria B, Vighnesh L, Agre VC, Jagadeeshwari U, Sasikala C and Ramana CV (2023) Revisiting the taxonomy of the genus *Rhodopirellula* with the proposal for reclassification of the genus to *Rhodopirellula* sensu stricto, *Aporhodopirellula* gen. nov., *Allorhodopirellula* gen. nov. and *Neorhodopirellula* gen. nov. Antonie Van Leeuwenhoek 116:243–264.

Stackebrandt E (2011) Reports of Ad Hoc Committees for the Reevaluation of the Species Definition in Bacteriology. Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches. doi: 10.1002/9781118010518.ch12

Stackebrandt E, Frederiksen W, Garrity GM, Grimont PADD, Kämpfer P, Maiden MCJJ, Nesme X, Rosselló-Mora R, Swings J, Trüper HG et al. (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52:1043–1047.

Touzel JP, O'Donohue M, Debeire P, Samain E and Breton C (2000) *Thermobacillus xylanilyticus* gen. nov., sp. nov., a new aerobic thermophilic xylan-degrading bacterium isolated from farm soil. Int J Syst Evol Microbiol 50:315–320.

Trujillo ME and Oren A (2018) Avoiding 'salami slicing' in publications describing new prokaryotic taxa. Int J Syst Evol Microbiol 68:977–978.

Tu Q, He Z and Zhou J (2014) Strain/species identification in metagenomes using genome-specific markers. Nucleic Acids Res 42:e67–e67.

Vandamme P, Pot B and Gillis M (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiological and molecular biology reviews. doi: 10.1007/s12088-007-0022-x

Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH and Whitman B (2009) Bergey's Manual of Systematic Bacteriology, 2nd Edn, Vol. 3: The Firmicutes. New York, NY: Springer doi 10:978.

Wang Y, Chen Y, Zhou Q, Huang S, Ning K, Xu J, Kalin RM, Rolfe S and Huang WE (2012) A Culture-Independent Approach to Unravel Uncultured Bacteria and Functional Genes in a Complex Microbial Community. PLoS One 7:e47530.

Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E et al. (1987) Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. Int J Syst Evol Microbiol 37:463–464.

Woese CR and Fox GE (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Natl Acad Sci U S A 74:5088–5090.

Xu G tang, Xue H, Piao C gen, Guo M wei and Li Y (2019) *Ancrocorticia populi* gen. nov., sp. nov, isolated from the symptomatic bark of *Populus* × euramericana canker. Microbiology Open. doi: 10.1002/MBO3.792

Yamano R, Yu J, Jiang C, Haditomo AHC, Mino S, Sakai Y and Sawabe T (2022) Taxonomic revision of the genus *Amphritea* supported by genomic and in silico chemotaxonomic analyses, and the proposal of *Aliamphritea* gen. nov. PLoS One. doi: 10.1371/JOURNAL.PONE.0271174

Yao R, Wang R, Wang D, Su J, Zheng S and Wang G (2014) *Paenibacillus selenitireducens* sp. nov., a selenite-reducing bacterium isolated from a selenium mineral soil. Int J Syst Evol Microbiol 64:805–811.

Yoon SH, Ha S min, Lim J, Kwon S and Chun J (2017) A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110:1281–1286.

Zaitsev GM, Tsitko I V., Rainey FA, Trotsenko YA, Uotila JS, Stackebrandt E and Salkinoja-Salonen MS (1998) New aerobic ammonium-dependent obligately oxalotrophic bacteria: Description of *Ammoniphilus oxalaticus* gen. nov., sp. nov. and *Ammoniphilus oxalivorans* gen. nov., sp. nov. Int J Syst Bacteriol 48:151–163.

Zuckerkandl E and Pauling L (1965) Molecules as documents of evolutionary history. J Theor Biol 8:357–366.

(2010) List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol 60:469–472.

(1994) Validation of the Publication of New Names and New Combinations Previously Effectively Published Outside the IJSB: List No. 51†. Int J Syst Evol Microbiol 44:852.